Automatisierte Textanalyse in der Rechnungslegungsforschung: Erkenntnisstand und Methodenfragen

Automated textual analysis (ATA) in accounting research
<table>
<thead>
<tr>
<th></th>
<th>Agenda</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Introduction</td>
<td>What is automated textual analysis (ATA hereafter), and why should we care?</td>
</tr>
<tr>
<td>2. Applications</td>
<td>What does ATA allow us to do?</td>
</tr>
<tr>
<td>3. Evidence</td>
<td>Which research questions are being addressed using ATA, and what has been found?</td>
</tr>
<tr>
<td>4. Methods</td>
<td>How does ATA work?</td>
</tr>
<tr>
<td>5. Critical discussion</td>
<td>What are the challenges of working with ATA?</td>
</tr>
<tr>
<td>6. Conclusion</td>
<td>What would we like you to take away from this talk?</td>
</tr>
</tbody>
</table>
What is automated textual analysis (ATA)?

- ATA extracts information by **parsing texts for patterns** using quantitative and automated methods rather than hand collection.

**Content analysis**

- **Of text**
  - manual
  - automated

- **Of other media**
  - Video
  - Audio
  - Other

- Synonyms and related terms: Quantitative content analysis, (statistical) natural language processing (NLP), information retrieval, computational linguistics, quantitative semantics, stylometrics, text mining.

1) For example, Coval Shumway (2001 JoF); Hobson Mayew Venkatachalam (2012 JoF)
ATA methods promise new insights into long-standing questions that occupy *disclosure and capital-markets* research:

- Does narrative disclosure (i.e., text) contain valuation-relevant information – beyond that inherent in the numbers?  
  — Tetlock (*2007 JoF*): “linguistic .. content captures otherwise hard-to-quantify aspects of firms’ fundamentals”

- Can information in narratives detect/predict important economic conditions/events, e.g., fraud or bankruptcy?

- Do firms “manage” text attributes strategically – like earnings?

- Does the way in which disclosures are written affect the way in which users process the signals being communicated?  
  — IASB’s concern about “information overload”: *Disclosure Initiative*
ATA allows harvesting quantitative information from large bodies of text

- Facilitated by growing data processing, storage, transmission capacities
- Technical progress in computational linguistics and artificial intelligence (e.g., search engines, plagiarism software)
- Online availability of large text archives (e.g., SEC EDGAR, analysts’ reports, conference call transcripts, press articles, social media)

Growing focus on reproducibility of research – ATA less subjective?

Increasingly applied in practice

- SEC’s Accounting Quality Model software (“RoboCop”) 2)
- BlackRock fund „Global Long/Short Equity“ uses algorithms to extract sentiment from analyst reports, earnings releases, and social media

---

1) See, for example, Loughran McDonald (2015 SSRN); Li (2008 JoAccLit) | 2) Eaglesham (2013 WSJ)
1. Introduction
2. Applications
3. Evidence
4. Methods
5. Critical discussion
6. Conclusion
ATA methods use… diverse sources of text… to generate different “outcomes”

1. Text attributes
   — “Sentiment” conveyed?
   — How readable?
   — Redundant or informative?

2. Disclosure scores
   — Overall quantity?
   — Content-specific quantity?

3. Variables and datasets
   — Risk factor disclosures?
   — Questions on fair value?
Text corpora generally lend themselves for ATA when they are:

- Highly machine-readable
  - HTML (+)
  - PDF (-)
- Large
- Correct orthography
  - Twitter, OCR‘ed documents (-)
  - Spoken language (-)

Other aspects specific to research question pursued

- Firm-originated vs external
- Spontaneous vs rehearsed/scripted
- Written vs spoken
- Voluntary vs mandatory disclosure
- Author known vs unknown
- Free vs costly
- Structured vs unstructured
- Standardized vs free-form
- Recurring vs ad-hoc
- Audited vs unaudited
- Regulated vs discretionary
Firm-originated disclosures

- 10-Ks or annual reports

- Specific 10-K sections

- 8-Ks or ad-hoc reports
  Cooper He Plumlee (2015 SSRN)

- Earnings announcements/press releases

- Proxy statements
  Mukhopadhyay Shivakumar (2015 SSRN)

- Conference calls

- IPO prospectuses
  Weiss-Hanley Hoberg (2010 RFS)

Multiple channels

Kothari Li Short (2009 TAR): Corporations, analysts, business press

External sources

- Analyst reports

- News stories

- Twitter posts

- Investor message boards
  Antweiler Frank (2004 JoF)
Key challenge: **Construct validity**

- Defining appropriate empirical constructs/proxies to capture a given theoretical concept
<table>
<thead>
<tr>
<th>Theoretical concepts</th>
<th>Examples of common empirical constructs</th>
</tr>
</thead>
</table>
| **Readability**      | ▪ Gunning-Fog, Flesch Reading Ease, and Flesch-Kincaid indices  
                       ▪ File size, word count  
                       ▪ Measures of “plain English” writing style |
| (complexity 2),      |                                           |
| understandability 3) |                                           |
| **Similarity**       | ▪ Variants of Jaccard similarity coefficients based on the intersection of words, bigrams or higher-order n-grams across documents/texts  
                       ▪ Vector Space Model |
| (redundancy, staleness, comparability, consistency, boilerplate) |                                           |
| **Tone**             | ▪ Relative frequency of pessimistic (vs optimistic) words, based on some word list/dictionary, e.g., the Harvard psychosocial dictionary  
                       ▪ Holding tone constant: Vivid vs pallid language |
| (sentiment, affect, optimism/pessimism) |                                           |
| **Deceit**           | ▪ Deceptive language  
                       ▪ Vocal markers of cognitive dissonance  
                       ▪ Disclosure that is ‘abnormal’ in comparison |
| (lying, manipulation) |                                           |
| **Other** (selected) | ▪ Concrete vs abstract language  
                       ▪ Vivid vs pallid language  
                       ▪ Ethics-related terms |

1) Incomplete; terms vary  
2) Challenging to distinguish economic complexity from linguistic complexity (see backup)  
3) Readability ≠ understandability (e.g., Smith Taffler 1992 AAAJ)
**Readability**: “ability of individual investors and analysts to assimilate valuation-relevant information from a financial disclosure” ¹)

Three main approaches:

- **Fog index**  
  — For example, Li (2008 JAE)

- **Measures of “Plain English” writing style**  
  — Prevalence of common style weaknesses, e.g., measured using text editing software “StyleWriter—The Plain English Editor”  
  — For example, Miller (2010 TAR)

- **Text quantity** (word count, file size)  
  — For example, Li (2008 JAE) and Loughran McDonald (2014 JoF)

¹) Loughran McDonald (2014 JoF: 1649)
Gunning-Fog Index \(^1\) = \(0.4 \left[ \frac{\text{words}}{\text{sentences}} \right] + 100 \left( \frac{\text{complex words}}{\text{words}} \right)^{2}\)

- Original purpose: Selecting suitable texts for students/readers of different age classes and education levels
  - Fog Index = 12 indicates 12 years of schooling needed for understanding a text upon first reading
- Applied to 10-Ks \(^3\), analyst reports \(^4\), and media articles \(^5\) etc.
- Suitable for assessing readability of *financial* texts? \(^6\)
  - First component: Average sentence length
    - In financial texts, as “.” does not necessarily mark the end of a sentence
      - Lists, tables of contents, headers, decimal signs, abbreviations
  - Second component: Complex words
    - Do you find the following words particularly hard to understand?
      - Corporation, agreement, management, telecommunications, liability

---

1) Similar: Flesch Index and Flesch-Kincaid Index  
2) Complex words: > 2 syllables  
3) Examples include Li (2008 *JAE*); Lawrence (2013 *JAE*); Miller (2010 *TAR*); Lehavy Li Merkley (2011 *TAR*); Biddle, Hilary Verdi (2009 *JAE*)  
4) De Franco Hope Vyas Zhou (2015 *CAR* )  
5) For example the WSJ “Abreast of the market” column: Dougal Engelberg Garcia Parsons (2012 *RFS*)  
6) Loughran McDonald (2014 *JoF*); Jones and Shoemaker 1994 (*JAL*)
Attempts to measure overlap of two or more texts

- Intersections of $n$-grams across texts (Jaccard similarity coefficients)
- Vector Space Model (VSM)

Labels and interpretations

- **Positive/neutral** notion
  - Accounting “consistency”: Peterson Schmardebeck Wilks (2015 TAR)
  - Audit client “similarity”: Brown Knechel (2016 JAR)

- **Negative** notion
  - Redundancy: Cazier Pfeiffer (2015 AccHor)
  - Staleness of news: Tetlock (2011 RFS)
  - Boilerplate: Lang Stice-Lawrence (2015 JAE)
  - Lack of “modifications”: Brown Tucker (2011 JAR)
1. Text attributes: Tone

- **Tone** (sentiment, affect) capture the relative frequency of pessimistic and optimistic words in a narrative

- Made popular by Tetlock (2007 *JoF*) and Tetlock Saar-Tsechansky Macskassy (2008 *JoF*) in studies on the sentiment of news stories

- Tone measures reflect the underlying **word lists**), or dictionaries, used to classify words or phrases as optimistic or pessimistic

  — **Harvard University General Inquirer IV-4 psychosocial dictionary**

  — **Diction optimism/pessimism word lists (dictionsoftware.com)**
    - Examples: Davis Piger Sedor (2012 *CAR*); Davis Ge Matsumoto Zhang (2014 *RAS*)

  — **Combinations of different word lists**
    - Examples: Rogers Van Buskirk Zechman (2011 *TAR*); Li (2010 *JAR*); Davis Tama-Sweet (2012 *CAR*)

---

1) Other word lists have been developed especially for financial texts, e.g., by Henry (2008 *JBusComm*) and Loughran and McDonald (2011 *JoF*)
“Optimistic” *Diction* words common in financial texts

- outstanding
- respect
- determined
- power
- trust
- security
- authority

Examples of these words in non-optimistic contexts

- “common shares outstanding”
- “… with respect to …”
- “discount rate is determined by …”
- “electric power generation”
- “Contractual Trust Arrangement”
- “asset-backed security”
- “fiscal authority”

83% of the most common “optimistic” *Diction* words do not appear in Loughran McDonald’s (2011 *JoF*) list of positive words custom-made for financial texts.

1) For this slide and the next, see Loughran McDonald (2015 *JBehavFin*)
“Pessimistic” *Diction* words common in financial texts

- not
- no
- gross
- lynch
- death

Examples of these words in non-pessimistic contexts

- “EBITDA before special items is not defined in IFRS”
- “sales growth shows no sign of slowing”
- “gross profit”
- “Merrill Lynch”
- “employee fluctuation includes terminations, retirements and deaths”
- “plaintiffs allege bleeding and death”

70% of the most common “pessimistic” *Diction* words do not appear in Loughran McDonald’s (2011 JoF) list of negative words custom-made for financial texts.
Objective: Detect cues for lying in firms’ communications that can complement other known ‘red flags’ in detecting and predicting manipulation, accounting fraud, and financial misreporting

Example: Deceptive language in earnings conference calls

— Larcker Zakolyukina (2012 JAR)

— Based on four theoretical perspectives on behavior during lying/deceit based on theory developed in Vrij (2008 Detecting Lies and Deceit: Pitfalls and Opportunities): Emotions, cognitive effort, control, and lack of embracement
My personal favorite: Enron conference call of April 17, 2001

“Operator: Richard Grubman of Highfield Capital (a hedge fund)

…

Richard Grubman: You’re the only financial institution that cannot produce a balance sheet or cash flow statement with their earnings.

Jeff Skilling (Enron CFO): Thank you very much, we appreciate that.

Grubman: We appreciate that.

Skilling: A%%-hole.”
2. Disclosure scores

- Using ATA methods to condense large bodies of text documents into 'objective' proxies for the quality and/or quantity of disclosure

- Example of a relatively **abstract** proxy
  - Cooper He Plumlee (2015 *SSRN*)
  - Extract 8K*_Vdisc*, a measure of corporate voluntary disclosure \(^1\) from 8-Ks filed with the SEC
  - 8K*_Vdisc* = count of *voluntary* reportable items for a firm within a calendar quarter, regardless of the number of 8Ks filed

- Example of a more **content-related** proxy
  - Grüning (2011 *EAR*)
  - Artificial Intelligence Measure of Disclosure (AIMD) captures the extent of corporate disclosure in English-language documents along 10 information dimensions

\(^1\) Note how the authors avoid referring to 8K*_Vdisc* as a measure of voluntary disclosure *quality*. We believe not doing so is entirely adequate.
Idea of topic extraction: Using ATA to identify (and extract) specific information from large text corpora to build variables of interest

Example 1
— Campbell Chen Dhaliwal Lu Steele (2014 RAST)
— Extract quantitative measures of risk factor disclosures from 10-K filings

Example 2
— Daske Bischof Sextroh (2014 JBFA)
— Identify fair value-related text passages in conference call Q&A sessions

Example 3
— Lundholm Li Minnis (2013 JAR)
— Extract scaled number of references to competition in 10-K filings as a measure of a firm’s competitive environment
1. Introduction
2. Applications
3. Evidence
4. Methods
5. Critical discussion
6. Conclusion
I Disclosure attributes: Determinants

II Disclosure attributes = signals or predictors

- **Disclosure attributes**
  - Tone
  - Deceptive language
  - Readability

  - Stock market outcomes
  - Subsequent financial restatements
  - Subsequent performance / type of news

  **Economic outcomes**

III Disclosure attributes = moderators or mediators

- Fundamental signal
  - Readability/complexity
  - Disclosure tone
  - Media/press coverage

  Market outcomes
**Readability:** Li (2008 JAE)
- Interested in „management obfuscation hypothesis“
- 10-Ks of poor-performing firms less readable/more complex (effect statistically, *but not* economically, significant)
- Profits of firms with more readable 10-Ks are more persistent (effect statistically *and* economically significant)
- Concludes: “managers may be opportunistically structuring the annual reports to hide adverse information from investors”

**Tone:** Huang Teoh Zhang (2014 TAR)
- Estimate “abnormal positive tone” as a measure of discretionary tone
- Positively associated with upward perception management, such as just meeting/beating thresholds, or future earnings restatements
- Conclude: “evidence is consistent with managers using strategic tone management to mislead investors about firm fundamentals”
Does tone have information content?

- **Tetlock (2007 JoF)**
  - Negative tone in “Abreast of the Market” column (WSJ)
  - Predicts downward pressure on stock prices, reversion to fundamentals
  - Unusually high or low pessimism predicts high market trading volume

- **Tetlock Saar-Tsechansky Macskassy (2008 JoF)**
  - Negative tone in WSJ and Dow Jones News Service (DJNS) stories
  - Conveys negative information about future earnings above and beyond analysts’ forecasts and historical accounting data
  - Stock prices respond to the information in tone with a one-day delay

- **Kothari Li Short (2009 TAR)**
  - Analyze disclosure by firms, analysts, and the business press
  - Favorable (unfavorable) disclosures are accompanied by significant decreases (increases) in risk measures
Do text attributes help detect fraud?

- Larcker Zakolyukina (2012 JAR)
- Deceptive language in earnings conference calls
- Fraud indicated by:
  - More references to general knowledge (“you know”)
  - More first-person plural pronouns (“we”)
  - More general statements (e.g., “everybody”, “anybody”)
  - Extreme positive emotional words (“fantastic”)
  - More tentative (“possibly”), less certain words (“definitely”)
  - Negative statements (negation, anxiety, swear words, and anger)
Tone of earnings press releases 1), earnings conference calls 2), and MD&A disclosures 3) has information content that triggers market reactions and helps predict future performance.

Compensation disclosures affect say on pay voting 4)

Greater target 10-K length enhances M&A efficiency 5)

Unusually optimistic disclosure tone in earnings announcements enhances shareholder litigation risk 6)

Frequency of 184 constraining words (e.g., required, obligations, requirements, permitted, comply, and imposed) has incremental diagnostic power for financial constraints 7)

Disclosure complexity can affect investors' trading behavior

- Underreactions to 10-K filings increase in 10-K complexity, indicating complexity hampers the processing of information contained in 10-Ks 1)
- Small investors' trading around filings decreases in 10-K complexity 2)
- More readable disclosures yield stronger reactions from small investors 3)
  - Interestingly, less so when potential readability differences are pointed out

Two competing interpretations

- Information processing costs explanation
  - „consistent with … more complex filings being too costly for small investors to process in the short window surrounding the filing date.” 4)

- Processing fluency explanation
  - „processing fluency from a more readable disclosure acts as a subconscious heuristic cue and increases investors’ beliefs that they can rely on the disclosure”, even if amount of information acquired is constant. 5)

Processing fluency is subjective, and represents how easy it feels to process information.

Individuals like messages that feel easy to process.

Processing fluency is affected by, for example,

- Font: Easy to read – hard to read
- Font size: Easy to read – hard to read
- Color: Easy to read – hard to read
- Rhyming: “What sobriety conceals, alcohol reveals / unmasks”
- Simple versus complex synonyms: “deterioration“ vs „decline“

Processing fluency has been associated with higher ratings of truth, preference for the message and the messenger, willingness to rely on information, and confidence in judgments.

1) Taken, sometimes literally, from Rennekamp (2012 JAR: 1325-26)
10-K tone moderates market reactions to earnings announcements

- **Henry Leone (2016 *TAR*)**
  - Tone positively associated with market reaction to earnings announcements
  - Tone intensifies post-earnings announcement drift

- Similar results obtain for tone *change* in the MD&A of 10-Ks/10-Qs
  - Feldman Govindaraj Livnat Segal (2010 *RAS*7)
1. Introduction
2. Applications
3. Evidence
4. Methods
5. Critical discussion
6. Conclusion
The field of natural language processing is like the field of econometrics:
- vast
- with a rich body of theory
- full of philosophical details, disputes, and nuances

We viewed our task as being similar to explaining OLS while breezing over the underlying theory and proofs.

As you would with practical problems of inference with OLS (bias, collinearity, precision, etc.), we will focus on practical issues when using ATA.
“You shall know a word by the company it keeps” (Firth 1957, p. 11)  
— E.g., the word “race”

- What are common patterns that occur in a language?
- “The major tool we use to identify those patterns is to count things, otherwise known as statistics…” (Manning Schütze 1999, p. 4)
- Statistical methods deal better with the ambiguity in natural languages (Manning Schütze 1999)
- Provides large sample methods to search and learn from word patterns.
**Accountants**

- Assumes that transactions and their amounts characterize a firm / national economy.
- By properly aggregating transactions into accounts and relating them to each other as well as other firms, we conduct financial analysis, etc. to measure different attributes of a firm (performance, financial health etc.)
- Measurement problems (e.g., economic performance vs. accounting performance)

**Text Analysts**

- Assumes that terms and their frequencies characterize a text
- By properly aggregating term-frequencies and relating them to each other and other texts, one can potentially measure text attributes (sentiment, topic, readability, etc.)
- Measurement problems (e.g., readability vs. FOG Index)

“Accounting is the language of business.”
Two areas of application in accounting research (so far)

- **Information Extraction**
  - Disclosure scores
  - Novel datasets
  - Extract certain entities from (unstructured) data
  - E.g., firm names, audit partner, management forecasts, etc.

- **Text Attribute Computation**
  - Quantify certain properties/attributes of a given text numerically
  - E.g., sentiment, readability, complexity, similarity
General approaches to information extraction

Rules based

- E.g., Gate Software for Named Entity Recognition
- Simplest rule: keyword lists
- Precision varies a lot with tangibility of the desired entity/information and inherent structure of the texts.

Unsupervised Machine Learning

- Essentially sophisticated clustering methods. Searches for term-cluster pattern that repeat across texts.
- Clusters need to be interpret by researchers afterwards.

Supervised Machine Learning

- “Trains” a model based on hand-coded training samples.
- Trained model is used to classify or predict new observations.
- Model tries to find the „best rules“ for prediction
- Most widely used form in industrial applications today
Application example: extracting risk factors

Extracted data and computed three content analysis measures per subsection:
1. total word count
2. total key word count (predefined keyword list (literature and clustering))
3. key word count by risk subcategory (predefined by literature)

Source: “Fig. 1 Analysis steps in constructing qualitative measures of risk factor disclosure for each company”, Campbell, J. L., Chen, H., Dhaliwal, D. S., Lu, H. M., & Steele, L. B. (2014). The information content of mandatory risk factor disclosures in corporate filings. Review of Accounting Studies, 19(1), 396-455., p. 406
- **Using term frequencies to classify documents.** For example:
  - Does an earnings release contain Non-GAAP pro-forma items?
  - Business versus non-business related news

- **Finding discerning key terms.** For example, what terms are associated with/predict:
  - Risk factors
  - Restatements
  - Qualified audit opinions
  - Management forecasts
Often we are interested in certain attributes of a text (aka of a document, paragraph, sentence, etc.).

Examples

- **Similarity**: How similar are two texts? (For example, Tetlock 2011 uses text similarity to identify “old“ news.)
- **Readability**: How readable is a text?
- **Tone**: What sentiment does a text express?

In each case, a text needs to be made machine readable in some form for further quantification.

“The major tool we use to identify those patterns is to count things” (Manning Schütze 1999, p.4)
Idea: Frequency of certain terms will be high or low, depending on tone, readability, topic, etc.

Identify grammatical structure often necessary for identifying terms

Count how often a “term” (depending on application: word, phrase, etc.) appears in a text

Transform text into a pattern of terms and their frequencies in the text.

Disclaimer: There are quite a few non-trivial issues involved in such a transformation. For now we abstract from these issues on purpose, but will return to some later (e.g., “good” and “not good” are obviously different “terms” for sentiment detection)
**International Financial Reporting Standards**

*From Wikipedia, the free encyclopedia (Redirected from IFRS)*

**International Financial Reporting Standards (IFRS)** are designed as a common global language; company accounts are understandable and comparable across international boundaries. They are an international standard for accounting and trade and are particularly important for companies that have progressively replacing many different national accounting standards. They are the result of the work of the International Accounting Standards Committee (IASC), which is responsible for setting International Accounting Standards. They adopted existing IAS and Standing Interpretations Committee standards (SICs). The IASB's standards are the basis for the IASB's standards for financial reporting.

IFRS began as an attempt to harmonize accounting across the European Union but the very concept is attractive around the world. However, it has been debated whether or not to factor in the fact that was issued by IASC (the predecessor of IASB) and are still within use today go by the Standards (IAS), while standards issued by IASB are called IFRS. IFRS were issued between International Accounting Standards Committee (IASC). On 1 April 2001, the new International Accounting Standards Committee (IASC) took over from the IASC the responsibility for setting International Accounting Standards.

In the absence of a Standard or an Interpretation that specifically applies to a transaction, accounting policy that results in information that is relevant and reliable. In making the recognition criteria, and measurement concepts for assets, liabilities, income, and expenses.

Critics of IFRS are (1) that they are not being adopted in the US (see GAAP); (2) a number of criticisms of IFRS were made during 6 years in Zimbabwe's hyperinflationary economies. It had no positive effect at all during 6 years in Zimbabwe's hyperinflationary economies. It offered no response to the last criticism while IAS 29 is currently (March 2014) being implemented.


### 20 most frequent terms:

<table>
<thead>
<tr>
<th>Term</th>
<th>Frequency</th>
</tr>
</thead>
<tbody>
<tr>
<td>the</td>
<td>383</td>
</tr>
<tr>
<td>of</td>
<td>260</td>
</tr>
<tr>
<td>in</td>
<td>137</td>
</tr>
<tr>
<td>and</td>
<td>124</td>
</tr>
<tr>
<td>to</td>
<td>119</td>
</tr>
<tr>
<td>ifrs</td>
<td>95</td>
</tr>
<tr>
<td>financial</td>
<td>80</td>
</tr>
<tr>
<td>is</td>
<td>73</td>
</tr>
<tr>
<td>for</td>
<td>72</td>
</tr>
<tr>
<td>companies</td>
<td>55</td>
</tr>
<tr>
<td>that</td>
<td>54</td>
</tr>
<tr>
<td>are</td>
<td>53</td>
</tr>
<tr>
<td>standards</td>
<td>53</td>
</tr>
<tr>
<td>accounting</td>
<td>50</td>
</tr>
<tr>
<td>capital</td>
<td>49</td>
</tr>
<tr>
<td>as</td>
<td>42</td>
</tr>
<tr>
<td>statements</td>
<td>37</td>
</tr>
<tr>
<td>not</td>
<td>35</td>
</tr>
<tr>
<td>be</td>
<td>35</td>
</tr>
<tr>
<td>or</td>
<td>33</td>
</tr>
</tbody>
</table>

Length of unfiltered vocabulary: 1166
Identify sentence structure using “part-of-speech” tags (POS tags):*

"International Financial Reporting Standards (IFRS) are designed as a common global language for business affairs so that company accounts are understandable and comparable across international boundaries."

('International', 'NNP') ('Financial', 'NNP') ('Reporting', 'NNP') ('Standards', 'NNP') ('(', '(') ('IFRS', 'NNP') (')', ')') ('are', 'VBP') ('designed', 'VBN') ('as', 'IN') ('a', 'DT') ('common', 'JJ') ('global', 'JJ') ('language', 'NN') ('for', 'IN') ('business', 'NN') ('affairs', 'NNS') ('so', 'IN') ('that', 'DT') ('company', 'NN') ('accounts', 'NNS') ('are', 'VBP') ('understandable', 'JJ') ('and', 'CC') ('comparable', 'JJ') ('across', 'IN') ('international', 'JJ') ('boundaries', 'NNS') ('.', '.')

- e.g., “International Financial Reporting Standards (IFRS)” is labeled as a sequence of “NNP” tags—a proper noun.
  — Can be used to identify it as one, and not 5 separate terms
- Can be used to find negations (“not … good”), etc.

* Using the UPenn tagset here (e.g., JJ = adjective or numeral, ordinal)
Automated POS taggers trained on (hopefully) representative texts, but not without error

Essential for important preprocessing tasks such as:

- Assessing word sense (e.g., are we talking about “race” in the sense of a contest?)
- Identifying terms consisting of many words (e.g., “financial statements”)
- Identifying negations (“not … good”) for sentiment analysis and similar applications
- Stemming (count swimming, swam, swim as one term where appropriate)
“We delivered a strong quarter and are well underway in executing our Vision 2020. Therefore, we will raise our earnings outlook for 2016, even though the macroeconomic and geopolitical developments remain a concern for our markets. We continue to focus on addressing our structural challenges in the company and invest into further developing our markets and strengthening our innovation power…”

*Q1/2016 Earnings Release of Siemens AG*

*Transform (Term = combination of words + POS tag*)

*Using the UPenn tagset here (e.g., JJ = adjective or numeral, ordinal)*
“We delivered a strong quarter and are well underway in executing our Vision 2020. Therefore, we will raise our earnings outlook for 2016, even though the macroeconomic and geopolitical developments remain a concern for our markets. We continue to focus on addressing our structural challenges in the company and invest into further developing our markets and strengthening our innovation power…”

Q1/2016 Earnings Release of Siemens AG

Transform (Term = combination of words + POS tag*)

* Using the UPenn tagset here (e.g., JJ = adjective or numeral, ordinal)
Can we somehow identify which terms convey sentiment?

**Approach 1: Computer-derived classification**

- Collect a hand-coded training sample
- Label training sample documents as positive/negative: 1/0
- Run/test a classification model (e.g., logistic regression) with all terms (word plus position in sentence) as variables
- Model parameters tell you which terms have the most discriminatory power in this sample.
- Extract a list of most discriminating positive/negative words, or use model directly for prediction
- Especially useful if sentiment contextual (e.g., Twitter sentiment)

**Approach 2: Human-generated classification**

- Example: Harvard General Inquirer
- Widely used in accounting research
<table>
<thead>
<tr>
<th>Entry</th>
<th>Positiv</th>
<th>Negativ</th>
<th>...184 other classes ...</th>
<th>Othtags</th>
<th>Defined</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>A</td>
<td></td>
<td></td>
<td>DET ART</td>
<td>...</td>
</tr>
<tr>
<td>2</td>
<td>ABANDON</td>
<td></td>
<td></td>
<td>SUPV</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>ABANDONMENT</td>
<td></td>
<td></td>
<td>Noun</td>
<td></td>
</tr>
<tr>
<td>...</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>213</td>
<td>ADVANCE#1</td>
<td>Positiv</td>
<td></td>
<td>SUPV</td>
<td>47% verb: To move or bring forward, improve, promote</td>
</tr>
<tr>
<td>214</td>
<td>ADVANCE#1</td>
<td>Positiv</td>
<td></td>
<td>Noun</td>
<td>20% noun-adj: A moving forward, improvement, approach, in front, prior</td>
</tr>
<tr>
<td>215</td>
<td>ADVANCE#1</td>
<td>Positiv</td>
<td></td>
<td>Modif</td>
<td>20% adj: &quot;Advanced&quot;--forward, in front, progressive</td>
</tr>
<tr>
<td>216</td>
<td>ADVANCE#1</td>
<td></td>
<td></td>
<td>LY</td>
<td>13% idiom-adv: &quot;In advance&quot;--before, beforehand</td>
</tr>
<tr>
<td>...</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

“The General Inquirer is basically a mapping tool.”

- Maps each text file with counts on dictionary-supplied categories.
- Most of processing time spent on identifying commonly used word senses. (e.g., "race" as a contest, "race" as moving rapidly, "race" as a group of people of common descent’).
- Applies word stemming: Swimming -> swim
- “The 182 General Inquirer categories were developed for social-science content-analysis research applications, not for text archiving, automatic text routing, automatic text classifying, or other natural-language processing objectives... People, not computers, created these categories, although some category developers drew upon cluster analyses produced by computers. Many categories were initially created to represent social-science concepts of several grand theories that were prominent at the time the system was first developed...”

Therefore Vision We a addressing and are challenges company concern continue delivered developing... power quarter raise remain strengthening strong structural the though to underway we well will

Transform, word sense disambiguation, stemming, etc

Neg = \frac{\text{negative words}}{\text{total words}}

Proxy for the degree of negative sentiment
Fog Index

- A weighted measure, scaled to give the number of years of education needed to read the text.
- Count of "complex" words a significant determinant. Especially in financial texts
- Professional programs label terms as complex if consisting of three or more syllables.
- But there are exceptions (depending on how readability is defined in context)
  - Don’t count proper nouns (e.g., “International Financial Reporting Standards”),
  - Omit familiar jargon,
  - Sometimes split compound words (e.g., “skyscraper”).
  - Ignore common suffixes (such as -es, -ed, or -ing) as a syllables (e.g., reporting)
“International Financial Reporting Standards (IFRS) are designed as a common global language for business affairs so that company accounts are understandable and comparable across international boundaries”

Fog Index

\[
0.4 \left( \frac{\text{words}}{\text{sentences}} + 100 \frac{\text{complex words}}{\text{words}} \right)
\]

\[
0.4 \left( \frac{25}{1} + 100 \frac{11}{25} \right) = 27.6
\]
“International Financial Reporting Standards (IFRS) are designed as a common global language for business affairs so that company accounts are understandable and comparable across international boundaries.”

Fog Index

\[
0.4 \left( \frac{\text{words}}{\text{sentences}} + 100 \frac{\text{complex words}}{\text{words}} \right)
\]

\[
0.4 \left( \frac{25}{1} + 100 \frac{7}{25} \right) = 21.2
\]
X = "This is a text."
Y = "This is a text. This is a text."
Z = "This is a more elaborate text containing more words"

<table>
<thead>
<tr>
<th>Term</th>
<th>X</th>
<th>Y</th>
<th>Z</th>
</tr>
</thead>
<tbody>
<tr>
<td>a</td>
<td>1</td>
<td>2</td>
<td>1</td>
</tr>
<tr>
<td>containing</td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>elaborate</td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>is</td>
<td>1</td>
<td>2</td>
<td>1</td>
</tr>
<tr>
<td>more</td>
<td>0</td>
<td>0</td>
<td>2</td>
</tr>
<tr>
<td>text</td>
<td>1</td>
<td>2</td>
<td>1</td>
</tr>
<tr>
<td>this</td>
<td>1</td>
<td>2</td>
<td>1</td>
</tr>
<tr>
<td>words</td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
</tbody>
</table>

**Compute angle between vectors:**
- Similarity(X,Y): {'cos': 1.00, 'Angle': 0.0}
- Similarity(X,Z): {'cos': 0.55, 'Angle': 56.8}
- Similarity(Y,Z): {'cos': 0.55, 'Angle': 56.8}

Text attribute 3: Similarity (cont'd)

- Textual similarity of documents, paragraphs, passages, etc.
- Used in older search engine applications and other information retrieval systems
- Used in accounting and finance research to:
  - Identify old news, redundant text
  - “Boilerplate” text
- Other commonly used similarity measure: Jaccard distance
  - Measures how many common elements two sets contain
  - Ratio of the intersection of set elements to the union of set elements.
  - E.g., as a measure of stale news (Tetlock 2011)
1. Introduction
2. Applications
3. Evidence
4. Methods
5. Critical discussion
6. Conclusion
After this brief overview, we need to settle a few things

- How valid are these proxies?
- How can we put structure on this question?
- Example:

**Concept**
- e.g., Readability of annual reports

**Proxy**
- e.g., Fog Index of 10-Ks
- We are mostly concerned about correct inferences in regressions.
- So, let’s frame the problem as a form of measurement error:

\[ \text{Proxy} = \text{Concept}^* + \text{MErr} \]

- For example:

\[ \text{Fog}_{10K} = \text{Readability}_{AR}^* + \text{MErr} \]
Problems with constructs

True relation:

\[ y = \alpha_0 + \alpha_1 x_1 + \alpha_2 \text{Readability}_{AR} + \epsilon \]

What we measure:

\[ y = \alpha_0 + \alpha_1 x_1 + \alpha_2 \text{Fog}_{10K} + (\epsilon - \alpha_2 \text{MErr}) \]

- If \( \sigma(\text{Fog}_{10K}, \text{MErr}) = 0 \), “only” increases standard errors.
- **Problem** if \( \sigma(\text{Fog}_{10K}, \text{MErr}) \neq 0 \).
  - For example, if \( \text{Fog}_{10K} \) measures annual report readability worse for texts with high \( \text{Fog}_{10K} \) than for texts with low \( \text{Fog}_{10K} \).
  - Extreme case: \( \hat{\alpha}_2 \) will be zero if \( \text{Fog}_{10K} \) measures only noise.
  - Direction of bias tricky, if \( \text{Fog}_{10K} \) also correlated with omitted variables \( (V^*) \), i.e. picks up something unrelated to readability.
    \[ y = \alpha_0 + \alpha_1 x_1 + \alpha_2 \text{Fog}_{10K} + (\epsilon - \alpha_2 \text{MErr} + \alpha_3 V^*) \]
Sources of measurement error 1: What is a term?

- Highly context specific and large leeway ("financial statement" versus "financial" and "statement"

- Words can have multiple meanings (e.g., "race")

- "run", "ran", "running" one term? "company", "companies"? (stemming or lemmatization methods).
  - Sometimes important to count terms properly
  - Importance depends on the language

- Most sophisticated programs:
  - Define "term" as a word plus its position in the sentence (e.g., "absent" could be a verb or adjective)
  - Try some kind of word sense disambiguation.
Sources of measurement error 2: Zipf’s law

20 most frequent terms:

<table>
<thead>
<tr>
<th>Term</th>
<th>Frequency</th>
</tr>
</thead>
<tbody>
<tr>
<td>the</td>
<td>383</td>
</tr>
<tr>
<td>of</td>
<td>260</td>
</tr>
<tr>
<td>in</td>
<td>137</td>
</tr>
<tr>
<td>and</td>
<td>124</td>
</tr>
<tr>
<td>to</td>
<td>119</td>
</tr>
<tr>
<td>ifrs</td>
<td>95</td>
</tr>
<tr>
<td>financial</td>
<td>80</td>
</tr>
<tr>
<td>is</td>
<td>73</td>
</tr>
<tr>
<td>for</td>
<td>72</td>
</tr>
<tr>
<td>companies</td>
<td>55</td>
</tr>
<tr>
<td>that</td>
<td>54</td>
</tr>
<tr>
<td>are</td>
<td>53</td>
</tr>
<tr>
<td>standards</td>
<td>53</td>
</tr>
<tr>
<td>accounting</td>
<td>50</td>
</tr>
<tr>
<td>capital</td>
<td>49</td>
</tr>
<tr>
<td>as</td>
<td>42</td>
</tr>
<tr>
<td>statements</td>
<td>37</td>
</tr>
<tr>
<td>not</td>
<td>35</td>
</tr>
<tr>
<td>be</td>
<td>35</td>
</tr>
</tbody>
</table>

Length of unfiltered vocabulary: 1166
Term-Frequency distribution in texts:

- Zipf’s Law: some words make up most of a text.
- E.g., 52 complex words of ca. 45,000 complex words in 10-Ks make up ca. 25% of all complex words occurring in that sample (Loughran McDonald 2014, p. 1645)
- Often need to filter for “stopwords“ („and“, „the“, etc.) and other common non-informative but context specific words.

Really make sure the most frequent words measure what you want.

Variation in highly frequent terms usually dominates variation in the proxy (sentiment, Fog, etc.).

If that is the case, even small measurement error (i.e. 2 out of 300 terms), when arising in highly frequent terms, often leads to severe, correlated measurement error.

Consider weighting terms

- Sdee, for example, Henry and Leone (2016 TAR)
“If the Combination completes, the existing Shell Shareholders and the former BG Shareholders will own a smaller percentage of Shell than they currently own of Shell and BG, respectively. Existing Shell Shareholders and former BG Shareholders will own approximately 81% and 19% respectively of the outstanding Shell Shares. As a consequence, the number of voting rights which can be exercised and the influence which may be exerted by them in respect of the Combined Group will be reduced.” 1)

Example: Measuring tone

“If the Combination completes, the existing Shell Shareholders and the former BG Shareholders will own a smaller percentage of Shell than they currently own of Shell and BG, respectively. Existing Shell Shareholders and former BG Shareholders will own approximately 81% and 19% respectively of the outstanding Shell Shares. As a consequence, the number of voting rights which can be exercised and the influence which may be exerted by them in respect of the Combined Group will be reduced.”

- outstanding (i.e. “excellent”) vs. (shares) outstanding
- How frequent is “common shares outstanding”
  - ... in an IPO prospectus?
  - ... in M&A deal announcements?
- Very frequent in an IPO prospectus. Does not help explain variation? Consider reweighting it?
- In M&A deals, frequency maybe tied to whether share or cash deal.
  - Sentiment measure does not (only) pick up tone, but also deal type?

Is representative text available for the question at hand?
Not so much a problem with the method, but an important aspect for validity.
Can affect the accuracy of the automated methods (like POS tagging)

Example:
- Twitter sentiment
- What is the demographic that tweets, and why?
- Is Twitter sentiment, even if appropriately capturing the sentiment of that demographic, always a good indicator for broader sentiment in the stock market? In the economy? Society?
- As an aside, measuring sentiment on 140 character Twitter feeds posts very unique challenges.
1. Introduction
2. Applications
3. Evidence
4. Methods
5. Critical discussion
6. Conclusion
Opportunities

- Researchers share word lists and code
- Extract innovative datasets to address questions of interest using large-sample approaches
- Create finer, more tailored measures of phenomena of interest

Challenges

- Correlated omitted variables
  — For example: Separating (intentional) linguistic complexity from underlying economic complexity
  — Measuring a causal effect of readability holding complexity fixed might not be possible (Loughran MacDonald 2014 JoF, p. 1646)
- Online filing requirements currently put US researchers at an advantage
- Crowding out of subtle, subjective, but potentially more accurate measures not readily captured by algorithms
Challenging initial interpretations and elucidating causal mechanisms

- Disclosure readability and investor responses: Information processing costs or perceived reliability due to processing fluency? ¹)

- Disclosure tone and investor responses: Signaling of private information, manager traits or tone management? ²)

New, more nuanced measures – endless possibilities, but do they all make sense?

„Under-researched“ disclosures

Trying to understand „who writes“

Goal of this talk: Introduce developments in Automated Textual Analysis (ATA) and related methods; maintaining critical distance

ATA is the examination (and retrieval) of document content using computer algorithms

Increasingly relevant in practice and research

ATA is being applied to a diverse range of documents

To *measure* attributes of disclosure, and to *construct* disclosure proxies, novel variables and large datasets in a replicable way

Facilitates research into (1) disclosure properties as fundamental signals and (2) as moderators of the effect of such signals

ATA subject to challenges of construct validity and causal inference

ATA holds great promise, but biggest potatoes have been gathered

Construct validity and transparency key requirements for progress
Thank you
References


Bartov, Eli/Faurel, Lucile/Mohanram, Partha (2015): Can Twitter Help Predict Firm-Level Earnings and Stock Returns?


Cooper, Michael J./He, Jing/Plumlee, Marlene A. (2015): Voluntary Disclosure and Investor Sentiment.


Hoberg, Gerard/Lewis, Craig (2015): Do Fraudulent Firms Produce Abnormal Disclosure?


