

# 上場会社の開示資料におけるテキスト品質の測定指標に関するレビューと体系化

A Review and Synthesis of Textual Quality Measures in Listed Company Disclosures



東京大学 大学院工学系研究科 博士後期課程

**土井 惟成**

2015 年より国内金融機関に勤務。2022 年に東京大学大学院 工学系研究科 先端学際工学専攻博士後期課程に入学。

✉ doinobushige@gmail.com

## 1 はじめに

上場会社の財務報告は、財務数値だけでなく、経営成績、財政状態、将来の見通し、リスク及び経営者の評価を伝達する文章によって構成される。Fisher et al.<sup>[1]</sup> は、会計、監査及びファイナンスの領域では、会計基準、財務諸表、監査結果、過去及び将来の企業業績などを伝達するために多様な文書が用いられており、自然言語処理にはこれらの文書から情報を抽出し、分類または予測に利用する可能性があるとして整理している。Bochkay et al.<sup>[2]</sup> は、2010 年から 2021 年までの主要会計学術誌を調査した結果、テキスト分析（注 1）を利用した研究は 207 本に達し、年間の平均掲載数は 2010 ～ 2013 年の 6 本から、2018 ～ 2021 年には 32 本へ増加していた。

一方、テキストを数量的な変数へ変換する過程には測定上の不正確さが伴う。Loughran and McDonald<sup>[3]</sup> は、テキスト分析が会計・ファイナンス分野の伝統的な数量分析よりも相当程度不正確であり、文書を解析し、テキストを単語及び数値指標へ変換する過程における透明性と再現可能性が重要であると指摘している。Bochkay et al.<sup>[2]</sup> も、テキストから作成された指標が、研究対象となる概念を適切に表しているかという、構成概念妥当性（Construct validity）を重視している。

そこで本稿では、先行研究で用いられてきたテキスト品質の測定指標を整理し、それぞれが測定する属性と測定上の留意点を明らかにすることを目指す。

本稿の構成は次のとおりである。第 2 章ではテキス

ト品質の測定指標を列挙し、第 3 章では測定指標の体系化に係るディスカッションを述べ、第 4 章で結論を述べる。

なお、本稿は、既存研究で用いられた測定指標のレビューと体系化を目的とするものであり、各指標の測定性能を同一のデータ上で実証的に比較するものではない。また、本稿におけるテキスト品質とは、開示資料に含まれる文章が利用者に情報を伝達する際に有する測定可能な属性の総称をいう。したがって、本稿は、テキスト品質を単一の優劣尺度として捉えるのではなく、可読性やトーンなどの複数の側面から整理する。

## 2 テキスト品質の測定指標

本章では、各先行研究に記載された指標の定義及び主張に沿って、テキスト品質に関連する測定指標を整理する。レビュー対象は、会計・ファイナンス分野において、主として米国上場会社の 10-K をはじめとする、SEC（米国証券取引委員会）への提出書類を対象にテキスト指標を構築した研究である。従って、対象の言語は英語である。

### 2.1 可読性

Li<sup>[4]</sup> は、年次報告書の可読性を Fog Index と文書長を用いて測定している。Fog Index は、1 文当たりの平均単語数と 3 音節以上の単語の割合の合計に 0.4 を乗じて算定され、値が高いほど可読性が低い。文書長には総単語数の自然対数が用いられ、長い文書ほど情報処理費用が高くなることが指摘されている。

Loughran and McDonald<sup>[5]</sup> は、財務文書では company、management、liability などの多音節語が一般的に使用されるため、音節数に基づく Fog Index には測定上の問題があると指摘し、10-K のファイルサイズを代替指標として提示している。もっとも、同研究は、ファイルサイズが企業の事業構造の複雑性も反映し得るとしている。

Loughran and McDonald<sup>[6]</sup> は、SEC の Plain English 規則に基づき、平均文長、平均単語長、受動態、法律用語、人称代名詞及び冗長な表現から Plain English 指標 (Plain English measure) (注 2) を作成している。Bonsall et al.<sup>[7]</sup> は、文長、受動態、隠れた動詞、冗長表現、専門用語及び単語の親近性等を反映する Bog Index を提案している (注 3)。また、10-K のファイルサイズには HTML、XML、画像及び添付資料等が含まれるため、その変動が本文の可読性を表さない場合があると指摘している。

## 2.2 トーン・文章構造

Loughran and McDonald<sup>[8]</sup> は、Harvard Dictionary を 10-K に適用した場合、否定語と判定される語の 73.8% が財務文脈では通常否定的でないと報告している。同研究は、財務文書における実際の用法に基づき、否定、肯定、不確実性、訴訟関連、強いモダリティ (確信度の高い表現) や弱いモダリティ (推測や可能性を表す表現) の単語リストを作成し、各分類に属する語の出現数または割合によってトーン (テキストにおける感情的なニュアンス) を測定している。

トーンの測定は、単語の出現数や割合だけでなく、文書内における配置にも拡張されている。Allee and DeAngelis<sup>[9]</sup> は、肯定語または否定語が文章内にどの程度均等に配置されているかを tone dispersion と定義し、Average Reduced Frequency (ARF) によって測定している。ARF は、文書をトーン語数と同数の区間に分割し、トーン語を含む区間の割合を対象に、開始位置を変えて反復計算した平均値であり、値が高いほどトーン語が文書全体に分散している。また、同研究では、トーン語数との機械的な関係を調整した指標も用いられている。

## 2.3 類似性・更新性

Brown and Tucker<sup>[10]</sup> は、当年度と前年度の

MD&A を単語頻度ベクトルとして表し、TF-IDF で加重したコサイン類似度を算定している。文書の相違度は 1 から類似度を控除した値であり、さらに、文書長と相違度との機械的な関係を調整した MD&A modification score を使用している。

Cohen et al.<sup>[11]</sup> は、前期の 10-K または 10-Q との類似性を、コサイン類似度、Jaccard 類似度、最小編集距離及び文書の並列比較によって測定している。コサイン類似度は単語の出現頻度、Jaccard 類似度は共通する単語集合、最小編集距離は文書を変換するために必要な操作数、並列比較は追加、削除及び変更された語数に基づく。

Lang and Stice-Lawrence<sup>[12]</sup> は、同一業種・同一年度の比較対象企業とのコサイン類似度の中央値を、企業間の比較可能性の指標としている。

## 2.4 ボイラープレート・粘着性・冗長性

Lang and Stice-Lawrence<sup>[12]</sup> は、多数の企業に共通する 4 単語句を含む文の単語数が、年次報告書の総単語数に占める割合を boilerplate (ボイラープレート) としている。

Dyer et al.<sup>[13]</sup> は、同一年度の企業の 75% 以上に共通する 4 単語句を含む文の単語を Boiler Words、前年度の 10-K と同一の 8 単語句を含む文の単語を Sticky Words、同一 10-K 内の別の箇所ですべて反復される文の単語を Redundant Words とし、それぞれを対象部分の総単語数で除すことにより、boilerplate (ボイラープレート)、Stickiness (粘着性)、Redundancy (冗長性) を算出している。

## 2.5 具体性

Hope et al.<sup>[14]</sup> は、リスク要因開示に含まれる人名、地名、組織名、割合、金額、時刻及び日付の数を総単語数で除した値を Specificity としている。これらは Named Entity Recognition によって抽出され、値が高いほど企業固有の詳細を含む開示である。同研究は、人による評価との比較によって指標の妥当性を検証している。

Dyer et al.<sup>[13]</sup> も同様の固有表現を用いて 10-K 全体の Specificity を測定している。また、日付や節番号等を除く情報性のある数値の数を総単語数で除した HardInfoMix を、定量情報の密度を表す指標としている。

表 1 テキスト品質指標の体系と解釈上の留意点

| 上位属性   | 第2章の区分           | 代表的な指標                                                          | 主な測定対象                      | 解釈上の留意点                                        |
|--------|------------------|-----------------------------------------------------------------|-----------------------------|------------------------------------------------|
| 伝達容易性  | 可読性              | Fog Index、文書長 [4]、ファイルサイズ [5]、Plain English指標 [6]、Bog Index [7] | 文長、語彙、文体及び情報処理量             | 指標ごとに測定対象が異なり、文書長やファイルサイズは企業の複雑性や非テキスト情報も反映し得る |
| 意味内容   | トーン              | 財務文書向け単語リスト [8]                                                 | 肯定、否定、不確実性、訴訟関連及びモダリティ      | 財務文脈に適合した辞書が必要であり、否定的トーンの多さが直ちに低品質を意味するものではない  |
| 文章構造   | 文章構造             | Tone dispersion、ARF [9]                                         | トーン語の文書内における分布              | トーン語が分散している程度を表すが、情報の提示順序や因果関係は測定しない           |
| 更新・反復  | 類似性・更新性          | コサイン類似度、Modification score [10]、Jaccard類似度、最小編集距離 [11]          | 過年度からの変更または企業間の類似性          | 同じ類似度であっても、比較対象によって更新性または比較可能性を表す              |
|        | ボイラープレート・粘着性・冗長性 | Boilerplate [12]、Stickiness、Redundancy [13]                     | 企業間の定型表現、過年度から継続する表現、文書内の反復 | 共通句の長さや判定基準によって測定値が変化する                        |
| 情報の具体性 | 具体性              | Specificity [13]、HardInfoMix [14]                               | 固有名詞、金額、割合、日付及び情報性のある数値     | 値が高いほど具体的または定量的であるが、開示内容の完全性を直接示すものではない        |
| 意味内容   | リスク情報            | リスク文数とその変化 [15]、リスク関連語数及び種類別語数 [16]                             | リスク情報の量及び種類                 | リスク情報の増加は、開示品質だけでなく企業が直面する実際のリスクを反映し得る         |
|        | トピック             | LDAによるトピック構成割合 [13]                                             | 文書を構成する潜在的な主題               | トピック数の設定及び各トピックの名称には研究者の判断が伴う                  |

出典：先行研究を踏まえて筆者作成

## 2.6 リスク情報・トピック

Kravet and Muslu<sup>[15]</sup> は、risk、uncertain、may、could 等のリスク関連語を一つ以上含む文をリスク文とし、その文数を Risk Disclosure としている。新たなリスク情報を測定するため、当年度と前年度のリスク文数の差も用いている。

Campbell et al.<sup>[16]</sup> は、10-K のリスク要因セクションについて、総単語数、リスク関連語数及びリスクの種類別の関連語数を測定している。同研究は、既存研究の関連語に、文書内で共起する単語に基づいて潜在的なトピックを識別する潜在的ディリクレ配分法 (Latent Dirichlet Allocation: LDA) によって抽出した語を加えた 300 語超のリストを作成し、関連語を財務、税務、訴訟、その他のシステムティック及びその他の固有リスクに分類している。

リスク関連語のように特定の語句をあらかじめ設定する方法に対し、文書集合から潜在的な主題を推定する方法としてトピックモデルが用いられている。Dyer et al.<sup>[13]</sup> は、LDA を用い、10-K における各トピックの構成割合を測定している。同研究は、モデルの適合度

と人による解釈可能性を検討して 150 トピックを採用し、これらを 13 のカテゴリーに集約している。また、1996 年から 2013 年までの 10-K の長文化の大部分が、公正価値・減損、内部統制及びリスク要因に関する三つのトピックの増加によると報告している。

## 3 テキスト品質指標の体系化

第 2 章で整理した指標は、測定対象に基づき、(1) 伝達容易性、(2) 意味内容、(3) 文章構造、(4) 更新・反復、(5) 情報の具体性という五つの上位属性に体系化できる。可読性は伝達容易性、トーン、リスク情報及びトピックは意味内容、tone dispersion は文章構造、類似性・更新性ならびにボイラープレート・粘着性・冗長性は更新・反復、Specificity 及び HardInfoMix は情報の具体性に対応する。もっとも、Fog Index、Plain English 指標、Bog Index、文書長及びファイルサイズは、いずれも可読性に関連する指標であるが、それぞれ測定する文章属性が異なる。

表 1 では、各指標について、上位属性、第 2 章の区分、

代表的な指標、主な測定対象および解釈上の留意点を整理する。可読性、トーン、具体性及びリスク情報は、原則として単一文書から算定される。文章構造と冗長性は同一文書内の配置または反復を、更新性と粘着性は同一企業の過年度文書との関係を、ボイラプレートと比較可能性は他企業の文書との関係を測定する。トピックモデルは、文書集合における単語の共起関係から各文書の内容構成を測定する。

この整理から、各指標が一律にテキスト品質の高低を示すものではないことが分かる。可読性、具体性、更新性及び反復性は、主として情報がどのように伝達されているかを測定する。一方、トーン、リスク情報及びトピックは、主として何が記載されているかを測定する。例えば、否定的トーンやリスク情報の多さは、文章の品質ではなく企業の経済状態を反映する可能性がある。また、過年度文書との高い類似性は更新の少なさを示す一方、同業他社との高い類似性は比較可能性を表すため、同じ計算方法であっても比較対象によって指標の意味が異なる。

したがって、指標を選択する際には、測定対象となる構成概念、分析単位及び比較対象をあらかじめ明示する必要がある。また、文書の取得範囲、表や添付資料の除外、文及び単語の識別方法を開示し、人手評価や代替指標との比較によって妥当性を検証することが求められる。これは、Bochkay et al.<sup>[2]</sup> が重視する構成概念妥当性と、Loughran and McDonald<sup>[3]</sup> が重視する測定過程の透明性及び再現可能性に対応する。

以上から、本稿では、テキスト品質を単一の指標ではなく、伝達容易性、意味内容、文章構造、更新・反復及び情報の具体性から構成される多次元的な概念として整理する。

## 4 おわりに

本稿では、上場会社の開示資料に用いられてきたテキスト品質の測定指標をレビューし、可読性、トーン、文章構造、類似性・更新性、ボイラプレート・粘着性・冗長性、具体性、リスク情報及びトピックに整理した。さらに、これらを伝達容易性、意味内容、文章構造、更新・反復及び情報の具体性という五つの上位属性に体系化した。

レビューの結果、テキスト品質を単一の指標によって包括的に測定することは困難であることが示された。各

指標は、文章の読みやすさ、記載内容、情報の配置、過年度または他社の文書との関係、企業固有情報の程度など、異なる属性を測定している。また、否定的トーン、リスク情報または文書間の類似性のように、指標の値が高いことを直ちに高品質または低品質と解釈できない場合もある。したがって、指標の利用に当たっては、測定対象となる構成概念を明確にする必要がある。

テキストから作成される指標には、文書の取得範囲、表や添付資料の処理、文及び単語の識別、辞書の選択など、測定結果に影響する多くの判断が含まれる。このため、人手による評価や代替指標との比較を通じて構成概念妥当性を確認するとともに、前処理を含む測定手続を開示し、透明性と再現可能性を確保することが重要である<sup>[2,3]</sup>。

本稿には、レビュー対象が選定した先行研究に限られていること、各指標の測定性能を共通のデータによって比較していないことなどの限界がある。また、レビューした指標の多くは、英語で作成された米国の開示資料を対象としている。例えば、Plain English 指標は SEC が示した英語表現上の要素に基づいており、Fog Index は文長と音節数を用いているため、日本語の開示資料にそのまま適用することはできない。

特に日本語では、英語と異なり単語境界が明示されないため、分かち書きや形態素解析の方法が単語数、専門語、固有表現及びトーン語の測定値に影響する。また、音節数に基づく Fog Index や、英語の受動態・人称代名詞を前提とする Plain English 指標は、日本語の文体評価に対応する形で再定義する必要がある。さらに、日本語の開示資料では、単語単位の処理だけでなく、文、段落、項目または開示セクションの境界設定が、可読性、冗長性、類似性及びトピック構成の測定値に影響するため、分析単位の設計自体を検証対象に含める必要がある。

今後は、日本語の語彙、文法、分かち書き及び開示慣行を踏まえた指標を構築し、人による評価との対応を検証する必要がある。さらに、有価証券報告書、決算短信、適時開示資料及び統合報告書など、文書の種類や利用者によって適切な指標が異なる可能性についても検討が求められる。こうした検証を積み重ねることにより、上場会社の開示資料におけるテキスト品質を、より妥当かつ再現可能な形で測定することが可能になると考えられる。

## 注

(注 1) Bochkay et al.<sup>[2]</sup> は、自然言語処理をテキストデータに適用し、情報を自動的に抽出または測定することをテキスト分析と定義している。

(注 2) Loughran and McDonald<sup>[6]</sup> は、Plain English 指標を、SEC の Plain English 規則において示された文章上の要素を集約して算定している。同指標は、次に示す六つの構成要素からなる。

(1) 1 文当たりの平均単語数

(2) 単語の平均文字数

(3) 受動態の使用割合

(4) 法律用語及び法律的表現の使用割合

(5) 一人称複数及び二人称の人称代名詞の使用割合

(6) 否定的な複合表現、冗長表現及び「respectively」の使用割合

各構成要素を平均 0、標準偏差 1 に標準化した上で、人称代名詞については正の符号、その他の五つの構成要素については負の符号を付して合計する。したがって、Plain English 指標の値が高いほど、文書が SEC の Plain English に関する文章基準に適合していることを示す。なお、同研究では、SEC の Plain English Handbook が「簡潔さ」ではなく「明瞭さ」を目的としていることから、文書長は同指標には含めず、回帰分析における統制変数として扱っている。

(注 3) Bog Index は、Sentence Bog と Word Bog の和から Pep を引いた値として定義される。Sentence Bog は平均文長に基づく。Word Bog は、受動態、名詞化された動詞、冗長表現、法律用語、常套句、抽象語及び回りくどい表現等の文章上の問題と、一般語彙、略語及び専門用語の難易度から構成される。単語の難易度は、音節数ではなく、20 万語を超える独自の語彙リストに基づく、親しみやすさ及び正確性によって評価される。Pep は、人名、興味を引く語、会話的表現及び文長の変化等に基づく控除項目である。

## 参考文献

[1] Fisher, I. E., Garnsey, M. R. and Hughes, M. E. (2016) "Natural Language Processing in Accounting, Auditing and Finance: A Synthesis of the Literature with a Roadmap for Future Research," *Intelligent Systems in*

*Accounting, Finance and Management*, 23, 157-214.

[2] Bochkay, K., Brown, S. V., Leone, A. J. and Tucker, J. W. (2023) "Textual Analysis in Accounting: What's Next?" *Contemporary Accounting Research*, 40, 765-805.

[3] Loughran, T. and McDonald, B. (2016) "Textual Analysis in Accounting and Finance: A Survey," *Journal of Accounting Research*, 54, 1187-1230.

[4] Li, F. (2008) "Annual Report Readability, Current Earnings, and Earnings Persistence," *Journal of Accounting and Economics*, 45, 221-247.

[5] Loughran, T. and McDonald, B. (2014) "Measuring Readability in Financial Disclosures," *Journal of Finance*, 69, 1643-1671.

[6] Loughran, T. and McDonald, B. (2014) "Regulation and Financial Disclosure: The Impact of Plain English," *Journal of Regulatory Economics*, 45, 94-113.

[7] Bonsall, S. B. IV, Leone, A. J., Miller, B. P. and Rennekamp, K. (2017) "A Plain English Measure of Financial Reporting Readability," *Journal of Accounting and Economics*, 63, 329-357.

[8] Loughran, T. and McDonald, B. (2011) "When Is a Liability Not a Liability? Textual Analysis, Dictionaries, and 10-Ks," *Journal of Finance*, 66, 35-65.

[9] Allee, K. D. and DeAngelis, M. D. (2015) "The Structure of Voluntary Disclosure Narratives: Evidence from Tone Dispersion," *Journal of Accounting Research*, 53, 241-274.

[10] Brown, S. V. and Tucker, J. W. (2011) "Large-Sample Evidence on Firms' Year-over-Year MD&A Modifications," *Journal of Accounting Research*, 49, 309-346

[11] Cohen, L., Malloy, C. and Nguyen, Q. (2020) "Lazy Prices," *Journal of Finance*, 75, 1371-



1415.

- [12] Lang, M. and Stice-Lawrence, L. (2015) "Textual Analysis and International Financial Reporting: Large Sample Evidence," *Journal of Accounting and Economics*, 60, 110–135.
- [13] Dyer, T., Lang, M. and Stice-Lawrence, L. (2017) "The Evolution of 10-K Textual Disclosure: Evidence from Latent Dirichlet Allocation," *Journal of Accounting and Economics*, 64, 221–245.
- [14] Hope, O.-K., Hu, D. and Lu, H. (2016) "The Benefits of Specific Risk-Factor Disclosures," *Review of Accounting Studies*, 21, 1005–1045.
- [15] Kravet, T. and Muslu, V. (2013) "Textual Risk Disclosures and Investors' Risk Perceptions," *Review of Accounting Studies*, 18, 1088–1122.
- [16] Campbell, J. L., Chen, H., Dhaliwal, D. S., Lu, H. and Steele, L. B. (2014) "The Information Content of Mandatory Risk Factor Disclosures in Corporate Filings," *Review of Accounting Studies*, 19, 396–455.