

非財務報告書の類型化の試み

大坪 史治
黄 海 湘

1. はじめに

わが国の非財務報告書は、1990年代初頭に登場した環境報告書からはじまり、拡充と再編が繰り返されるなかで、多種多様に分化している。最近では、非財務報告書とアニュアルレポートの結合や IIRC (International Integrated Reporting Council) の牽引する統合報告フレームワークの実践など国際的に著しく進展しており、財務情報と非財務情報との距離がますます接近している。わが国においても統合報告を意識する企業は、1994年から2015年にかけての期間で158社を確認している¹。

本稿で示す非財務報告書とは、環境報告書、レスポンシブル・ケア報告書、環境・社会報告書、持続可能性報告書、CSR 報告書²、統合報告を意識した報告書 (IIRC 統合報告書および広義の統合報告書³) などであり、営利企業においては任意開示されている報告書である。それ故、非財務報告書の形態、内容、ボリューム、参考とするガイドラインなどは、組織により異なる。

今後、組織の規模や業種によりますます報告書の形態が多様化することを想定すれば、非財務報告書の類型化は、非財務報告書の変遷、トレンド、枠組みの理解はもとより、将来予測や企業ごとの位置関係を把握し迅速に報告書作成に反映することができる点で大きな意味がある。報告書の作成者側にとっても、非財務報告書の類型と自らの組織を照らし合わせてどのような報告書を設計し、ディスクローズしていくべきかを探る重要な手掛かりとなるのである。

しかしながら報告書の形態から非財務報告書を分類しようとする、必ずしも報告書の形態とその内容が一致している訳ではないので、単純に分類することは困難である。本稿は、報告書の形態やボリュームなどからではなく、報告書の内容から報告書の類型を導くことを目指し、報告書の内容分析をおこな

ううえで観測される用語とその出現頻度を分析データに用いている。使用される用語とその出現頻度を分析データに用いることにより、記述のようなナラティブ情報を定量的に捉えて分析することが可能である。

2. これまでの本研究の分析経緯と特徴

本研究は、非財務報告書の作成者が意識的に（あるいは無意識に）どのようなステイクホルダーを想定し、どのようなことを重要課題に取り挙げ、そして重要課題はどのように経年的に変化しているのかという関心からはじまった。このような重要課題とその変化を解明することは、非財務報告書の内容、目的、役割や非財務情報開示の在り方を考えていくうえで重要な作業であり、研究分野においても大きな関心事である。

非財務報告書がどのように変遷し重要課題が変化してきたかを探るためには、内容を精査する必要がある。しかし報告書内のほとんどがナラティブ情報であるため分析が難しく、さらに報告書が膨大な数になると人手による分析は限界である。一方、大量のテキスト文書を対象にする情報処理分野の成果⁴から報告書を分析するために、情報処理技術は有効な手段であることと推測できる。

海外の先行研究では、コンピュータサイエンスの進歩により既にあらゆる分野で言語分析技術を活用する研究が多くある⁵。わが国においても言語分析技術 (テキストマイニング) を活用する研究がみられるようになり、記述情報のようなナラティブ情報を定量的に捉えて実証を導く研究がある⁶。しかし、本研究のように、各種の非財務報告書を横断して、特定用語や特定の分析箇所限定せず、報告書全体の用語分布を分析し、類型することを目的とする試みはまだ存在しない。

まず、本研究は統合報告を意識する日本企業34社の報告書（総冊数：135冊程度）に使用されている用語とそれらの出現頻度を観測し、非財務報告書から統合報告を意識した報告書への移行期における変化を分析したことから始まる。さらに非財務報告書（廃止）からアニュアルレポート（継続）への移行期の変化について、出現した用語に属性を与え、経年的動向を観察した。属性の種類には、ステイクホルダー属性、地域属性、テクニカル属性、プロダクト属性、環境負荷インベントリ属性、財務属性、および一般用語・その他属性の7つの属性に分け、特にステイクホルダー属性と財務属性にスポットをあてて分析をおこなっている。報告書で頻繁に使用される用語とその用語の経年的な変化を属性別に観察することで、作成者側の意図や重要課題を明確化することを狙いとしている。

分析手法として、形態素解析 (morphological analysis) を活用して報告書の文章に含まれている用語とその用語の品詞情報を解析する。なお、専門用語辞書および不要語リストを作成し、対象用語に対してフィルタリングをおこなうことで、用語の傾向がより鮮明に浮き彫りになった。結果として、統合報告を意識した報告書は、従来のアニュアルレポートとしての機能や役割を保ちながら、CSR 報告書等に掲載していた非財務情報の一部を組み込んだ報告書として結論付けている⁷。同時に、報告書の内容を把握するには有効であることも分かった。

次に分析対象を1,366組織の1994年から2013年にかけて公表された CSR 報告書等（総冊数：8,629冊）に広げ、全体的に用語の使われ方がどのように変化しているかを分析した。専門用語辞書および不要語リストを精緻化しつつ上記と同様の分析プロセスをおこなうとともに属性ごとに区分し、年度別に並べ替えたリストを作成した。考察して得られた知見は、作成者側がより将来的思考に立ち、リスク管理、戦略や経営ビジョンを強調している点、とりわけ財務的持分関係にあるステイクホルダーにとって情報価値が高い情報である点、環境負荷情報や環境に関わる専門用語は著しく後退している点などの傾向を示したことである⁸。この結果では、報告書の内容の違いや強調する情報の相違などによる報告書の分類が可能であることを示唆している。報告書の

分類方法については、今回の分析の主眼であり、リサーチデザインで詳細を示したい。

前述したように、本研究の特徴の一つは、膨大かつ長期にわたるデータを取り扱い、特定用語や特定箇所の分析に限定しない点である。その理由は、ある一定時点の断片的な現象を確認するのではなく、分析数による客観性を担保しながら継続的に全体動向を捉えるためである。これは本研究が継続かつ経年的な分析により非財務報告の在り方や傾向を予測することを目的としているためである。

そのため本研究は、言語分析技術による記述情報の定量化に加え、より結果情報を鮮明にするために独自の分析プロセスを構築している。先行研究では、研究目的が異なるため、言語分析技術による記述情報の定量化に限定されるが、本研究はこれを出発点に段階を踏んだ分析ステップと複数のデータ解析手法を用いている。さらに、より結果情報を可能な限り精緻化するために、不要語や専門用語や類語などの各種リストを継続して蓄積している点も特徴の一つであることを強調したい。

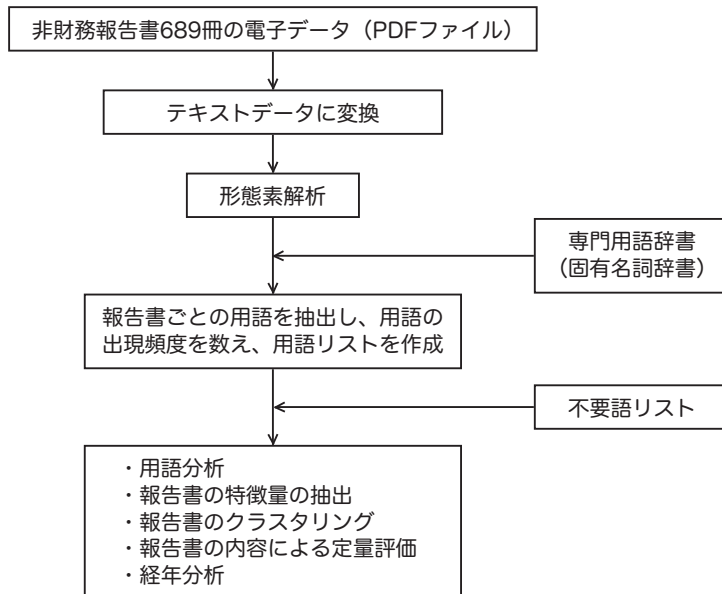
3. リサーチデザイン

今回の分析の対象となるデータは、2015年度に公表された PDF ファイル形式の非財務報告書（フルレポート）689冊であり、すべて営利企業である。図表1は分析手法の概要を示している。まず、対象となる PDF ファイル形式の報告書を分析可能な形式であるテキストデータに変換する。

報告書から用語を抽出するためには、対象となる報告書を形態素解析が可能なテキストデータに変換する必要がある。現在、公開されている報告書は PDF 形式が主流となっている。先行研究のように、フリーソフト Adobe Reader の「高度検索機能」を利用して、特定の用語の出現回数と所在を検索できる。しかし、この方法では人手で操作しなければならないので、大量文書の分析に不向きである。従って、各報告書を文字レベルのテキストデータに変換し、分析可能な状態に加工する必要がある。

次に、変換したテキストデータに対して、自然言語分析の基礎技術である形態素解析を使用して、報告書に使用されている用語と品詞情報を解析する。形態素解析とは、コンピュータで文書中の意味を担

図表1 リサーチデザイン



図表2 形態素解析の例

EOS	環境	マネジメント	を	実践	し	持続	可能	な	社会	の	発展	に	貢献	し	ます	EOS	
	環境	マネジメント	を	実践	し	持続	可能	な	社会	の	発展	に	貢献	し	ます		
	名詞,一般,* * *,* * *,* * *,* * *	名詞,一般,* * *,* * *,* * *,* * *	助詞,格助詞,一般,* * *,* * *,* * *	名詞,サ変接続,* * *,* * *,* * *,* * *	動詞,自立,* * *,* * *,* * *,* * *,* * *	記号,読点,* * *,* * *,* * *,* * *	名詞,サ変接続,* * *,* * *,* * *,* * *	名詞,形容動詞語幹,* * *,* * *,* * *,* * *	助動詞,* * *,* * *,* * *,* * *,* * *	名詞,一般,* * *,* * *,* * *,* * *	助詞,連体化,* * *,* * *,* * *,* * *	名詞,サ変接続,* * *,* * *,* * *,* * *	助詞,格助詞,一般,* * *,* * *,* * *	名詞,サ変接続,* * *,* * *,* * *,* * *	動詞,自立,* * *,* * *,* * *,* * *,* * *	助動詞,* * *,* * *,* * *,* * *,* * *	
	環境,カンキョウ,カンキョー	マネジメント,マネジメント,マネジメント	を,ヲ,ヲ	実践,ジッセン,ジッセン	し,シ,シ	読点,、,、,、,、	持続,ジゾク,ジゾク	可能,カノウ,カノー	特殊・ダ,体言接続,だ,ナ,ナ	社会,シャカイ,シャカイ	の,ノ,ノ	発展,ハッテン,ハッテン	に,ニ,ニ	貢献,コウケン,コーケン	し,シ,シ	ます,マス,マス	

う最小の言語要素（形態素）を自動的に制定する処理である。図表2は形態素解析の結果の一例を示している。

本研究では、フリーソフトのMeCabを使用して形態素解析をおこなっている。形態素解析で得られた用語のうち、最終的な分析目的によってデータマイニング分析に用いる用語は異なる。例えば、今回の分析に置いて図表2で示した結果のように、「名詞、一般」以外の品詞は、分析上あまり意味がない場合があり、文中の動詞、形容詞、副詞、助詞など、

目的と関係ない用語を抽出対象から除外している。また、図表2のような用語の過分割を防ぐために、専門用語辞書と固有名詞辞書を独自に作成して解析結果にフィルタリングをかけている。例えば、専門用語辞書を利用することにより、「環境」と「マネジメント」の二つの用語に分断してカウントすることを防ぎ、「環境マネジメント」という一つの意味を形成する用語を結果に残すことが可能となる。固有名詞も同様である。なお、辞書の作成については、繰り返し形態素解析の結果を目視しながら過分割さ

れた専門用語を追加して精度を高めている。

非財務報告書689冊から抽出される一般名詞は約14万語にのぼるが、明らかに不要な用語に対して、独自に作成した不要語リストでフィルタリングをかけて除外する。最終的に形態素解析により得られた品詞情報に加え、全ての用語の出現回数を集計し、序列したリストを作成する。最終的に変数として残す用語は専門用語とその周辺用語を中心とする956

語とし、真船洋之助編著『環境マネジメントハンドブック』、日本社会関連会計学会編『企業情報ディクショナリー辞典』などのハンドブックや辞典を活用して可能な限り意図性を排除し、分析の透明性と客観性のある結果が得るように努めている。図表3は、956語の上位50の使用されている用語とその出現頻度のリストを示している。

図表3 用語と出現頻度のリスト（フィルター後）

順位	用語	出現頻度	順位	用語	出現頻度
1	環境	77684	26	水	4583
2	社会	32795	27	エコ	4580
3	CSR	20005	28	地球環境	4562
4	廃棄物	15593	29	顧客	4452
5	従業員	15368	30	コスト	4304
6	CO ₂	14928	31	生物多様性	4113
7	リスク	11738	32	収益	4009
8	エネルギー	11481	33	CO	3945
9	コンプライアンス	11077	34	地域社会	3917
10	資源	10936	35	財務	3816
11	災害	8285	36	利益	3728
12	環境負荷	8202	37	資本	3724
13	株主	7756	38	省エネルギー	3681
14	化学物質	6464	39	銀行	3597
15	省エネ	6392	40	環境マネジメント	3476
16	人権	6280	41	投資家	3460
17	戦略	6242	42	地球温暖化	3459
18	マネジメント	5849	43	費用	3455
19	売上高	5692	44	内部統制	3415
20	リサイクル	5659	45	業績	3274
21	原単位	5438	46	CSR 報告書	3261
22	社会貢献	5185	47	森林	3223
23	ステークホルダー	5123	48	経営計画	3151
24	安全衛生	5058	49	温室効果ガス	2975
25	資産	4690	50	リスク管理	2960

出所：わが国企業の非財務報告書689冊（2015年度版）の品詞情報と用語の出現頻度をもとに作成

図表4 企業ごとの使用用語とその出現頻度

	word ₁	word ₂	word ₃	・・・	word _n
企業 a					
企業 b					
企業 c					
・					
・					
・					
企業 n					

図表3の用語と出現頻度のリスト（フィルター後）をもとに、次に図表4のように企業ごとの使用する用語とその出現頻度の表に置き換える。この表は、どの企業がどのような用語をどの程度使用しているかの一覧を示しており、用語一つ一つが報告書の類型を決定する変数となる。図表3の用語と出現頻度のリスト上位50のなかにもみられるように、15位「省エネ」と38位「省エネルギー」などの表記ゆれは、類義語リストを作成し、一つにカウントする調整をおこなっている。その他、結果情報を眺めなが

ら表記ゆれの課題に対して少しずつ対応しているが完全とは言えない現状にある。その理由は、結果情報が膨大であること、そして若干のニュアンスの違いや厳密には異なる意味を持つ用語（例えば売上高と売上）など、一つにカウントすることの判断が難しいケースが多くあるためである⁹。しかし取り扱うデータ量が膨大なため、多少の誤差は、結果に大きく左右されないことを確認している。そのため、判断が困難な類義語は、一つにカウントすることなく留保している。最後に、各用語の出現頻度をその報告書の用語総数で除すことで用語の出現頻度の標準化をおこなう。

4. インプリケーション

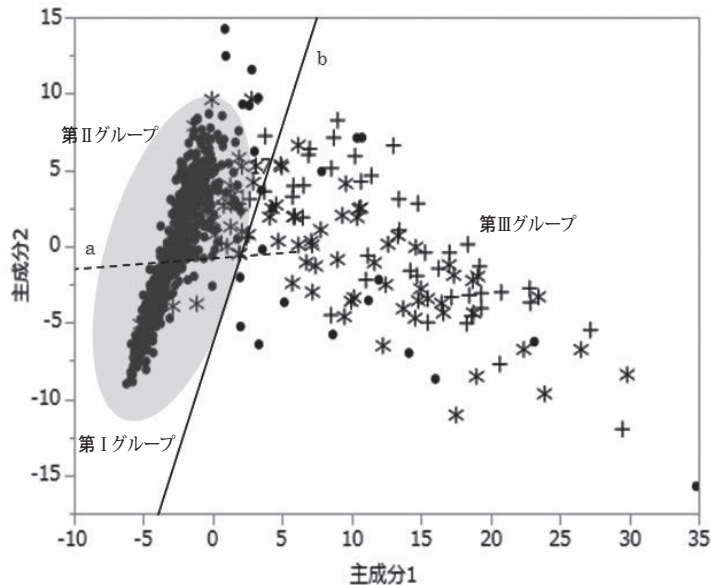
図表5は、報告書内に使用される用語とその出現頻度からその類似性をプロットした表であり、プロットした各点が企業の非財務報告書を表している。「+」はアニュアルレポートと非財務報告書の一つにする広義の統合報告書を展開する企業、「*」はIIRC 統合報告フレームワークを意識する企業、「●」はその他従来型の非財務報告書を公表する企業を示している。

全体を概観すると、網掛けで示している部分に一つの大きなグループが存在し、そのグループから離れていくような分布を示している。会社の規模および非財務報告書の形態から分布の内容を観察すると、概ね3つのグループに分類される。

第Iのグループは、点線aより下、直線bより左側に位置しているグループである。このグループは、内容を確認すると中小企業が多く、環境報告書を中心とした非財務報告書の公表が支配的である。比較的規模の小さい企業では環境報告書が依然として主流であり、ひとつのグループを形成している。これは、環境省が推進する「エコアクション21」が大きく影響している。エコアクション21は、環境マネジメント、環境パフォーマンス評価、および組織外部への環境報告に関する認証登録制度である。エコアクション21ガイドラインでは、環境マネジメントの成果報告の一環として「環境活動レポート」の作成・公表を求めており、環境活動レポートに記載すべき要求事項を示している¹⁰。

IIのグループは、点線aより上、直線bより左側に位置しているグループであり、このグループは、大企業が多く、環境報告書から持続可能性報告書、CSR 報告書など、幅広い形態の報告書の公表が混在

図表5 2015年度に公表された非財務報告書の類型



「+」は広義の統合報告書、「」はIIRC 統合報告書、「●」はその他従来型の非財務報告書を公表する企業
* k-means clustering による。

出所：わが国企業の非財務報告書689冊（2015年度版）のクラスター分析

する。レスポンシブル・ケア報告書、環境・社会報告書、持続可能性報告書、CSR 報告書の内容に大きな差異はなく、比較的近い範囲に集まる。

第Ⅲのグループは、直線bより右側に位置するグループであり、大企業であり、従来型の非財務報告書ではなく統合報告を意識した報告書が大部分を占める。統合報告を意識した報告書は、従来の非財務報告書からかけ離れ、網掛けで示す集団から距離を置く。さらに IIRC 統合報告と広義の統合報告の間に大きな差異は認められない。両者とも、依然として GRI ガイドラインや国内のガイドラインを継承しつつ統合報告に部分的に移行している状況であるため¹¹、極端な差異は内容からは区分できない。

その他、金融、証券、保険業界において独自性が認められ、第Ⅰグループおよび第Ⅱグループからは離れて点在する。

5. 本研究の課題と展望

本研究の課題は、まず各種辞書とリストの改良である。分析結果をより鮮明にするために、かつより精緻な結果が得られるために、専門用語辞書、固有名詞辞書、不要語リスト、属性リスト、ならびに類義語リストをより精度の高いものに継続して改良していく。専門用語辞書については、『経営学辞典』、『会計学辞典』、『企業情報ディスクロージャー辞典』、『環境マネジメントハンドブック』などの専門用語辞典・辞書などを活用して客観性の高い辞書の完成を目指している。固有名詞辞書、不要語リストならびに属性リストについては、用語とその出現頻度のリストを観察しながら継続して更新している。

用語の表記ゆれ問題については認識しているものの、先に述べたように判断が困難なケースが多くある。類義語リストを慎重に整理することにより対応している。変数の選択も含め各種リストの不完全性は、膨大なデータ量を取り扱うため結果にさほど大きな影響を与えていないとはいえ、十分に注視する課題である。

本稿では、2015年度に公表された非財務報告書の類似性を示すプロット図を作成し、組織の規模と報告書の形態を視座に考察するに留まるが、今後は業種、消費者との距離、参考とするガイドラインなどの視点から各グループの共通点の探索を継続して

おこないたい。また、他の年度も同様に作成し、経年的比較を試みることでどのように非財務報告書が展開してきたかを明らかにしたい。

並行して、日本企業を対象にこれまでにおこなってきた同様の分析プロセスを海外先進企業に適用して分析する作業を進めている。言語が異なるため、他言語のための各種リストの作成や技術的な課題も山積するが、日本企業と海外先進企業の比較分析をおこない、変遷、重要課題、類型の違いを明らかにしていきたい。

[注]

- 1 大坪 (2016), p127
- 2 わが国においては、非財務報告書のうち CSR 報告書が最も採用されている。
- 3 統合報告には二つの思考があり、非財務情報の拡充と再編の延長で実践される広義の統合思考と、国際統合報告評議会 (IIRC: International Integrated Reporting Council、以下 IIRC) が提唱する統合思考とある。広義の統合思考は、いわゆる IIRC 統合報告より先行して研究・実践されており、非財務報告書とアニュアルレポートを一本化して移行するケースが多い。また二つの統合思考は、報告目的、想定するステイクホルダーなど細かい点で異なる。詳しくは大坪(2016)を参考にされたい。
- 4 奥村 (2010)
- 5 言語分析技術 (内容分析およびテキストマイニング) を活用する先行研究の動向については、Aureli(2017), p6 を参照。なお、Aureli(2017)では、内容分析およびテキストマイニングの二つの技術を用いて 4 社の比較分析をおこなっている。
- 6 会計領域で言語分析技術を用いる先行研究として、中野他 (2009), 澤登 (2010), 前田他 (2012), 川上他 (2013) などが挙げられる。
- 7 大坪, 黄 (2014a), p81
- 8 大坪, 黄 (2014b), pp21-23
- 9 同じ用語でも時代により意味や捉え方が変化している用語がある。そのため単純に一つの用語としてカウントすることの問題点も指摘されている。

- 10 エコアクション21ガイドライン（2009年版）は、業種ごとにも公表されている。また「エコアクション21大学等高等教育機関向けガイドライン2009年版」では、環境配慮促進法の記載事項等、エコアクションガイドライン要求事項、および環境報告ガイドライン（2007年度版）の3つの環境報告に関するガイドラインの対照表が示されている。
- 11 大坪（2016），pp129-131

- 日本社会関連会計学会編（1991）『企業情報ディスクロージャー辞典』中央経済社
- 真船洋之助編著（2005）『環境マネジメントハンドブック』日本工業新聞社
- Selena Aureli（2017）“A comparison of content analysis usage and text mining in CSR corporate disclosure”, The International Journal of Digital Accounting Research, Vol. 17, pp1-32.
- その他多数の企業の報告書を参考とした。

参考文献

- 大坪史治（2016）「非財務報告の新たな展開—二つの統合思考とわが国企業実践における基礎調査—」『獨協経済』第98号
- 大坪史治，黄海湘（2014a）「統合報告を意識したわが国企業における報告書の現状分析—社会関連会計の視点から—」『SUSTAINABLE MANAGEMENT』（特定非営利活動法人環境経営学会）第13号，1・2合併号
- 大坪史治，黄海湘（2014b）「CSR報告書等の変遷と重要課題の探索—単語に着目した属性分析—」『社会関連会計研究』（日本社会関連会計学会）第26号
- 奥村学（2010）『自然言語処理の基礎』コロナ社
- 川上直哉，中條良美，朴恩芝，前田利之（2013）「テキストマイニングによる環境コスト支出要因の時系列分析」2013年経営情報学会秋季全国研究発表大会
- 環境省（2011）『エコアクション21ガイドライン2009年版』および業種ごとのガイドライン
- 澤登千恵（2010）「テキストマイニングを活用した財務報告実務の展開の考察—London and Birmingham 鉄道会社の財務報告実務を中心に—」『研究紀要』（高松大学），52・53号合併号，pp39-70
- 中野常男，橋本武久，清水泰洋（2009）「わが国における会計史研究の過去と現在：テキストマイニングによる一試論」『国民経済雑誌』200(4)，pp1-23
- 前田利之，朴恩芝，中條良美，村井孝行（2012）「経営者メッセージから読み取る環境投資—テキストマイニングによる分析—」日本社会関連会計学会西日本部会2012年6月2日報告資料