

# 自然言語処理を用いた大手総合建設会社の R&D 活動の分析 経営戦略への活用を目的とした R&D トレンドの観測と可視化

## Analysis of R&D Activities of Major General Construction Companies Using Natural Language Processing Observation and visualization of R&D trends for use in management strategies

○齋藤 寛彰\*<sup>1</sup>  
Hiroaki Saito\*<sup>1</sup>

\*<sup>1</sup> 戸田建設株式会社 イノベーション推進センター 主任 修士 (工学)  
東京工業大学環境・社会理工学院 技術経営専門職学位課程  
Chief, Innovation Center, Toda Corporation, M. Eng.  
Grad. Student, Dept. of Technology and Innovation Management, School of Environment and Society,  
Tokyo Institute of Technology

### Summary:

In this study, it is analyzed that to visualize a company's R&D strategy using text mining. As a case study, this paper analyzes the natural language of securities reports of 14 major general contractors in Japan in 2009 and 2019.

As a result of the case study, it was quantitatively clarified that the R&D activities of general contractors in Japan have changed greatly in 10 years under the influence of not only the market environment but also the macroeconomic and competitive environment. From the viewpoint of R&D strategy, it was suggested that during this period, the strategy have shifted from product innovation to process innovation.

We also found that text mining can summarize large amounts of data and co-occurrence analysis can help us understand trends of R&D activities in construction industry. This paper shows the possibility that these visualization techniques can be utilized in corporate R&D strategy.

キーワード: R&D 戦略, 建設産業, テキストマイニング, 自然言語処理

Keywords: R&D strategy, Construction industry, Text mining, Natural language processing

## 1. はじめに

### 1.1 研究背景

企業は、限られた経営リソースを合理的に分配しながら競争優位性を確保していく必要がある。そのために、企業は他社の活動や業界のトレンドを注視しながら、自社の経営戦略に反映させる場合がある。財務諸表等に現れる構造化された量的データの分析(財務分析など)は、その目的で多くの企業で分析対象として認識されているが、公開情報のうち非構造化データとして入手可能な自然言語で書かれた情報は、従来経営戦略への反映を目的に分析されることは少なかった。しかし、近年自然言語処理が普及する中で計量的に扱うことが容易となり、業界のトレンドを観測する目的でも用いることが可能となりつつある。なかでも、企業の研究開発 (R&D) に関する情報の重要な部分は自然言語で記述されることが多く、自然言語処理を用いて計量的に分析することができれば、

企業は自社の経営戦略に反映させる目的で分析・活用すること容易になると考えられる。

そのような背景から、本研究では、有価証券報告書に記載されている「研究開発活動」(R&D 活動)に関する記述を対象として、R&D 活動の定量的測定、及び業界トレンドの可視化を検討することを目的に研究を行った。

### 1.2 既往の研究

日本の大手建設会社の R&D 活動を対象とした研究には、インタビューを基に日豪間での R&D に関する意識の差に着目した Fraser らの研究<sup>1)</sup>や、入札制度と研究開発投資に着目した Konno らの研究<sup>2)</sup>、製品アーキテクチャの視点で分析した志手らの研究<sup>3)</sup>がある。一方、本研究で扱う自然言語処理を用いた研究は、様々な分野で行われている。特に建築に関連するものとしては、建設プロセスの自動化に関する論文のキーワードを分析対象とした Kim らの研究<sup>4)</sup>がある。また、日本建築学会情

報・システム・利用・技術シンポジウムにおいても福田の研究<sup>5)</sup>をはじめとした複数の研究事例がある。

しかし、日本の総合建設会社の R&D 活動の内容を俯瞰的に捉える目的で自然言語処理を用いた研究は見られず、本研究により新たな視座が得られる可能性がある。

## 2. 分析対象の選定

本研究では、R&D 活動の変化を捉えるために、分析対象として金融商品取引法第 24 条に規定される有価証券報告書を選定した。有価証券報告書は、企業が事業年度ごとに事業内容や経営情報をまとめた報告書であり、同報告書「第 2 事業の状況」には非財務データの 1 つとして研究開発活動に関する記載項目が存在する。企業毎に記載の仕方は若干異なるものの、多くの場合企業における R&D 活動の総括と、個別 R&D 活動の目的や成果等が記載されている。

研究開発活動の成果物としてはニュースリリースや学術論文、特許なども挙げられるが、これらの文書は企業の方針や担当者の裁量により公開の有無やテキストの量が影響されやすい。一方で、有価証券報告書は、各事業年度終了後に金融庁への提出が求められ、財務会計基準機構が公開する作成要領に従って記載されるため一定の型がある。また、企業間の比較検討等が行いやすいように留意すること、簡潔かつ明瞭に記載することが求められるため、分析対象として適していると考えられる。

本研究のケーススタディとして、分析対象企業には総合建設会社のうち活発に R&D 活動を行っていると考えられる 14 社<sup>注1)</sup>を選定した。選定の基準として、2019 年 3 月期で単体の売上が 3,000 億円を超え、かつ有価証券報告書を開示している企業とした。

分析の期間は、2011 年を境に建設投資額が上昇トレンドに変化した<sup>注2)</sup>ことから、2011 年を境に R&D 活動の内容に大きな変化があった可能性がある。そこで、本研究では R&D 活動の定量的測定とトレンドの可視化を検討するために、大きな変化があった可能性がある 2019 年 3 月期とその 10 年前の 2009 年 3 月期のデータを用いて試験的に分析を行った。なお、分析対象のうちの 1 社は 2012 年に合併しているため、2009 年の分析対象は合併前の 2 社を含む 15 社となる。

## 3. データ抽出プロセス

「研究開発活動」に記載されている自然言語データは、非構造化データであるため、分析対象とする建築セグメントに関連する項目を抽出する必要がある。まず、分析対象企業の有価証券報告書の「研究開発活動」に記載されている自然言語を、記載項目ごとに分離する作業を行った。その後、建築に関する記述を抜き出すことを目的に、Figure.1 に記載のフローに沿って、項目 (item) 単位

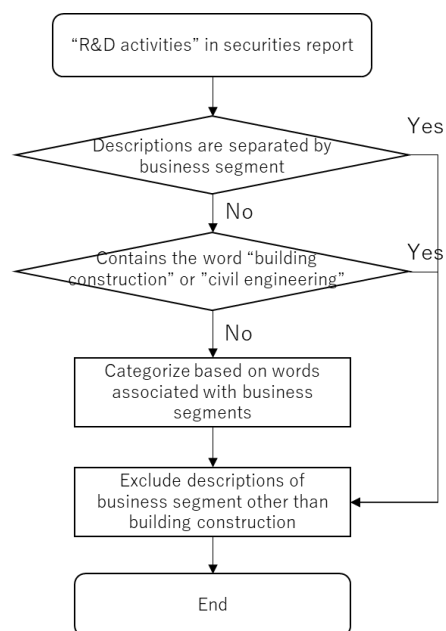


Figure 1. Flow of Data Extraction

Table1 List of compound words and words to forced extraction or exclusion

Compound word	List of forced extraction (48words)	Exclude
研究開発 R&D	RC PCa 5G RFID Apps	連結 Consolidated
技術開発 Technological development	CO <sub>2</sub> RCS 3D WEB Connected	会計 Accounting
生産施設 Production facility	CFT CASBEE JIS Sustainable Stud	当社 Our company
研究機関 Research institution	ICT LCC LEED Heat island Soil Cement	年度 Fiscal year
生産性 Productivity	IT LED EV Smart Solution	
集合住宅 Apartment house	PC CLT MR Precast Heat pump	
	BCP BIM AR Alliance Prestress	
	ALC AI VR Biotope Resilient	
	BCM IoT CAD Bracing	
	LAN ZEB LGS Isolator	

で抽出を行った。事業セグメント毎に分けて記載のある企業はそれによって項目を抽出し、それ以外のケースは項目内の文章に事業セグメントを想起させる単語(「トンネル」「建物」等)が有るか無いかによって筆者が分離した。また、研究開発活動全体に対する記述と、環境や地盤分野など事業横断的に関連する記述は、事業領域を限定する記述がない限りは分析対象とした。なお、グループ会社の建築関連の研究開発活動の項目は分析対象とし、注釈内の記述は分析対象外とした。2019年3月期の場合、抽出前は179項目約55,000字であったのに対し、抽出後は121項目、約36,400字<sup>注3)</sup>となった。抽出前と抽出後の項目数と文字数の差分が、建築以外の事業セグメントに関するものである。

## 4. 頻出語の経時的変化

### 4.1 形態素解析

テキスト分析には、KH Coder3を用いた。分析の前に、まず分析対象とする単語の品詞について絞り込みを行った。本研究では、分析対象とする単語の品詞として、名詞、サ変名詞、形容詞、形容動詞を採用した。形態素解析を行う上で注意しなければならないこととして、同じ

Table2. Frequent Words and rate of change between 2 points

2009				2019				Rate of Change	
Word	FoA	CR	Word	FoA	CR				
1	開発	Development	249	3.5%	開発	Development	187	2.4%	-32%
2	技術	Technology	238	3.4%	技術	Technology	179	2.3%	-32%
3	システム	System	133	1.9%	システム	System	101	1.3%	-31%
4	工法	Construction method	99	1.4%	施工	Construction	82	1.0%	0%
5	研究開発	R&D	92	1.3%	工法	Construction method	76	1.0%	-30%
6	施工	Construction	74	1.0%	構造	Structure	76	1.0%	23%
7	環境	Environment	69	1.0%	建築	Architecture	69	0.9%	52%
8	適用	Apply	60	0.8%	研究開発	R&D	66	0.8%	-35%
9	事業	Business	58	0.8%	コンクリート	Concrete	61	0.8%	0%
10	構造	Structure	56	0.8%	適用	Apply	58	0.7%	-12%
11	コンクリート	Concrete	55	0.8%	建設	Construction	57	0.7%	32%
12	共同	Collaboration	54	0.7%	性能	Performance	57	0.7%	52%
13	建物	Building	49	0.7%	環境	Environment	56	0.7%	-26%
14	可能	Possible	48	0.7%	可能	Possible	55	0.7%	4%
15	高層	High rise	48	0.7%	活用	Utilization	54	0.7%	75%
16	設計	Design	48	0.7%	建物	Building	53	0.7%	-2%
17	施設	Facilities	46	0.6%	設計	Design	50	0.6%	-6%
18	地震	Earthquake	44	0.6%	事業	Business	49	0.6%	-23%
19	実用	Practical application	43	0.6%	共同	Collaboration	48	0.6%	-19%
20	建築	Architecture	41	0.6%	向上	Improvement	48	0.6%	28%
21	活動	Activities	39	0.6%	実現	Realization	47	0.6%	77%
22	建設	Construction	39	0.6%	工事	Construction	44	0.6%	90%
23	評価	Evaluation	39	0.6%	研究	Research	42	0.5%	6%
24	対策	Countermeasure	38	0.5%	耐火	Fire resistance	42	0.5%	1168%
25	基礎	Foundation	36	0.5%	管理	Management	39	0.5%	152%
26	研究	Research	36	0.5%	現場	Site	33	0.4%	130%
27	向上	Improvement	34	0.5%	推進	Promotion	33	0.4%	-4%
28	性能	Performance	34	0.5%	基礎	Foundation	32	0.4%	-19%
29	関連	Related	32	0.5%	取得	Acquisition	32	0.4%	-6%
30	取得	Acquisition	31	0.4%	品質	Quality	32	0.4%	26%
31	集合住宅	Apartment house	31	0.4%	ロボット	Robot	30	0.4%	806%
32	推進	Promotion	31	0.4%	実証	Demonstration	30	0.4%	2618%
33	利用	Utilization	31	0.4%	CLT	CLT	29	0.4%	new
34	耐震	Earthquake resistance	30	0.4%	活動	Activities	29	0.4%	-33%
35	低減	Reduction	30	0.4%	技術開発	Technology development	29	0.4%	14%
36	物件	Property	30	0.4%	施設	Facilities	29	0.4%	-43%
37	RC	RC	28	0.4%	関連	Related	28	0.4%	-21%
38	活用	Utilization	28	0.4%	作業	Work	28	0.4%	69%
39	対応	Response	26	0.4%	自動	Auto	28	0.4%	182%
40	効率	Efficiency	25	0.4%	利用	Utilization	28	0.4%	-18%
41	効果	Effect	24	0.3%	地震	Earthquake	27	0.3%	-44%
42	実現	Realization	24	0.3%	評価	Evaluation	27	0.3%	-37%
43	実施	Implementation	24	0.3%	安全	Safety	26	0.3%	39%
44	積極	Positive	24	0.3%	社会	Society	26	0.3%	24%
45	提案	Proposal	24	0.3%	展開	Deployment	26	0.3%	2%
46	技術開発	Technology development	23	0.3%	高い	High	25	0.3%	13%
47	展開	Deployment	23	0.3%	遮音	Sound Insulation	25	0.3%	224%
48	土壌	Soil	23	0.3%	生産	Production	25	0.3%	183%
49	品質	Quality	23	0.3%	積極	Positive	25	0.3%	-6%
50	エネルギー	Energy	22	0.3%	対策	Countermeasure	25	0.3%	-40%
51	処理	Treatment	22	0.3%	木造	Wooden	25	0.3%	1032%
52	補強	Reinforcement	22	0.3%	省エネ	Energy saving	25	0.3%	51%
53	ビル	Building	21	0.3%	効率	Efficiency	24	0.3%	-13%
54	強度	Strength	21	0.3%	材料	Materials	24	0.3%	45%
55	工事	Construction	21	0.3%	対応	Response	24	0.3%	45%
56	対象	Target	21	0.3%	高層	High rise	23	0.3%	-57%
57	ニーズ	Needs	20	0.3%	実験	Experiment	23	0.3%	39%
58	空間	Space	20	0.3%	低減	Reduction	23	0.3%	-31%
59	計画	Plan	20	0.3%	分野	Field	23	0.3%	30%
60	高い	High	20	0.3%	確保	Secure	22	0.3%	66%

※FoA: Frequency of appearance  
※CR: Composition Ratio

語でも異なる文脈で用いられる場合がある点が挙げられる。例えば、本研究の場合「生産」という語は「生産性」という文脈で用いられる場合と「生産施設」で用いられる場合があり、研究の目的上混同しない方が望ましい語については予め配慮しておく必要がある。そこで、本研究の目的を考慮した上で複合語として扱う6語を設定した。また、未知語として分類された語のうち、「CLT<sup>注4)</sup>」「ICT」「IoT」「BIM」等48語を分析対象として加える処理、及び頻出語の中から本研究の目的上不要となる語の除外処理を行った (Table1)。なお、除外対象の4語は有価証券報告書のなかで定型的な記述で使われる語である。

以上の前処理を行った上で、形態素解析を実行した。

#### 4.2 頻出語の比較

形態素解析の結果として、2009年3月期と2019年3月期における頻出語の上位60語とその頻度(Frequency of appearance)、構成比(Composition Ratio)、構成比の変化率(Ratio of change)をまとめたものをTable 2に示す。なお、構成比は、抽出語の総数を分母にした比率で定義し、変化率は2019年3月期と2009年3月期の構成比の差を2009年3月期の構成比で除した値とした。

まず、単純に2009年3月期と2019年3月期の頻出語を比較することで、各年のR&D活動のテーマのトレンドの把握を行う。

30位までの語の2009年3月期と2019年3月期の変化量を見ると、「環境」「事業」「研究開発」等が順位を下げている一方で、「構造」「建築」「建設」「性能」「工事」「管理」「現場」等の語の出現頻度が増えていることがわかつ

た。また、全体では、「耐火」「ロボット」「実証」「CLT」「自動」「遮音」「木造」等の語は2009年3月期には圏外であったが、2019年3月期には出現するようになり、上位60語までの特徴的変化となっている。一方で、2009年3月期には48回使用されていた「高層」が2019年3月期には23回に減少していること等が確認された。

#### 5. R&D活動の変化の可視化

形態素解析の結果は、大量のデータを扱う性質上、情報として直観的に理解することが難しい。そこで、解析結果を直観的に表現する方法として縦軸に構成比の変化率、横軸に直近年の構成比をとる散布図による可視化を検討した。上側にプロットされる語は、比較対象年と比べて構成比を増加させている語、つまり上昇トレンドに関する語と判別でき、下側は同様に下降トレンドにある語と言える。また、右側にプロットされる語は近年頻度が相対的に高い語、つまり重要度が高い語と判別でき、左側は重要性が相対的に低いと解釈できる。

本研究で分析対象とした2009年3月期と2019年3月期に当てはめると、横軸は2019年3月期の語の構成比、縦軸は2009年3月期と2019年3月期の構成比の変化率が相当し、プロットするとFigure.3のようになる。紙面の都合上、横軸1.2%以上、縦軸250%以上の語はFigure.3内の別図で示す。右上にプロットされた語ほど2019年3月期に頻出し、かつ2009年3月期よりも構成比を伸ばしている語と見ることができる。なお2009年3月期には出現しなかったが2019年3月期には出現した語は図の中に

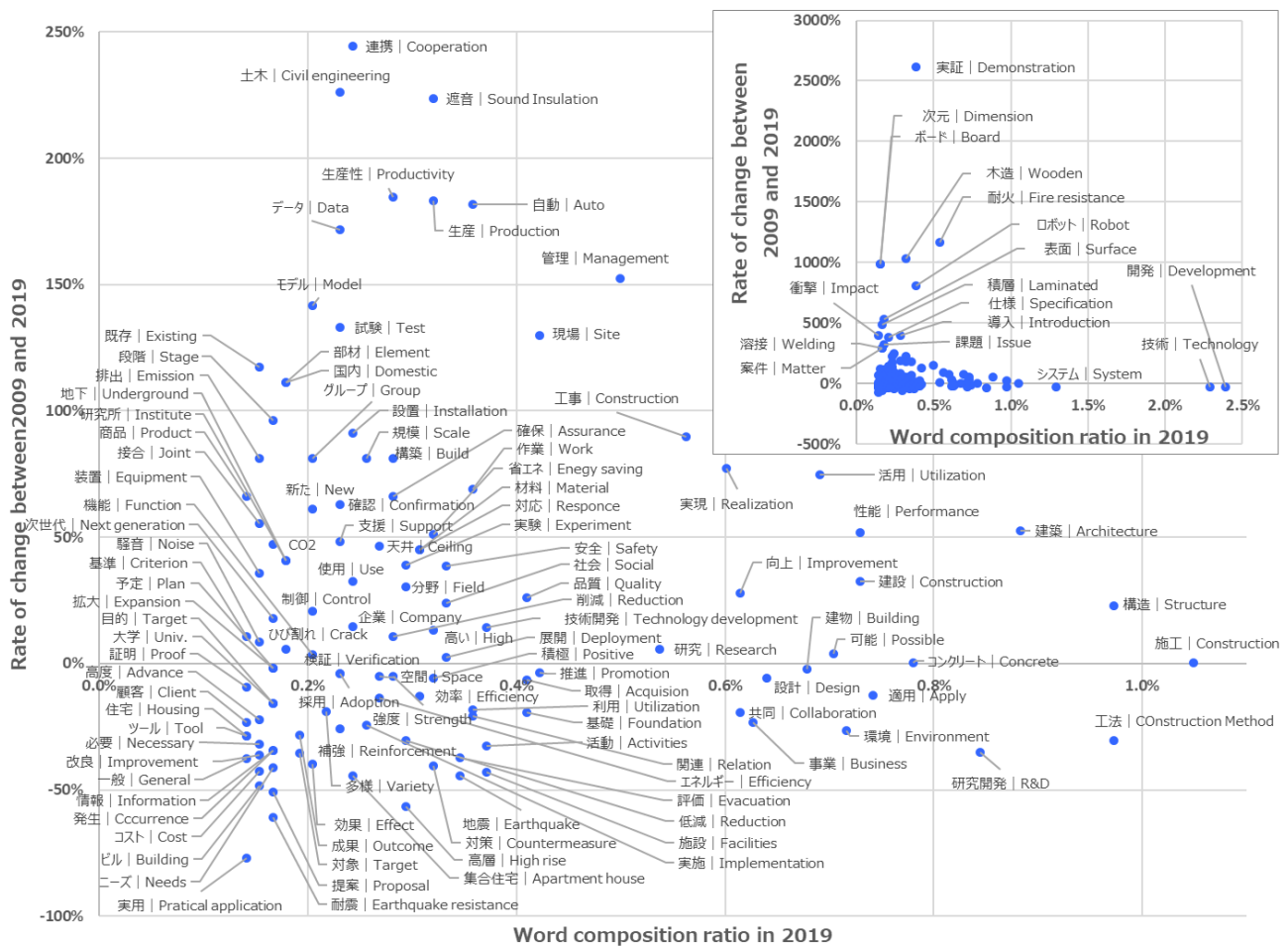


Figure 2. Relationship between word composition ratio in FY ended Mar. 2019 and rate of change between 2 points

プロットされないため、Table3に別途示す。

まず Figure.2 に挙げられた語のうち、構成比の正の変化率が大きかった語及びTable3に挙げられた語の特徴を見ると、木造に関する語（「耐火<sup>注5)</sup>」「木造」「CLT」）、建築プロセスに関する語（「ロボット」「溶接」「PC」「計測」「検査」「生産」「工事」「管理」「現場」等）、情報処理・先端技術に関する語（「BIM」「AI」「ICT」「IoT」「AR」「データ」「モデル」「自動」等）が多く見られた。以上より、2019年3月期は2009年3月期と比べて、建築プロセスに関する R&D 活動が急速に進んだと同時に、情報処理・先端技術が発展したことで、建築プロセスにおける適用が活発に模索されたと考えることができる。また、2010年に国産木材の使用促進を目的とした法律等<sup>注6)</sup>が公布・施行された影響により、総合建設会社の R&D 対象として木造関係の技術が注目されはじめ、2019年頃に R&D の成果として現れ始めたと考えられる。

一方、下側にプロットされた集合住宅に関する語（「高層<sup>注7)</sup>」「集合住宅」）や「環境」は、これらが2009年3月期には上位で出現していることから R&D のトレンドであったことが伺えるが、2019年3月期には大きく減少していることが確認できた。また、「提案」は2009年3月期には24回使用されていたのに対して、2019年3月

Table3. List of words appeared only in FY ended Mar. 2019

	New word	FoA	CR
CLT	CLT	29	0.4%
BIM	BIM (Building Information Modeling)	19	0.2%
操作	Control	17	0.2%
AI	AI (Artificial intelligence)	16	0.2%
PC	PC (Precast Concrete)	16	0.2%
ZEB	Net zero energy building	16	0.2%
計測	Measure	15	0.2%
ICT	Information and Communication Technology	14	0.2%
ラボ	Laboratory	14	0.2%
遠隔	Remote	14	0.2%
IoT	IoT (Internet of Things)	13	0.2%
運用	Operation	13	0.2%
検査	Inspection	13	0.2%
造形	Modeling	13	0.2%
AR	AR (Augmented Reality)	11	0.1%

期には13回に減少している他、大きな変化ではないものの、関連する語として「顧客」「ニーズ」「コスト」も減少傾向にある。これらから、2009年から2019年にかけて R&D 活動は顧客企業のニーズ志向が相対的に低下した可能性がある。

以上を総合的に考察すると、大手総合建設会社の R&D 活動は、顧客ニーズを満たすためのプロダクトイノベーションから、より効率的に生産していくためのプロセスイノベーションにその戦略がシフトした可能性が示唆される。その背景には、2011年に建設投資が増加に転じた



「CLT—木造」が強い共起関係として抽出された。同時に木造の弱点である「遮音」「耐火」「規模」が、R&Dの課題として挙げられたと考えられる。

さらに、2009年3月期に挙げられていたRC、耐震等の建築に関する専門用語は、2019年3月期では木造関係の語を除き減少している他、2009年3月期にはなかった「生産性」が出現したことも特徴として確認できた。

## 6.2 関係性に経時的変化が見られなかった語

関係性に経時的変化が見られる語が多く挙げられた一方で、2009年3月期と2019年3月期で共起関係に目立った変化が見られないものもあった。

例えば、最も多く出現する語で構成される「開発—技術—システム—共同」の群は、2009年3月期2019年3月期で大きな変化は見られなかった。他にも、2009年3月期で比較的強い共起関係が見られた「コンクリート—強度」は、2019年3月期にも依然として高い共起関係を確認できた。これは、コンクリート強度が建築領域のR&Dにとって比較的普遍的なテーマである可能性を示唆するものとして考えられる。

一方で、2009年3月期には「コンクリート」と共起されていた「取得」が、2019年3月期では共起されなくなっていることも分かった。「取得」は、2009年3月期2019年3月期に共通して、建築基準法に基づく国土交通大臣認定を取得するための性能評価に関する文脈で使われていることが多い。2009年3月期は、コンクリート杭やコンクリート材料の性能評価に関する性能評価に関する記述が多かったのに対して2019年3月期はその記述が減少していることが共起ネットワークの変化に影響を与えていることが分かった。

## 7. 結語

本研究では、大手総合建設会社のR&D活動を対象として、有価証券報告書の自然言語データを用いたR&D活動の定量的測定、及びトレンドの可視化を検討した。

また、ケーススタディとして日本の建設業界で外部環境に変化のあった2009年と2019年3月期の有価証券報告書のテキストマイニングにより、分析と考察を試みた。

分析の結果、歴史が長く成熟産業に数えられる建設産業において、一見大きく変化していないように見える総合建設会社のR&D活動が、マクロ環境、市場環境、競争環境の影響を受けながらその活動内容を変化させていることを定量的に観測することができた。また、R&D戦略の視点では、この期間にプロダクトイノベーションを重視する戦略からプロセスイノベーションを重視する戦略に変化した可能性が示唆された。

今回実施したケーススタディから多くの知識を獲得できたことから、本手法を企業の経営戦略やR&D戦略へ活用していくことも可能であると考えられる。一方で、

本手法の限界としては、有価証券報告書に現れる語の多くはR&Dの成果としてアウトプットされたものである性質上、R&Dに着手した時期とは数年のギャップが生じてしまう点である。これは、学術論文等の他の資料を分析対象としても同様の制約があり、企業活動を外部から観測することの共通の限界と言える。経営戦略・R&D戦略への活用を目的とする場合は、このギャップを認識した上で活用することが求められる。

最後に、2020年はCOVID-19の感染拡大が世界的危機を引き起こし、社会・経済活動に対して多面的な影響を与えている。日本の大手総合建設会社にも経営戦略及びR&D戦略に対して新たなパラダイムシフトをもたらす可能性があり、今後も注目していく必要がある。

## 【注】

- 注1) 鹿島建設、清水建設、大成建設、大林組、竹中工務店、長谷工コーポレーション、五洋建設、戸田建設、前田建設工業、三井住友建設、西松建設、安藤・間、東急建設、熊谷組の14社。前年12月期決算の企業を含む。
- 注2) 例えば、国土交通省総合政策局情報政策課建設経済統計調査室が公開している建設投資見通しに関する資料に、過去の推移が記載されている。
- 注3) 文字数は誤差を考慮し10の位を四捨五入した。
- 注4) CLT (Cross Laminated Timber) は、ひき板を繊維方向が直行するように積層接着した木質系材料のことを言う。
- 注5) 「耐火」は木造に関連する固有の語ではないものの、後の分析 (Figure.3) において木造と高い共起関係であることから、ここでは木造に関連する語として述べた。
- 注6) 2010年に公布・施行された国土保全及び国産材の需要喚起を目的とした「公共建築物における木材の利用促進に関する法律」、2013年のJAS規格化、及び2016年に公布・施行されたCLTに関連する建築基準法告示等がある。
- 注7) 「高層」は集合住宅に関連する固有の語ではないものの、注5)と同様の理由から、ここでは集合住宅に関連する語として述べた。

## 【参考文献】

- 1) Fraser, C. and Zarkada, A.: The philosophy, structure and objectives of research and development in Japan, *Construction Management and Economics*, Vol.19, pp.831-840, 2001.
- 2) Konno, Y. and Itoh, Y.: Empirical analysis of R&D in the Japanese construction industry based on the structure conduct performance model, *Cogent Business & Management*, Vol.5, 2018.1
- 3) 志手一哉, 安藤正雄: 総合建設業の研究開発活動における対顧客ポジショニングの変化 -集合住宅と半導体工場と比較-, *日本建築学会計画系論文集*, 第668, pp.1929-1935, 2011.10
- 4) Kim, T., Lee, D., Lim, H., Lee, U., Cho, H. and Cho, K., Exploring research trends and network characteristics in construction automation and robotics based on keyword network analysis, *Journal of Asian Architecture and Building Engineering*, 2020.8.
- 5) 福田知弘: テキストマイニングを用いた建築情報分野の論文タイトル分析 -日本建築学会 情報・システム・利用・技術シンポジウム論文集を対象として-, *日本建築学会・情報システム技術委員会 第41回情報・システム・利用・技術シンポジウム論文集 (報告)*, pp.150-153, 2018.