

有価証券報告書における研究開発活動の記載内容の分析

石 光 裕

要 旨

本稿は、企業の研究開発活動について、有価証券報告書の「研究開発活動」の文章からその内容特性を捕捉できないかを検討したものである。具体的には、記載の総単語数、総単語数に占める異なる単語の割合を情報内容の多様性の代理変数とした場合、後者の方が優れている可能性が高いことが示された。

1. はじめに

本稿は、企業の研究開発活動について、有価証券報告書の文章からその特性を捕捉できないかを検討したものである。

企業の研究開発活動の内容が、企業価値を見積もるうえで大きな位置を占めていることは従前より広く知られている。1999年には、企業会計審議会は、研究開発活動の様子が企業の将来予測において重要であると位置づけ、研究開発投資の総額を開示することを要求する会計基準を公表した。これにより研究開発投資の額は発生時の費用として処理することとされた。また研究開発に関する情報の公開は財務諸表にとどまらず、企業のホームページや記者会見の場などを通しても行われている。

これまで研究開発を対象とした会計研究では、研究開発投資の額をもとに、その特性を探ろうとするものが多かった。例えば Lev and Sougiannis (1996) は、本来は費用処理されている研究開発投資の額を資産計上し、それが5年で償却されたと仮定したうえで、将来利益の額との関連性を検証し、両者にプラスの関係があることを見出した。また Kothari et al. (2002) は、研究開発支出が将来利益の変動に与える影響が、資本的支出のそれよりも3倍大きいことを示した。これ以外にも研究開発費の額と将来業績との関連性を検証した研究は数多く存在する。

有価証券報告書の「第2 事業の状況」には、研究開発活動について述べた箇所があり、文章によって企業の研究開発活動の様子が記載されている。近年の情報開示の拡大の流れと研究開発活動の重要性の認識が高まるにつれて、当該箇所の記載量も増える傾向にあり、将来業績を予測するうえで有用な情報となっていると考えられるが、これまであまり検証の対象とはされてこなかった。文字(テキスト)情報の分析があまり行われてこなかった理由として、技術的な制約が大きかったと考えられるが、データの整備や解析プログラムが充実してきたこともあって、近年公表される論文にテキスト情報を用いた検証が一定数含まれるようになり、レビュー論文も相次いで出されている。

また、近年の証券取引は、コンピューターが事前にプログラミングされた投資戦略にもとづいて

売買を行うアルゴリズム取引や極めて短い時間の間に繰り返し売買を行う高頻度取引が主流となってきた。これら取引に用いられるプログラミングの内容は企業秘密とされるが、企業が公表する財務数値だけではなくネット上のニュースや経営者の発言などのテキスト情報も取り込んでいるとされ、証券投資の現場においてもテキスト情報に注目が集まっている。ただし現在の技術では、文章の内容を判断する技術は未だ完全とは言えず、プログラムでテキスト内容をどのように解釈するのかについては、研究が進んでいる状況である。

本稿ではこのような現状のもと、有価証券報告書のテキストの記載内容を検討することによって、どのような内容が述べられているのか、そして記載内容の特性を捕捉するためにはどのような方法があるのかを研究開発活動の記載箇所をもとに考察している。

本稿の構成は以下のとおりである。第2節では、有価証券報告書の一般的な記載内容とそのなかで研究開発活動がどのように記述されているのかを確認する。第3節では、記載内容の特性を捉える方法として単語数、異なり語数の比率に着目し、それらの推移や頻出単語ランキングを検討する。第4節では、3社×13年の有価証券報告書を対象に記載内容を詳細に検討し、それが先の特性値にどのように反映されているのかを考察する。第5節ではまとめと今後の課題を示している。

2. 有価証券報告書の記載内容と研究開発活動

(1) 有価証券報告書の記載内容

有価証券報告書とは、有価証券を発行する企業が事業年度ごとに、当該会社の商号、当該会社の属する企業集団及び当該会社の経理の状況その他事業の内容に関する重要な事項その他の公益又は投資者保護のため必要かつ適当なものとして内閣府令で定める事項を記載した報告書を指す（金融商品取引法第24条）。有価証券報告書は当該事業年度の経過後3ヶ月以内に提出する必要がある、提出された報告書は金融庁が運営するホームページ（Electronic Disclosure for Investors' NETwork: EDINET）で閲覧することができる。

その具体的な記載内容については「企業内容等の開示に関する内閣府令」において示されており、おおよそ図表1のような項目に分けて記載される。まず有価証券報告書は、「第一部 企業情報」、第三者の保証によって社債を発行している会社が保障会社の情報について述べた「第二部 提出会社の保証会社等の情報」の部分に分けることができ、「第一部 企業情報」がその内容の大部分を占める。さらに企業情報は、「第1 企業の概況」、「第2 事業の状況」、「第3 設備の状況」、「第4 提出会社の状況」、「第5 経理の状況」、「第6 提出会社の株式事務の概要」、「第7 提出会社の参考情報」に分けて記載がなされる。企業情報のなかでも中心となるのが、損益計算書、貸借対照表、株主資本等変動計算書といった財務諸表が収容されている「第5 経理の状況」であり、主に定量的な情報としての記載が多いという特徴がある。一方、「第1 企業の概況」、「第2 事業の状況」、「第3 設備の状況」、「第4 提出会社の状況」などは比較的多くの文章を含んでいる。

図表 1 有価証券報告書の記載内容

| | | |
|-------------------|-----------------|--|
| 第一部 企業情報 | 第1 企業の概況 | ①主要な経営指標等の推移, ②沿革, ③事業の内容, ④関係会社の状況, ⑤従業員の状況 |
| | 第2 事業の状況 | ①業績等の概要, ②生産, 受注及び販売の状況, ③経営方針, 経営環境及び対処すべき課題等, ④事業等のリスク, ⑤経営上の重要な契約等, ⑥研究開発活動, ⑦財政状態, 経営成績及びキャッシュ・フローの状況の分析 |
| | 第3 設備の状況 | ①設備投資等の概要, ②主要な設備の状況, ③設備の新設, 除却等の計画 |
| | 第4 提出会社の状況 | ①株式等の状況, ②自己株式の取得等の状況, ③配当政策, ④株価の推移, ⑤役員の状況, ⑥コーポレート・ガバナンスの状況等 |
| | 第5 経理の状況 | ①連結財務諸表等, ②財務諸表等 |
| | 第6 提出会社の株式事務の概要 | 事業年度, 定時株主総会, 基準日, 株券の種類などの項目 |
| | 第7 提出会社の参考情報 | ①提出会社の親会社等の情報, ②その他の参考情報 |
| 第二部 提出会社の保証会社等の情報 | 第1 保証会社情報 | ①保証の対象となっている社債, ②継続開示会社たる保証会社に関する事項, ③継続開示会社に該当しない保証会社に関する事項 |
| | 第2 保証会社以外の会社の情報 | ①当該会社の情報の開示を必要とする理由, ②継続開示会社たる当該会社に関する事項, ③継続開示会社に該当しない当該会社に関する事項 |
| | 第3 指数等の情報 | ①当該指数等の情報の開示を必要とする理由, ②当該指数等の推移 |

図表1からも明らかなように、有価証券報告書の記載内容には一定のフォーマットがあり¹⁾、企業はそれに従い有価証券報告書を作成している。そのため、有価証券報告書の記載内容は定量、定性といった情報の種類に関わらず、企業間の比較可能性が高いと言える。特に定性的な情報は、形式や記載方法に一定の制約がなければ記載内容に多くのバラつきがみられ、情報間での比較が難しくなると考えられる。そのため有価証券報告書の統一された定性的な情報は投資家にとって貴重であるといえる。

これまで、研究開発に関連する研究では、研究開発活動の文章中の金額部分や「第5 経理の状況」の財務数値を主たる分析対象としていたのに対して、文章についての分析はほとんど行われてこなかった。そこで、本稿では文章部分の分析を行うことを前提に、文章内容の特性を取り扱う方法を検討する。

(2) 研究開発活動の記載内容

研究開発活動の記載上の注意点として、「企業内容等の開示に関する内閣府令」の「第二号様式記

1) 財務会計基準機構は、決算年度ごとに、その時点で利用可能な金融商品取引法、開示布令、各種のガイドラインにもとづいた「有価証券報告書の作成要領」を公開している。

載上の注意(35)」では、以下のように述べられている。

最近連結会計年度等(連結財務諸表を作成していない場合には最近事業年度等)における研究開発活動の状況(例えば、研究の目的、主要課題、研究成果、研究体制等)及び研究開発費の金額を、セグメント情報に関連付けて記載すること。

ここでは、研究開発活動の状況を示す具体的な項目として、研究の目的、主要課題、研究成果、研究体制が例示されている。またこれら情報とセグメント情報とを関連させて述べることも要求されている。

では実際の記述は、どのようになっているのだろうか。武田薬品工業の2016年3月決算の有価証券報告書の「第5 経理の状況」研究開発活動の内容をもとに検討してみよう。分量は5ページほどであり、研究開発の様子が非常に詳細に記載されている。まず、冒頭に全体の研究開発費が3,459億円であること、うち医療用医薬品事業は3,382億円、コンシューマーヘルスケア事業は16億円であり、研究開発費のほとんどが医療用医薬品事業に向けられていることが述べられている。そのため、コンシューマーヘルスケア事業に関連する記述は非常に少ない。

医療用医薬品事業については、研究開発活動の内容が領域(オンコロジー、消化器系疾患、中枢神経系疾患、ワクチン、その他)ごとに時系列で示されている。例えば、ワクチンの箇所では、組織体制、製品(季節性インフルエンザワクチン、ヴァクセムヒブ、ノロウイルスワクチン)、パートナーシップ・事業開発活動の項目ごとに時系列でその内容を挙げるという構成になっている。組織体制の箇所では、ワクチン事業の成長と開発を促進するために、世界各地の拠点を統廃合することが書かれている。次に製品については、研究開発に必要な細胞培養技術の買収や取り扱うワクチンが厚生労働省より製造販売承認を取得したこと、研究開発が現在どのような段階にあるのかについての記載がある。そしてパートナーシップ・事業開発活動については、ポリオワクチンの開発のために、ある財団から資金助成を受ける事業提携を結んだことが述べられている。

武田薬品工業の記載内容は、先に挙げた「記載上の注意」の内容と整合的である。まず、全体として事業セグメント(医療用医薬品事業、コンシューマーヘルスケア)ごとに記載がなされている。またセグメントごとの記載内容は、主力となる製品別に、どのような研究がなされ研究成果が挙げられているかが述べられている。さらに、研究を進めるための組織体制についても触れている。

3. 研究開発活動の記載内容の特性

前節では、研究開発活動の箇所の具体的な記載内容を、武田薬品工業を例に確認した。本節では、東京証券取引所に上場する3月決算企業を対象を拡大し、記載内容の特性を計量的な指標を用いて検討する。計量的な指標を用いる意義は、多くの企業サンプルを対象に、客観性を担保したうえで、

時系列とクロスセクションでの比較を行うことができる点にある。なお、本稿に用いるテキスト情報は、プロネクサス社の提供する企業情報データベースサービス eol から取得した。テキスト形式によるデータ取得が可能であった2004年3月期から2016年3月期までを対象とし、期間を通してデータが入手可能な企業を対象とした。

業種によって研究開発投資の占める重要性は異なり、多くの企業が研究開発費を計上している業種から、ほとんどの企業の研究開発費の額がゼロである業種まで様々である。本稿では、研究開発投資が活発な業種に焦点をあてて、研究開発活動の箇所の記載の特性を検討していく。対象となる業種は、2016年3月期において、研究開発集約度（研究開発費 / 売上高）の業種平均が2%を超えた化学、医薬品、ゴム、窯業、機械、電気機器、自動車、精密機器の8つである²⁾。

またテキスト分析には、統計分析パッケージである R 上で、文章を単語レベルに分解する手続き³⁾を行う MeCab を動作させることができる RMeCab パッケージを用いている⁴⁾。

(1) 記載単語数

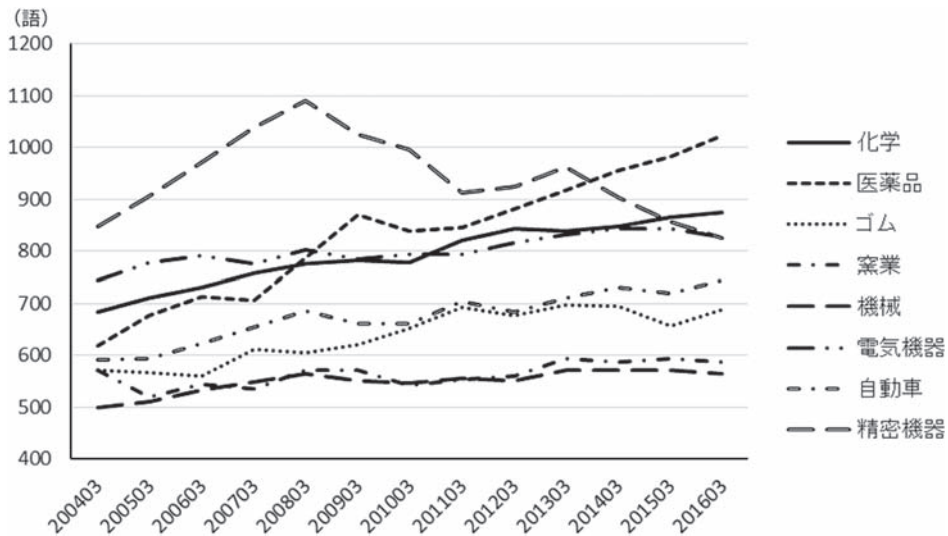
有価証券報告書における研究開発活動の箇所の情報量は、企業間でどれほど異なっているのか、また時系列でどのように変化しているのだろうか。そのことを観察するための最も単純な指標が記載されている単語数である。単語数が多ければ多いほど、その文章はたくさんの情報を含んでいると考えられる。図表2は、「研究開発活動」の箇所に記載されている総単語数の時系列変化を8つの業種別に示したものである。

図表から、2004年3月期から2016年3月期にかけて、精密機器を除いて、概ね記載されている単語数が増加している傾向を観察することができる。これは近年の情報開示拡大の要請に依るものであると考えられる。特に大きな伸びを示しているのが医薬品であり、2004年3月期では600語程度であったものが、2016年3月期では1,000語ほどに増加している。変動の大きな精密機器は、2008年3月にピークである1,100語程度に到達したのち減少し、2016年では電気機器と同程度の水準となっている。2016年においては、600語前後の2業種（機械、窯業）、700語前後の2業種（ゴム、自動車）、850語前後の3業種（電気機器、精密機器、化学）と1,000語程度の医薬品という分布になっている。

2) サンプルに含まれる業種ごとの企業数は、化学136社、医薬品31社、ゴム10社、窯業38社、機械160社、電気機器182社、自動車65社、精密機器30社である。

3) この手続きは形態素解析と呼ばれるものであり、最小の意味の単位である形態素に文章を分割し、品詞情報を付け加える作業を指す。単語の定義が論者によって異なるため、単語と形態素は完全に同一のものとはいえない。ただし分析にあたってはほぼ同様のものと捉えられていることが多く、本稿では単語という用語に統一している。

4) Rのバージョンは3.2.3 (2015-12-10)、RMeCabのバージョンは0.99993 for R-3.2.3を用いている。RMeCabで文章を形態素に分割するために用いる辞書は、メインとなるシステム辞書に naist-jdic for mecab (naist-jdic-0.6.3b-20111013) を選択し、ユーザー辞書には wikipedia の項目とはてな日記の word list with hurigana を追加した。ユーザー辞書の追加により、人名、企業名など固有名詞や専門分野の用語に関する語彙が拡張されている。



図表2 業種別の総単語数の推移

(2) 記載内容の多様性

単語数は記載されている情報内容の多さを測る目安の1つと考えられるが、弱点がある。それは同じ文章が繰り返されたときであり、このとき単語数は増えるが、そこに記載されている情報内容は変わらない。本稿では、この弱点を克服するため、文書中に異なった単語がどれだけ出現しているかを測定した TTR (types token ratio) という指標に着目している。TTR 値は、単語総数の中でどれだけ異なる名詞が用いられているのかの割合を表すものであり、以下のように計算される。

$$TTR = \frac{\text{Types}}{\text{Tokens}}$$

Types は、出現した単語の種類 (異なり語数) であり、*Tokens* は、出現した総単語数 (延べ語数) である。TTR 値の解釈は直感的である。例えば、ある年度の「研究開発活動」の項目に出現する単語の総数が 10 であったとする。それが全て異なる単語であれば、異なり語 (*types*) は 10、TTR の値は 1 となり、これは文章の多様性が高い (内容が豊かである) ことを示していると考えられる。逆に全て同一の単語が用いられていたとすると、TTR の値は 1/10 になり、文章の多様性が低いことを意味する。ただし、上記の計算のままだと、文章が長くなるにつれ TTR 値が低く計算されてしまう。そこで本稿では、Carroll (1967) で提唱されたように *Tokens* を 2 倍したものの平方根をとって、この影響を緩和している。また本稿では、分子の異なり語には一般名詞のみを用いているのに対し、分母は全ての品詞を含んだ文字通りの総単語数としている。



図表3 産業別のTTRの推移

図表3は産業別のTTRの時系列推移を表したものである。増加の傾向を示しているのは、化学、医薬品、自動車の3つであり、逆に減少傾向にあるのはゴムである。それ以外の業種はおおよそ同じ水準を維持しており、多様性はそれほど変化していない。

(3) 出現単語のランキング

本節では総単語数、TTRといった計量的な側面から情報の特性を捉えようと試みているが、実際にはどのような単語が多くみられるのであろうか。図表4は8つの業種における出現単語の30位までのランキングである。各業種とも1位から5位では、おおよそ研究開発、技術、製品、事業、製品といった単語が共通して出現しているが、6位以降になると業種を反映した単語が出てくる。例えば、医薬品においては、治療(6位、以下括弧内の数字は順位を表す)、フェーズ(9)、申請(12)、遺伝子(17)、医薬品(19)、DNA(22)、治療薬(25)、臨床(26)、細胞(27)が業種の特性を表す単語といえ、特に17位の「遺伝子」以降の単語がより業種を代表するものとなっている。この傾向は他の業種においても同様であることが図表からも分かる。

各業種において特性を表す単語はその出現頻度に応じて、5位から15位までと16位以降に分けられる。なかでも高い頻度で業界の特有語を含んでいるのは16位以降ということになる。これはもし業界の特徴をつかむ単語をピックアップしようとするならば、16位以降の単語に注目すればよいということを意味している。

図表4 8業種の出現単語ランキング

| 順位 | 化学 | 医薬品 | ゴム | 窯業 | 機械 | 電気機器 | 自動車 | 精密機器 |
|----|------|------|--------|--------|-------|-------|------|------|
| 1 | 開発 | 開発 | 開発 | 研究開発 | 開発 | 開発 | 開発 | 開発 |
| 2 | 研究開発 | 研究開発 | 研究開発 | 開発 | 研究開発 | 研究開発 | 研究開発 | 研究開発 |
| 3 | 技術 | 試験 | 技術 | 技術 | 技術 | 技術 | 技術 | 技術 |
| 4 | 製品 | 当社 | 製品 | 事業 | 事業 | 製品 | 製品 | 事業 |
| 5 | 事業 | 承認 | 事業 | 製品 | 活動 | 事業 | 活動 | 製品 |
| 6 | 分野 | 治療 | ゴム | 活動 | 当社 | 当社 | 当社 | システム |
| 7 | 材料 | 活動 | 当社 | 当社 | 製品 | 活動 | 事業 | 当社 |
| 8 | 当社 | 米国 | 活動 | 年度 | 年度 | システム | 部品 | 活動 |
| 9 | 機能 | フェーズ | 年度 | 連結会計 | 連結会計 | 対応 | 商品 | 年度 |
| 10 | 活動 | 研究 | 研究 | 分野 | 対応 | 機器 | 対応 | 装置 |
| 11 | 研究 | 事業 | 部品 | 商品 | 装置 | 市場 | 年度 | 連結会計 |
| 12 | 新規 | 申請 | 分野 | グループ | システム | 年度 | 連結会計 | 機能 |
| 13 | 年度 | 販売 | 連結会計 | 関連 | 市場 | 分野 | 自動車 | 機器 |
| 14 | 対応 | 年度 | 材料 | 研究 | グループ | 連結会計 | システム | 分野 |
| 15 | 市場 | 日本 | 量産 | 材料 | 研究 | 機能 | 環境 | 関連 |
| 16 | 連結会計 | 連結会計 | グループ | 環境 | 関連 | グループ | グループ | 可能 |
| 17 | 関連 | 遺伝子 | 樹脂 | 製造 | 商品 | 装置 | 性能 | 実現 |
| 18 | 環境 | 対象 | 自動車 | 対応 | 分野 | 実現 | 技術開発 | 対応 |
| 19 | グループ | 医薬品 | 機能 | 機能 | 機能 | 関連 | エンジン | グループ |
| 20 | 塗料 | 製品 | 商品 | 耐火物 | 加工 | 可能 | 研究 | 市場 |
| 21 | 用途 | 技術 | 中心 | 性能 | 環境 | 商品 | 採用 | 研究 |
| 22 | 樹脂 | DNA | 環境 | コンクリート | 機器 | 部門 | ニーズ | シリーズ |
| 23 | 部門 | 領域 | 対応 | 市場 | シリーズ | 環境 | 量産 | 使用 |
| 24 | 新製品 | 一般名 | セグメント | 中心 | ニーズ | シリーズ | 燃費 | 検査 |
| 25 | 研究所 | 治療薬 | メーカー | 工法 | 部門 | 研究 | 安全 | 成果 |
| 26 | 商品 | 臨床 | 市場 | 部門 | 機械 | 性能 | 機能 | 商品 |
| 27 | ニーズ | 細胞 | 関連 | セラミックス | セグメント | 制御 | 材料 | 製造 |
| 28 | 展開 | グループ | 新規 | 施工 | 製造 | セグメント | 市場 | 測定 |
| 29 | 製造 | 適応 | コンパウンド | セグメント | 次 | ニーズ | 実現 | 発売 |
| 30 | 顧客 | 製造 | 次 | 技術開発 | 実現 | 成果 | コスト | 加工 |

4. 企業別の記載内容の特性比較

先の総単語数とTTRは、企業の研究開発活動の記述のどのような側面を捉えているのだろうか。本稿では、研究開発活動についての情報内容の量に焦点を当てており、画一的なフォーマットに基づく記載を前提とするとき、記載の項目数がそれにあたりと考えられる。そこで総単語数とTTRを記載項目数と比較することによって、どちらが情報内容の量の指標として妥当かを検討している。

ここではまず研究開発/売上高比率の最も高い医薬品業界に注目し、業界内において比率の高い武田薬品工業と低い沢井製薬とを比較する。武田薬品工業は新薬も開発する総合製薬会社であり、沢井製薬はジェネリック医薬品を取り扱う企業である。よって両者における研究開発の位置づけは

当然異なり、記載の内容も大きく違うと考えられる。次に他業種の検討対象として電機産業からパナソニックを取り上げる。

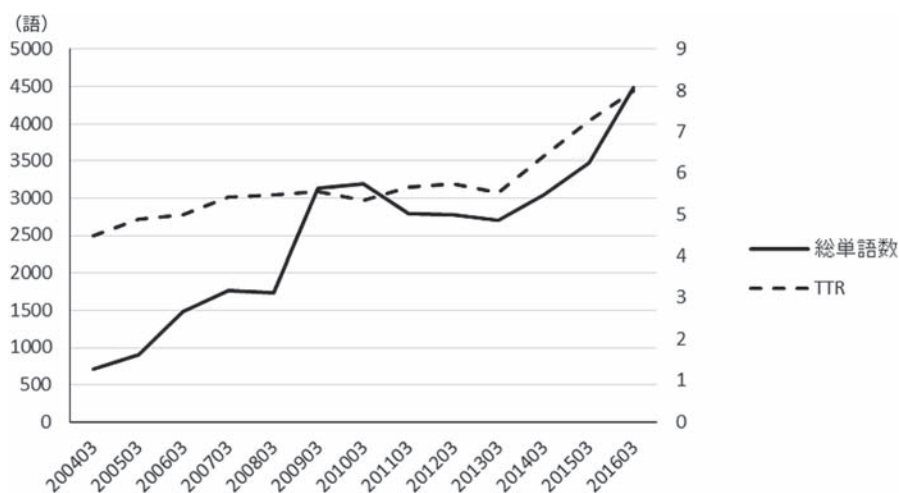
(1) 武田薬品工業

武田薬品工業は、医薬品業界では国内の売上高第1位であり、日本を代表する製薬メーカーである。2016年3月期の事業セグメントは、医療用医薬品事業、コンシューマーヘルスケア事業、その他事業となっており、売上高構成比で見ると90%を医療用医薬品事業が占めている。武田薬品工業は研究開発に積極的であり、売上収益の19%にあたる3,459億円を研究開発に投じている。

図表5 武田薬品工業の記載内容の概要

| 決算年 (総項目数) | 記載内容の概要と項目数 |
|---------------|--|
| 2004 (10) | ・重点領域 (1) ・研究開発活動の主な成果: ①自社研究開発 (1), ②製品付加価値の最大化 (2), ③導入・アライアンス活動 (4), ④研究体制の整備・強化 (2) |
| 2005 (11) | ・重点領域 (1) ・研究開発活動の主な成果: ①自社研究開発 (1), ②製品付加価値の最大化 (2), ③導入・アライアンス活動 (5), ④研究体制の整備・強化 (2) |
| 2006 (22) | ・重点領域 (1) ・研究開発活動の主な成果: ①自社研究開発 (7), ②製品付加価値の最大化 (4), ③導入・アライアンス活動 (10), ④研究体制の整備・強化 (0) |
| 2007 (19) | ・重点領域 (1) ・研究開発活動の主な成果: ①自社研究開発 (2), ②製品付加価値の最大化 (5), ③導入・アライアンス活動 (9), ④研究体制の整備・強化 (2) |
| 2008 (25) | ・重点領域 (1) ・研究開発活動の主な成果: ①自社研究開発 (8), ②製品付加価値の最大化 (5), ③導入・アライアンス活動 (9), ④研究体制の整備・強化 (2) |
| 2009 (32) | ・重点領域 (1) ・研究開発活動の主な成果: ①自社研究開発 (11), ②製品付加価値の最大化 (6), ③導入・アライアンス活動 (11), ④研究体制の整備・強化 (3) |
| 2010 (35) | ・重点領域の変更 (1) ・研究開発活動の主な成果: ①自社研究開発 (15), ②製品付加価値の最大化 (5), ③導入・アライアンス活動 (13), ④研究体制の整備・強化 (1) |
| 2011 (37) | ・重点領域の変更 (1) ・研究開発活動の主な成果: ①自社創成品に関する取組 (10), ②製品付加価値の最大化 (4), ③導入・アライアンス活動 (20), ④研究体制の整備・強化 (2) |
| 2012 (34) | ・重点領域 (1) ・研究開発活動の主な成果: ①自社創成品に関する取組 (15), ②製品付加価値の最大化 (3), ③導入・アライアンス活動 (11), ④研究体制の整備・強化 (4) |
| 2013 (31) | ・重点領域 (1) ・研究開発活動の主な成果: ①自社創成品に関する取組 (8), ②製品付加価値の最大化 (4), ③導入・アライアンス活動 (13), ④研究体制の整備・強化 (5) |
| 2014 (30) | ・重点領域 (1) ・研究開発活動の主な成果: ①自社創成品に関する取組 (11), ②導入・アライアンス活動 (13), ③研究体制の整備・強化 (5) |
| 2015 (36) | ・重点領域 (1) ・研究開発活動の主な成果: ①自社創成品に関する取組 (10), ②導入・アライアンス活動 (16), ③研究体制の整備・強化 (9) |
| 2016 (36) | ・重点領域 (1) ・研究開発活動の主な成果: ①自社創成品に関する取組 (17), ②導入・アライアンス活動 (14), ③研究体制の整備・強化 (4) |

武田薬品工業の記載は、企業が取り組む重点領域を示したうえで、研究開発活動の主な成果を①自社創成品に関する取組、②製品付加価値の最大化、③導入・アライアンス活動、④研究体制の整備・強化の項目別に記載するという形式をとっており、これは2006年から2016年までほとんど変わっていない。2004年と2005年は、①から④の項目を設定してはいないが、当該内容を含んでおり、図表では項目ごとにカウントしている。図表では、項目ごとに記載件数を示しているが、期間を通して記載の項目名が変わらないことによって件数を比較することが容易になっている。2010年までは年を追うごとに項目数が増加しており、その後は2013年と2014年を除いて、35項目前後となっている。



図表6 武田薬品工業の総単語数およびTTRの推移

図表6は武田薬品工業の総単語数（左軸）とTTRの値（右軸）の推移を表したものである。2004年では714語であった総単語数も、2016年には4,491語に達している。文章の豊かさを表すTTRはこれと同様の動きをしている。両者の特性を探るために変化の方向が異なる2つの期間に着目して検討する。まず2007年から2008年にかけては、総単語数は1,766語から1,736語へと30語減少しているのに対して、TTRは5.418から5.482へと0.064増加している。情報内容のベンチマークとして項目数を用いるとき、同時期に19項目から25項目へと6項目増加していることから判断すると、TTRの方が情報内容を適切に描写している。また2010年から2011年にかけては、総単語数が3,190語から2,797語へと393語減少する一方、TTRは5.358から5.656へと0.297増加している。この間に項目数は35項目から37項目へと2つ増加しており、ここでもTTRの方がこの動きを適切に補足しているといえる。

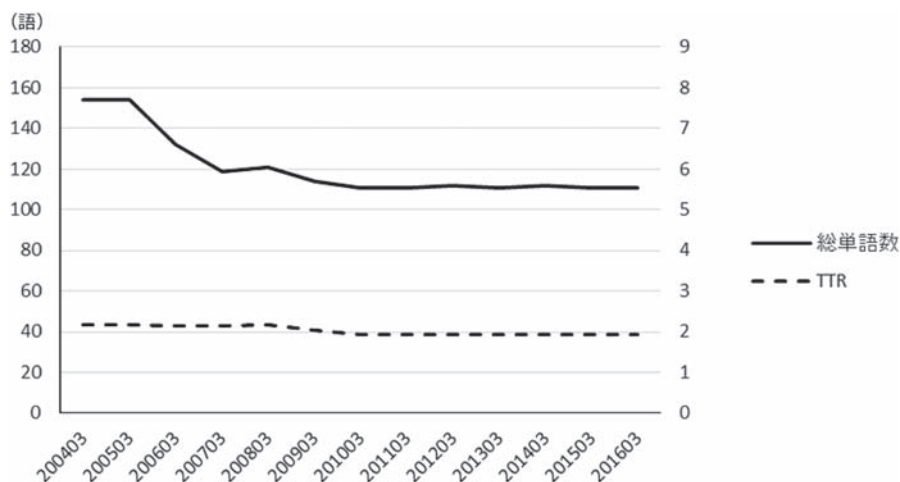
(2) 沢井製薬

沢井製薬は、すでに販売されている薬品と同じ有効成分をもつジェネリック医薬品の製造および販売を行う製薬会社である。そのため、新薬の研究、開発を手掛ける武田薬品工業に比べて、研究開発活動の規模も大きく異なる。2016年では80億円の研究開発費を支出しており、売上高に占める研究開発費の割合は6.4%となっている。

図表 7 沢井製薬の記載内容の概要

| 決算年 (総項目数) | 記載内容の概要と項目数 |
|---------------|---|
| 2004 (3) | ・研究開発体制：当社に研究開発本部、メディスン新薬に研究開発部 (1) ・製造承認 28 品目 (1)、製造承認申請 26 品目 (1) |
| 2005 (3) | ・研究開発体制：当社に研究開発本部、メディスン新薬に研究開発部 (1) ・製造承認 35 品目 (1)、製造承認申請 48 品目 (1) |
| 2006 (3) | ・研究開発体制：当社に研究開発本部、メディスン新薬に研究開発部 (1) ・製造承認 47 品目 (1)、製造承認申請 55 品目 (1) |
| 2007 (2) | ・研究開発体制：当社に研究開発本部、メディスン新薬に研究開発部 (1) ・製造承認 46 品目 (1) |
| 2008 (2) | ・研究開発体制：当社に研究開発本部、メディスン新薬に研究開発部 (1) ・製造承認 63 品目 (1) |
| 2009 (2) | ・研究開発体制：当社に研究開発本部 (1) ・製造承認 65 品目 (1) |
| 2010 (2) | ・研究開発体制：当社に研究開発本部 (1) ・製造承認 53 品目 (1) |
| 2011 (2) | ・研究開発体制：当社に研究開発本部 (1) ・製造承認 44 品目 (1) |
| 2012 (2) | ・研究開発体制：当社に研究開発本部 (1) ・製造承認 40 品目 (1) |
| 2013 (2) | ・研究開発体制：当社に研究開発本部 (1) ・製造承認 32 品目 (1) |
| 2014 (2) | ・研究開発体制：当社に研究開発本部 (1) ・製造承認 23 品目 (1) |
| 2015 (2) | ・研究開発体制：当社に研究開発本部 (1) ・製造承認 28 品目 (1) |
| 2016 (2) | ・研究開発体制：当社に研究開発本部 (1) ・製造承認 26 品目 (1) |

沢井製薬の記載内容の概要を示した図表 7 から、2004 年から 2016 年まで、沢井製薬の記載のフォーマットはほとんど変化がないことが分かる。記載内容はコンパクトであり、製造承認または製造承認申請は件数のみについて開示されており、詳細は示されていない。そのため項目数のカウントしてはそれぞれ 1 としている。



図表8 沢井製薬の総単語数およびTTRの推移

図表8は沢井製薬の総単語数（左軸）およびTTR（右軸）の時系列の推移を表したものである。総単語数で見た場合には、2004年の154語から徐々に減少し、2010年には111語となり、以後その水準を維持している。またTTRについても、2004年の2.165から2010年には1.964へとわずかに減少し、その後は同様の水準となっている。両者の動きはほぼ同じである。

2004年、2005年、2006年に着目すると、情報内容の多さを示す項目数は3のみであるが、総単語数は2005年の154語から2006年の132語と14%の減少となっている。一方、TTRは2.165から2.154と0.5%の減少であり、ほぼ同様の水準を保っていることから、より適切に情報内容の多さを描写しているといえる。

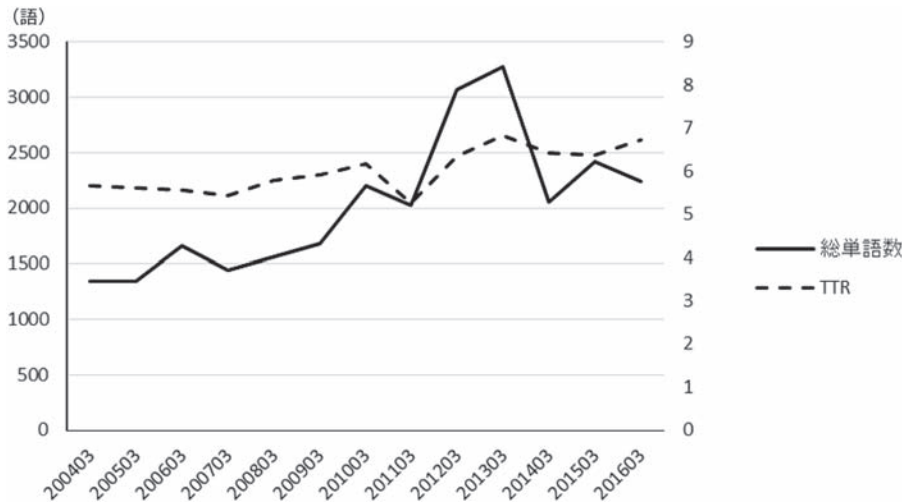
(3) パナソニック

電気機器業界のなかでも大手のパナソニックは、自社での研究開発に多額の資金を投じている。2016年3月決算の投資額は4,498億円であり、これは売上高の6%に相当する。

図表 9 パナソニックの記載内容の概要

| 決算年 (総項目数) | 記載内容の概要と項目数 |
|---------------|---|
| 2004 (30) | <ul style="list-style-type: none"> ・全社的な研究開発体制 (1) ・事業セグメント別の成果: ① AVC ネットワーク: 研究体制, 研究成果 (7), ② アプライアンス: 研究体制, 研究成果 (7), ③ デバイス: 研究体制, 研究成果 (7), ④ 日本ビクター: 研究体制, 研究成果 (5), ⑤ その他: 研究体制, 研究成果 (3) |
| 2005 (16) | <ul style="list-style-type: none"> ・全社的な重点開発テーマ (1) ・事業セグメント別の成果: ① AVC ネットワーク: 研究体制, 研究成果 (3), ② アプライアンス: 研究体制, 研究成果 (2), ③ デバイス: 研究体制, 研究成果 (3), ④ 電工・パナホーム: 研究体制, 研究成果 (3), ⑤ 日本ビクター: 研究体制, 研究成果 (2), ⑥ その他: 研究体制, 研究成果 (2) |
| 2006 (16) | <ul style="list-style-type: none"> ・全社的な重点開発テーマ (1) ・事業セグメント別の成果: ① AVC ネットワーク: 研究体制, 研究成果 (3), ② アプライアンス: 研究体制, 研究成果 (2), ③ デバイス: 研究体制, 研究成果 (3), ④ 電工・パナホーム: 研究体制, 研究成果 (3), ⑤ 日本ビクター: 研究体制, 研究成果 (2), ⑥ その他: 研究体制, 研究成果 (2) |
| 2007 (18) | <ul style="list-style-type: none"> ・全社的な研究開発方針 (1) ・事業セグメント別の成果: ① AVC ネットワーク: 研究体制, 研究成果 (4), ② アプライアンス: 研究体制, 研究成果 (3), ③ デバイス: 研究体制, 研究成果 (4), ④ 電工・パナホーム: 研究体制, 研究成果 (3), ⑤ 日本ビクター: 研究体制, 研究成果 (2), ⑥ その他: 研究体制, 研究成果 (1) |
| 2008 (16) | <ul style="list-style-type: none"> ・全社的な研究開発方針 (1) ・事業セグメント別の成果: ① AVC ネットワーク: 研究体制, 研究成果 (4), ② アプライアンス: 研究体制, 研究成果 (3), ③ デバイス: 研究体制, 研究成果 (4), ④ 電工・パナホーム: 研究体制, 研究成果 (3), ⑤ 日本ビクター: 研究開発費の額のみ表示, ⑥ その他: 研究体制, 研究成果 (1) |
| 2009 (13) | <ul style="list-style-type: none"> ・全社的な研究開発方針 (1) ・事業セグメント別の成果: ① デジタル AVC ネットワーク: 研究体制, 研究成果 (3), ② アプライアンス: 研究体制, 研究成果 (3), ③ 電工・パナホーム: 研究体制, 研究成果 (3), ④ デバイス: 研究体制, 研究成果 (2), ⑤ その他: 研究体制, 研究成果 (1) |
| 2010 (15) | <ul style="list-style-type: none"> ・全社的な研究開発方針 (1) ・事業セグメント別の成果: ① デジタル AVC ネットワーク: 研究体制, 研究成果 (3), ② アプライアンス: 研究体制, 研究成果 (3), ③ 電工・パナホーム: 研究体制, 研究成果 (3), ④ デバイス: 研究体制, 研究成果 (2), ⑤ 三洋電機: 研究体制, 研究成果 (2), ⑥ その他: 研究体制, 研究成果 (1) |
| 2011 (13) | <ul style="list-style-type: none"> ・全社的な研究開発方針 (1) ・事業セグメント別の成果: ① デジタル AVC ネットワーク: 研究体制, 研究成果 (3), ② アプライアンス: 研究体制, 研究成果 (2), ③ 電工・パナホーム: 研究体制, 研究成果 (2), ④ デバイス: 研究体制, 研究成果 (2), ⑤ 三洋電機: 研究体制, 研究成果 (2), ⑥ その他: 研究体制, 研究成果 (1) |
| 2012 (16) | <ul style="list-style-type: none"> ・全社的な研究開発方針 (1) ・事業セグメント別の成果: ① デジタル AVC ネットワーク: 研究体制, 研究成果 (2), ② アプライアンス: 研究体制, 研究成果 (2), ③ システムコミュニケーションズ: 研究体制, 研究成果 (2), ④ エコソリューションズ: 研究体制, 研究成果 (3), ⑤ オートモーティブシステムズ: 研究体制, 研究成果 (1), ⑥ デバイス: 研究体制, 研究成果 (2), ⑦ エナジー: 研究体制, 研究成果 (1), ⑧ その他: 研究体制, 研究成果 (2) |
| 2013 (11) | <ul style="list-style-type: none"> ・組織横断的な取り組み (1) ・事業セグメント別の成果: ① AVC ネットワークス: 研究体制, 研究成果 (2), ② アプライアンス: 研究体制, 研究成果 (2), ③ システムコミュニケーションズ: 研究体制, 研究成果 (2), ④ エコソリューションズ: 研究体制, 研究成果 (2), ⑤ オートモーティブシステムズ: 研究体制, 研究成果 (2) |
| 2014 (16) | <ul style="list-style-type: none"> ・組織横断的な取り組み (4) ・事業セグメント別の成果: ① アプライアンス: 研究体制, 研究成果 (3), ② エコソリューションズ: 研究体制, 研究成果 (3), ③ AVC ネットワークス: 研究体制, 研究成果 (3), ④ オートモーティブ & インダストリアルシステムズ: 研究体制, 研究成果 (3) |
| 2015 (13) | <ul style="list-style-type: none"> ・組織横断的な取り組みと成果 (1) ・事業セグメント別の成果: ① アプライアンス: 研究体制, 研究成果 (3), ② エコソリューションズ: 研究体制, 研究成果 (3), ③ AVC ネットワークス: 研究体制, 研究成果 (3), ④ オートモーティブ & インダストリアルシステムズ: 研究体制, 研究成果 (3) |
| 2016 (15) | <ul style="list-style-type: none"> ・組織横断的な取り組みと成果 (3) ・事業セグメント別の成果: ① アプライアンス: 研究体制, 研究成果 (3), ② エコソリューションズ: 研究体制, 研究成果 (3), ③ AVC ネットワークス: 研究体制, 研究成果 (3), ④ オートモーティブ & インダストリアルシステムズ: 研究体制, 研究成果 (3) |

パナソニックも基本的なフォーマットは武田薬品工業と同様である。はじめに全社的な研究開発の体制や方針に触れたあと、事業セグメントごとに、研究開発活動を担当している部署と具体的な成果が述べられている。パナソニックでは、事業セグメントがある程度細かく分かれ、それぞれ研究開発費の額も示されており、セグメント別の売上高や利益といった情報と対比することができるという特徴がある。



図表 10 パナソニックの総単語数および TTR の推移

パナソニックの総単語数（左軸）は2004年の1,350語から、2013年の3,275語まで増加し、その後2,000語程度の記述に落ち着いている。TTR（右軸）の動きについても、その変動は穏やかながら、総単語数と同様の動きをしていることが分かる。

総単語数とTTRのどちらが記載内容を示す指標として妥当なのかを、両者の動きが異なる2つの期間を対象に検討する。2014年から2015年にかけては、総単語数が356語増加、TTRが0.061の減少となっている。同期間での項目数は3項目の減少である。また、2015年から2016年にかけては、総単語数が176語減少、TTRが0.350の増加となっている。この期間、項目数は2項目の増加となっており、先の例とあわせるとTTRが情報内容の量を適切に捉えていることが分かる。

3社の記載内容の分析から分かることは、以下のとおりである。まず、有価証券報告書の作成にあたって一度用いられたフォーマットはそのまま引き継がれる傾向が強いという点がある。これは情報利用者にとっては、時系列での比較が行いやすいという特性となっている。また記載内容については、概ねガイドラインの例示を踏襲しているといえる。

次に、研究開発活動の記載内容の多さを表す値として、総単語数およびTTRともに一定の妥当性を有するが、記載内容の項目数を基準に判断した場合には、TTRの方が記載内容を適切に補足している可能性が高いことが分かる。

5. まとめと今後の課題

本稿では有価証券報告書の研究開発活動の状況について、総単語数、TTRのどちらが記載内容の量を適切に表しているかを、時系列およびクロスセクション、具体的なケースを通じて検討した。

各企業の記載内容からは、おおよそガイドラインで示された内容にしたがって記載がなされていることが分かった。具体的には、各企業ともガイドラインに例示されていた、研究の目的、主要課題、研究成果、研究体制等の観点から記載を行っていることが分かる。また研究開発活動の内容や研究開発費の金額が、セグメント情報に関連付けて開示されていた点もガイドラインの通りであった。これらの特徴を有しているため、各企業の記載内容に関しては比較可能性は高く、とくに各企業とも一度採用した記載フォーマットを用いる傾向があることから、同一企業の時系列比較は行いやすいといえる。

研究開発の記載が前述のような特徴をもつことをふまえ、本稿では情報内容のベンチマークとして、記載項目数を用いて検証を行った。記載内容の情報量を表すと考えられる総単語数およびTTRの分析からは、記載内容のベンチマークである項目数と比較した場合、TTRのほうが総単語数よりも記載内容の情報量を適切に捕捉している可能性が高いことが示された。

残された課題は以下のとおりである。まず記載内容のベンチマークの問題がある。本稿では記載項目数に焦点を当てているが、企業によっては、セグメントごとに同数の項目を設定したうえで記述を行っている可能性が高いものがある。また同じ項目数であっても、記述の深さに違いがある場合も存在するだろう。後者については、例えば単に異なり語のみを対象とするのではなく、第3節でも取り上げたように、研究開発により関連する頻出単語の出現率を用いるなどの工夫が考えられる。今後はより多くのベンチマーク指標も用いて多角的にチェックすることが有用であると考えられる。

最後に、文章の内容は、単語より大きな区分である節や文さらには段落単位で意味をなすものであり、分析対象をこれらのレベルに拡大する必要がある。これまではPCの性能やプログラムの制約があり取り上げることが難しかったが、PCの処理速度の向上や形態素解析、形態素の関連を分析する構文解析といった、テキスト分析とその周辺の技術の進歩は目覚ましいものがある。これら技術を用いることによって更なるテキスト分析研究の進展が期待される。

(付記) 本研究はJSPS 科研費(若手B: 研究課題番号17K13833)の助成を受けたものです。

引用・参考文献

- Carroll, J. B. 1967. On sampling from a lognormal model of word-frequency distribution. In *Computational analysis of present-day American English*, edited by H. K. & W. N. Francis, 406–424.
- Kothari, S. P., T. E. Laguerre, and A. J. Leone. 2002. Capitalization versus expensing: evidence on the uncertainty of future earnings from capital expenditures versus R&D outlays. *Review of Accounting Studies* 7 (4): 355–382.
- Lev, B., and T. Sougiannis. 1996. The capitalization, amortization, and value-relevance of R&D. *Journal of Accounting and Economics* 21 (1): 107–138.
- 伊藤雅光. 2002. 『計量言語学入門』. 大修館書店.
- 金明哲. 2009. 『テキストデータの統計科学入門』. 岩波書店.
- 田島ますみ・深田淳・佐藤尚子. 2008. 「語彙多様性を表す指標の妥当性に関する研究：日本人大学生の書き言葉コーパスの場合」. 『中央学院大学社会システム研究所紀要』9 (1): 51–62.

The Content Analysis of R&D Activities in YUKASHOUKENHOUKOKUSHO

Yu ISHIMITSU

ABSTRACT

This paper examines how to capture a feature of sentence which manager describe about R&D. From a comparison between tokens (number of words) and TTR (types token ratio), the latter is better proxy of richness of sentence.