

テキストマイニングを活用したリスク概念の分析

Analysis of Risk Concept through Text Mining

小室 達章

Tatsuaki KOMURO

1. 問題意識

リスクという言葉は、多種多様な意味を有しており、日常生活においても、また、学問領域においても、必ずしも厳密な意味で用いられていない。リスクは、直訳すれば「危険」という意味になり、日常で使用する言葉としては、厳密に定義する必要はないのかもしれない。しかし、学問の対象として、もしくは、実務の対象として、リスクという現象を取り扱う場合、リスクという言葉について何らかの定義をおこなうことが求められる。リスクマネジメントとは、企業経営や組織運営において、何らかにリスクに対応する業務のことを意味する。そのため、マネジメントすべきリスクを特定する場合に、どのようにリスクが定義されているかは重要な意味を持つと考えられるからである。

リスクマネジメントの研究において、リスクの定義についての分類・整理、および統一化について多くの考察がなされている¹。ここでは、既存研究において、どのような定義が採用されているのか、また、それらの定義がどのように分類・整理されているのかを概観

してみよう。例えば、Vaughn (2007) は、リスクマネジメントの既存研究におけるリスクの定義を、以下のように整理している。1) 損失のチャンス (the chance of loss), 2) 損失の可能性 (the possibility of loss), 3) 不確実性 (uncertainty), 4) 期待された結果と現実の差異 (the dispersion of actual expected results), 5) 期待したものと異なった結果が生じる可能性 (the probability of any outcome different from the one expected), 6) 損失の可能性が存在するための条件 (a condition in which a possibility of loss exists) である。

ここでいう1) 損失のチャンスおよび2) 損失の可能性は、将来の損失を意味している。もう少し詳細に言えば、将来発生するかもしれない損失に対して、それが発生する機会(確率)や可能性をリスクとしている。3) 不確実性、4) 期待された結果と現実の差異、5) 期待したものと異なった結果が生じる可能性は、予想(思い描いた将来)と、実際に起こりうる事象との差異を意味している。これも、もう少し詳細に言えば、何かを期待して(予想して)意思決定をおこなう場合、当然ながら、その期待を裏切る結果となる可能性が存在し、その期待に対する裏切りをリスクとするのである。6) 損失の可能性が存在

¹ 多様なリスク概念、リスクの定義の整理については、Vaughn (2007)、石名坂 (1994)、亀井 (1992) などが詳しい。

するための条件は、損失をもたらす状況を意味する。つまり、損失そのものよりも、損失の原因となる事故、災害、事件などをリスクと考えるのである。このように、既存研究におけるリスク概念を簡単に整理しただけでも、そのとらえ方が多種多様であることがわかる。統一的概念として、リスクを定義するのであれば、これらの意味を包括するような表現が必要となる。

しかしながら、リスクマネジメント研究者自身も、このようなリスク概念の多様性を目の当たりにして、リスク概念をいかに定義するかについて議論の不毛さを指摘する。例えば、亀井（1992）は、仮に統一的概念を規定したとしても、それ自身が自己満足的で、孤立するだけであり、後の理論展開にほとんど機能しないことを指摘する。つまり、リスクの定義に関する議論は、結論に到達していないのである。また、仮に統一的概念をおこなったとしても、その定義からはずれた現象をリスクとみなさなければならない場合、定義と現実の記述との間で齟齬が生じることにもなる。このことから、リスク概念の定義

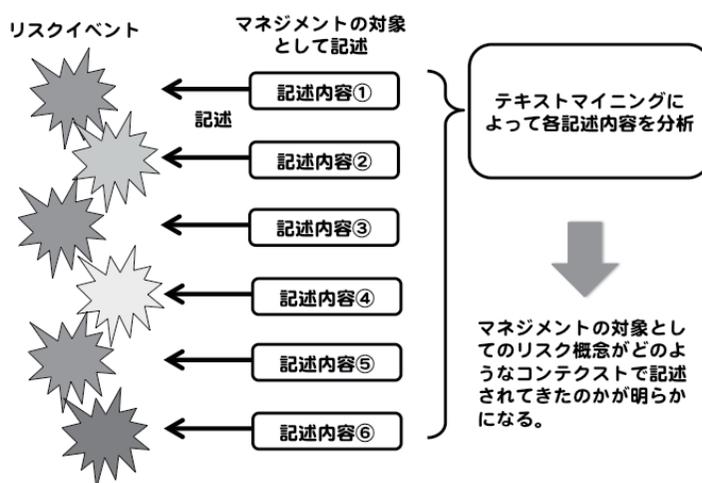
において重要なことは、統一的概念を目指すのではなく、リスク概念のどのような特性を研究対象とするか、どのようなマネジメントのあり方を考察の対象とするかを把握しておくことであり、また、それに合わせてリスクという概念を認識しておくことである。

そこで、本研究では、リスクという概念を统一的に定義することを目指すのではなく、これまでリスクという現象に直面したときに、どのようにリスクを意味づけてきたのか、リスクという言葉をどのようなコンテキスト（文脈）において用いてきたのか、について考察することを目的とする。具体的には、リスクマネジメント（リスク管理）について書かれた文章（本研究では、新聞記事）において、リスクという言葉がどのような意味合いで用いられてきたのかを明らかにすることで、リスク概念の意味や、それが使用されるコンテキストを考察する。

2. 研究アプローチ

（1）研究枠組み

本研究では、リスクという概念がどのよう



図表 1：本研究の枠組み

な意味・コンテキストで用いられているのかを、テキストマイニングという手法を用いて分析する。テキストマイニングとは、文字列を対象としたデータマイニングのことで、通常の記事からなるデータを、「単語」や「文節」で区切り、それらの出現の頻度や共出現の相関、出現傾向などを解析し、共時的・時系列的に比較分析することで有用な情報を取り出す研究方法である²。

本研究において、テキストマイニングという分析手法を用いる理由は、以下の仮説を検証するためである。それは、1) 組織や社会は、新聞記事などに取り上げられるリスクイベント（企業事故、災害、不祥事など）が発生すると、それをマネジメントすべきリスクとして認識する、そして、2) さまざまな種類のリスクイベントが発生すると、その結果、マネジメント（管理）すべきリスクが、特定のコンテキストだけでなく多様なコンテキストにおいて注目されるようになる、という仮説である。そのため、現実には起こったイベントをふまえながら、リスクマネジメントについて書かれた新聞記事のテキストマイニングをおこなうことで、この仮説を確認できると考える（図表1参照）。つまり、テキストマイニングをおこなうにあたり、テキストデータを解析するだけでなく、そのテキストにおいて記述の対象となっている現実をも考察の射程に入れるということである。

本研究におけるテキストマイニングの活用方法について、いくつかの既存研究の枠組みを参考にした。喜田（2006）は、1976年から1998年にかけてのアサヒビールの有価証券報告書に記載された内容を時系列的に分析し、組織変革に先行して概念変化が存在すること、また、企業の業績が悪い時に概念変化

が起りやすいことを明らかにしている。久保（2008）は、1939年から2007年にかけてのアメリカ会計基準における「risk」という用語のコンテキストを数量的に分析し、1980年代以降に「risk」の使用頻度が増加し、純粹リスクから投機的リスクへとコンテキストが移行していることを明らかにしている。内藤（2011a）は、1989年、1991年、1997年、2002年における新聞記事に記述された「企業倫理」概念について、共起関係に基づく意味ネットワークを分析し、「社会的責任」「法令順守」「社会貢献」という概念も合わせて複数の中心からなる意味ネットワークを年代ごとに抽出している。吉田他（2011）は、業績の高い企業の有価証券報告書に記載された内容の共通項を分析し、好業績企業は有価証券報告書の各項目においていくつかの特徴語が存在することを明らかにしている。また、正田他（2012）は、東証一部上場企業、三重県企業、米ビジョナリーカンパニーの経営理念に使用されている言葉の違いを分析し、それぞれ使用されている言葉を文脈カテゴリーに位置付けて、それぞれの経営理念の違いを明らかにしている。

これらの既存研究におけるテキストマイニングの活用方法の特徴は、以下のように整理できる。それは、1) 時系列的な分析、2) 意味カテゴリーへの位置付け、である。時系列的な分析では、新聞記事や有価証券報告書など同種のテキストを年代ごとに分析し、それぞれの年代における記述の特徴を明らかにする。特に、分析対象となるキーワードが、どのような言葉と共起されるかについて、年代ごとの相違を考察の対象とする。また、意味カテゴリーへの位置付けでは、文章の中で使用されている言葉や、特定のキーワードに共起される言葉を、いくつかの意味カテゴリーに位置付けることで、分析対象となる

² テキストマイニングの分析手法を用いた研究については、内藤（2011b）を参照。

図表 2：分析対象となる新聞記事³

	総記事数	人事情報	図表・箇条のみ	セミナー等情報	著作権関係	社名	分析対象数
1994	142	1	5	4	1	1	130
1995	213	9	3	4	0	0	197
1996	174	14	4	1	0	0	155
1997	184	31	5	3	0	0	145
1998	270	41	4	4	1	0	220
1999	278	65	4	3	4	0	202
2000	287	70	5	6	1	0	205
2001	310	109	7	9	0	0	185
2002	256	100	8	5	2	2	139
2003	319	123	8	4	0	1	183
2004	303	127	5	7	0	1	163
2005	325	150	3	2	0	1	169
2006	316	155	7	2	0	8	144
2007	384	180	4	2	0	8	190
2008	418	188	9	2	0	4	215
2009	368	183	3	1	0	6	175
2010	337	202	1	4	0	8	122
2011	329	176	2	2	0	7	142
2012	357	201	3	3	0	5	145
2013	367	202	3	1	0	7	154

キーワードが、どのような意味合い（意味カテゴリー）の言葉として用いられているのかを明らかにしている。

本研究におけるリスク概念の分析に活用するテキストマイニングも、1) 時系列的な分析、2) 意味カテゴリーへの位置付けという枠組みに沿っておこなう。具体的には、以下の通りの手続きをとる。時系列的な分析は、テキストデータを年代ごとに整理して、その年代ごとのテキストデータの解析結果を比較する。本研究のテキストデータは、1994年から2013年にかけてのリスクマネジメントについて書かれた新聞記事である。特に、本研究

では企業経営やビジネスという分野におけるリスク概念の分析に主眼を置いているため、日本経済新聞（全国版／朝刊）の記事を用いる。日経テレコンの検索機能を利用し、「リスクマネジメント」および「リスク管理」というキーワードで、当該年代の新聞記事のタイトルと本文に検索をかけて、テキストデータとなる記事を抽出する。これらの記事から、人事・企業情報、書誌・セミナー情報、図表・箇条書き、文脈不統一（雑多な情報の中の1つとしてリスクが掲載されているもの）、著作権により本文掲載不可、企業名のみ記載された記事は削除する（図表2参照）。このように、テキストデータを整理した上で、それぞれの記事を年代ごとに1つのファイルにまとめる。年代ごとにテキストデータを1つのファイルにまとめて、そのファイルごとに解析して比較検討することで、時系列的な分析

³ リスクマネジメントに関する新聞記事の総数が年代を追うごとに増えていること、また、今回は削除された人事情報の総数も同様の傾向にあることから、リスクという言葉の社会的認知・関心が年々高まっていることや、企業の部署・担当者レベルにおいてリスクへの対応が具体的業務になっていることなどがうかがえる。

をおこなうことが可能となる。

意味カテゴリーとしては、以下の2つの分類軸を用いる。それは、純粹リスクと投機的リスクという区分の軸と、経済性、信頼性、正当性という区分の軸である。純粹リスクとは、それが実際に発生した時には損失のみをもたらす、利益を得る可能性のないリスクである。具体的には、地震や洪水などの災害、死傷者などを出す各種事故、過失による法的賠償責任などである。投機的リスクとは、損失をもたらすが、利益となることもあるリスクである。具体的には、株価や為替の変動、景気の好不況、戦争・動乱などの社会的変動、新製品開発等の企業活動に起因するリスクである。

また、経済性、信頼性、正当性という区分は、多面的な視点で追究されるリスクマネジメント研究を、研究対象の違いというよりは、研究者の学術的背景や、その研究者の属する学派の考え方の違いに焦点を当てた区分である⁴。経済性という視点では、経済的な損失をもっとも効率的に最小にする方法を選択することが強調される。経済性の考えの下では、リスクが現実化したときの損失だけでなく、リスクをマネジメントするためのコストも経済的損失と捉え、その損失を最小限にするための手続きに焦点が当てられる。信頼性という視点では、損害をもたらす可能性をできる限りゼロに近づけることが重要となるという考え方を強調する。損害を出さないこと、つまり、業務の安全性や組織の信頼性を高めていくことが、リスクマネジメントの本質と考える。そのため、事故の発生を未然に防止するための仕組みや、事故が発生したとしても被害を食い止めるための仕組みに焦点を当

てる。正当性という視点では、どのようにリスクに対応するかについて、社会的な合意を得ることが重要であると考えられる。リスクマネジメントの仕組みや手続きが経済的に割に合うものでも、信頼性や安全性を高めるものであったとしても、それが利害関係者（ステークホルダー）によって合意が得られていなければ、それを遂行することは実質的に困難となる。そのため、社会的合意が得られた手続きをおこなうことこそが重要となる。経済性、信頼性という客観的な指標で求められるリスクというよりは、主観的に認識されるリスクを主眼に置く一方で、規格や基準を制定するなど制度的に対応することで、リスクの存在およびマネジメントのあり方を正当化する。

以上のように、本研究におけるテキストマイニングを活用したリスク概念の分析では、1) 20年間に渡っての新聞記事を年代ごとに比較検討するという時系列分析と、2) 純粹リスク・投機的リスクという区分、および経済性・信頼性・正当性という視点の区分という意味カテゴリーへの位置付け、という枠組みに沿って分析をおこなう。

(2) テキストマイニングによる分析方法

テキストマイニングでは、まずテキスト上での言葉の解析を済ませ、続いてデータ解析をおこなう。テキスト上での言葉の解析では、まず、言葉が連なる文章を、意味のある単語に区切り、辞書などを利用して品詞や内容を判別する。これを形態素解析という。形態素とは、文章の要素のうち意味を持つ最小の単位である。データ解析では、この形態素ごとに分けられたテキストデータを用いて、上記の時系列的な分析および意味カテゴリーへの位置付けをおこなう。

これらの処理をPC上で一括しておこなうソフトウェアがいくつか存在している。本研

⁴ 経済性、信頼性、正当性という区分については、小室(2013)を参照。

究においては、Rを用いたテキストマイニングをおこなう⁵。形態素解析は、RのパッケージであるRMeCabと、辞書mecab-ipadicを使用する。テキストの解析結果は、Rで使用されるデータ形式に変換されているので、それらをそのままRによるデータ解析や、共起ネットワークのグラフィックス化のインプットデータとして用いる。Rでのグラフィックス化は、同じくRのパッケージであるiGraphを用いる。

テキストの形態素解析をおこなう前処理として、テキストデータを整理する。それぞれのテキストデータの中の文字列を整え、使用可能なデータを分析に適した形式に整形する。具体的には、記事テキストにおける、括弧つきの地名、人名、部署名、記者名、年齢、電話番号、ふりがな等を削除、もしくは、句読点に変換する。これらは、明らかに分析上余計なテキストデータであり、それをあらかじめ除去しておくことで、分析の精度を高めることができる。また、同様の理由により、記号、特集名、写真、記事参照、問い合わせ等の付記的な記述も除去して、データの精度を高める。また、略語・略字の使用統一、同義語・類義語の同一化、表記ぶれ等を統一させる⁶。これは、明らかに同じ意味で使用されているにもかかわらず異なる表記をしている

言葉を、統一的な表記に整理する作業である。以上のテキストデータの整理をおこなった上で、形態素解析を実施する。

テキストの解析では、単語を、内容語と機能語に分ける。内容語とは、名詞、形容詞、動詞であり、機能語とは、助詞などの文書を文法的に成立させるために必要な単語である。今回は、「リスク」という概念がどのような意味で用いられているかを明らかにすることを目的としているため、内容語のみの抽出をおこなう。また、分析を簡素化するために、内容語のうち「名詞」のみを、特に、一般名詞、サ変接続名詞、形容動詞語幹名詞に限定し、その使用頻度を計測する⁷。

また、リスクという言葉に共起される言葉の抽出によって、リスクがどのような意味として用いられているか、どのようなコンテキストで用いられているかを判断する。ビジュアル的にわかりやすくするために、その抽出された言葉のつながりを共起ネットワークとしてグラフィックス化する。ただし紙幅の都合上、すべての年代で共起ネットワークを描くのは困難であるため、今回は、1994年～1998年(第I期)、1999年～2003年(第II期)、2004年～2008年(第III期)、2009年～2013年と、5年間ごと4時期に分割して比較する。つまり、年代ごとにリスクという言葉に共起される言葉を比較検討することで、リスクに紐付けされている言葉が時系列的にわかることになり、そこから、リスクがどのような意味として用いられているかを推察することが可能となる。

また、意味カテゴリーへの位置付けでは、年代ごとに、特定の意味カテゴリーに属する

⁵ Rを用いたテキストマイニングの方法については、石田(2008)を参照。

⁶ 例えば、「リスク管理」を「リスクマネジメント」に、「供給網」「部品供給網」を「サプライチェーン」に統一させるなどである。すべての表記統一をここで示すことはできないが、同様の意味を持つ言葉は、統一的な表記にするという原則で、テキストデータの整理をおこなった。略語・略字の使用を統一させるべきか、また、同義語・類義語を同一にすべきかどうか、表記ぶれを修正すべきかどうかについては、確かに議論の余地はある。特に、厳密な意味で、どのような「表現」として用いられているかを分析するのであれば、修正すべきではないかもしれない。しかし、本研究では、言葉としての厳密な「表現」よりも、その言葉が持っている「意味」を分析対象としているため、これらの修正をおこなうという判断をした。

⁷ 名詞のうち、場所、人名、企業名を表す固有名詞を分析の対象に入れるかどうかの判断について、今回は、概念の意味を考察するという目的から、固有名詞は外すことにした。もちろん、この判断には議論の余地はある。

言葉の出現頻度を測定する。この特定語の出現頻度の測定は、すべての年代でおこなう。以下、その意味カテゴリーとしての純粹リスクと投機的リスクという区分に属する言葉、および、経済性、信頼性、正当性という3つの視点という区分に属する言葉について解説する。

純粹リスクは、上述したように、それが実際に発生した時には損失のみをもたらす、利益を得る可能性のないリスクである。具体的には、事故や災害が発生することを示す。そのため、純粹リスクの意味カテゴリーに属する言葉としては、「事故」「災害」「地震」の3つの単語を選択する。投機的リスクは、損失をもたらすが、利益となることもあるリスクであり、株価や為替の変動、景気の好不況、戦争・動乱などの社会的変動、新製品開発等の企業活動に起因するリスクである。そのため、投機的リスクの意味カテゴリーに属する言葉として、「ベンチャー」「デリバティブ」「不良債権」「サブプライム」の4つの単語を選択する⁸。

経済性という視点では、リスクが現実化したときに発生する損失・損害だけでなく、リスクに対応する際に発生するコストを含めて、最も効率的に損失を最小限にする方法を選択するという考え方が強調される。そのため、経済性という意味カテゴリーに属する言葉として、「コスト」「損失」「損害」の3つの単語を選択する。信頼性という視点では、損害を出さないこと、つまり、業務の安全性や組織の信頼性を高めていくことが、リスクマネジメントの本質と考え、事故発生の防止・予防、また、安全というものを強調す

図表3：意味カテゴリーとキーワード

意味カテゴリー	キーワード
純粹リスク	事故、災害、地震
投機的リスク	ベンチャー、デリバティブ、不良債権、サブプライム
経済性	コスト、損失、損害
信頼性	発生、防止、予防、安全
正当性	法令順守、コーポレートガバナンス、ステークホルダー、CSR

る。そのため、信頼性という意味カテゴリーに属する言葉として、「発生」「防止」「予防」「安全」の4つの単語を選択する。正当性という視点では、どのようにリスクに対応するかについて、社会的な合意を得ることが重要となり、社会すなわち企業をとりまくステークホルダーからの期待や要請に応える形でリスクマネジメントを行っているかどうか焦点となる。そのため、正当性という意味カテゴリーに属する言葉として、「法令順守」「コーポレートガバナンス」「ステークホルダー」「CSR」の4つの単語を選択する。

以上のように、純粹リスクと投機的リスクという区分および経済性、信頼性、正当性に区分に属する言葉を選択し（図表3参照）、それぞれの年代のテキストデータにおけるその言葉の出現頻度を測定する。

3. 分析結果

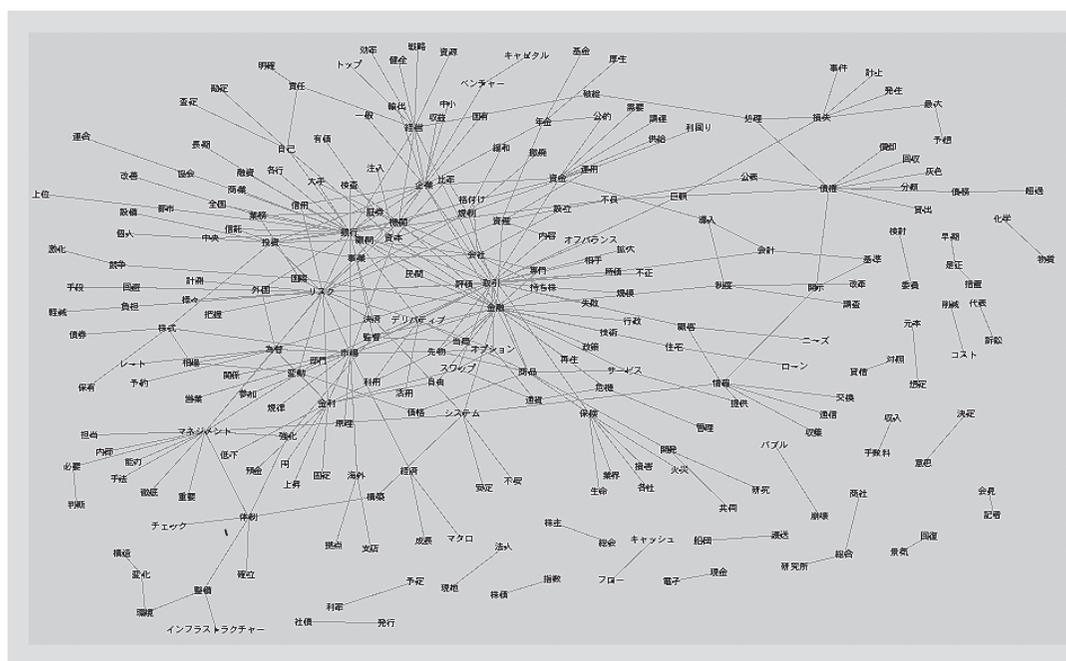
(1) 共起ネットワークの時系列的分析

ここでは、1994年～1998年（第I期）、1999年～2003年（第II期）、2004年～2008年（第III期）、2009年～2013年（第IV）の時期において、それぞれリスクという言葉に共起される言葉を抽出することによって、リスクがどのような意味で用いられているかを推察する。

1994年～1998年の第I期のテキストデータにおいて、リスクという言葉に共起して出現

⁸ 形態素解析の結果、「サブプライム」は「サブ+プライム」を2つの単語から構成されるが、ここでは、1つの単語として取り扱う。また、正当性のキーワードである「法令+順守」「コーポレート+ガバナンス」「ステーク+ホルダー」も同様の扱いである。

図表4：第I期の共起ネットワーク



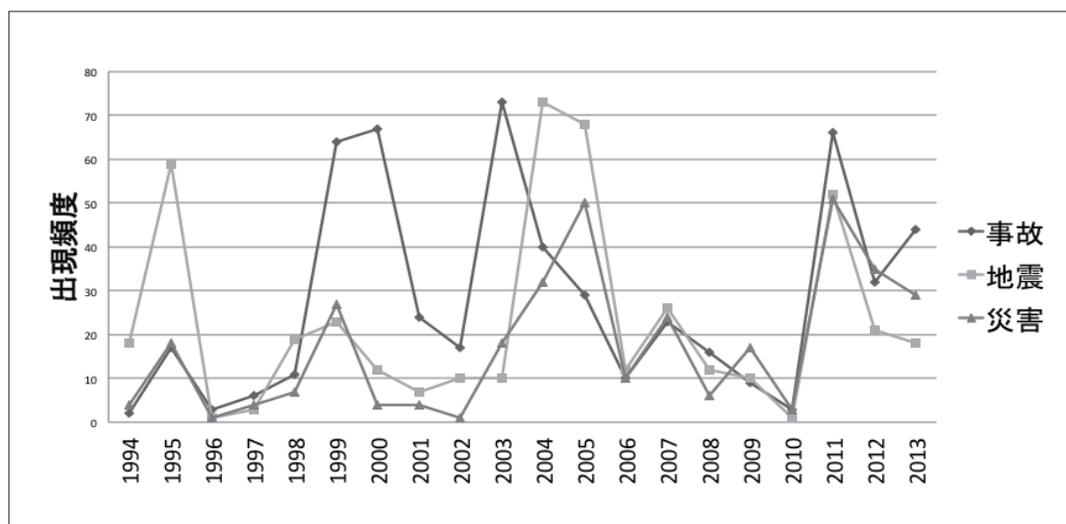
した言葉の連なりを图示したものが共起ネットワークである(図表4参照)。ネットワークを簡素化するために、共起頻度が20以上のものを選抜して图示してある(以下、第II期、第III、第IV期も同様)。第I期の共起ネットワークの特徴は、リスクという単語に共起される言葉として、金融系のものが多いことである。また、「デリバティブ+取引」「スワップ+取引」「先物+取引」「金利+上昇」「巨額+損失+事件」のように、投機的リスクとしての意味合いが強い言葉のつながりが見てとれる。この時期のリスクという言葉は、金融や投機のコンテキストで用いられていたと推察できる。また、「海外+拠点」「現地+法人」など、海外を意味する言葉のつながりも登場し、いわゆるカントリーリスクを想起させる。

1999年～2003年の第II期の共起ネットワークの特徴も、金融系の言葉が安定的に使用されていることである(図表5参照)。し

かし、金融や投機といった特定のコンテキスト以外の用語が占める割合も増えている。「同時+テロ」「内部+告発」「食品+安全」「法令+順守」などといった言葉のつながりである。これは、リスクという言葉が多様な意味として用いられるようになった表れと捉えることができる。また、ネットワークそのものも使用される言葉の種類と、そこから連なる分岐が増えていることも見てとれる。これまでは「リスク」という言葉として捉えられてなかった現象が、当時のさまざまなイベントの発生によって、リスクと捉えられるようになってきたことの表れといえる。

2004年～2008年の第III期の共起ネットワークでも、金融系の言葉が少なからず使用されていることがわかる(図表6参照)。しかしその一方で、第I期、第II期のような金融関連の言葉のネットワークの広がりには縮小している。また、新たに「サブプライム」といった言葉も登場しているのが

図表8：純粋リスクの出現頻度



	1994	1995	1996	1997	1998	1999	2000	2001	2002	2003	2004	2005	2006	2007	2008	2009	2010	2011	2012	2013
事故	2	17	3	6	11	64	67	24	17	73	40	29	10	23	16	9	3	66	32	44
地震	18	59	1	3	19	23	12	7	10	10	73	68	12	26	12	10	1	52	21	18
災害	4	18	1	4	7	27	4	4	1	18	32	50	10	24	6	17	3	51	35	29
リスク	730	887	586	377	599	576	563	494	384	507	516	555	350	663	578	463	317	367	390	550

て表現されるようになる一方で、東日本大震災のような未曾有の大災害は、リスクと表現してマネジメントの対象としなければならないという考えの表れだと推測できる。しなしながら、これは、共起頻度が20以上の言葉を抽出した簡素化された共起ネットワークを分析した結果であり、少ないながらも共起していること自体が重要であると考えらるならば、より詳細な共起ネットワークで分析することが求められる。そして、ここで示したことはあくまでも仮説の域を出ず、学術的な含意を持つためには、さらなる分析を必要とする。

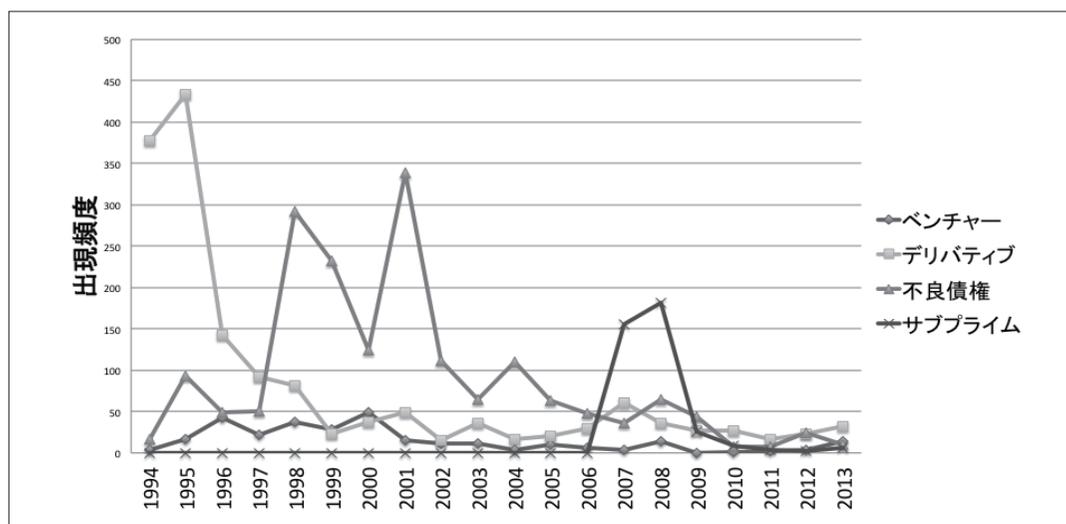
(2) 意味カテゴリーへの位置付け

次に、純粋リスクと投機的リスクという意味カテゴリーに属する言葉と、経済性、信頼性、正当性という意味カテゴリーに属する言葉が、テキストデータの中で、どれくらいの出現頻度なのかを年代ごとに測定する。

純粋リスクの意味カテゴリーでは、事故、地震、災害というキーワードの出現頻度を測定する(図表8参照)。これらのキーワードの出現頻度の絶対数は、それほど多くなく、出現頻度にばらつきがある。一見すると、出現頻度の傾向が見えにくいですが、それぞれの年に発生したリスクイベントと照らし合わせてみると、その出現頻度の傾向が見えてくる。1995年の阪神大震災、1999年のJCO臨界事故、2003年のコロンビア号爆発事故、2004年の中越地震、2011年の東日本大震災など、事故、地震、災害の言葉の出現頻度が高くなる年には、大規模な事故、地震、災害が発生しているのである。このように、純粋リスクは、現実に発生している事故、地震、災害を記述するときに、それを表現する言葉として用いられていることがわかる。

投機的リスクの意味カテゴリーでは、ベンチャー、デリバティブ、不良債権、サブプラ

図表 9：投機的リスクの出現頻度



	1994	1995	1996	1997	1998	1999	2000	2001	2002	2003	2004	2005	2006	2007	2008	2009	2010	2011	2012	2013
ベンチャー	4	17	42	22	38	28	49	15	11	12	4	10	6	3	14	0	1	2	4	14
デリバティブ	378	434	143	92	82	23	37	49	15	36	17	21	30	61	36	27	27	17	23	32
不良債権	16	93	49	51	292	232	124	339	111	65	110	64	48	36	65	44	7	8	25	10
サブプライム	0	0	0	0	0	0	0	0	0	0	0	0	0	156	182	26	9	3	2	6

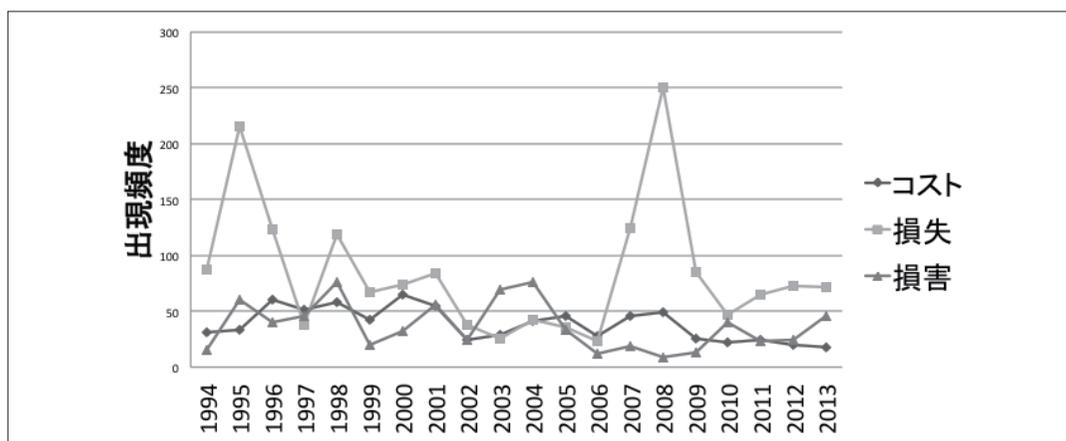
イムというキーワードの出現頻度を測定する（図表9参照）。これらのキーワードの出現頻度の絶対数は比較的多く、キーワードごとによって出現傾向は明確である。特に、デリバティブは、1994年、1995年での出現頻度が極めて高く、その後も、一定数の出現をみせている。不良債権問題は、バブル崩壊後、多くの金融機関が抱えるリスクであり、どの年代においても高い出現頻度を示している。また、サブプライムローン問題は、2007年までまったく問題にされていなかったが、それがリスクとして認識されるようになると、急激に出現を高めるといった傾向が見てとれる。

経済性という意味カテゴリーでは、コスト、損失、損害というキーワードの出現頻度を測定する（図表10参照）。これらのキーワードの出現頻度の傾向は、上述の純粹リスクと同様に、その年に発生したリスクイベントと照らし合わせてみると見えてくる。1995年の大

和銀行ニューヨーク支店巨額損失事件、ベアリングス銀行の破綻、2007年以降のサブプライムローン問題、2008年のリーマンショックなど、巨額な損失を想起させる事象とともに出現頻度は高くなる。そして、その多くが金融や投機と関連するものであるため、その出現頻度の傾向は、上述の投機的リスクのキーワードの出現頻度と類似する。

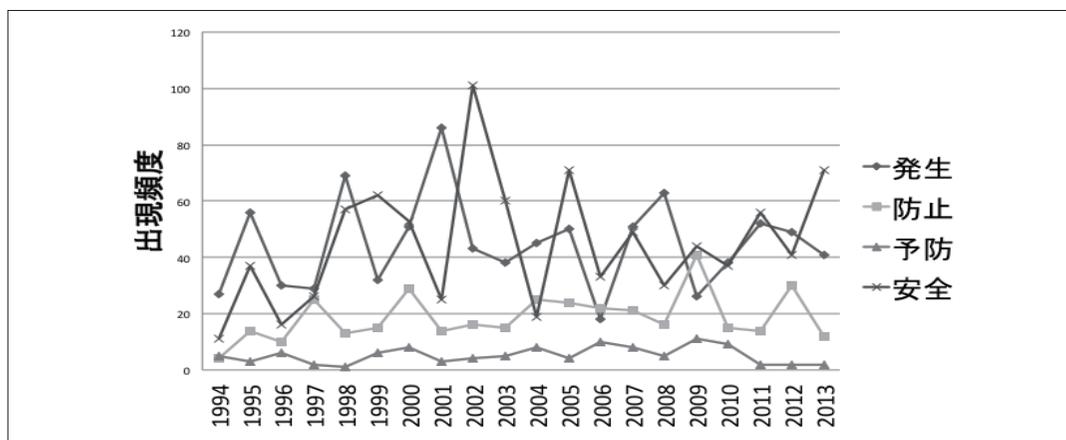
信頼性の意味カテゴリーでは、発生、防止、予防、安全というキーワードの出現頻度を測定する（図表11参照）。これらのキーワードの出現頻度の絶対数はそれほど多くはないが、どの年代でも一定の出現を見せている。これは、選択したキーワードがいずれも個別事象を意味するものではなく、抽象的な行動や現象を示すものであるため、使用される範囲が広いと考えられる。また、キーワードごとに出現傾向が類似しており、この4つのキーワードの相互関連性が高いことを示し

図表10：経済性の出現頻度



	1994	1995	1996	1997	1998	1999	2000	2001	2002	2003	2004	2005	2006	2007	2008	2009	2010	2011	2012	2013
コスト	31	34	60	51	58	42	65	55	25	29	41	46	28	46	49	26	22	24	20	18
損失	88	216	123	38	119	67	74	84	38	26	43	36	23	125	251	85	47	65	73	72
損害	15	61	40	46	76	20	32	56	24	70	76	33	12	19	9	13	40	23	24	46

図表11：信頼性の出現頻度



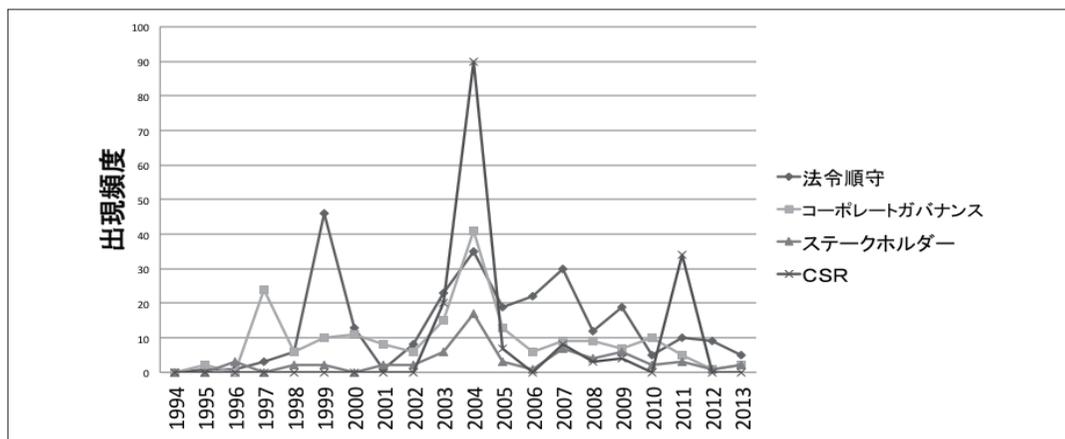
	1994	1995	1996	1997	1998	1999	2000	2001	2002	2003	2004	2005	2006	2007	2008	2009	2010	2011	2012	2013
発生	27	56	30	29	69	32	51	86	43	38	45	50	18	51	63	26	38	52	49	41
防止	4	14	10	25	13	15	29	14	16	15	25	24	22	21	16	41	15	14	30	12
予防	5	3	6	2	1	6	8	3	4	5	8	4	10	8	5	11	9	2	2	2
安全	11	37	16	26	57	62	53	25	101	60	19	71	33	49	30	44	37	56	41	71

ている。

正当性の意味カテゴリーでは、法令順守、コーポレートガバナンス、ステークホルダー、CSRというキーワードの出現頻度を測定する(図表12参照)。これらのキーワードの出現頻度の絶対数は多くない。特徴的なのは、それ

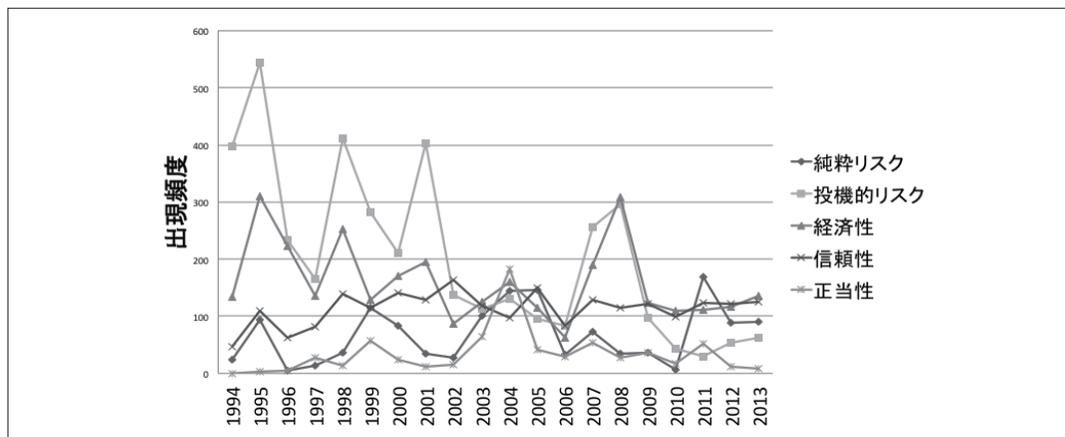
ぞれのキーワードの出現頻度が極端に高くなる年代が存在するということである。特に、CSR元年と言われた2003年に、CSRというキーワードの出現頻度が極端に高くなり、それ以降も少しずつ登場している。これは、リスクマネジメントが、CSRという言葉とともに

図表12：正当性の出現頻度



	1994	1995	1996	1997	1998	1999	2000	2001	2002	2003	2004	2005	2006	2007	2008	2009	2010	2011	2012	2013
法令順守	0	1	1	3	6	46	13	1	8	23	35	19	22	30	12	19	5	10	9	5
コーポレートガバナンス	0	0	2	0	24	6	10	11	8	6	15	41	13	6	9	7	10	5	1	2
ステークホルダー	0	0	3	0	2	2	2	0	2	6	17	3	1	7	4	6	2	3	1	2
C S R	0	0	0	0	0	0	0	0	0	20	90	7	0	8	3	4	0	34	0	0

図表13：すべての意味カテゴリーの出現頻度



	1994	1995	1996	1997	1998	1999	2000	2001	2002	2003	2004	2005	2006	2007	2008	2009	2010	2011	2012	2013
純粹リスク	24	94	5	13	37	114	83	35	28	101	145	147	32	73	34	36	7	169	88	91
投機的リスク	398	544	234	165	412	283	210	403	137	113	131	95	84	256	297	97	44	30	54	62
経済性	134	311	223	135	253	129	171	195	87	125	160	115	63	190	309	124	109	112	117	136
信頼性	47	110	62	82	140	115	141	128	164	118	97	149	83	129	114	122	99	124	122	126
正当性	0	3	4	27	14	58	24	11	16	64	183	42	29	54	28	36	17	52	11	9

に語られるようになってきたことを示しているといえる。そして、他のキーワードもCSRとは深い関わりのある言葉なので、2003年、2004年を中心に出現頻度が高くなっている。

以上、純粹リスクと投機的リスクという意

味カテゴリーに属する言葉と、経済性、信頼性、正当性という意味カテゴリーに属する言葉が、テキストデータの中で、どれくらいの出現頻度なのかを年代ごとに測定してきたが、最後に、それぞれの意味カテゴリーのキー

ワードの出現頻度の合計を1つにまとめて比較をする(図表13)。出現頻度の合計を比較すると、以下の特徴が見えてくる。

第一に、投機的リスクと経済性は相関しているということである。投機的リスクは、損失をもたらすが利益となることもあるリスクであり、その一方で、経済性という視点は、もっとも効率的に損失を最小限にする方法を選択するという考え方が強調される。そのため、この2つの意味カテゴリーには共通性があるともいえる。第二に、純粹リスクと信頼性には相関がないということである。純粹リスクは、それが実際に発生した時には損失のみをもたらす利益を得る可能性のないリスクであるため、それを未然に防止するという信頼性の考え方と親和性が高いように思える。しかし、純粹リスクの出現頻度が低い年代でも、信頼性の意味カテゴリーに属するキーワードは、一定の出現を示している。

4. 考察

ここで、もう一度、本研究においてテキストマイニングという分析手法を用いる理由について振り返ってみる。それは、以下の仮説を検証するためであった。1)組織や社会は、新聞記事などに上げられるリスクイベント(企業事故、災害、不祥事など)が発生すると、それをマネジメントすべきリスクだと認識する、2)さまざまな種類のイベントが発生すると、その結果、マネジメントすべきリスクが、特定のコンテキストだけでなく、多様なコンテキストにおいて注目されるようになる、という仮説である。

上述したように、時系列的な共起ネットワークの分析および意味カテゴリーへの位置付けをおこなった結果、これらの仮説について、ある程度確認することができたと考えられる。テキストデータが新聞記事ということ

もあり、当然ではあるが、社会的関心の高いリスク現象が取り上げられている。そのため、共起ネットワークの分析および意味カテゴリーへの位置付けをおこなう中で、その年代で発生したリスク現象と照らし合わせてみると、その年代のリスク現象と関連性の高い言葉が用いられていることが、共起ネットワークおよび出現頻度というデータで確認できた。また、時系列的な共起ネットワークの分析から、第I期・第II期のような金融・投機のコンテキストで使用されていたリスクという言葉が、第III期・第IV期になるにつれて、そのコンテキストの中だけで用いられるなくなったことを確認した。これは、さまざまな種類のイベントが発生すると、特定のコンテキストだけでなく、多様なコンテキストにおいて用いられるようになるという仮説の一部を確認できたと考えられる。

そして、テキストマイニングの結果、以下のように考察を広げることも可能となる。それは、確かに時代とともに、特定のコンテキストだけではなく、多様なコンテキストでリスクをとらえることになるが、投機的リスクのようにどの年代でも一定の出現頻度を有するコンテキストも存在することである。つまり、根強いリスクマネジメントのコンテキストが存在するということである。そして、リスクは、リスクイベントの発生によって認識されるということは、逆にいえば、イベントが発生しないと沈静化する可能性があるということでもある。多様なイベントが発生すると、リスクのコンテキストは多様化するが、そのコンテキストに当てはまるリスクが発生しないと、そのコンテキストでリスクは語られなくなるという可能性が存在するということである。

5. まとめ

本研究は、リスクという概念がどのような意味・コンテキストで用いられているのかを、20年間の新聞記事をテキストデータとしたテキストマイニングという手法を用いて分析した。特に、リスクというキーワードに共起される言葉のネットワークを時系列的に描き、それを比較検討した。また、「純粹リスク・投機的リスク」「経済性・信頼性・正当性」という意味カテゴリーを用いて、それぞれのカテゴリーに属する言葉の出現頻度を時系列的に分析した。その結果、特定のカテゴリーのリスクは、特定のイベントの発生によって認識される可能性を確認するとともに、年代によっては、多様なコンテキストにおいて用いられるようになることを確認した。

最後に、研究上の課題を示したい。第一に、本研究におけるテキストマイニングの稚拙さである。一定の手続きに則ってテキストマイニングをおこなっているものの、その方法が最適かどうかは議論の余地がおおいにある。また、テキストマイニングに用いるツールも多様に存在しているため、どれを用いるのかについても考慮すべき点である。第二に、リスクイベントの明確化と、その分類の精緻化である。新聞記事に掲載されるようなリスクイベントは社会的関心も高く、テキストデータだけでなく、現実をも考察の射程に入れるには適している。しかし、同じイベントであっても、地震や大規模事故のように突発的に発生するイベントもあれば、法改正のように時間をかけてそれが認識されるイベントも存在する。この区別は、今後しなければならない課題である。第三に、カテゴリー間の関係性の検討である。本研究においても、投機的リスクと経済性という視点の関連性については一定の考察を加えることができたが、あ

くまで試論的な検討であった。詳細に突き詰める必要がある。また、経済性、信頼性、正当性という視点については、別個に議論されるのではなく、相互に関連した関係にあると考えられるため、視点間の関係性を検討する必要もでてくる。第四に、テキストデータの妥当性である。今回は、日本経済新聞という媒体を用いた。日本経済新聞は、社会の視点というよりは、企業でも特に金融機関の視点が強い。そのため、投機的リスクや経済性のコンテキストでリスクという言葉が用いられている結果が導かれた。もう少し社会全般において認識されるリスク概念を明らかにするのであれば、一般紙からテキストデータを抽出する必要もでてくる。

付記

本研究は、JSPS科研費26380485の助成を受けたものである。

参考文献

- 林俊克 (2002) 『Excelで学ぶテキストマイニング入門』 オーム社。
- 正田真也・萩原克幸・鶴岡信治 (2012) 「組織研究におけるテキストマイニングを用いた系統的分析法」『日本情報経営学会誌』 Vol.32, No.3, pp.97-109。
- 石田基広 (2008) 『Rによるテキストマイニング入門』 森北出版。
- 喜田昌樹 (2006) 「アサヒの組織革新の認知的研究—有価証券報告書のテキストマイニング—」『組織科学』 Vol.39, No.4, pp.79-92。
- 小室達章 (2013) 「リスクマネジメント研究における経済性、信頼性、正当性」(日本経営学会第87回大会自由論題)。
- 久保淳司 (2008) 「アメリカ会計基準におけるリスク概念—内容分析によるコンテキストの検討—」『会計』 Vol.174, No.2, pp.228-243。
- 内藤勲 (2011a) 「企業倫理を巡る意味ネットワーク分析」『経営学研究』 Vol.20, No.4, pp.283-296。

- 内藤勲 (2011b) 「テキストマイニングを巡る問題
点—古民家概念の意味ネットワークの分析に寄
せて—」『経営学研究』Vol.21, No.1, pp63-84。
- 吉田慎一郎・中藤哲也・御手洗秀一・廣川左千男
(2012) 「利益伸び率に着目した有価証券報告書
のテキストマイニング」『情報処理学会研究報
告』 pp.1-5。
- 上田太郎 (2008) 『事例で学ぶテキストマイニ
ング』共立出版。