

〈7〉 アンケート調査における自由回答記述内容の構造化と活用手法に関する研究－EBPMに資する記述内容の分析－

市政研究センター 専門研究嘱託員 坪井 壺太郎

要旨 本研究は、これまで主として定量データによる解釈を補完するものとされてきたアンケート調査の自由回答記述内容を形態素（意味を持つ最小の言語単位）に分類したうえでその構造を明らかにし、EBPMに資する計量テキスト分析の手法を提示するものである。分析にあたってはパブリックコメント「第6次宇都宮市総合計画（素案）」を対象として、意見が「ハード-ソフト」および「理念・理想-具体」の意味をもつ空間で構成されることを示した。本手法が調査担当者の熟練度に依拠せず客観的に分析できることから、アンケート調査の自由回答や議事録等の分析においても応用可能な手法であると考えられる。

キーワード：アンケート調査、自由回答記述、形態素、品詞、計量テキスト分析

1 はじめに

(1) 研究の背景と目的

近年、特に企業ソリューションの現場において、アンケートの自由回答やコールセンター、ウェブサイトなどに寄せられるユーザーからの意見や書き込み等の定性的な文字情報を解析し、自社製品へのニーズや不満点等を分析する「テキストマイニング」手法が着目されつつある（那須野2001）。テキストマイニングとは、定型化されていない文章の集まりを自然言語解析の手法を援用して単語やフレーズに分割し、それらの出現頻度や相関関係进行分析して情報を抽出するテキストデータ分析手法やシステムの総称である。

テキストマイニングでは、定型化されていないデータから知識・情報を見つけ出すことに特徴をもち、その利点は、その発話者・発信者・書き手のもつ思考世界への接近が期待できることにある。

一般に、質問紙（アンケート）を用いた調査の多くは、あらかじめ設定された尺度（選択肢回答法）への回答をもとにデータの解析が行われる場合が多く、これまで「自由回答」は、そのデータ解釈を補完するものとして位置付けられてきた。そのため、自由回答が分析対象そのものになることは少なかったことが指摘できる。

一方、ワークショップ等において広く実践されてきている「KJ法」（川喜田1967）は、あるテーマに対する意見等を付箋等へ書き出し、類似した内容をグループ化することにより、最終的に複数の意見を構造化、明確化する手法である。KJ法における作業の重要な点は「直感」であり、配置やグループ化に際してはあらかじめ仮定した理論に基づくものではなく、情報に対する直感的な判断が重視される。そのため、調査担当者の経験や分類された項目に対する理解度などが影響する限界点も一定程度存在するほか、集約された意見の方向性や階層性を示すことは可能ではあっても、構成された集塊の「意見」同士がもつ意味の構造を論理的に解釈することは必ずしも容易ではない。

本研究では、上述の課題に着目し、得られた自由回答記述内容（テキスト）を可能な限り客観的に解析するために、日本語を形態素（意味を持つ最小の言語単位）に分類した上で、「自由回答」のもつ意味の構成や構造を明らかにする。

日本語を形態素に分類するための手法は、単語ごとに「分かち書き」をする英語に比べて、日本語文法が有する特殊性から、文章を単語ごとに切り分けるための処理技術（形態素解析）の遅れが指摘されてきた。しかし、近年では言語解析プロ

グラム技術の進化に伴い、岩見ほか（2011）による行政・政策分野での事例研究等も徐々に蓄積されてきており、本研究には一定の意義があるものと考えられる。

宇都宮市（以下「本市」という）においても、計画策定の際などに各部局でアンケート調査が実施されている。しかし、既往の調査では自由回答記述内容を主体的に分析する取組は充分には行われておらず、政策等に活かしていくまでには至っていない。

本研究では、自治体が実施する質問紙（アンケート）調査票における自由回答の記述内容について客観的かつ計量的に分析を試みることで、論拠に基づく政策立案（EBPM）に資する方法論を検討し、行政における活用手法を提案することを目的とする。

本研究は、調査対象者の「自由回答記述（テキスト）」を主体的に扱うことを主眼に置き、定型化されていない文章の集まりを自然言語解析の手法を用いて単語やフレーズの形態素に分解したうえで、集合的意味を考究（マイニング）し、あわせて、分析にあたり統計学的解析を援用した「計量テキスト分析」を行う。本研究による論及方法の特徴は、記述された用語の傾向から課題への接近を図る「質的性格」と、文法と出現法則に基づいた用語間の集塊性から課題への接近を図る「量的性格」をあわせ持つことにある。

(2) 研究対象

本研究では、本市において、既に公表されているテキストデータであり、一定の文章量を持つ「パブリックコメント」を事例に、市民からの「意見提出」の分類に際して客観的に自由回答内容を分析するための手法を提示する。

パブリックコメントは、行政の施策・計画などを策定していく中で、その素案を公表し、広く市民に意見や情報を求め、提出された意見や要望等

を考慮して決定していくものであり、本市においてもこれが制度化されている。本制度は、政策決定過程における公正性の確保と透明性の向上を図り、市民参加による開かれたまちづくりの実現をめざす目的を有する。

表1に、平成26年度から平成30年度におけるパブリックコメントの件数とその概要を示す。本市では、応募者の意見はあらかじめ意見別に分類され「意見数」として集計が行われている。

本研究では、平成29年度に実施された本市の行政施策の全般に係る内容が含まれる「第6次宇都宮市総合計画（素案）」に関するパブリックコメント（平成29年12月8日～平成30年1月8日実施）を事例に分析を行った。同パブリックコメントの応募者は、8者（男性3名、女性4名、法人団体1者）であり、35件の意見の内訳は「全体（5件）、序論（2件）、基本構想（10件）、基本計画（18件）」であった。

表1 年度別パブリックコメント件数と概要

		2018 (平成30) 年度	2017 (平成29) 年度	2016 (平成28) 年度	2015 (平成27) 年度	2014 (平成26) 年度
基本項目	件数	8	11	8	11	14
	応募者数	23	45	18	35	46
	意見数	46	113	34	107	110
意見提出方法	郵送	0	3	2	6	7
	FAX	6	14	5	9	22
	E-mail	11	14	6	11	4
	持参	6	13	5	9	13
	電話	0	5	0	0	0
市の対応区分	A（意見反映）	2	11	3	15	3
	B（反映済み）	18	54	18	40	32
	C（計画参考）	7	12	1	25	24
	D（反映なし）	2	6	0	1	4
	E（要望意見）	18	35	12	26	49

本市ホームページ¹から作成

¹ 宇都宮市「パブリックコメント」、
<https://www.city.utsunomiya.tochigi.jp/shisei/koho/publiccomment/index.html>, 2020年1月31日取得

2 研究方法

形態素を対象とする形態論研究 (morphology) は、言語学分野における研究対象の1つであり、単語や品詞といった言語が意味を持つ最小の単位の語形変化や並び方などの考察、理論から発達してきたものである。この理論枠組みを引き継ぎ、日本語の品詞間の接合規則性を分類するための解析プログラム「Juman」, 「Chasen」, 「KAKASI」, 「すもも」, 「Breakfast」, 「Sudachi」等が開発されている。本研究では、基礎的な記述内容の構造解析に重点を置くと同時に、品詞の細分類化に利点を持つ「Mecab」²によるプログラムを用いて解析を行った。

日本語形態素は、品詞とその前後の接続関係をもとに分類され、分析においてどの品詞を採用するかについては分析者の判断に依拠しており、全用語を含む形態素を投入したうえで探索的に品詞を決定する方法や、あらかじめ投入品詞を設定する方法がある。

本研究では後者の方法を採用し、先行研究 (林 2003) での知見をふまえ、名詞と形容詞を中心に、行動や意思や要望を指す「-する」, 「-したい」等のサ行変格活用動詞に結節する「名詞-サ変接続」のほか、「名詞-形容動詞語幹」, 「名詞-一般」, 「名詞-複合名詞」および、「形容詞-自立」「形容詞-接尾」「形容詞-非自立」の7品詞を分析に採用した (表2)。解析の結果、469用語、343形態素が抽出された。分析にあたっては、あらかじめ表記のゆれの確認を行ったうえで、出現頻度により意味空間が決定されるという形態素分析の前提に基づき、出現回数2回以上の35用語を抽出した (表3)。

表2 形態素の種類と本分析における採否

品詞	品詞細分類	採否	品詞	品詞細分類	採否
連体詞	—	△	動詞	自立 接尾 非自立	△ × ×
接頭詞	形容詞接続 数接続 動詞接続 名詞接続	× × × ×	形容詞	特殊 自立 接尾 非自立	× ◎ ◎ ◎
名詞	サ変接続 ナイ形容詞語幹 形容動詞語幹 動詞非自立的 副詞可能 一般 数 接続詞的 固有名詞 接尾 代名詞 非自立 特殊 複合名詞	◎ △ ◎ × × ◎ × × △ × × × × ◎	副詞	一般 助詞接続類	△ △
			助詞	格助詞 係助詞 終助詞 接続助詞 副詞化 副助詞 並立助詞	× × × × × × ×
			助動詞	—	×
			感動詞	—	×
未知語	—	×	記号	—	×

注) ◎=採用, △=検討 (本研究では不採用), ×=不採用
筆者作成

表3 本研究での採用用語 (品詞別) と出現頻度

用語	品詞	品詞細分類	全体	序論	基本構想	基本計画	合計
分野 別 計画	複合名詞	複合名詞	0	0	0	13	13
追加	名詞	サ変接続	0	0	0	10	10
施策	名詞	サ変接続	0	0	0	7	7
構成 事業	複合名詞	複合名詞	0	0	0	6	6
整備	名詞	サ変接続	0	0	2	4	6
市民	名詞	一般	0	0	1	4	5
推進	名詞	サ変接続	0	0	1	3	4
人	名詞	一般	0	0	3	0	3
不便	名詞	形容動詞語幹	0	0	1	2	3
必要	名詞	形容動詞語幹	1	1	0	1	3
状況	名詞	一般	1	0	1	1	3
生活	名詞	サ変接続	0	0	3	0	3
宇都宮 市	複合名詞	複合名詞	0	0	1	2	3
表記	名詞	サ変接続	1	0	0	2	3
期待	名詞	サ変接続	0	1	1	1	3
姿	名詞	一般	0	0	1	2	3
計画	名詞	サ変接続	3	0	0	0	3
現状	名詞	一般	0	1	0	1	2
文言	名詞	一般	0	0	0	2	2
移動	名詞	サ変接続	0	0	0	2	2
検討	名詞	サ変接続	0	0	1	1	2
削除	名詞	サ変接続	1	0	1	0	2
活動	名詞	サ変接続	0	0	1	1	2
向上	名詞	サ変接続	0	0	1	1	2
行政	名詞	一般	0	0	1	1	2
将来	名詞	副詞可能	0	0	1	1	2
公共 交通	複合名詞	複合名詞	0	0	0	2	2
まち づくり	複合名詞	複合名詞	0	0	2	0	2
交通	名詞	一般	0	0	1	1	2
変化	名詞	サ変接続	1	1	0	0	2
環境 保全 行動	複合名詞	複合名詞	0	0	0	2	2
交通 インフラ	複合名詞	複合名詞	0	0	2	0	2
重要	名詞	形容動詞語幹	0	0	1	1	2
提唱	名詞	サ変接続	0	0	2	0	2
目標 年次	複合名詞	複合名詞	2	0	0	0	2

筆者作成

² 工藤拓「MeCab: Yet Another Part-of-Speech and Morphological Analyzer」, <http://taku910.github.io/mecab/>, 2020年1月31日取得

3 分析結果と解釈

(1) 出現用語の意味の空間構造

既往の自然言語処理・解析研究においては、文章中に同時に出現する用語を、ベクトル空間において依存関係を明示的にモデル化するLSI分析が知られている。本研究では、この理論枠組みのもと、用語の出現の有無により行列を作成し、数量化理論Ⅲ類による解析を行った。本手法は、外的基準を持たないカテゴリ変数行列における変数相互間の関連から、データ間の類似性を明示することができ、固有値と相関比より適合基準を判断することが可能である。分析で得られる固有値は、個体の数量とカテゴリ数量との相関の二乗あるいは個体間のカテゴリ数量の相関比を表しているため、0から1の間をとる固有値の大きさにより、各次元でどの程度個体とカテゴリの関係が説明されているかを評価することが可能である。本分析の結果、第1軸の固有値(0.9059)・相関比(0.9518)・寄与率10.45%および第2軸の固有値(0.8837)・相関比(0.9401)・寄与率10.20%が示されたことから、高いモデル適合度が得られた(図1)。

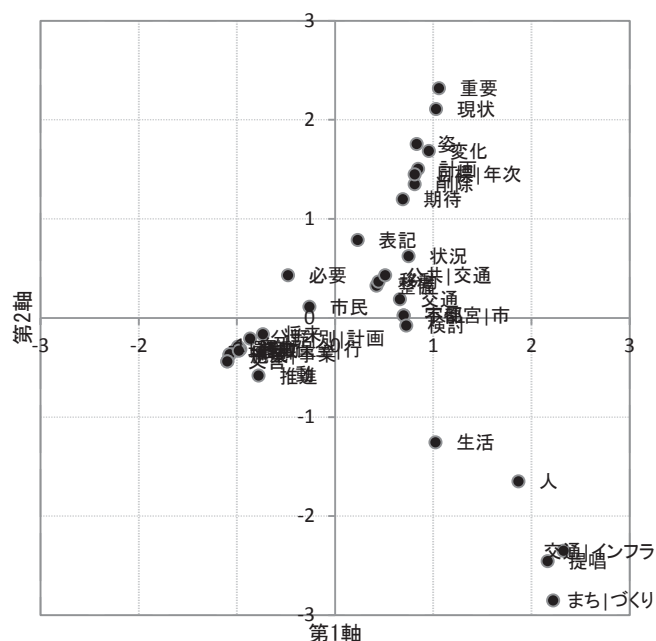


図1 カテゴリスコアの布置図

筆者作成

表4 第1軸カテゴリスコア (ハードーソフト)

用語	スコア	用語	スコア
環境 保全 行動	-1.096	期待	0.690
文言	-1.096	不便	0.700
施策	-1.075	宇都宮 市	0.700
構成 事業	-1.073	検討	0.728
追加	-1.013	状況	0.751
向上	-0.987	削除	0.812
行政	-0.978	目標 年次	0.812
活動	-0.964	姿	0.832
分野 別 計画	-0.865	計画	0.846
推進	-0.780	変化	0.957
将来	-0.736	生活	1.024
必要	-0.478	現状	1.030
市民	-0.261	重要	1.059
表記	0.231	人	1.868
整備	0.423	提唱	2.167
移動	0.443	まち づくり	2.223
公共 交通	0.510	交通 インフラ	2.329
交通	0.661		

筆者作成

表5 第2軸カテゴリスコア (理念・理想—具体)

用語	スコア	用語	スコア
まち づくり	-2.852	不便	0.025
提唱	-2.456	市民	0.112
交通 インフラ	-2.350	交通	0.188
人	-1.652	整備	0.324
生活	-1.258	移動	0.366
推進	-0.584	公共 交通	0.429
文言	-0.441	必要	0.430
環境 保全 行動	-0.441	状況	0.621
施策	-0.376	表記	0.786
構成 事業	-0.366	期待	1.199
行政	-0.331	削除	1.349
追加	-0.329	目標 年次	1.449
活動	-0.318	計画	1.504
向上	-0.290	変化	1.685
分野 別 計画	-0.208	姿	1.755
将来	-0.166	現状	2.107
検討	-0.076	重要	2.318
宇都宮 市	0.025		

筆者作成

本分析の結果、相関適合度の高い2軸を抽出し、これを採用した(表4, 5)。各々の軸の解釈にあたっては、採用した単語の出現頻度から算出したカテゴリスコアの数値を対象としてクラスター分析を補完的に行い用語の集塊性を考慮して実施した。

第1軸では、正(+)側に「まちづくり」、 「交通インフラ」のほか、不満を意味する用語の

「不便」にかかる原文では、「市民の移動はとても不便になると考える」、「車椅子で生活をしている人にとっては不便を感じる」など、ハード整備の内容に対する意見があげられている。一方、負（－）側には、「環境保全行動」のほか「施策」、「活動」など、ソフトの取組に関する用語があげられている。このことから、第1軸を「ハード－ソフト」軸であると解釈する。第2軸では、正（＋）側に「計画」「重要」等の理念に関する用語が、負（－）側に、「人」（人と自転車と自動車の分離）や「生活」（生活基盤である日照権の確保）などの具体的な施策内容を含む用語が含まれることから、「理念・理想－具体」の軸であると解釈する。

以上よりパブリックコメントの意見は、両軸の持つ意味の中で構成されることが示された。

(2) 対応分析による属性と出現用語の可視化

本研究で取り上げた「第6次宇都宮市総合計画（素案）」に関するパブリックコメントに寄せられた意見を、あらかじめ分類された『全体』、『序論』、『基本構想』、『基本計画』の4カテゴリに基づき、出現用語との近接性をもとにその特性を検討する。分析にあたっては、4カテゴリと採用した35用語のクロス集計表を用い、行要素と列要素の相関関係が最大になるよう数量化を行った。

本分析の特徴は、項目間の類似性と同時に、カテゴリの近くに布置されることで、集塊性を二次元グラフで可視化できるほか、偏りの小さい項目（全体において共通している項目）は原点付近に、偏りの大きい項目（特徴的な項目）は原点から遠くに布置される特徴を持つ。本分析の結果、『基本構想』には、「まちづくり」、「交通インフラ」、「人」、「生活」等の理念や理想に関する用語が、『基本計画』には、具体的な施策に関する要望を意味する「追加」（－に追加してほしい）の用語

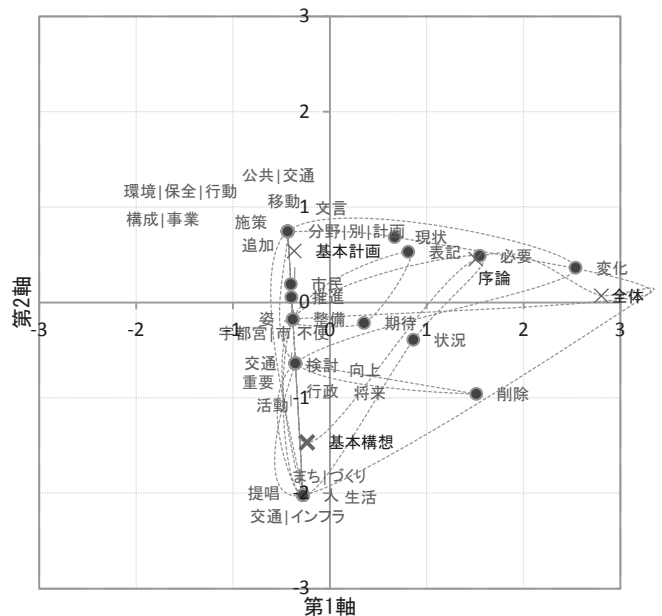


図2 分類カテゴリと用語の対応

注) 分類カテゴリは図中の「全体」、「序論」、「基本構想」、「基本計画」の4カテゴリを指す。

筆者作成

が近接している。また、『全体』に近接する「変化」の用語は、「市の現状や時代潮流の変化と展望にも各種学校や企業への外国人の転入が期待できることを記載していただきたい」や、「時代の変化に即さない状況が多々出てくると感じる」等、時代の変化への対応と要請を含むものとなっている。

このように、対応分析では、各属性と出現用語の近接性を明示することができることから、アンケート調査の自由回答記述内容を、属性（年齢・性別・地域等）で分析を行うことが可能であり、本市における各種調査の記述内容の分析においても意見を構造的に可視化できるため、課題を共有したうえでの議論展開に向けた有用性のある技術であると考えられる。

(3) 係り受け分析による出現用語の検討

形態素分析では、文章の分かち書きから「単語」を対象として検討が行われるが、語と語の修辞関係や主従関係を明示することができない。例えば「必要」の用語に係る単語が「ない」と「あ

る」では意味が異なることがあげられる。そこで、本節では、文法規則によって、文の構造を「句」を単位として検討する係り受け分析を行う。

句とは2つ以上の語が集まって1つの品詞と同様の働きをしながら文を構成する語の集塊を指し、名詞句・形容詞句・副詞句・動詞句がある。日本語は語順が比較的自由的な言語であるが、文節内の構造および語順には強い制限を有している。そのため、日本語は文節の連続として解析するためのさまざまな解析手法が開発されてきており、本研究ではこのなかで代表的なシステムCaboCha³を用いて解析を行った。分析にあたっては、名詞句を対象とし、出現回数の多い順に「追加」、「施策」、「推進」に係る用語を抽出した(表6)。

本表より、これらの用語はいずれも『基本計画』の中で出現する割合が高く、「追加」に係る用語では、「構成事業」、「分野別計画」のほか、「－という文言を追加してほしい」とする具体的な要望が含まれていることがわかる。

このことから、単語単位の形態素と組み合わせることで、より文章内容の全体構造や特性を把握することが可能となる。

4 分析結果と解釈

本研究で行った自由回答の形態素に基づく計量テキスト分析では、分析担当者の主観に依拠せず、客観的に解析が可能である点において利点をもち行政における各種アンケート調査の自由回答記述の分析において、有効な手法となることが示唆された。しかし、本手法ではさまざまな意見を内包している自由回答に対し、統計学的な作業上、出現頻度の少ない用語を排除して集約をしているため、「少数ではあっても重要な意見」を捨象し

³ 工藤拓「CaboCha/南瓜: Yet Another Japanese Dependency Structure Analyzer」, <https://taku910.github.io/cabocha/>, 2020年1月31日取得

表6 係り受け分析と出現用語数

	係り受け	品詞	品詞細分類	全体	序論	基本構想	基本計画	合計
追加	構成事業+追加する	名詞+名詞	サ変接続+サ変接続	0	0	0	5	5
	分野別計画+追加する	名詞+名詞	一般+サ変接続	0	0	0	4	4
	文言+追加する	名詞+名詞	一般+サ変接続	0	0	0	2	2
	正規化+追加する	名詞+名詞	一般+サ変接続	0	0	0	1	1
	全面整備+追加する	名詞+名詞	一般+サ変接続	0	0	0	1	1
	共同対応+追加する	名詞+名詞	サ変接続+サ変接続	0	0	0	1	1
	雪害防止+追加する	名詞+名詞	一般+サ変接続	0	0	0	1	1
	設置+追加する	名詞+名詞	サ変接続+サ変接続	0	0	0	1	1
	農業危害防止+追加する	名詞+名詞	一般+サ変接続	0	0	0	1	1
	育成+追加する	名詞+名詞	サ変接続+サ変接続	0	0	0	1	1
施策	整備+追加する	名詞+名詞	サ変接続+サ変接続	0	0	0	1	1
	目標+追加する	名詞+名詞	一般+サ変接続	0	0	0	1	1
	施策+構成事業	名詞+名詞	サ変接続+サ変接続	0	0	0	6	6
	推進+施策	名詞+名詞	サ変接続+サ変接続	0	0	0	3	3
	行政経営システム確立+施策	名詞+名詞	一般+サ変接続	0	0	0	1	1
	施策+目標	名詞+名詞	サ変接続+一般	0	0	0	1	1
	参画促進+施策	名詞+名詞	サ変接続+サ変接続	0	0	0	1	1
	確保+施策	名詞+名詞	サ変接続+サ変接続	0	0	0	1	1
	向上+施策	名詞+名詞	サ変接続+サ変接続	0	0	0	1	1
	分野別計画+推進	名詞+名詞	一般+サ変接続	0	0	0	3	3
推進	推進+施策	名詞+名詞	サ変接続+サ変接続	0	0	0	3	3
	環境保全行動+推進	名詞+名詞	一般+サ変接続	0	0	0	2	2
	免震高層化等+推進する	名詞+名詞	一般+サ変接続	0	0	1	0	1
	推進+文言	名詞+名詞	サ変接続+一般	0	0	0	1	1
	推進する+図る	名詞+動詞	サ変接続+自立	0	0	1	0	1
	宇都宮市グリーン調達推進方針+推進	名詞+名詞	固有名詞+サ変接続	0	0	0	1	1
	活動+推進	名詞+名詞	サ変接続+サ変接続	0	0	0	1	1
	地域別総合的開発+推進する	名詞+名詞	一般+サ変接続	0	0	1	0	1

筆者作成

てしまう危険性を内包している。そのため実際の解釈や政策決定への本手法の援用に際しては、原文を参照することで慎重に解釈を行っていくことに留意する必要がある。

参考文献

- 那須野哲哉, 2001, 「コールセンターにおけるテキストマイニング」, 『人工知能学会誌』16 (2), 219-225
 川喜田二郎, 1967, 『発想法』中公新書
 岩見麻子・大野智彦・木村道徳, 2011, 「公共事業計画策定過程の議事録分析のための変動係数を用いた対象語選定手法の開発」, 『環境情報科学』25, 55-60
 林俊克, 2003, 「テキストマイニングの現在」, 『マーケティングリサーチ』94, 16-25