

機関番号：22604

研究種目：若手研究(B)

研究期間：2009～2010

課題番号：21792322

研究課題名(和文) 生体肝移植ドナーの移植医療に対する評価のテキストマイニング導入による検討

研究課題名(英文) Analysis of evaluations of living donors for liver transplantation utilizing text mining approach

研究代表者

清水 準一 (SHIMIZU JUNICHI)

首都大学東京・大学院人間健康科学研究科・准教授

研究者番号：40381462

研究成果の概要(和文)：

本研究では、近年、コンピュータを用いた自然言語処理技術の進歩により活用され始めたテキストマイニングを導入することで、2004年に行った「生体肝移植ドナーに関する調査」の自由回答への記述から収集されたテキストデータの分析を行うことを目的とした。

分析にはTrue Teller(野村総研：以下TT)とSPSS Text Analysis for Surveys(IBM：以下STaFS)の双方を用い、単語の出現頻度や共起性を中心に分析を実施し、データの解析と共にソフトウェアの特徴や分析上の限界についても明らかにすることとした。

今回、生体肝移植ドナーが抱える健康不安について分析を行ったところ、「漠然」「将来」「不安」「現在」「ない」といった用語が、高頻度かつ同時に用いられており、現在は無症候でも漠然と将来への不安を感じていたが、「体力」や「疲労感」、「腸」と「閉塞」などは、それらとは別のグループを構成していると考えられたため、具体的な不安を感じている者がいることも分かった。その一方で「閉塞」と「閉そく」は別の語として認識されるため、非医療専門職を対象とした調査データを用いる場合、同義語としての登録など辞書機能の管理が分析結果に影響を及ぼし重要であることも明らかになった。

また双方のソフトウェアとも、分析上、最も時間を有するのは分析用の辞書(類義語等)の作成であり、分析結果について大きな相違は見られなかった。TTは、データのインポートから出力までが一体化されており、図表の作成等の操作性に優れているが、STaFSはIBM SPSS Statisticsで利用できる形態にデータを加工するツールであり、その後の分析はどのような統計手法も研究者側で選択できるため、自由度が高いと考えられた。

テキストマイニングは、大量のテキストデータから判別しづらい単語の頻度や関係性を客観的な手続きを用いて視覚的に把握することができ、今後の自由回答などの分析に用いる意義があると考えられた。

研究成果の概要(英文)：

The purpose of this study was to analyze the free answers of the investigation about the living donor liver transplantation (LDLT) donor which collected in 2004, utilizing text mining which had become popular in late years.

We used two software for text mining; True Teller (TT: Nomura Research Institute Ltd.) and IBM SPSS Text Analysis for Surveys (STaFS: IBM). We mainly analyzed frequencies and co-occurrences of words in free answers, and evaluated usability and limitations of software in analyzing procedures.

Text mining about anxiety about health showed that “vague”, “future”, “anxiety”, “recent” and “nothing” had high frequencies and co-occurrences. It suggested that they felt a vague uneasiness for future health without present symptoms. “Vitality” and “fatigue”, and “intestine” and “obstruction” also have them, and these three groups were recognized different from each other. Their anxieties included not only vague but specific.

On the other hand, the dictionary function including the synonym registration had an influence on the analysis when survey data from the non-health profession which includes various expressions of nearly same phenomenon.

In both analytical procedures of TT and STAfS, the making of dictionaries for analyses took a lot of time, and results of them were similar. But TT had automatic user interface from data import to output graph, STAfS was a tool for preparing dataset for statistic analysis using IBM SPSS Statistics and had high flexibility for following analysis.

Using an objective procedure, text mining could visually grasp frequency and the relationship of the word that was hard to distinguish from a large quantity of text data, and there was significance to use it for future analyses of free answers.

交付決定額

(金額単位：円)

	直接経費	間接経費	合計
2009年度	1,300,000	390,000	1,690,000
2010年度	800,000	240,000	1,040,000
年度			
年度			
年度			
総計	2,100,000	630,000	2,730,000

研究分野：看護学

科研費の分科・細目：看護学・地域・老年看護学

キーワード：①看護学 ②移植・再生医療 ③ドナー

#### 1. 研究開始当初の背景

2004年に研究代表者が中心メンバーの一人として実施した日本肝移植研究会による「生体肝移植ドナーに関する調査」<sup>1)</sup>は、2003年までに日本で行われた生体肝移植のすべてのドナー（臓器提供者）を対象とし、1480名からの回答が得られた。この調査は生体ドナーの死亡事故を受ける形でそれまで実態が明らかでなかった生体肝移植のドナーの健康状態や手術時のインフォームド・コンセントの状況、家族関係や就労の状況、移植に対する総括的評価などについて多面的に解明した。

この大規模調査からの提言を受けて、この数年間で移植医療にかかわる医師・看護師などのドナーに対するケアや権利擁護に対する意識は大きく変わったと考えられる。

これにより今後ドナーとなる場合への対応は改善されるものと思われるが、既に5,000人を超える患者の家族がドナーとして手術を受けており、調査結果から重篤な症状は少ないものの傷の違和感や悪心、疲労感の

増強といった症状を伴い4割の回答者が健康不安を抱えている現状、レシピエント（患者）を亡くし遺族となったドナーのうち、移植施設との継続的な関わりを持っている者が13%にとどまっていることなどを考えると、移植施設との連携の下でこのような健康に障害を抱える者や遺族への対応は地域医療・在宅看護の対象として視野に入れるべきであろう。

また本調査では、調査票の大項目ごとに関連する内容への自由回答欄を設けたところ、多数の回答者から、それらの多くの項目について大量の記述がみられたことが特徴として挙げられる。それらの多くは、これまで回答者が抱えてきた困難や苦悩を研究者に投げかけるものであった。一部の回答を報告書に掲載することはできたが、特に移植医療に対する評価に関する項目では、多くの回答者から、あまりにも大量で複雑な内容が書き込まれており、一般的な内容分析では全体を分類・整理することは困難であったため、その当時は貴重なデータでありながら十分に報

告することはできなかった。

このような状況の中、自然言語処理技術（特に日本語への対応）とコンピュータの性能の向上を受けて、テキストマイニングと呼ばれる手法が近年注目を集めている。テキストマイニングとは大量のテキストデータを形態素解析により単語や文節単位で区切り、コンピュータを用いた様々な計量的な方法により、それらの意味や品詞の類似性や発生頻度の関連などを明らかにする手法である。

2)

医学中央雑誌データベースでの“テキストマイニング”をキーワードとした検索では、この5年間で60件余の文献が抽出されており、それらの研究の多くは診療録や医療系学生の記録物などをデータとして用いたものであった。この技術を導入することにより、上述した懸案の解決につながる可能性が見えてきた。

しかし、データ解析に取り組む前に検討すべき課題があった。1)最も現実的な問題として、市販されているテキストマイニングソフトは価格が数百万円台で、研究目的での購入が困難であったこと、2)先行研究は診療録などの比較的定型的で、専門職が記述するものであるため用語が比較的統一されているが、ドナーを対象とした質問紙調査の回答は、非定型で用語も統一されていないこと、3)文章を単語・文節単位に分割することで文脈の把握が困難になること、といった点である。

現在1)については、Word Miner(日本電子計算株式会社 製)という比較的廉価なソフトが誕生したためこれを購入し、2)の問題に取り組むため、学生を対象とした在宅看護学教育に対する評価に関する調査を実施し、小規模なサンプルで試行的に分析を行ってきた。

これまでに、利点として1)分析のプロセスが明確で、研究者の根拠の乏しい恣意的なデ

ータの選択や強調を防ぐことができる点、2)分析結果を図として表示可能であるため視認性が高い点、3)データの効率的な管理が可能である点などがある一方、類義語の取り扱いの困難さや研究報告が少なく分析において基準となる指標が不明確である点が明らかになった。また大量データの解析においては類義語辞書の作成機能などが貧弱な廉価なソフトでの限界も明らかになった。

## 2. 研究の目的

本研究では以上のような状況や研究の取り組みを踏まえ、「生体肝移植ドナーに関する調査」における自由回答の分析を、テキストマイニング導入により行い、これまで十分に明らかにすることができなかった生体肝移植ドナーの想いの全体像を明らかにすることを第1の目的とした。

また、これらの分析結果や分析過程を踏まえ、同一のデータを用いて別のテキストマイニングソフトでの分析を行うことを通じて、非専門職の回答者によるテキストデータの分析における課題や限界、参考とすべき指標の発見などの分析方法の体系化を第2の目的として実施した。

## 3. 研究の方法

本研究では、研究代表者が委員を務めていた日本肝移植研究会ドナー調査委員会が2004年に実施した「生体肝移植ドナーに関する調査」において得られたデータを用いた。

これまで方法論的に分析が困難であった生体肝移植ドナーによる移植医療への評価や困難さなどに関する自由回答の大量のテキストデータについて、テキストマイニングソフト「True Teller」(野村総合研究所 製)による分析を行った。

また今後、テキストマイニングの利用が増

えることを考慮し、同一のデータを用いて別のテキストマイニングソフトによる分析を加え、それらの結果や分析過程の比較を通じて、専門職が記載する診療録などの比較的定型的で用語が統一されているテキストデータではなく、一般の患者・住民等が記載した非定型で用語も不統一なテキストデータという特徴に焦点を当て、テキストマイニングを行う上での課題や留意すべき指標などを明らかにする。

使用するデータ：2004年に日本肝移植研究会ドナー調査委員会が実施した「生体肝移植ドナーに関する調査」の回答を用いた。(使用許可を調査委員会委員長であった里見進：東北大学病院病院長より得ている。)

本調査は2003年末までに国内で実施されたすべての生体肝移植手術のドナー2,667例を対象としており、個人情報の保護のため各移植施設からそれぞれの名簿に基づき依頼状と調査票等の一式を送付した。そのうち256票が宛先不明として返送され、1480票の回収があった。(回収率61.4%)

本調査の実施については、2004年6月に開催された日本移植学会の倫理委員会にて承認を得た。

本研究では、「あなたはドナー手術を受けたことによって、今後、健康に影響が出るのではないかと不安に感じることはありますか。」という設問に対して、「1. はい」と回答した者(560名、38.9%)に対して自由回答の記入を求めた設問に対する503の回答を分析に用いた。

回答の例を以下に示す。(例は無作為に抽出した。プライバシーへの配慮のため一部改編してある。) 回答の文章の長さは10数文字のものから、200字程度まで違いがある。

- 肝臓がほんとうに元通りの機能をはたしているのかどうか。一生何もなくすごせるのか。

るのか。

- 腹部の膨満感が治らない。
- 特に根拠はないが、体にメスを入れているため不安に感じることもある。
- 新婚で手術をしたので、これから子供を作ろうと思っているが、本当に妊娠、出産時に問題がないのか不安。
- 精神的な苦痛がある。メスを入れることにより、寿命が縮まった気がする。
- 胃腸がとても弱くなり、体調をくずしやすくなった。
- 術後、腹筋と使いづらい環境になったので体の筋肉がだいぶおちてしまい、基礎代謝が落ちたせいかわどく体が冷えるようになってしまい体のバランスがとても崩れたと思う。体力も落ち、度々感染症になる様になったので、体にストレスを多くかけていることを考えると、確実に寿命は短くなったのだろうと思っている

分析手順：ソフトの標準的な機能を用いて文章の形態素分析・構文解析を行った後、単語やその係り受けの頻度の比較、レスポンス分析・主成分分析などを行い、自由回答の全体的な傾向を明らかにし、内容的な妥当性を検証するため生データの記述との比較、形態素分析の方法や精度、辞書機能の優劣といった日本語処理上の項目や、同一の分析を行う際の所要時間や、結果表示の柔軟性などの検討を行った。

分析ソフト：テキストマイニングソフトとして「True Teller v6.5」(野村総合研究所製)および、SPSS Text Analysis for Surveys 3.0 (STaFS) と SPSS Statistics 19 (IBM社製)を用いた。

両製品の特長を挙げるとすれば、前者は、テキストマイニングを行うために開発されたソフトウェアであるのに対し、後者は、IBM SPSS Statistics等の統計パッケージソフトで

統計的分析を行うためのデータセットを作るためのソフトウェアと呼べる。

分析に用いたコンピュータの仕様：

Panasonic 社製 Let's Note premium edition CF-S9: CPU Intel® Core™ i7 M620 2.67GHz メモリ 4 GB OS : Windows 7 Professional 64bit

Windows エクスプレィエンス インデックス プロセッサ 6.9 メモリ 5.9 グラフィックス 3.5 ゲーム用グラフィックス 3.8 プライマリハードディスク 5.9

#### 4. 研究成果

##### 1) 生体肝移植ドナーの抱える健康不安

###### ① True Teller を用いた分析

データをインポートした後、一度「グループの再構築」を行い、「単語ランキング」を表示させ、「腸閉塞」のような専門用語で分割されない方が望ましい単語をユーザー辞書に登録(表 1-1)し、「腸閉そく」、「腸へいそく」のように明らかに同義と思われるものは同義語辞書に登録(表 1-2)をした。また、特に出現頻度の低い単語において、文脈においては若干使われ方が異なるが、ほとんどのケースにおいて同義語として用いられている用語も意味内容を吟味し同義語として登録した。たとえば「出る」はほとんどのケースにおいて「症状が出る」という意味で用いられているため、「起こる」などと共に「発症する」に統一した。

最終的に作成した辞書を用いて、「グループの再構築」により、テキストの分析を行ったところ、処理時間は 16 秒であった。

表1-1 ユーザー辞書登録単語の一覧

登録語	品詞	活用
これから	名詞	
腸閉そく	名詞	
腸閉塞	名詞	
メスを入れる	動詞	上一段・下一段
生活	名詞	
おきる	動詞	上一段・下一段
この先	名詞	
脂肪肝	名詞	
健常	名詞	

表1-2 同義語辞書登録単語の一覧

統一語	品詞	単語	品詞
できる	動詞	出来る	動詞
悪化する	動詞	くずす	動詞
以前	名詞	前	名詞
異常だ	形容詞	おかしい	形容詞
癌	名詞	がん	名詞
経過する	動詞	たつ	動詞
経過する	動詞	とる	動詞
経過する	動詞	経る	動詞
経過する	動詞	重ねる	動詞
現在	名詞	今	名詞
言う	動詞	いう	動詞
酒	名詞	アルコール	名詞
術後	名詞	手術後	名詞
術前	名詞	手術前	名詞
将来	名詞	これから	名詞
将来	名詞	今後	名詞
将来	名詞	将来的だ	形容詞
少ない	形容詞	すこし	名詞
少ない	形容詞	少し	名詞
心配だ	形容詞	心配	名詞
身体	名詞	体	名詞
切る	名詞	切除する	名詞
切る	動詞	メスを入れる	動詞
切除する	動詞	切り取る	動詞
胆嚢	名詞	胆のう	名詞
腸閉塞	名詞	腸へいそく	名詞
腸閉塞	名詞	腸閉そく	名詞
痛み	名詞	痛い	形容詞
低下する	動詞	おちる	動詞
低下する	動詞	落ちる	動詞
低下する	動詞	低下	名詞
年齢	名詞	年	名詞
年齢	名詞	年令	名詞
漠然たる	形容詞	ばくぜんたる	形容詞
発症する	動詞	おきる	動詞
発症する	動詞	おこる	動詞
発症する	動詞	かかる	動詞
発症する	動詞	起きる	動詞
不安だ	形容詞	不安	名詞
腹部	名詞	お腹	名詞
閉塞	名詞	閉そく	名詞
戻る	動詞	もどる	動詞
癒着	名詞	癒着する	動詞
良い	形容詞	いい	形容詞

このデータを用いて、「単語ランキング」機能を用いて、単語の出現頻度・件数を確認したところ、表1-3の通りとなった。これらの比較的頻回に使用されている単語間の共

起性を明らかにするため、単語ランキングの上位26位までの単語と、関連のあるキーワード3個を使用してマッピング機能を用い、主成分分析により得られた第1軸（固有値

表1-3 出現件数別に見た単語の一覧

	単語	品詞	頻度	割合(%)	件数		単語	品詞	頻度	割合(%)	件数
1	不安だ	形容詞	128	23.26	117	28	術前	名詞	23	4.17	21
2	肝臓	名詞	90	14.71	74	29	傷	名詞	22	4.17	21
3	発症する	動詞	77	14.31	72	30	他	名詞	21	3.98	20
4	将来	名詞	68	13.12	66	31	経過する	動詞	20	3.78	19
5	肝機能	名詞	62	12.33	62	32	状態	名詞	20	3.78	19
6	手術	名詞	70	12.33	62	33	腸閉塞	名詞	19	3.78	19
7	ない	形容詞	68	11.73	59	34	以前	名詞	21	3.58	18
8	心配だ	形容詞	54	9.94	50	35	癌	名詞	23	3.38	17
9	影響	名詞	51	9.54	48	36	わかる(否定)	動詞	16	3.18	16
10	現在	名詞	47	8.75	44	37	異常	名詞	16	3.18	16
11	疲れる	動詞	43	8.35	42	38	自分	名詞	19	3.18	16
12	低下する	動詞	45	8.15	41	39	腸	名詞	17	3.18	16
13	術後	名詞	42	7.95	40	40	考える	動詞	16	2.98	15
14	身体	名詞	43	7.36	37	41	症状	名詞	16	2.98	15
15	病気	名詞	34	6.56	33	42	健康だ	形容詞	15	2.78	14
16	痛み	名詞	32	5.96	30	43	酒	名詞	15	2.78	14
17	言う	動詞	35	5.77	29	44	人	名詞	14	2.78	14
18	感じる	動詞	30	5.57	28	45	胆嚢	名詞	17	2.78	14
19	体力	名詞	29	5.57	28	46	下痢	名詞	14	2.58	13
20	ドナー	名詞	35	5.37	27	47	飲む	動詞	12	2.19	11
21	悪い	形容詞	28	5.17	26	48	健康	名詞	12	2.19	11
22	癒着	名詞	26	5.17	26	49	検査	名詞	11	2.19	11
23	年齢	名詞	25	4.77	24	50	体調	名詞	11	2.19	11
24	胃	名詞	25	4.57	23	51	内臓	名詞	11	2.19	11
25	切る	動詞	25	4.57	23	52	漠然たる	形容詞	11	2.19	11
26	腹部	名詞	24	4.57	23	53	負担	名詞	11	2.19	11
27	受ける	動詞	21	4.17	21	54	腹筋	名詞	12	2.19	11

注:上位のみ表示した

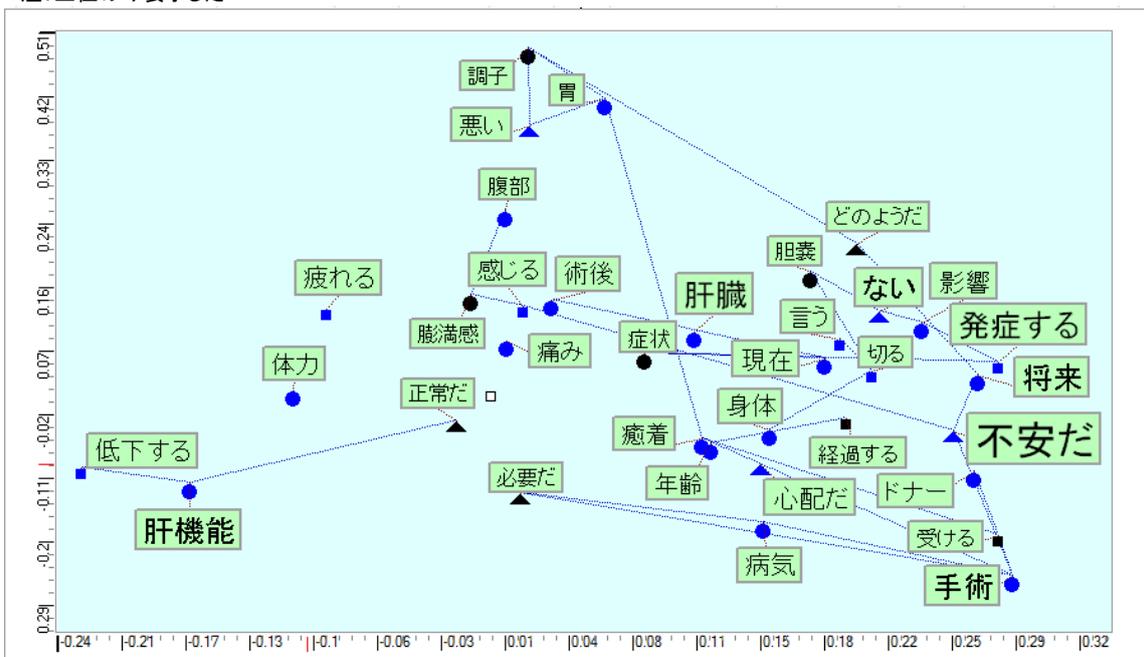


図 1-1 主成分分析を用いた頻出単語の共起性

2.23、寄与率 6.38) を横軸、第 2 軸(固有値 1.88、寄与率 5.38)を縦軸として図示した(図 1-1)。(図中では、件数が多くなるほど、文字サイズも大きくなるように設定した。)

布置図から共起性を見ると、生体肝移植ドナーの術後の健康不安の内容としては、図の左側に見られる肝機能の低下や体力面の心配や疲労、中央付近に見られる消化器症状や痛みに関する事、のような具体的な心配事と、「手術」-「受ける」や「胆嚢」-「ない」といったことを原因としつつ、加齢や術後の

経過の影響も加味しながら、将来的に何らかの症状が発症することへの不安、心配事が示されていることが分かった。

また、単語間の係り受け(フレーズ)を踏まえた検討のため、「主な話題」機能を用いて、単語間の関係を件数別に見たところ、表 1-4 に示した通りになった。同様に係り受け関係の共起性を明らかにするため、件数の上位 23 位までの係り受けを用い「マッピング機能」を用い、主成分分析により得られた第 1 軸(固有値 2.11、寄与率 9.16)を横軸、第

表1-4 出現件数順にみた主な話題の一覧

単語1	全体	単語2	タイプ	頻度	件数	単語1	全体	単語2	タイプ	頻度	件数
1 肝機能	62	低下する	名-動	24	24	19 肝臓	74	発症する	名-動	5	5
2 影響	48	発症する	名-動	16	16	20 身体	37	切る	名-動	5	5
3 手術	62	受ける	名-動	14	14	21 胆嚢	14	ない	名-形	5	5
4 不安だ	117	発症する	形-動	13	13	22 病気	33	発症する	名-動	5	5
5 影響	48	ない	名-形	13	11	23 負担	11	発症する	名-動	5	5
6 年齢	24	経過する	名-動	11	11	24 データ	5	ない	名-形	4	4
7 不安だ	117	感じる	形-動	11	11	25 ない	59	発症する	形-動	4	4
8 将来	66	不安だ	名-形	9	9	26 肝機能	62	発症する	名-動	4	4
9 何か	10	発症する	形-動	7	7	27 肝臓	74	悪い	名-形	4	4
10 肝臓	74	切る	名-動	7	7	28 肝臓	74	戻る	名-動	4	4
11 現在	44	ない	名-形	7	7	29 自信	4	ない	名-形	4	4
12 症状	15	発症する	名-動	7	7	30 手術	62	必要だ	名-形	4	4
13 体力	28	低下する	名-動	8	7	31 心配だ	50	発症する	形-動	5	4
14 酒	14	飲む	名-動	6	6	32 胆嚢	14	切る	名-動	4	4
15 将来	66	発症する	名-動	6	6	33 調子	6	悪い	名-形	4	4
16 体力	28	ない	名-形	6	6	34 痛み	30	感じる	名-動	4	4
17 ない	59	疲れる	形-動	5	5	35 不安だ	117	感じる(否定)	形-動	4	4
18 異常	16	発症する	名-動	5	5						

注:上位のみを示した。

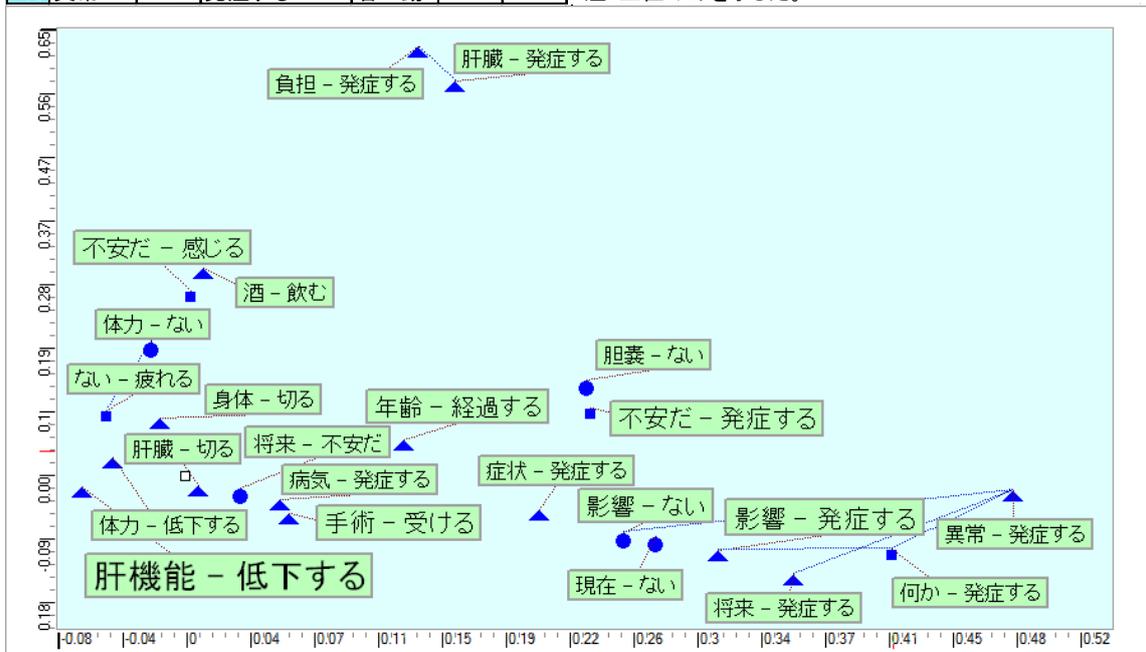


図 1-2 主成分分析を用いた単語の係り受け(フレーズ)の共起性

2 軸(固有値 1.81、寄与率 7.88)を縦軸として図示した(図 1-2)。(図中では、件数の多くなるほど、文字サイズも大きくなるように設定した)

布置図の左側では「肝機能-低下する」のように、相対的には具体的な不安の内容が「身体-切る」「肝臓-切る」といった原因となりうる項目との共起関係にあり、右側には「現在-ない」「将来-発症する」など、将来的な不安感を想起させる内容の共起性が示された。

単語としては、出現頻度が必ずしも高くなかった「酒」「飲む」という単語が組み合わせでよく用いられるため「酒-飲む」という係り受けでは「不安だ-感じる」「体力-ない」といった係り受けとの共起性が強い形で示されていた。その一方、単語レベルでは表示されていた比較的頻度の低い「胃」「腹部」などの単語は表示されなくなっていた。

② SPSS Text Analysis for surveys を用いた分析

本ソフトにおいても、Excel 形式のファイル等からデータをインポートし、「抽出」を行うと約 90 秒で、キーワードが作成され、その後「言語学に基づく分類手法」を用いてカテゴリーを作成した。

True Teller と同様、単語や類義語を登録することで、よりの確かな分類に近づくと考えられるため、同じ語句について登録することとしたが、「腸閉そく」は「腸閉塞」として同じカテゴリーに分類されるなど、登録の必要のない単語も多かった。

「カテゴリーの作成」により分類されたキーワードは、自動的に作成されたカテゴリーに含まれるものもあれば、自動的に振り分けられず、手動でこれらの語句をカテゴリーに追加する必要があった。例えば作成当初では、「肝臓」や「肝」には、これらの単語のほか



図 1-3 カテゴリー内のキーワードの例

表 1-5 カテゴリーの出現頻度

カテゴリー	ケース数	カテゴリー	ケース数
1 不安	171	32 他	21
2 ない	139	33 アルコール	20
3 肝臓	133	34 悪い	20
4 手術	114	35 易罹患性	18
5 発症する	105	36 妊娠	18
6 将来	64	37 悪化	17
7 病気	64	38 検査	17
8 胃腸	51	39 飲む	15
9 痛み	51	40 下痢	15
10 疲労	51	41 状態	15
11 現在	45	42 胆嚢	15
12 体力	39	43 精神面	14
13 以前	37	44 仕事	13
14 肝機能異常	36	45 食	13
15 腹部	35	46 障害	12
16 低下	34	47 運動	11
17 体	33	48 機能	11
18 手術後	31	49 子供	11
19 傷	31	50 正常	11
20 症状	30	51 漠然	11
21 体調	29	52 経過する	10
22 癒着	29	53 レシピエン	9
23 健康	27	54 必要	9
24 少し	27	55 感覚	8
25 腸閉塞	27	56 感染症	8
26 戻る	27	57 寿命	7
27 内臓	24	58 更年期	6
28 ドナー	23	59 胆汁	5
29 大丈夫	23	60 者	4
30 年	23	61 便秘	3
31 異常	22	62 経済的	2

にも、「先々」や「将来」といった単語も同時に含まれていることがわかる。(図 1-3) そ

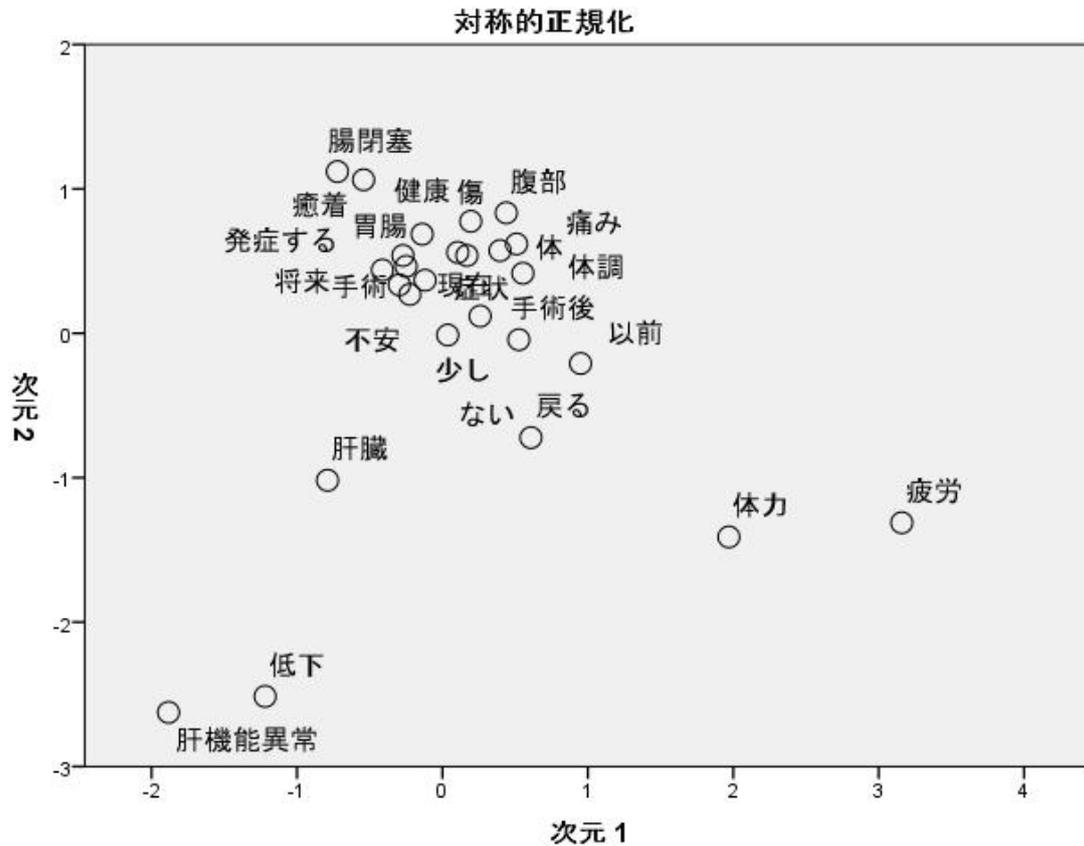


図 1-4 コレスポンデンス分析を用いた単語の共起性

ここで、ここでは、部位としての肝臓を意味するキーワードを「肝臓」に、肝臓の異常を意味するキーワードを「肝機能異常」として改めて割り振ることとした。

こうして、最終的に 62 個のカテゴリーを作成し、(表 1-5) このうち上位の 26 位までのカテゴリーを用い、コレスポンデンス分析を行い、布置図を示した。(図 1-4)

「肝機能異常」と「低下」、「体力」と「疲労」のカテゴリーの共起性が見られたほかには、「腸閉塞」「癒着」「痛み」といったカテゴリーは、True Teller による主成分分析による布置図(図 1-1)に比べ、他の単語と近い位置に置かれた。

## 2) 考察

### ①生体肝移植ドナーの健康不安

今回の分析において、二つのソフトによる解析の結果、双方とも生体肝移植ドナーが抱

える健康不安としては、肝機能低下や易疲労感、胃痛等の具体的な症状をあげている者と、現在はあまり症状がないが、漠然とした将来的な健康不安を有していると考えられる者に大別することができた。

今回の分析は、今後の経過観察やドナー外来等におけるドナーの不安の理解を助けることになると考えられる。

また本調査の報告書<sup>1)</sup>においては、「寿命が短くなるのではないか」といった回答例を印象的な例として示したが、実際には「寿命」という単語を記載した回答者は7名であることも分かり、印象だけではなく、回答の頻度といったテキストマイニングによって得られる情報を加味して検討することで、分析を行う研究者の主観によるバイアスを一定程度回避する効果が期待される。

### ②テキストマイニング手法の比較

## 1)形態素分析・構文解析と辞書機能

二つのソフトウェアとも、適切な類義語、単語を登録することにより、今回のデータの分析に適切な単語レベルでの形態素分析・構文解析を比較的短時間に行えた。

今回の分析では、STaFS はカテゴリーに該当するキーワードが含まれたケースの有無をエクスポートするため、カテゴリーの作成機能を用いたが、「肝臓」と「肝」の例で示したようにやや広範な内容が含まれる傾向が見られた。その一方結果には示さなかったが、口語的表現が類似する意味内容のカテゴリーの中うまく取り込まれている場合も見られ、単語ベースで分析を進める True Teller との違いが見られた。

STaFS のカテゴリーの変更自体は、マウスでドラッグするだけででき、それほど負担とはならないことや、容易に元のデータにアクセスして単語や前後の文脈を確認できること、また類義語や単語の登録は双方とも簡単に行え、ユーザーが登録した単語や類義語が他の分析にも活用できるようになっていることから、分析を続けていくほど操作の負担は軽減するものと考えられた。

## 2)分析結果

テキストマイニングと称される分析には、今回用いた主成分分析やコレスポンデンス分析を用いて共起性を布置図で図示するほかにも、クラスター分析など様々な方法があり<sup>3,4)</sup>、また、グループに分けて分析する機能などソフトウェアに標準装備されている機能も用いておらず、本研究結果のみで比較することは適切ではないが、True Teller はデータのインポートから布置図の出力や表示形式の変更までが一体的に行われるため、操作が容易である一方、ソフト上では分析の選択枝が限られるのに対し、STaFS はその後の分析は SPSS Statistics で行うため、さまざま

な分析方法が選択可能である。

実際には、True Teller でもマッピング機能で指定した上位の単語について、ケースごとに記載の有無を 1-0 の 2 値データでエクスポートすることができるため、使い慣れた統計パッケージで解析することも可能である。

## ②自由回答データへのテキストマイニングの導入

日本語は英語などと異なり、スペースで単語が区切られず、類似の表現が多いなど、テキストマイニングが困難な言語であるとされ<sup>5)</sup>、これまでは大量のデータが扱われる顧客分析<sup>6)</sup>などの分野で活用されてきた。

今回は、生体肝移植ドナーの経験者を対象とした調査データを用いたため、標記がひらがな混じりであったり、誤字があったりした場合や、口語表現が多いと、十分には形態素分析が行えない場面も見られた。

このようにテキストマイニングでは、単語を分類する上で若干の誤差が生じることが否めず、その他にも文脈的理解が限定的にしか行えないことや、少数意見を反映させるのには向いていないといった限界がある。こうした限界はあるものの、全体としての回答傾向を、辞書の公開などにより分析過程を明示的に示しながら、統計的指標を用いて客観的に示す機能や、他の変数によるグループ別の分類を容易にするといった点では、人間の手作業による分析に比べ秀でていけるといえる。フリーウェアの統計パッケージである R やそのパッケージである RMeCab なども活用され始めており<sup>7-9)</sup>、今後の積極的な活用が期待される。

また、テキストマイニングに適したデータ収集や質問の仕方などを検討することにより、さらに利便性が増すものと考えられる。

なお、今回は時間の都合により予定していた評価に関する自由回答の分析の全てを行

うことができなかつたため、今後も検討を続ける予定である。

【文献】

- 1) 日本肝移植研究会ドナー調査委員会：生体肝移植ドナーに関する調査報告書、2005年
- 2) 藤井美和他：福祉・心理・看護のテキストマイニング入門、中央法規、2005
- 3) 金明哲：テキストデータの統計科学入門、岩波書店、2009
- 4) 上田太郎監修：事例で学ぶテキストマイニング、共立出版、2008
- 5) 服部兼敏：テキストマイニングで広がる看護の世界、ナカニシヤ出版、2010
- 6) 三室克哉他：顧客の声マネジメントーテキストマイニングで本音を「見る」、オーム社、2007
- 7) 石田基広：Rによるテキストマイニング入門、森北出版、2008
- 8) 伊藤直枝：「甘えの心理」に迫る Rでテキストを分析、北樹出版、2010
- 9) 松浦真宏他：人文・社会科学のためのテキストマイニング、誠信書房、2009

5. 主な発表論文等

(研究代表者、研究分担者及び連携研究者には下線)

〔雑誌論文〕(計 0件)

〔学会発表〕(計 1件)

① 清水準一：生体肝移植ドナー調査に見るドナーの健康状態と長期予後把握の課題、パネルディスカッション 7「生体肝移植ドナー長期予後の問題点」、第 96 回日本消化器病学会総会、新潟、2010。(日本消化器病学会誌,107 Supple,A149,2010)

〔図書〕(計 0件)

〔産業財産権〕

○出願状況(計 0件)

名称：

発明者：

権利者：

種類：

番号：

出願年月日：

国内外の別：

○取得状況(計 0件)

名称：

発明者：

権利者：

種類：

番号：

取得年月日：

国内外の別：

〔その他〕

ホームページ等

<http://www.j-shimizu.net/modules/MyWorks/>

6. 研究組織

(1) 研究代表者

清水 準一 (SHIMIZU JUNICHI)

首都大学東京・大学院人間健康科学研究科・准教授

研究者番号：40381462

(2) 研究分担者

なし

(3) 連携研究者

なし