

看護・心理・福祉などの研究への テキストマイニングの活用

伊藤武彦（和光大学）

数理システムユーザーコンファレンス2009

2009年11月20日15:55-16:40 六本木ヒルズ 森タワー49F

Also in

第3回 軍事・環境・被害研究会

[日時] 11月23日(月)祝日 13時から

[場所] 法政大学市ヶ谷キャンパス

ボアソナードタワー19階 D会議室

[テーマ] 「被害」のくりかた、「被害」の定義、その批判的な見方。

伊藤研究室とテキストマイニング これまでの研究

- 学生奨励賞から(すべてダウンロード可能)

松上2006 異文化コンフリクトの解決 答案

八城2007 児童養護施設職員の調査

佐々木2007 偏見低減ビデオの感想文

大高2008 精神障害者のWebサイト

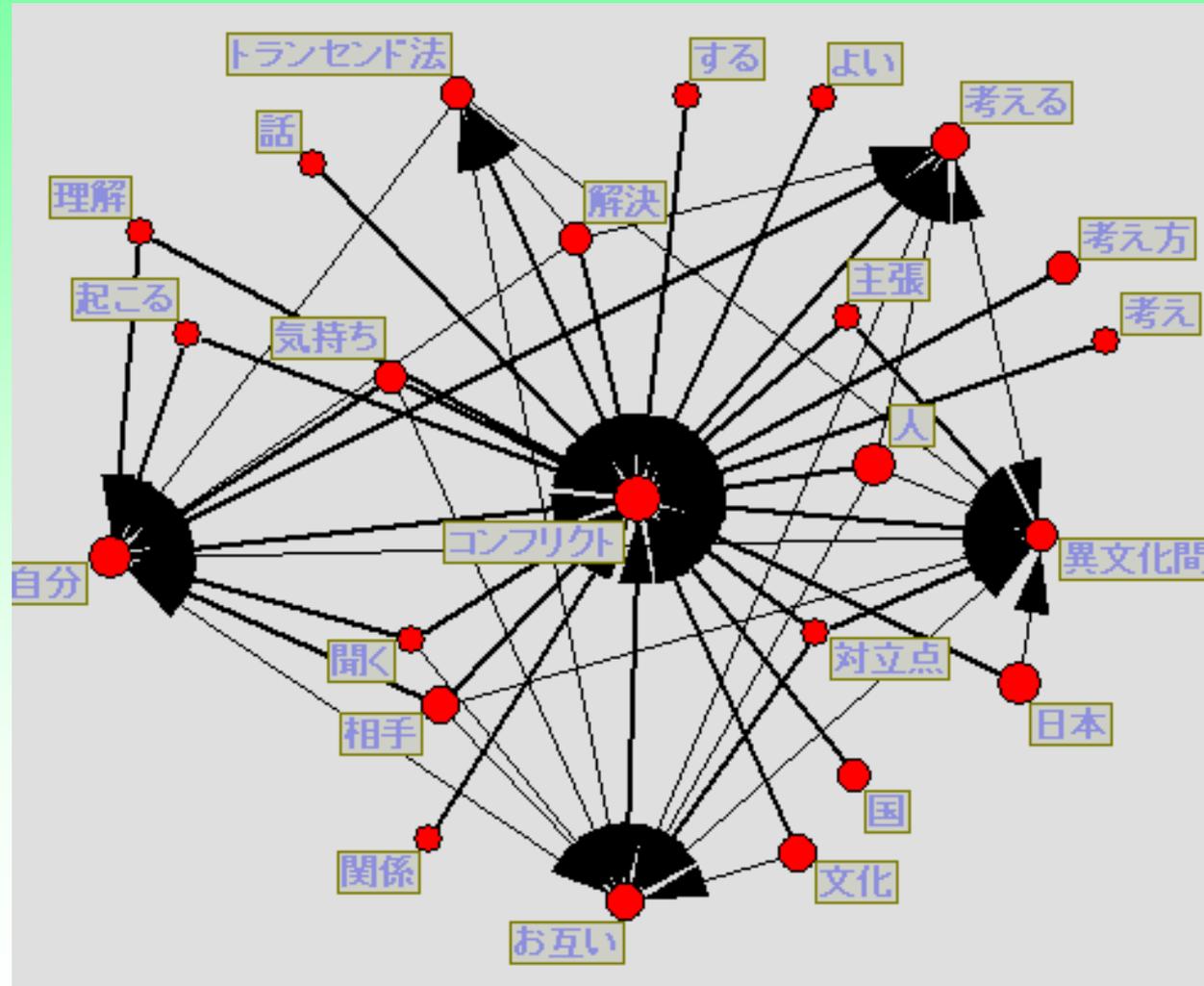
久木田2008 偏見のCiniiによる文献検索

孫2009 ターミナルケアの医中誌による文献検索

松上伸丈2006

異文化間コンフリクトの解決方法について

分析データとして、東京都内A大学の2006年度前期期末試験課題の回答者192人のうち33人分(17%)をサンプリングにて使用した。



児童養護施設職員の児童養護施設の改善に関するニーズと職務継続意思に関する一研究

八城 真里 2007

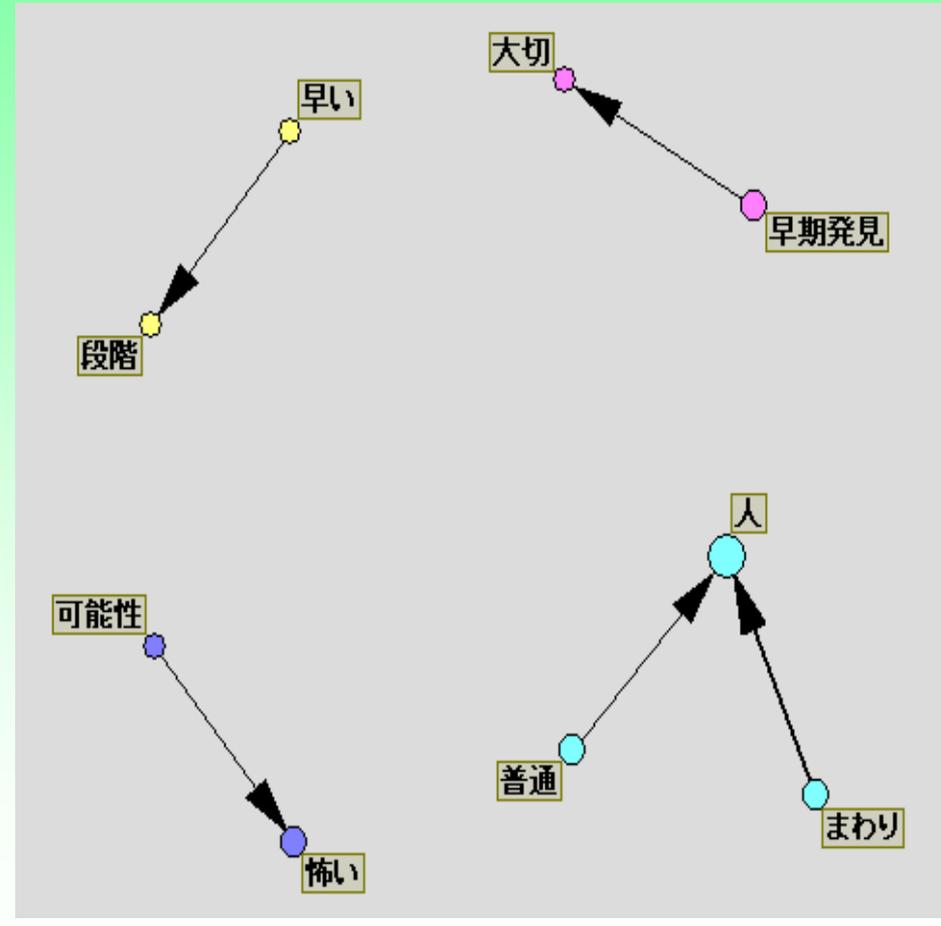
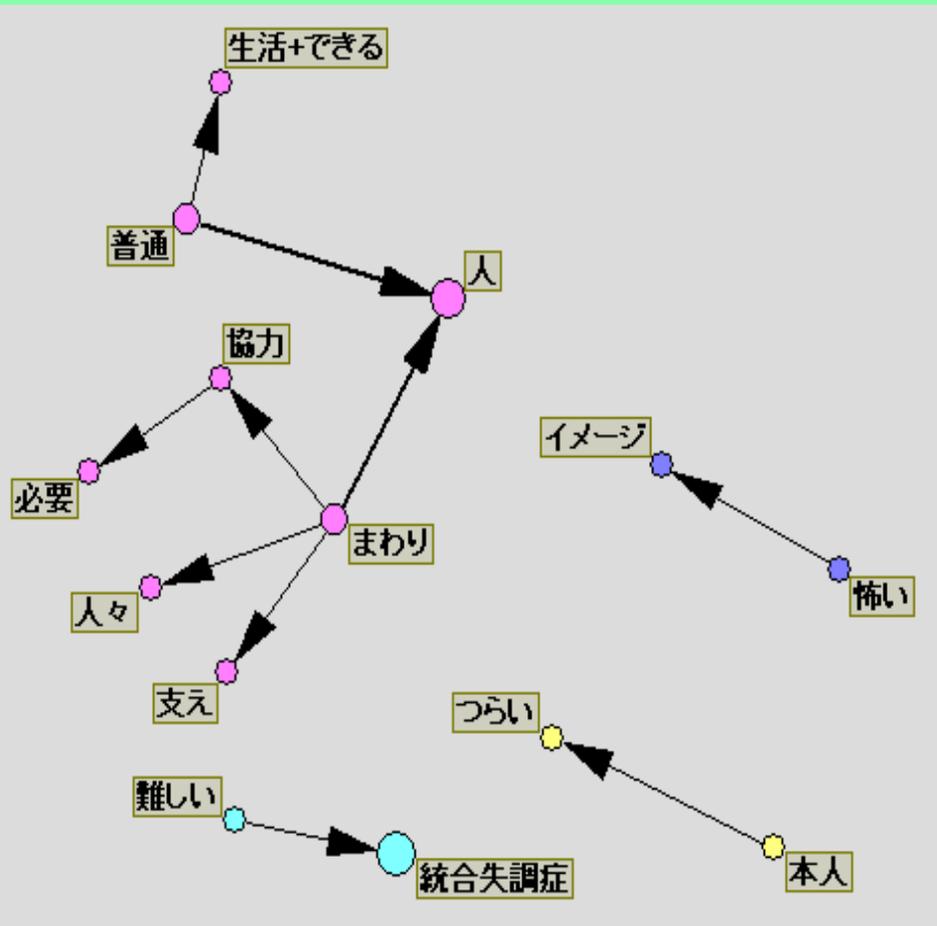
児童養護施設で直接処遇に従事する職員を対象に質問紙調査をおこない、826人の「現状をさらによくしていくために何が必要ですか」の質問の自由記述回答を分析した。

表5-1. 男女別特徴語抽出 [上位20]

男性		女性	
単語	指標値	単語	指標値
職員配置基準	44.976	子ども	25.111
見直し	27.388	大切	16.049
向上	18.58	支援	13.5
職員	17.194	理解	12.112
給料	16.672	知識	10.734
改善	15.842	現状	9.913
予算	13.557	良い	9.827
措置費	12.736	数	9.639
小規模化	9.8	職員同士	8.638
増やす	9.564	見る	8.346
質	7.779	心	7.892
確保	7.232	親	7.156
少ない	6.316	連携	6.523
改正	6.137	十分	6.42
人材	5.768	意見	6.146
人件費	5.589	大人	5.872
処遇	5.459	女性	5.778
解消	4.947	定期的	5.778
統一	4.947	入所	5.778
増える	4.853	上司	5.598

大学生の統合失調症に対するビデオ視聴後のテキスト分析

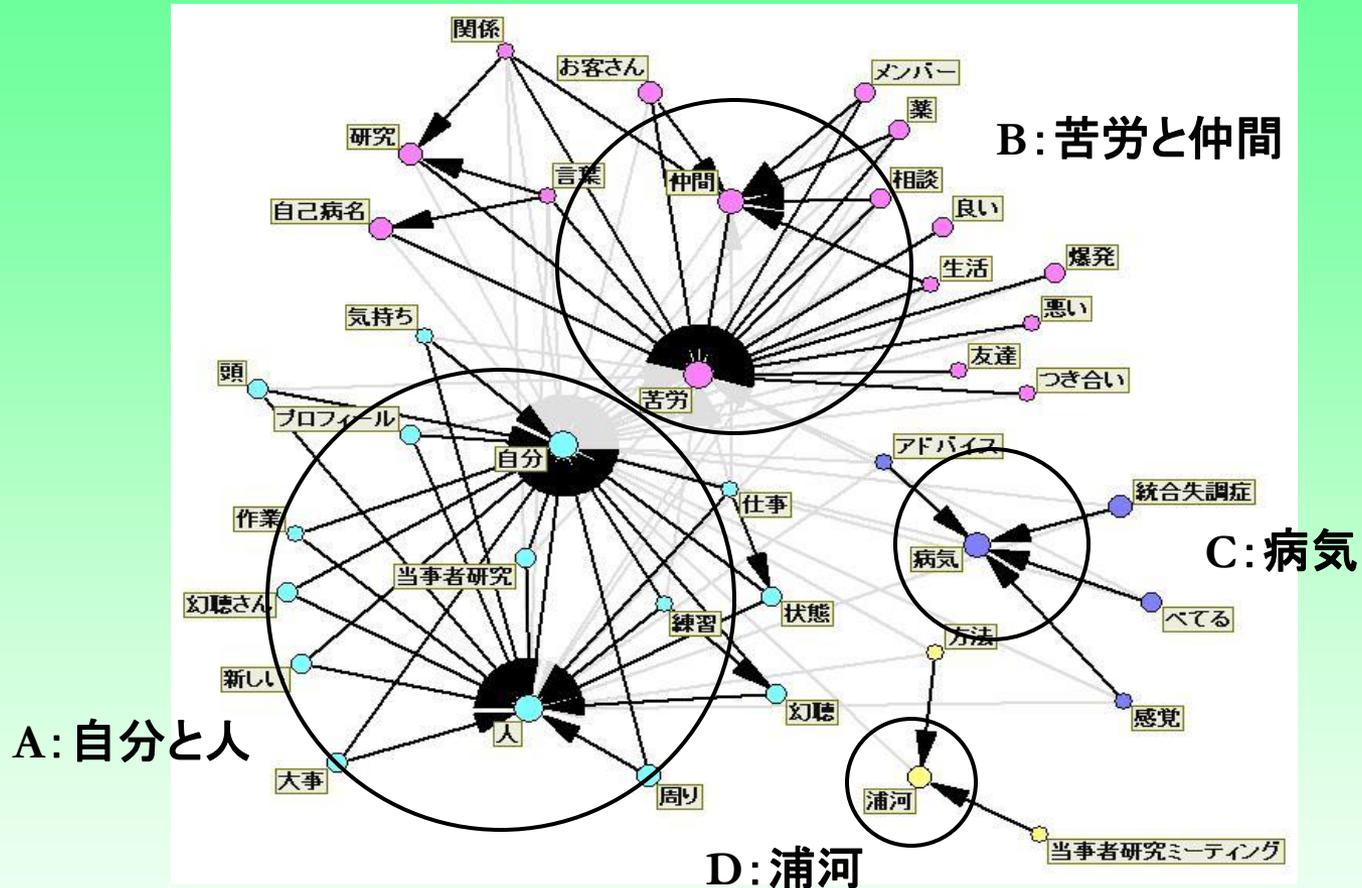
佐々木彩 2007



ことばネットワーク 患者談話ビデオ

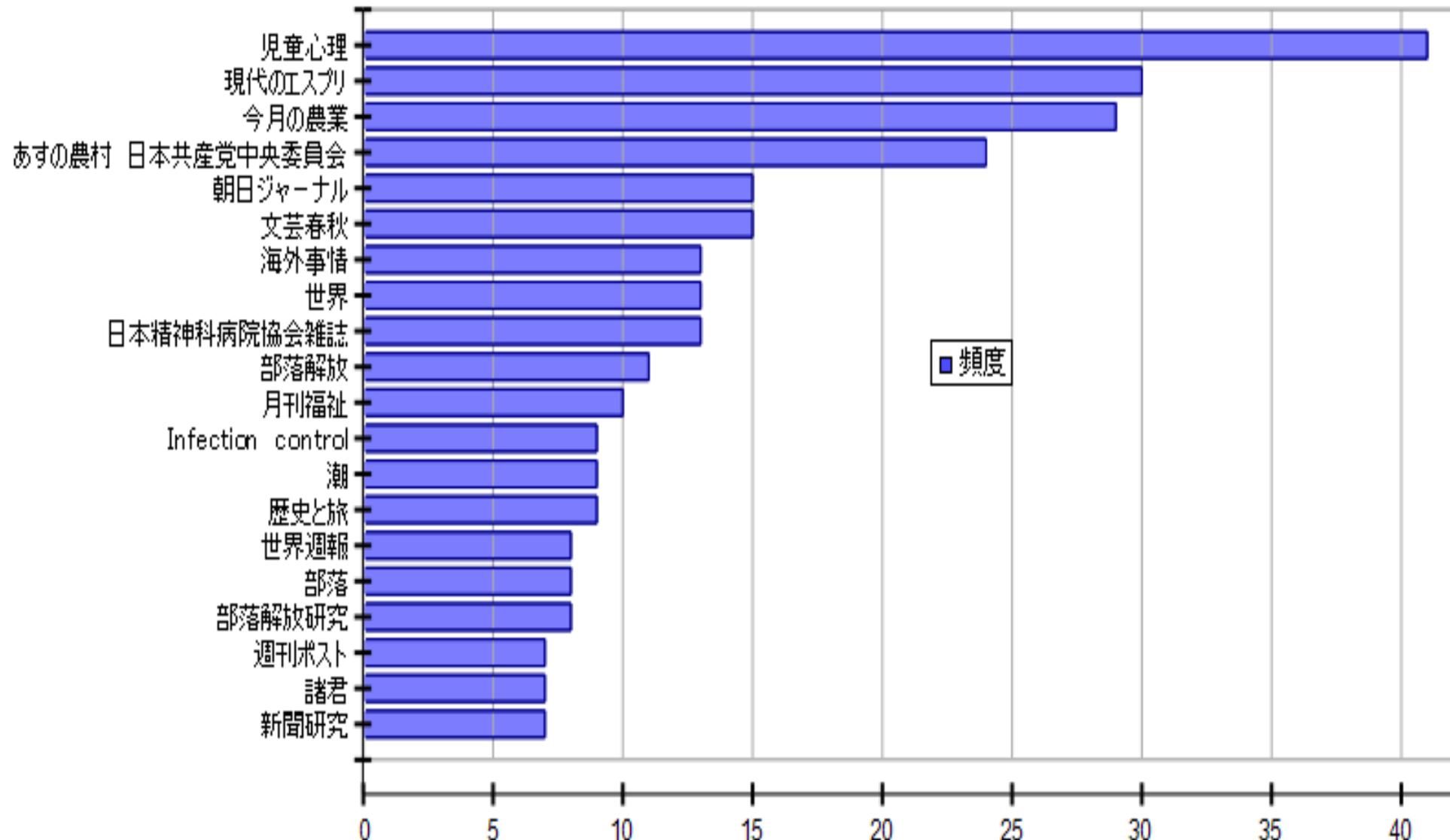
医者説明ビデオ

全体分析:ことばネットワーク(共起関係) イメージ抽出

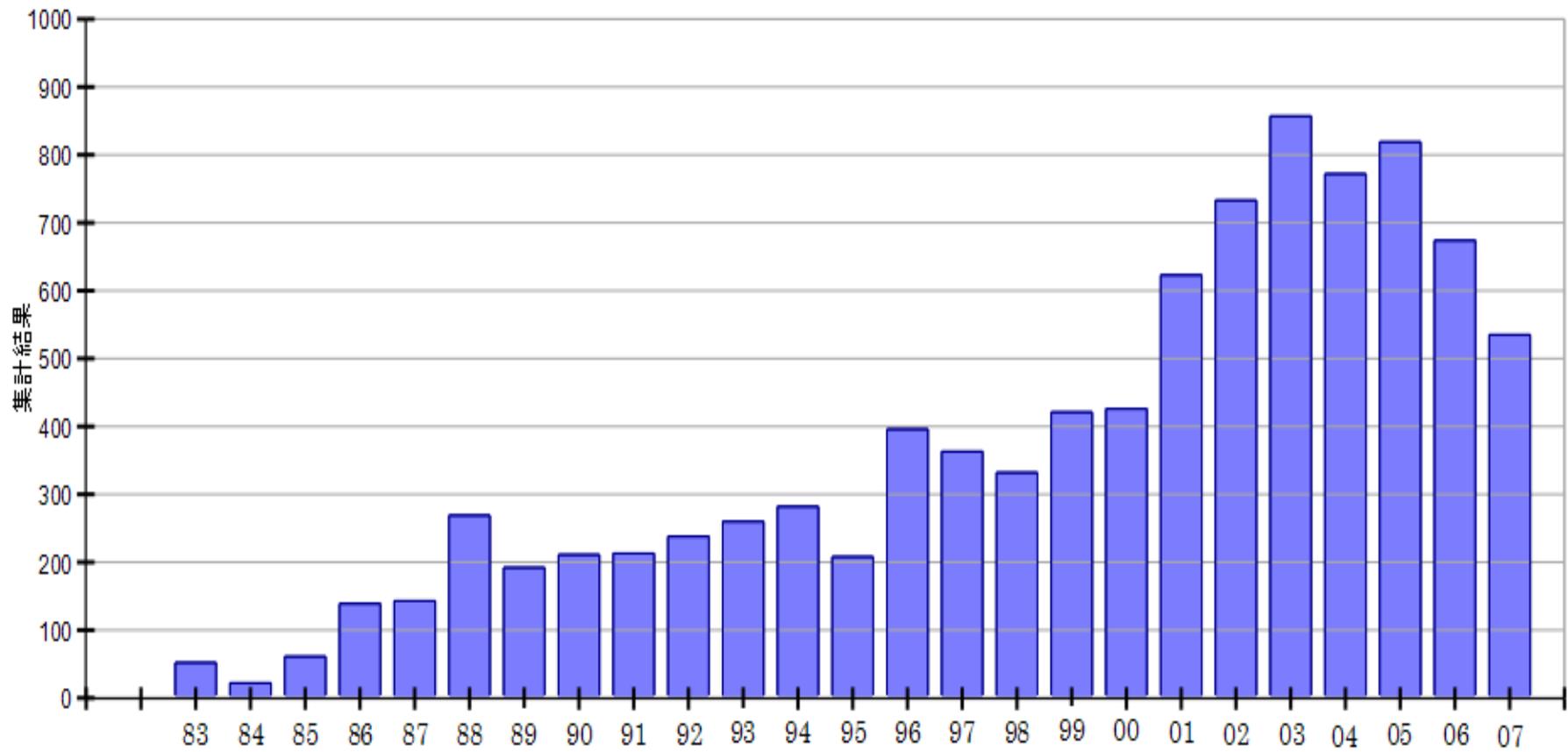


精神障害者のためのWebsite『当事者研究の部屋』
の記述の分析 大高 庸平2008

偏見の研究：CiNiiによる1949年-2007年の 文献の分析 久木田隼 2008



ターミナルケアについての文献研究：医中誌データベースによる1983から2007までのデータ分析 孫波 2007

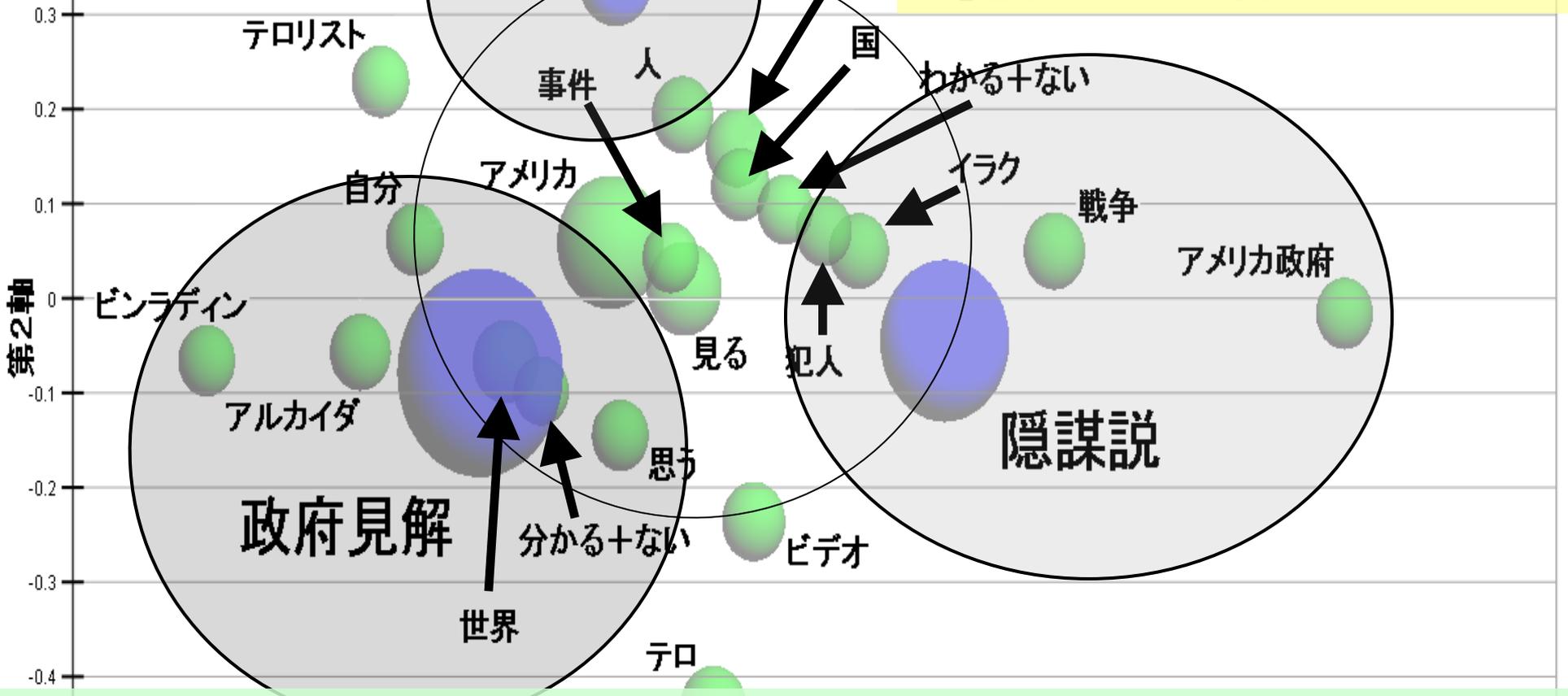


研究対象となるテキスト

- 論文データベース (Cinii, 医中誌など)
(孫,2008; 久木田,2008; 井上・伊藤,2009)
- 質問紙の自由記述回答 (佐々木,2007) (八城,2007)
- 調査報告書の自由記述再分析 (井川, 2009)
- 試験の答案 (松上,2006)
- 記述式の質問紙 (伊藤・川島,2009; 野中, 2009)
- Webサイト、ブログ (大高他,2008; 松上・伊藤,2008)
- 新聞記事データベース (目黒、2009)
- 単行本 (闘病記; コミックエッセイ(小平・伊藤、進行中))

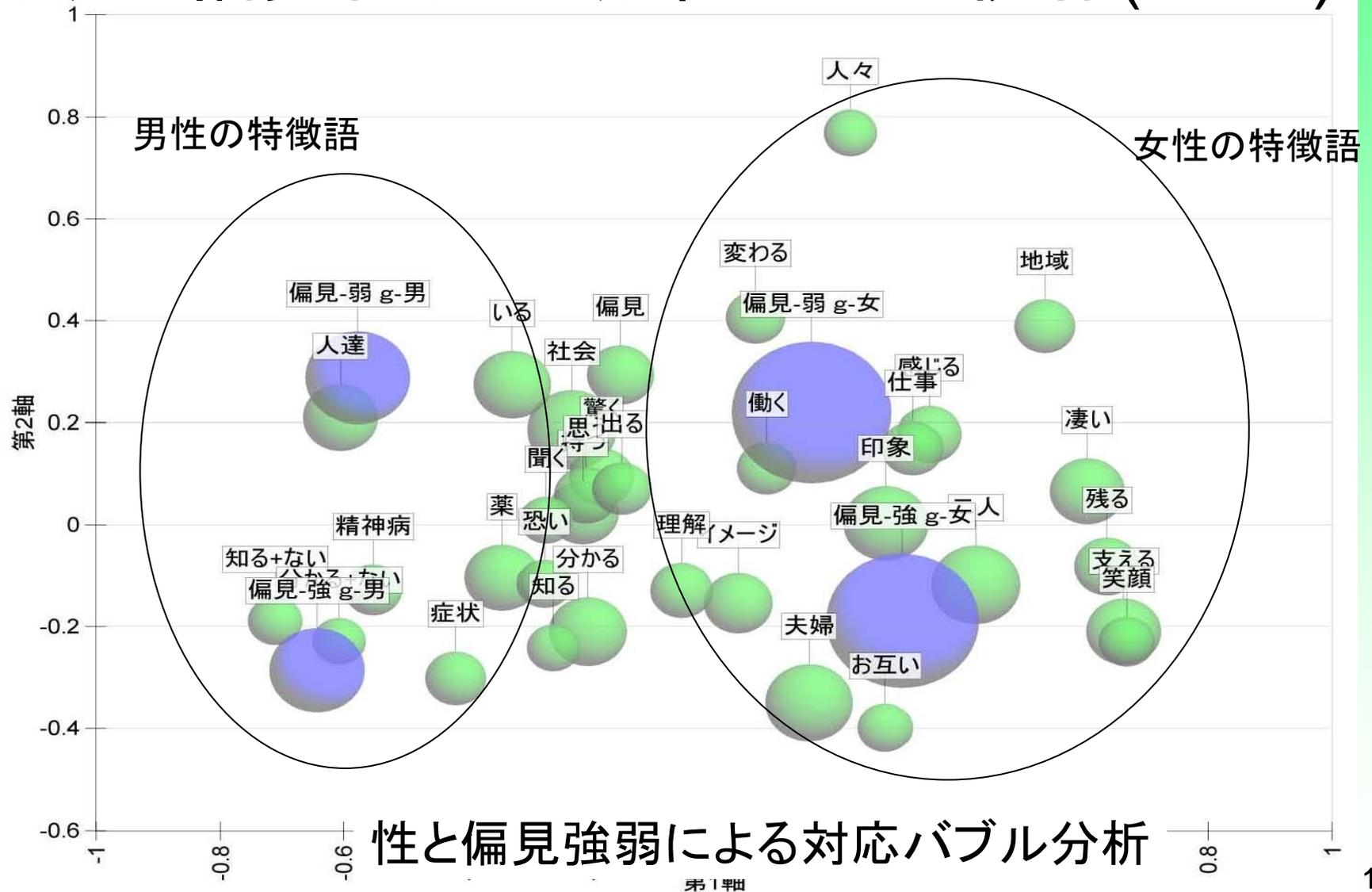
3群による感想文の単語の対応バブル分析

米国政府見解の群は、ビンラディンやアルカイダといった単語を多用し、陰謀説の群は「戦争」や「アメリカ政府」を多用している。



ビデオ視聴による政治的態度の変容:『911ボーイングを捜せ』視聴前から視聴後への「陰謀説」支持の増加はなぜおこったか? 伊藤・川島(2009)

テレビ番組視聴による統合失調症を持つ人に対する偏見低減の効果 小平・伊藤(2009)



★表1 典型的な質的研究と量的研究と テキストマイニングとの比較

方法	データ	分析方法
量的研究	数値(量的データ)	量的分析(統計)
質的研究	文字(質的データ)	質的分析
データマイニング	数値(量的データ)	量的分析(統計)
テキストマイニング	文字(質的データ)	量的分析(統計)

テキストマイニングは質的方法と量的方法との両方の特徴をもつ。この特性を生かした研究への活用法が期待される。

★那須川(2006)テキストマイニングを使う技術／作る技術

表 1.1 大量文書を活用する技術の比較

処理の深さ	処理の概要	機能(用途)	技術的要素	処理対象 (データ内容 表現形式)	自然言語 処理内容
レベル1	検索	目を通す対象を絞り込む	情報検索	文字列 単語列	単語の抽出 (語の原型・基本型への置換など)
レベル2	分類整理	文書の振り分け及び全体的にどのような内容が含まれているかの把握	クラスタリング クラシフィケーション	単語の集まり (Vector Space Model)	単語分布 状況の分析
レベル3	分析 (知識発見)	面白い内容・役に立つ知見を抽出	自然言語処理 データ・マイニング 視覚化	意味的概念の集まり	意味の分析 関係の分析

★佐藤郁哉(2008)質的データ分析法 :原理・方法・実践 新曜社 pp.52-58

質的データ分析

質的コーディング
量的内容分析
テキストマイニング

KJ法

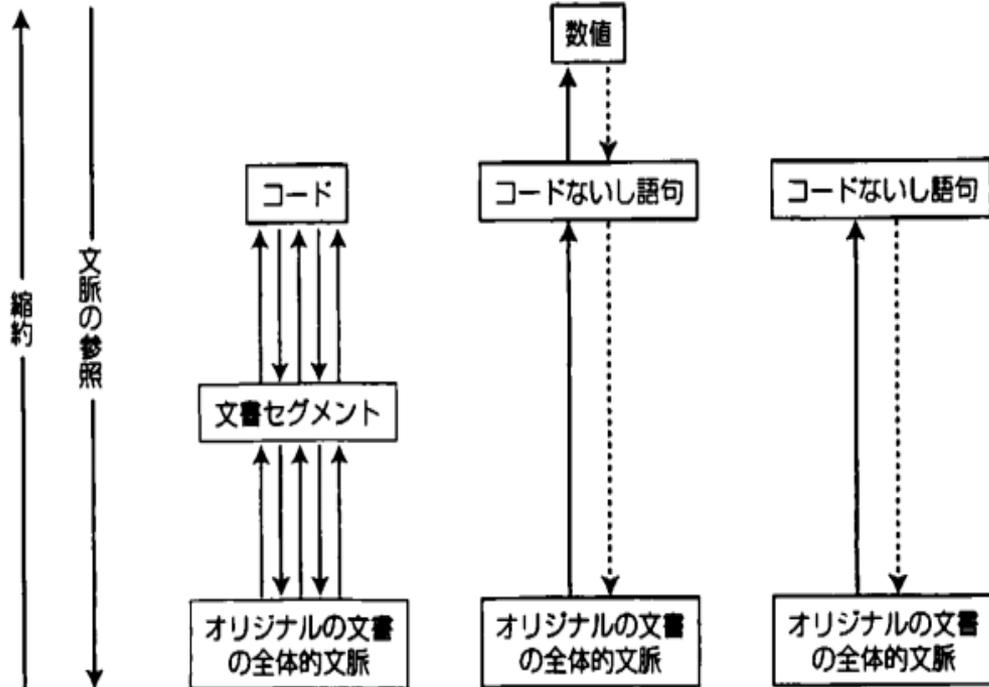


図4-4 質的データ分析と4つの技法の違い

本書で解説していく質的データ分析の場合には、データの縮約がおこなわれる一方でオリジナルの文脈への参照が何度となく繰り返されるのに対して、他の4種類の技法の場合には、縮約の手続きが終わった後で元の文脈に立ち帰るようなことは減多になされない。

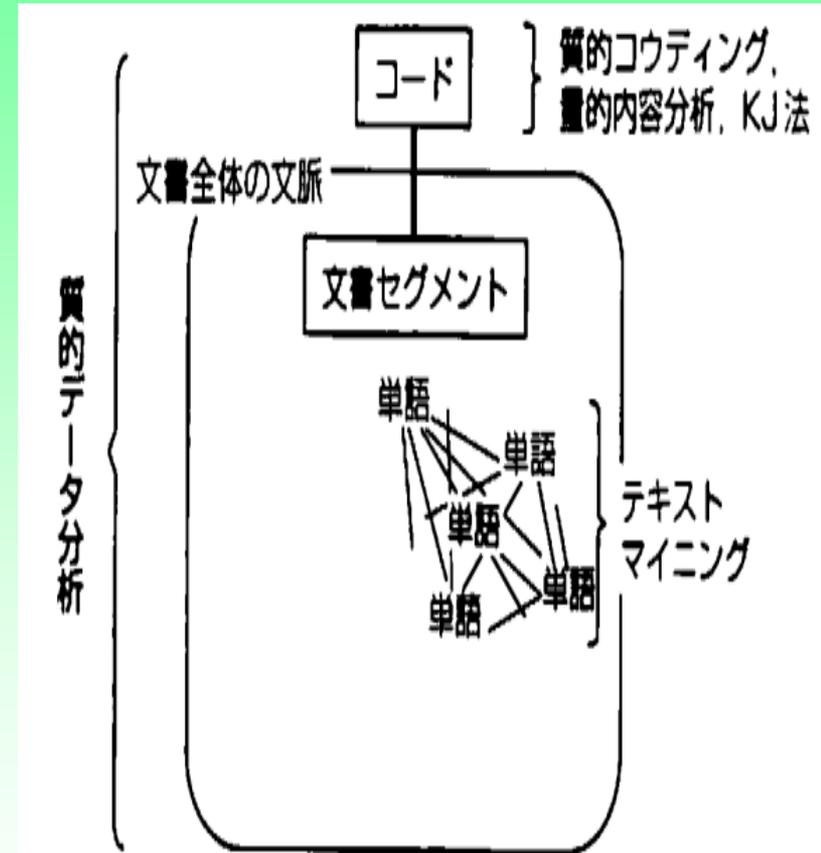


図4-5 さまざまな分析技法における基本的な分析単位

藤井美和2005テキストマイニングと質的研究 藤井美和・小杉孝司・李政元(編)福祉・心理・看護のテキストマイニング入門 中央法規Pp 13-28

- 社会科学において、テキストの分析は重要である。
- テキストマイニングは質的研究の問題点である、信頼性・妥当性・エビデンスに基づくアカウンタビリティを解決する手法である。
- KJ法の限界＝分類結果の信頼性の問題＋扱うデータ数の限界
- テキストマイニングは「数量化」と「視覚化」が可能

テキストマイニングの役割

- テキストマイニング中心の研究
- ミックス法による補完的使用
- 質的研究との連携
- 探索的研究vs.仮説検証的研究
(仮説生成的研究)

看護学においては量的研究として
心理学においては質的研究として

★(1)テキストマイニング単独の(中心の)研究

1 情動的・文字資料の分析:

例: Ciniiなどの文献データベース、公式的なWebサイト、ガイドブック、ハンドブック・・・等の情報中心の記事; ※医師説明条件のビデオ

2 物語的(ナラティブ)文字資料の分析

例: 闘病記、感想文、ブログ、小説、体験談、DIPEXなどのインタビューサイトなど、話者の語りを分析する場合 ※患者談話条件のビデオ

★(2)ミックス法①質的研究との連携

質的心理学や文化人類学や質的研究中心の看護学研究においては、テキストマイニングは量的研究として、エスノグラフィやグラウンデッドセオリーなどによる患者・当事者の物語り(ナラティブ)の質的研究と連携することによりミックス法を実現できる。

①順次的研究方略: 質的データ収集と分析の後に、テキストマイニングによる分析を行ない、解釈の段階で統合する。質的分析で仮説生成的に得られた仮説をテキストマイニングにより量的な裏付けをもって検証することができる。質的研究データの再分析としてのテキストマイニングである。

例: 精神障害者の居場所の研究(小平、未公刊)

②並行的トライアングレーション: テキストマイニング研究を質的研究と独立して研究; 並行する2つの方法で得た結果を分析の段階で統合する。質的研究とテキストマイニングの分析結果は対等のものとして統合される。質的研究者との共同研究プロジェクト。

★(3)ミックス法②量的研究との連携

- 心理学では(質的心理学・臨床心理学以外では)量的資料による量的分析が主流である。そのような量的研究と連携してテキストマイニングは次のように位置付く。
 - ① 並行的入れ子方略: 量的データと質的データを同時に収集する。質問紙の評定値と自由記述の結果のように、量的データ・量的分析が優位で支配的である。補足的に行なわれるテキストマイニングによる文字データ分析により、量的分析結果を裏付けたり、補足したり、矛盾点を指摘したりできる。
 - ② 並行的トライアングレーション: テキストマイニング研究を量的研究と独立して研究; 並行する2つの方法で得た結果を分析の段階で統合する。量的研究とテキストマイニングの分析結果は対等のものとして統合される。量的研究者との共同研究プロジェクト。

★テキストマイニングと研究目的のタイプ

探索的 仮説生成的 仮説検証的

個性記述的

法則定立的

- 論理科学的思考モード
- 物語的思考モード

★「水平社宣言」のテキストマイニング

長い間虐められて来た兄弟よ、過去半世紀間に種々なる方法と、多くの人々によってなされた吾らの爲の運動が、何等の有難い効果を齎らさなかった事實は、夫等のすべてが吾々によって、又他の人々によって毎に人間を冒瀆されてみた罰であったのだ。そしてこれ等の人間を勤るかの如き運動は、かえって多くの兄弟を墮落させた事を想へば、此際吾等の中より人間を尊敬する事によって自ら解放せんとする者の集團運動を起せるは、寧ろ必然である。

兄弟よ、吾々の祖先は自由、平等の渴迎者であり、實行者であった。陋劣なる階級政策の犠牲者であり、男らしき産業的殉教者であったのだ。ケモノの皮を剥ぐ報酬として、生々しき人間の皮を剥ぎ取られ、ケモノの心臓を裂く代價として、暖かい人間の心臓を引裂かれ、そこへ下らない嘲笑の唾まで吐きかけられた呪はれの夜の悪夢のうちにも、なほ誇り得る人間の血は、涸れずにあった。そうだ、そして吾々は、この血を享けて人間が神にかわらうとする時代にあつたのだ。犠牲者がその烙印を投げ返す時が来たのだ。殉教者が、その荊冠を祝福される時が来たのだ。

吾々がエタである事を誇り得る時が来たのだ。

吾々は、かならず卑屈なる言葉と怯懦なる行爲によって、祖先を辱しめ、人間を冒瀆してはならぬ。そして人の世の冷たさが、何んなに冷たいか、人間を勤る事が何であるかをよく知つてゐる吾々は、心から人生の熱と光を願求禮讃するものである。

水平社は、かくして生れた。

人の世に熱あれ、人間に光りあれ。

綱領

- 一、特殊部落民は部落民自身の行動によって絶対の解放を期す
- 一、吾々特殊部落民は絶対に経済の自由と職業の自由を社会に要求し以て獲得を期す
- 一、吾等は人間性の原理に覚醒し人類最高の完成に向つて突進す

「水平社宣言」テキストのパラドックス

歴史的な部落差別批判の声明であるが、「差別」「偏見」の文字は無い。

具体物としての現象(=単語)がないからといって、本質が存在しないということではない。

【文献】

- Baker, P. 2006 Using Corpora in Discourse Analysis.
London: Continuum
- Creswell (2003) 操華子, 森岡崇訳(2007)研究デザイン: 質的・量的・そしてミックス法 日本看護協会出版会
- 井上孝代・伊藤武彦(2009)ミックス法としてのPAC分析: テキストマイニングによる現状の展望、および今後の課題 内藤哲雄・井上孝代・伊藤武彦・岸 太一(編) PAC分析研究・実践集2 ナカニシヤ出版(印刷中)
- 伊藤武彦・川島 充 ビデオ視聴による政治的態度の変容: 『911 ボーイングを捜せ』視聴前から視聴後への「陰謀説」支持の増加はなぜおこったか? 日本教育心理学会第51回総会(静岡大学)発表論文集 456.
- 小平朋江・伊藤武彦 テレビ番組視聴による統合失調症を持つ人に対する偏見低減の効果 日本教育心理学会第51回総会(静岡大学)発表論文集 589.

- 小平朋江・伊藤武彦・松上伸丈・佐々木 彩(2007) テキストマイニングによるビデオ教材の分析:精神障害者への偏見低減教育のアカウンタビリティ向上をめざして マクロ・カウンセリング研究, 6, 16-31.
- Liu, B. 2007 Web Data Mining-Exploring Hyperlinks, Contents, and Usage Data, Berlin: Springer
- 松上伸丈・伊藤武彦 バイオ燃料をめぐる言説について:ネット上の記述を対象にしたテキストマイニング分析 マクロ・カウンセリング研究, 7, 22-29.
- 大高庸平・伊藤武彦・小平朋江 2009 精神障害者の自助の心理教育プログラム「当事者研究」による体験と回復の構造:「浦河べてるの家」のウェブサイト「当事者研究の部屋」の語りのテキストマイニング(投稿中)
- 下山晴彦(1997)臨床心理学研究の理論と実際:スチューデントアパシー研究を例として 東京大学出版会
- 八城真理・伊藤武彦・井上孝代 2008 児童養護施設職員のニーズと職務継続意思に関する研究:Text Mining Studio2.2.1によるテキストマイニング 日本応用心理学会第95回大会発表論文集, 95.