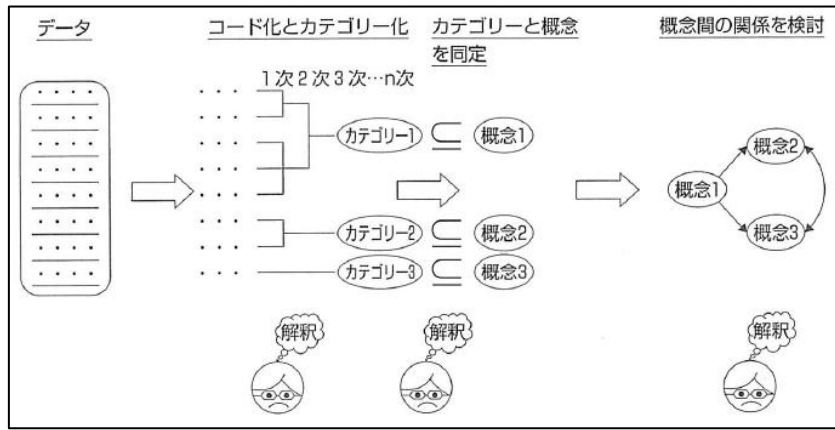


## フリーソフトを用いたテキストマイニング入門

### 1. 量的研究と質的研究

- 量的研究とは、一定の手続きに則り、問題を理論的に理解し、仮説を立て、それを測定可能なものに操作化し分析する。そして数値で表された結果から、科学的、客観的に仮説を証明することによって問題や効果を説明するもの。
- 質的研究は、いまだ説明されていない問題を探索する際の一手法として行われることが多い。インタビューや観察などで得られた結果をもとに理論を構築していくもの。質的研究の分析の流れは以下の通り。



- 質的研究の課題は、カテゴリーを抽出する際に研究者の洞察や直観による解釈が加わることである。そのため、質的研究によって導き出された結論については、実証研究の手続きを経る必要がある。

### 2. テキストマイニングとは

- テキストマイニングは、テキストデータからノイズを取り除いてルールやパターンを発見していく手法。データの数量化、及び統計分析を取り入れることで、質的研究でありながら、客観性の確保という量的研究の長所も併せもつ。
- 以下の流れで行われる。

#### ① テキストデータの収集



#### ② テキストデータの分析

- (1) テキストデータの数量化 (形態素解析等)
- (2) 統計解析 (対応分析、主成分分析、因子分析、クラスター分析、重回帰分析等)
- (3) 分析結果の視覚化 (布置図、デンドログラム等)



#### ③ 分析結果の解釈

- テキストマイニングの限界は、テキストデータのもつ曖昧さ、つまり同じ言葉でも様々な文脈

SPSS のソフトでは「係り受け解析」  
や「感性分析」も実行可能

によって様々な意味を持ちうることに起因する。また、形態素分析では複数の語で意味を成す言葉(例：介護保険)を不必要に分解してしまう(例：「介護」と「保険」)ことがあるので注意が必要である。したがって、分析前も分析中も、原文に戻りデータの見直しが常に必要な分析方法だと言える。

### 3. 研究事例の紹介(藤井, 2003の研究)

#### (1)目的

- 大学生が「死」に対してどのようなイメージをもっているか調べること。
- 探索的に死のイメージを構成する概念を抽出すること。

#### (2)調査対象とインフォームドコンセント

- 死生学を受講している大学生 102 名(学部は様々)。
- 調査の目的、自由参加であることなどを伝えた。

#### (3)調査方法

- 初回講義時に「死とは？」という質問に対して、思いつく概念や感情を記述してもらった。なるべく「死とは…である」という形で、箇条書きで記述してもらった。

#### (4)分析方法

- WordMiner を用いたテキストマイニングにより分析。以下の点に留意した。
  - ・ 構成要素の抽出にあたって、句読点・助詞・特殊記号を除いた。
  - ・ 同種の語を1つの語に置換した(終わり・終焉・最後→「終わり」など)。
- 頻度 2 以上の構成要素を対象に、対応分析を行った(対応分析では 15 成分が抽出され、累積寄与率は 53%であった)。
- 対応分析で得られた成分スコアをもとにクラスター分析を行い、構成要素の類型化を試みた。

#### (5)結果

- 抽出された全構成要素は 1883、句読点等を除いた後は 533、語の置換を行った後は 219 であった。
- 頻度 2 以上の構成要素は 65 で、最も頻度が高かったのは「終わり」で、31 名のサンプルによって 38 回出現していた。このほか 30 回以上出現した構成要素は「人間」「消滅」であった。
- 頻度 7 以上の構成要素は表 1 の通り。また対応分析の結果は図 1 の通りである。
- クラスター分析の結果、8 つの死観が浮かび上がった(表 2)。
- 8 つのクラスターの関係を示した図 2 によると、原点に近いクラスター 4 は、死に対する肯定的・否定的な感情を数多く含んでいる。そこでウォード法<sup>1</sup>、平方ユークリッド距離<sup>2</sup>を用いて再度クラスター分析を行った。その結果をデンドグラムで表したのが図 3 で、それぞれのグループの特徴は以下の通りである。
  - ・ グループ 1=死へのネガティブな感情、死は自然・現実、肉体的な側面

- ・ グループ 2=現実からの解放、現実から忘れられる
- ・ グループ 3=死に向き合おうとする積極的態度
- ・ グループ 4=死は新しい生のスタート(さらに積極的な態度)
- ・ グループ 5=死を肯定と否定両面から捉える姿勢

#### (6) 考察

- テキストマイニングにより、大学生の死観には 8 つの構成概念があることがわかった。Spilka らの研究で抽出された構成概念との重複は以下の通りである。

藤井 (2003)	Spilka ら (1977)
スピリチュアルな側面	浄福な来世
生命の終わり	自然な終焉
現実的・客観的側面	
死に対する感情	
大切な人との別離	家族との別離
孤独と未知	苦しみと孤独、未知
漠然	
行為の中断	

※Spilka ら (1977) では他に「無関心」「勇気」「挫折」が構成概念に挙げられている

- 丹下 (2002) の調査結果と同じように、成人よりも若者のほうが死への恐怖や否定的感情が多いことがわかった。しかし少数ながら、死を受容しようという前向きな態度も見られ、死への複雑な感情が窺えた。
- 死によって自分の存在が周りから忘れられる悲しさなど、死を個人のものでなく、他者との関係のなかで考える傾向もあった。
- クラスターの 1 つである「死に対する感情」は 5 つの構成概念に分かれ、それぞれ他のクラスターと関係があるようである (例：グループ 2 は「孤独と未知」と関係がありそう)。これら下位グループと各上位クラスターとの関係は今後の研究課題。
- 本研究の限界は、①死生学の受講生が対象であったため、死に何らかの興味を持った者が対象になっていたこと (Spilka らの研究では「無関心」という概念が現れていた)、②サンプル 102 は十分とは言えないこと、の 2 点である。

注 1：クラスター分析の時に使用される、クラスター間の距離を定義する距離関数のひとつ

注 2：座標を用いた 2 次元の距離の測定 (x, y) を、3 次元以上の空間における距離の測定 (x, y, z, …) に応用したもの

#### <参考文献>

藤井美和・小杉考司・李政元 (編著) (2005) 『福祉・心理・看護のテキストマイニング入門』 東京：中央法規出版