テキストマイニングを用いた自由記述データの有効活用に関する研究 －限界自治体の群馬県南牧村を対象として－

森田 哲夫[1]・諸岡 峻一[2]・塚田 伸也[3]・橋本 隆[4]

[1]正会員　東北工業大学工学部都市マネジメント学科（〒982-8577 仙台市太白区八木山香澄町35-1）
E-mail:ttmorita@tohtech.ac.jp
[2]宇都宮大学農学部農業環境工学科（〒321-8505 栃木県宇都宮市峰町350）
[3]正会員　前橋市建設部公園緑地課（〒371-8601 群馬県前橋市大手町2-12-1）
[4]正会員　伊勢崎市企画部企画調整課（〒372-8501 群馬県伊勢崎市今泉町2-410）

近年，日本の山間地域では過疎化・高齢化が急速に進行し，高齢化率50％を超える限界自治体が出現している．現在は少数の自治体であるが，近い将来において急速な増加が見込まれている．本研究では，限界自治体の中で最も高齢化率の高い群馬県南牧村を対象地域とする．本研究の目的は，テキストマイニングを適用しアンケート調査の自由記述欄のテキストデータを分析し，生活質評価のプリコードデータと合わせて用いることにより自由記述データの有効活用の方法を検討することである．分析の結果，アンケート調査の自由記述データとプリコードデータの関係を把握することができ，自由記述データからはプリコードデータに設定されていない内容を把握することができた．

Key Words : *text mining, free answer, marginal local community, quality of life*

1. はじめに

住民を対象としたアンケート調査の調査票には，自由記述欄が設けられている場合が多いが，十分定量的に分析されているとはいえない．一方で，最近では自然言語処理分野で研究が進められているテキストマイニングにより自由記述アンケートなどを定量的に分析できるようになってきた．土木計画学分野においても，自然言語処理分野の成果を用いた研究がみられるようになっている．

過疎化，高齢化が急速に進行している山間地域では，近年，高齢化率 50%を超える限界自治体 [1)]が出現している．現在は少数の自治体ではあるが，近い将来において急速な増加が見込まれる．これら自治体内の集落の多くは限界集落であり，日常生活や社会基盤の維持が問題となっている．

本研究は，テキストマイニングを用いアンケート調査票の自由記述欄に記入されたテキストデータ（以下，自由記述データと称す）を分析し，その結果を選択肢を予め設定した調査票から得られるデータ（以下，プリコードデータと称す）に補完することにより，生活質評価と居住意向との関係を明らかにすること目的とする．これにより，自治体などのアンケート調査における自由記述データの有効活用の方法を提案することをめざす．

対象地域は山間部の限界自治体とした．この理由は，深刻で切実な生活状況にある住民の意識がプリコードデータだけでは把握できていないと考えたためである．

2. 本研究の位置づけと研究の流れ

(1) 既存研究と本研究の位置づけ

本研究が着目する，テキストマイニングを適用した研究，山間自治体・限界自治体の生活質評価・居住意向を分析した研究の2点から既存研究を整理し，本研究の位置づけを示す．

1つめのテキストマイニングを適用した研究は，土木計画学分野においても，意見の分類・分析に自然言語処理技術を用いる研究がみられるようになった[2), 3)]．また，佐々木ら[4)]によるワークショップの討議内容の分析，佐々木・丸山[5)]によるワークショップの討議内容の可視化に関する研究，塚田ら[6)]による委員会の発言内容の分析，佐々木ら[7)]によるブログマイニングによる行動デー

タの抽出とアンケート調査との比較など，さまざまな場面で得られたデータを，マイニング手法により定量的に分析する研究が進められている．長ら[8]は，自由回答インタビュー調査データを用い，対象者の経験知識を抽出する方法を検討している．本研究では，佐々木ら[7]の研究では，ブログマイニングの結果と，行動アンケート調査との比較を行っており，マイニング結果を検証している点で注目している．森田・入澤ら[9]は，群馬県前橋市の居住者に対するアンケート調査データを用い，自由記述データとプリコードデータの関係をコレスポンデンス分析により把握している．

本研究は，近年の研究における会議での発言等のテキストデータを用いたデータマイニングの適用研究の流れに属する．本研究は，自由記述データだけではなく，プリコートデータを補完するデータであると考え，自由記述データとプリコードデータを用い分析する点に特徴がある．さらに，自由記述データをより積極的に活用することを検討する点で，森田・入澤らの研究[9]の延長線上に位置する．これにより，全国の自治体で実施している市民アンケート調査等の自由記述データを有効に活用できる可能性を検討できると考えられる．

2つめの山間自治体・限界自治体の生活質評価・居住意向を分析した研究は，交通サービス水準や交通特性に着目したものが多く，過疎地域の公共交通サービス水準に関する森山ら[10]の研究，交通サービス水準と生活の質の関連に関する宮崎ら[11]の研究がみられる．居住意向については，谷本ら[12]が，地方部の自治体を対象に，定住意向に影響を及ぼしている社会生活環境の要素を分析している．まとまった研究群としては，「「限界集落」を対象とした中山間地域のモビリティの確保と地域再編戦略に関する研究」[13]が進められ，住民の居住・移住意向には地域への愛着が大きく影響し，移動利便性が低いことにより直ちに移住意向につながらないことを明らかにした．森田ら[14]は，山間地域を対象に居住意向を分析し，生活に不便さを感じていながらも定住意向が高く，高齢者ほどこの傾向が高いことを明らかにしている．塚井・桑野[15]は，中山間地域の住民の移住意向と移住要件との関係を分析し，移住に係る費用等の条件が整えば移住意向を示す世帯が存在することを明らかにした．また，森田ら[16]は，限界自治体である群馬県南牧村を対象に生活質評価と居住意向の関係を分析しているが，統計的にはさらなる分析が必要であるとしている．

本研究は，既存研究における山間自治体・限界自治体の生活質評価・居住意向に関する研究系列上に位置し，高齢化の極端に進行した限界自治体を対象とする点が特徴である．過疎・高齢地域においては，人口減少と高齢化は免れ得ない状況である．この問題に対し対策を施さないと，自治体自体の存続が危ぶまれる．本研究においては，限界自治体を対象に，プリコードデータに加え自由記述データを用いるという点で，森田らの研究[16]の延長線上に位置する．

表-1　全国の限界自治体（2010年）

	自治体名	高齢者数	高齢化率%
1	南牧村　（群馬県）	1,387	57.2
2	金山町　（福島県）	1,356	55.1
3	天龍村　（長野県）	896	54.1
4	大豊町　（高知県）	2,549	54.0
5	昭和村　（福島県）	798	53.2
6	上勝町　（徳島県）	935	52.4
7	神流町　（群馬県）	1,231	52.3
8	大鹿村　（長野県）	598	51.6
9	川上村　（奈良県）	833	50.7
10	北山村　（和歌山県）	245	50.4
11	仁淀川町（高知県）	3,267	50.3

資料：2010年国勢調査

(2)　本研究の流れ

本研究は，次に示す流れの沿い進める．まず，次の3章において，限界自治体である対象地域を設定し，住民の生活質評価・居住意向に関するプリコードデータ，自由記述データを収集・整理する．4章では，自由記述データについて，自由記述データの記入特性の分析，テキストマイニングを用いた分析を行う．5章では，生活質評価・居住意向に関するプリコードデータを自由記述データで補完する分析方法について検討する．

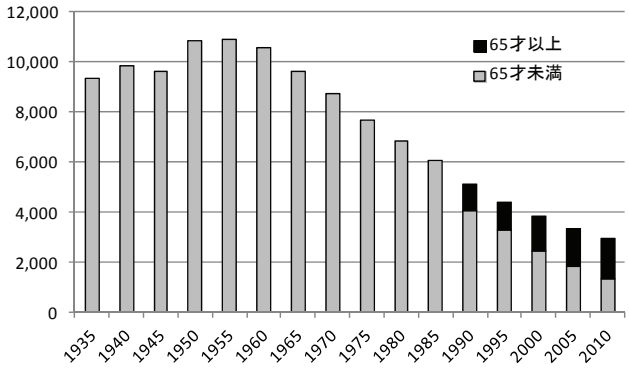
本研究では，自然言語処理の要素技術を用いたテキストマイニングのためのフリーウェアKH Coder[16), 17)]を利用する．KH Coderは，操作性が高く，多くの分野の研究論文で利用されている．形態素解析器として「茶筌」[18)]を用いており，精度の高い単語抽出を行うことができる．

3.　対象地域の設定とデータ収集

(1)　対象地域の設定

限界自治体，限界集落の概念を提唱した大野は[20)]，限界自治体を「65歳以上の高齢者が自治体総人口の半数を超え，税収入の減少と老人福祉・高齢者医療関連の支出増という状況の中で財政維持が困難な状態におかれている自治体」とし，全国の山間地域で限界自治体が発生すること，集落においては自治機能が低下し，高齢者の生活状態の悪化し，独居世帯が残されると指摘している．

2005年国勢調査によると，限界自治体（高齢化率50%以上）は全国に6町村存在した．5年後の2010年国勢調査では，限界自治体は11町村となり（**表-1**），大野が予見



注：65才以上の表示は1985年以降のみ

図-1　群馬県南牧村の人口推移

図-2　群馬県南牧村の位置

したように，限界自治体は着実に増加している．最も高齢化率が高いのは，群馬県南牧村であり，2005年の53.4%から2010年の57.2%に上昇した．0～14才人口は4.3%，15～64才は38.5%である．人口推移をみると（**図-1**），1955年に10,573人であった人口が，2010年に2,423人（国勢調査）と減少している．南牧村は，急傾斜地が多く災害危険性の高く，2007年9月の台風9号で道路の寸断，孤立集落の発生など大きな被害を経験している．

本研究では南牧村を対象とする．その理由は，1)高齢化率が最も高い自治体であるが，将来的に限界自治体の増加が見込まれ，早い時期に住民の意向を把握**する**ことは将来の政策検討のために重要な資料となると考えたため．2)今後，社会基盤整備・維持のための予算は減少すると予想され，災害危険性の高い自治体を対象とすることは，今後の社会基盤整備・維持に関する政策を検討する上で有用な知見を得られると考えたため．

南牧村位置を**図-2** に，60 集落別の高齢化率を**図-3** に示した．南牧村は群馬県南西部の県境，高崎市から約40km に位置する．集落別にみると，役場付近の中心部は高齢化率が低いが，中心部から離れた地区には高齢化率の高い集落が分布する．2007 年台風 9 号による河川の決壊地点の奥部にも集落があり，それら集落の高齢化率は高い．

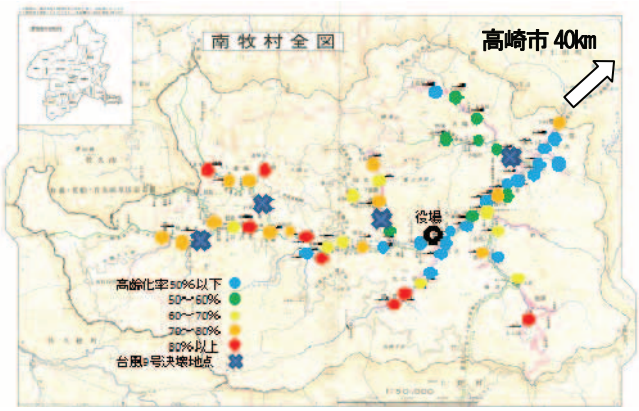


図-3　集落別高齢化率と台風による河川決壊箇所
（群馬県南牧村，2007 年台風 9 号）

(2)　分析データの収集

自由記述データ，プリコードデータを収集するアンケート調査の概要を**表-2**に示した．南牧村の全世帯を対象とし，分区長による戸別配布，郵送回収によりアンケート調査を実施した．南牧村には60の分区（本研究では「集落」とする）が存在し，全ての集落・世帯に配布・回収した．調査内容は，世帯属性，災害による被害経験生活質評価，居住意向，自由記述である．自由記述欄の設問文は「南牧村のイメージを，短い言葉や文で，自由に書いてください．」であり，よいイメージ，わるい

表-2　アンケート調査の概要

調査日	配布：2010 年 11 月 1 日 回収：2010 年 11 月 21 日（郵送投函期限）
対象地域	群馬県甘楽郡南牧村全域
対象者	全 1,117 戸の世帯主あるいは代表者（1 戸に複数の世帯が居住している場合があるため，住民基本台帳の世帯数とは異なる）
調査方法	配布：分区長による戸別配布（60 分区，本研究では「集落」とする） 回収：郵送回収
調査内容	1)世帯属性（世帯主属性，世帯構成，住宅の所有形態，自動車保有台数，居住年数） 2)災害による被害経験（2007 年 9 月台風 9 号による被害，それ以前の被害） 3)生活質評価（23 項目，総合評価） 4)居住意向（定住/転居意向，転居意向理由） 5)自由記述欄 設問文「南牧村のイメージを，短い言葉や文で，自由に書いてください．」
回収数	配布数：1,117 票　　回収数：637 票， 回収率：57.0% 有効回収数：631 票，有効回収率：56.5%
調査主体	群馬工業高等専門学校 環境都市工学科 群馬県 県土整備部 都市計画課

イメージの両面を捉えるものとした．本調査は，森田らの研究[16]に用いた調査と同一であり，自由記述データを新たに分析データとして加える．

役場，分区長，住民の協力を得ることができ，有効回収率は約 57%と良好な結果となった．

4. 自由記述データの分析

自由記述データについて，記入有無別の特性の分析，出現する語の分析，出現語間の関係の分析の 3 つを行うこととする．

(1) 自由記述欄の記入有無別の特性

自由記述欄に記入した人は，地域の問題・課題に興味がある，伝えたい事項がある，政策に対し意見があるということを想定し，記入有無別の特性を分析する．特性として世帯属性，災害による被害経験，居住意向，地区特性について分析した．

世帯属性別の特性について**図-4** から**図-7** に示した．世帯主の年齢階層については，記入している人の方が若い傾向がある（有意差あり）．世帯主の職業については記入している人の方が就業が多い（有意差あり）．世帯主の就業地については，記入している人では南牧村以外の人が多いが，有意な差は得られなかった．世帯人数については，記入している人の方が複数人の世帯である傾向があった（有意差あり）．以上より，記入をしている人は，家族をもつ比較的若い就業者である傾向ある．逆に，記入していない人は，無職の独居の高齢者である傾向があると考えられる．

災害による被害経験については，2007 年の台風 9 号の自宅建物被害（**図-8**）をみたが，記入有無別には差異はみられなかった．居住意向（**図-9**）については，記入している人は記入していない人に比べ，村外への転居を考えている人が多い（有意差あり）．地区特性については，最寄小売店までの距離（**図-10**）をみた．記入している人は記入していない人に比べ，小売店までの距離が短い傾向があった（有意差あり）．

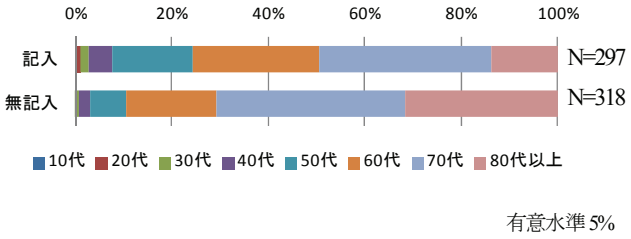


図-4 自由記述欄の記入有無別の世帯主の年齢階層

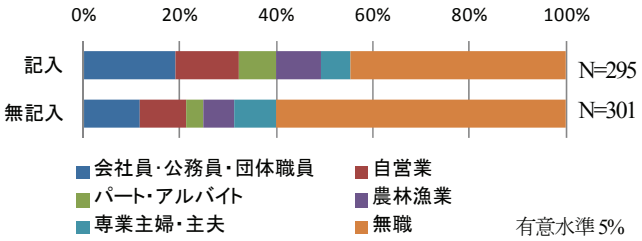


図-5 自由記述欄の記入有無別の世帯主の職業

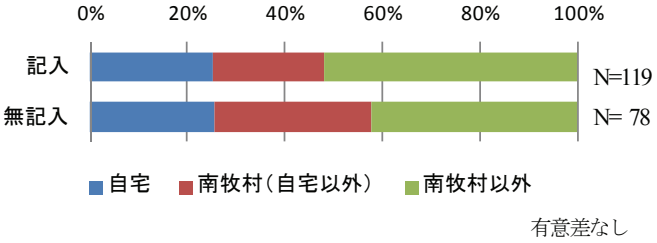


図-6 自由記述欄の記入有無別の世帯主の就業地

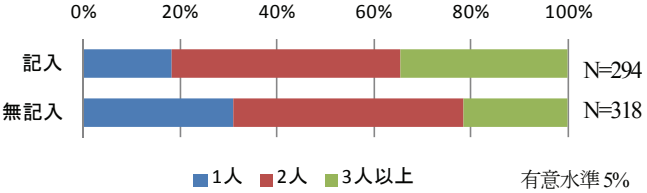


図-7 自由記述欄の記入有無別の世帯人数

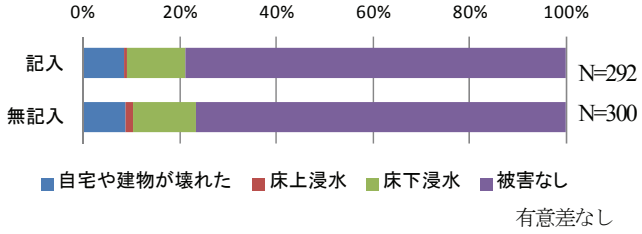


図-8 自由記述欄の記入有無別の災害被害経験（台風 9 号）

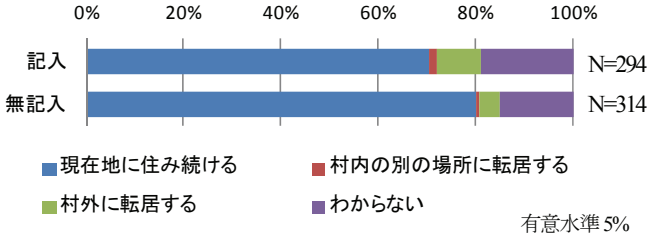


図-9 自自由記述欄の記入有無別の居住意向

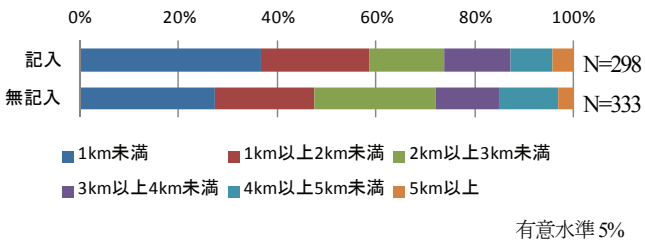


図-10 自由記述欄の記入有無別の最寄小売店までの距離

表-3　自由記述欄の世帯属性別の平均記入語数

世帯主の年齢階層	65 才未満	65 才以上
サンプル数	109	188
語数の合計	1918	4131
平均語数	17.6	22.0
世帯主の職業	就業者	無職
サンプル数	145	150
語数の合計	3113	2903
平均語数	21.5	19.4
世帯人数	1 人	2 人以上
サンプル数	54	240
語数の合計	1067	4940
平均語数	19.8	20.6

(2)　自由記述欄の出現語の特性

自由記述欄の出現語を分析することにより，記入者の意図をできると考えられる．本研究では．形態素解析器として「茶筌」[19]を用い，語を抽出する．

自由記述データにおいては，同一の内容を表す語であっても表記や表現が異なる場合がある．また，1つの語として使用されている語が分割して抽出される場合もある．そのため，本研究では，以下のようにコーディングルールを作成した．

1)「みどり」と「緑」など，同じ意味で表記のしかたが異なる語に同一のコードを与えた．
2)「良い」と「よい」など，同じ意味でも表現が異なる語（類義語）に同一のコードを与えた．
3)「南牧村」は，「南牧」，「村」というように単語が分割されたため，固有名詞のコードを与えた．

表-3 に世帯属性別の平均記入語数を整理した．語数は，記入者の回答に対する熱心さを表すと考えた．世帯主の年齢階層では，65 才以上の人の方が記入語数が多く，世帯主の職業では就業者の方が多く，世帯人数では複数人の世帯の方が記入語数が多い傾向があった．

次に，**表-4** に頻出 50 語を整理した．10 位までの体言は「村」「自然」「水」等の，南牧村の立地に関連する語が現れ，20 位までには「高齢者」「不便」「生活」等の過疎の進む地域に関する語がみられる．用言をみると，「ない」「多い」「良い」等が現れている．

前項までの分析において差異のみられた年齢階層別に出現語を分析する（**表-5**）．65 才未満と 65 歳以上で出現率が異なるのは，「良い」が 65 才以上 13%に対し 65 才以上 32%，「きれい」が 65 才以上 9%に対し 65 才以上 19%と，高齢者の方が肯定的な記述をしているものと想定されるが，体言との関係を把握できないため，記入者の意図は十分把握できない．そこで，次節において出現語間の関係を分析することとする．

表-4　自由記述データの頻出50語

順位	出現語	出現数	順位	出現語	出現数
1	ない	107	26	できる	18
2	村	104		過疎	
3	多い	82		不安	
4	良い	76	29	人口	17
5	自然	64	30	川	15
6	する	60	31	人情	14
7	人	47		道路	
8	水	46	33	場所	13
9	きれい	45		働く	
10	思う	42		おいしい	
11	空気	40		悪い	
12	住む	38	37	出る	12
13	高齢者	33	38	恵まれる	11
14	なる	32	39	とても	10
	南牧村			環境	
16	山	31		心	
17	ある	27		静か	
18	不便	26	43	近所	9
19	生活	25		進む	
20	少ない	24		日本一	
	いる			役場	
22	若い	23	47	今	8
23	高齢化	22		春	
	緑			冬	
25	子供	19		大変(な)	

注：網掛は体言

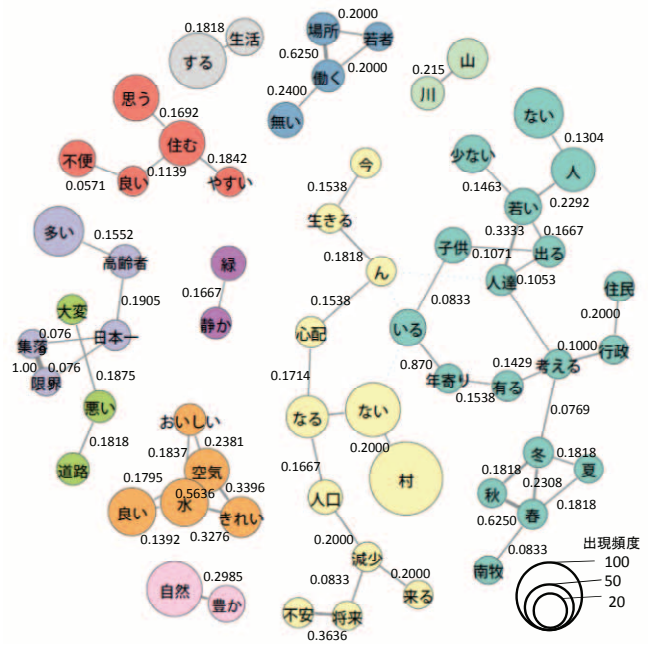
表-5　年齢階層別の頻出語

順位	出現語	65 才未満 N=109		65 才以上 N＝188	
		出現数	出現率 注	出現数	出現率 注
1	ない	33	30%	73	39%
2	村	41	38%	63	34%
3	多い	34	31%	48	26%
4	良い	14	13%	61	32%
5	自然	30	28%	34	18%
6	する	18	17%	41	22%
7	人	11	10%	35	19%
8	水	14	13%	32	17%
9	きれい	10	9%	35	19%
10	思う	16	15%	25	13%
11	空気	10	9%	30	16%
12	住む	10	9%	27	14%
13	高齢者	8	7%	25	13%
14	なる	12	11%	20	11%
15	南牧村	5	5%	27	14%

注：サンプル数に対する出現率

(3)　出現語間の関係

前節の分析では，出現語と出現語の関係について把握できていない．本節では，体言と用言，体言と体言など，出現語間の関係について分析する．これにより，記入者の意図を明確にできると考える．



数値はJaccard係数

図-11　出現語の共起関係（共起ネットワーク）

出現語間の関連性や語がどのような文脈で出現しているかといった語と語のつながりのパターンに着目した分析を行う．出現語間の関係については，自然言語処理で一般的に用いられる語と語の共起関係を分析する方法を用いることとする．本研究では，共起関係を確率的に表すJaccard係数[21]を用いることした．本研究で用いているKH Coderは，Jaccard係数による分析に対応しており，共起関係を視覚的に出力できる．

Jaccard 係数とは，語 ω と語 ω'の共起関係を示し，大きいほどつながりが強いことを示す．$p(\omega \cap \omega')$は語 ω と語 ω'の共起確率を表し，$p(\omega \cup \omega')$は語 ω と語 ω'のいずれかが出現する確率を表す．語 ω と語 ω'の Jaccard 係数は式(1)で与えられる．

$$Jaccad(\omega,\omega') = \frac{p(\omega \cap \omega')}{p(\omega \cup \omega')} \tag{1}$$

Jaccard 係数は，0 から 1 の値をとる．Jaccard 係数の適用方法については，自然言語処理分野において研究が続けられている．図中の語の配置は意味をもたない．

作成した共起ネットワークから，「空気・水ーきれい・おいしい・良い」というよいイメージ，「将来ー不安」，「若者ー働くー場所ー無い」というわるいイメージなど，プリコードデータには設定されていない内容を読み取ることができる．また，「高齢者ー日本一ー多い」「限界ー集落」など，限界自治体である南牧村ならではの出現語間の関係が明らかになった．

5.　自由記述データによるプリコードデータの補完方法の検討

森田らは[16]，南牧村を対象に，客観的変数である世帯属性，地区特性，主観的変数の生活質評価・居住意向，生活質の潜在変数の関係を統合的に検証した．しかし，生活質と居住意向の関係については十分には確認できなかった．この理由について，本研究では，深刻で切実な生活状況にある住民の意向がプリコードデータだけでは把握できていないと考えた．

本研究では，自由記述データを分析し，プリコードデータでは把握できない情報が存在することを明らかにした．これら情報をプリコードデータによる生活質の評価を補完するデータとし，図-12に示すモデル体系を検討することとする．この点については，今後の課題とする．

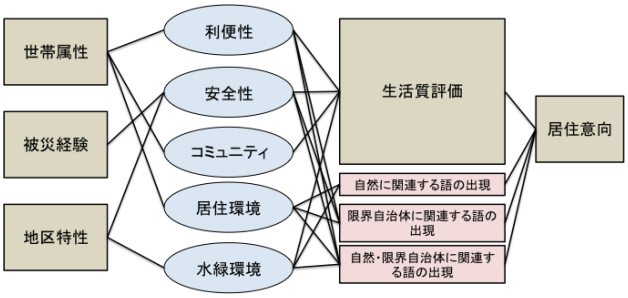


図-12　モデルの体系

6.　おわりに

本稿では，これまで定量的な分析が行なわれてこなかったアンケート調査の自由記述データを有効に活用するために，テキストマイニングを用い定量的な分析を行った．対象地域を限界自治体である群馬県南牧村とし，プリコードでは把握できない情報が自由記述データに存在することを明らかにした．

プリコードデータによる分析を自由記述データで補完する方法の検討については，今後の課題として取り組んでいく．

謝辞：本研究は，科学研究費補助金・基盤研究(c)（課題番号：23560637，課題名：「限界自治体」における生活質と居住意向からみた集約型地域モデルに関する研究，2011-2013年度）の助成を受けたものである．アンケート調査の実施にあたり，南牧村役場，住民の皆様の協力を得た．研究の遂行にあたり，南牧村出身の下風笑美子氏の協力を得た．本分析には，KH Coder[18]（著作権者，樋口耕一氏）を使用した．ここに記し，感謝の意を表す．

参考文献

1) 大野晃：山村環境社会学序説－現代山村の限界集落化と流域共同管理，社団法人農山漁村文化協会，2005.
2) 福田大輔，庭田美穂，屋井鉄雄：疑問型表現自由回答データを用いた社会資本整備に対する市民の関心の抽出方法に関する基礎的研究，土木計画学研究・論文集，Vol.24，pp.139-148，2007.
3) 鄭蝦榮，羽鳥剛史，小林潔司，白松俊：ファセット学習モデルを用いた公的討議のプロトコル分析，土木計画学研究発表会・講演集，Vol.36，2007.
4) 佐々木邦明，飯島陽介，鈴木猛康，秦康範，大山勲：ワークショップ運営支援のためのテキスト分析，土木学会論文集 F4，Vol.66，No.1，pp.57-64，2010.
5) 佐々木邦明，丸石浩一：テキストマイニングを用いたワークショップの討議内容の特徴把握と可視化に関する研究，都市計画論文集，Vol.46，No.3，pp.1039-1044，2011.
6) 塚田伸也，森田哲夫，湯沢昭：委員会の発言から捉えた歩行者の交通空間整備に関する検討，第 31 回交通工学研究発表会論文集，pp.503-506，2011.
7) 佐々木邦明，紀藤舞華，山崎慧太：ブログマイニングからの行動データ抽出・分析可能性とアンケート調査との比較，土木計画学研究・講演集，Vol.43，CD-ROM(145)，2011.
8) 長尚希，室町泰徳，板谷和也：計量的言語処理を利用した大規模交通プロジェクトに関する経験知識の抽出に関する研究，都市計画論文集，Vol.47，No.3，pp.793-798，2012.
9) 森田哲夫，入澤覚，長塩彩夏，野村和広，塚田伸也，大塚裕子，杉田浩：自由記述データを用いたテキストマイニングによる都市のイメージ分析，土木学会論文集 D3，Vol.68，No.5，I_315-I_323，2012.
10) 森山昌幸，藤原章正，杉恵頼寧：過疎地域における公共交通サービスの評価指標の提案，日本都市計画学会都市計画論文集，Vol.8，No.3，pp.475-480，2003.
11) 宮崎耕輔，徳永幸之，喜多秀行，谷本圭志，菊池武弘，高山純一：過疎地域におけるバス運行サービスの変化が地域住民の生活に与えた影響分析に関する研究，土木学会土木計画学研究・講演集，Vol.33，CD-ROM(78)，2006.
12) 谷本圭志，森健治：地方部における定住意向と社会生活環境の関係に関する考察－住民のライフステージに着目して－，環境システム研究論文集，Vol.35，pp.19-27，2007.
13) 日本交通政策研究会：日交研シリーズ A-473，「限界集落」を対象とした中山間地域のモビリティの確保と地域再編戦略に関する研究，2009.
14) 森田哲夫，塚田伸也，佐野可寸志：過疎・高齢地域における集約型居住に向けた人口動向・居住意識の分析－群馬県六合村におけるケーススタディー，都市計画論文集，Vol.45，No.3，pp.511-516，2010.
15) 塚井誠人，桑野将司：中山間地域住民の移住意向と移住要件に関する分析，都市計画論文集，Vol.45，No.3，pp.277-282，2010.
16) 森田哲夫，木暮美仁，塚田伸也，橋本隆，杉田浩：限界自治体の生活質と居住意向に関する研究，社会技術研究論文集，Vol.10，pp.86-95，2013.
17) 樋口耕一：テキスト型データの計量的分析－2 つのアプローチの峻別と統合－，理論と方法，No.19(1)，pp.101-115，2004.
18) KH Coder，http://khc.sourceforge.net/，2013.5.1 閲覧.
19) ChaSen「茶筌」形態素解析器，http://chasen-legacy.sourceforge.jp/，2013.5.1 閲覧.
20) 大野晃：山村環境社会学序説－現代山村の限界集落化と流域共同管理，社団法人農山漁村文化協会，2005.
21) 佐々木靖弘，佐藤理史，宇津呂武仁：関連用語収集問題とその解法，自然言語処理，Vol.13，No.3，pp.150-175，2006.

A STUDY ON EFFECTIVE USE OF THE FREE DESCRIPTIVE DATA USING TEXT MINING
- A CASE STUDY ON NANMOKU-MURA, GUNMA -

Tetsuo MORITA, Syunichi MOROOKA, Shinya TSUKADA
and Takashi HASHIMOTO