

原発避難者の手紙

ウェブサイト記事のテキストマイニング

和光大学 4 年 長谷川将太

目次

問題.....	3
目的.....	3
方法.....	3
1. 分析対象	3
2. 分析手順	4
結果.....	4
1. 基本情報	4
2. 単語頻度分析	5
2-1 名詞・動詞・形容詞	5
2-2 名詞	5
2-3 形容詞	6
2-4 動詞	7
2-5 願望・要望の表現	8
3. 係り受け頻度分析	9
4. ビジュアル集計	10
5. 特徴語抽出	11
6. ことばネットワーク	11
7. 評判抽出	12
考察.....	13
謝辞.....	15
引用文献	15

問題

私たちは、日本で地震が起こるのは当たり前だと思っている。いくら国策によってなされた原発政策があっても一度の巨大地震でその被害は甚大なものになり、復興にかかる時間やお金はとてつ大なものになる。2011年3月11日に発生した東北地方太平洋沖地震による災害および、これに伴う福島第一原子力発電所事故による災害を東日本大震災と呼ぶ。この震災における死者、行方不明者は1万8000人以上と言われており、2018年になった現在でも震災による避難者等は約7万3000人もいると聞く。今でも避難者の数が多いのは原発事故による放射能被害の影響によるものだろう。放射能被害は今後何十年にも渡り私たちの身体に害を与えるものとテレビで散々聞かされて今に至るが、当時の避難者は少ない情報と時間で何を信じ、何を捨て、何を守ったのか。

今回資料として使わせていただく原発避難者の手紙は『福島乳幼児妊産婦ニーズ対応プロジェクト(FnnnP)』という東日本大震災に伴う福島第一原子力発電事故後、放射能汚染による健康被害の不安を抱える乳幼児や妊産婦を含む家族と、避難中の乳幼児や妊産婦を含む家族を対象としたニーズに対応することを目的としているプロジェクトから、避難の苦悩や状況を一人でも多くの人に知ってもらうために募集した手紙である。震災から二年が経過し、誰に何を伝えたいか、何を思っているのかを紐解いていきたい。

本研究では『福島乳幼児妊産婦ニーズ対応プロジェクト(FnnnP)』の「お手紙プロジェクト」で紹介されている原発避難者の手紙全26通をテキストマイニングし、原発避難者の心情や思いを探ろうと考えている。

目的

『福島乳幼児妊産婦ニーズ対応プロジェクト(FnnnP)』の「お手紙プロジェクト」で紹介されている原発避難者の手紙を用いて避難者の願いや家族への思いを検討する。実際に原発の影響で現地から避難してきた人たちや、我が子の被爆を心配する親の声を理解、分析することによって原発被害の悲痛さや今後の日本の原発被害者への理解を助長するとともに、災害によって生まれた声の風化を防止することができるだろう。

方法

1. 分析対象

今回は、『福島乳幼児妊産婦ニーズ対応プロジェクト』のウェブで公開されている「原発

避難者の手紙」を分析の対象とした。

2. 分析手順

分析手順としては、『福島乳幼児妊産婦ニーズ対応プロジェクト』のウェブで公開されている「原発避難者の手紙」26通の文章を Word に貼り付け、誤字や脱字を修正し、タブ区切りテキストを Word で作成した。次に完成したものを Excel ファイルに変換し、ミスがなにか確認した後に保存した。そしてそのファイルを「Text Mining Studio ver.6.1.2」で読み込んで、テキストの基本統計量、単語頻度分析、係り受け分析、ビジュアル集計、特徴語抽出、ことばネットワーク、評判抽出の順に行った。その際、単語頻度分析では「名詞・動詞・形容詞」と「名詞」、「形容詞」、「動詞」、「願望・要望」、と5つに分けて分析を行った。

『福島乳幼児妊産婦ニーズ対応プロジェクト』 「原発避難者の手紙」は原発事故で被災した方々の手紙を「今、あなたに伝えたい」というテーマで平成23年7月15日まで募集し、ウェブへの掲載を可能と答えた方たちの手紙26通をまとめたものである。

結果

1. 基本情報

表1は手紙の基本情報である。ここでは総行数、平均行数、総文数、平均文長、延べ単語数、単語種別数を示す。まず、総行数は分析対象の原発避難者の手紙の数を表しており、26通であった。次に、一通あたりの文字数を表す平均行長(文字数)は1404.4文字であった。この全手紙の総文数は1109文、その平均文長(文字数)は32.9であった。内容語の延べ単語数は7779個、単語種別数は2856個だった。そして、語彙の豊かさを表す指標であるタイプ・トークン比(Type-Token Ratio, TTR)は、 $TTR = \text{単語種別数} / \text{延べ単語数}$ から算出したところ0.367と示した。

	項目	値
1	総行数	26
2	平均行長(文字数)	1404.4
3	総文章数	1109
4	平均文章長(文字数)	32.9
5	延べ単語数	7779
6	単語種別数	2856

表1 手紙の基本情報

2. 単語頻度分析

2-1 名詞・動詞・形容詞

図1は手紙を単語頻度分析し、名詞・動詞・形容詞の上位38の単語を横棒グラフで表したものである。この分析を行うことで、手紙の中ではどの単語が多く用いられているかを明らかにし、避難者の考えを汲み取る。図1を見ると「言う」、「思う」という単語が最も多く18回使われていた。次点で「家族」、「良い」という単語が17個と2番目に多く使われていることがわかった。また、「いる」、「福島」が15回、「人」、「生活」、「避難」が14回と同頻度扱われていることも分かった。

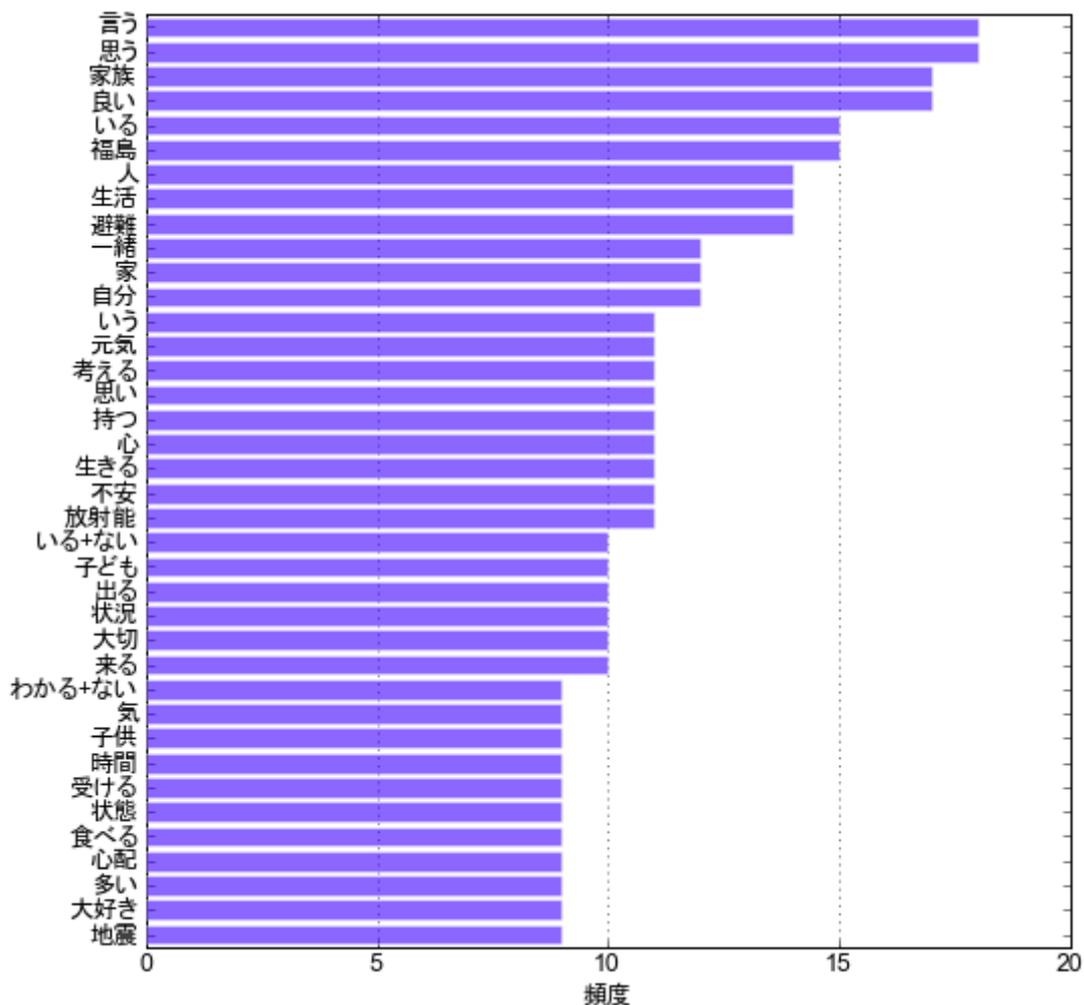


図1 手紙の単語頻度分析 (名詞・動詞・形容詞)

2-2 名詞

図2は手紙を単語頻度分析し、名詞上位42の単語を横棒グラフで表したものである。こ

の分析を行うことで、手紙の中ではどの名詞が多く用いられているかを明らかにし、避難者の考えを汲み取る。図2を見ると、名詞は「家族」という単語が最も多く17回使われていた。次点で「人」という単語が14個と2番目に多く使われていることがわかった。また、「家」、「自分」が12回、「元気」、「思い」、「不安」、「放射能」が11回と同頻度扱われていることも分かった。

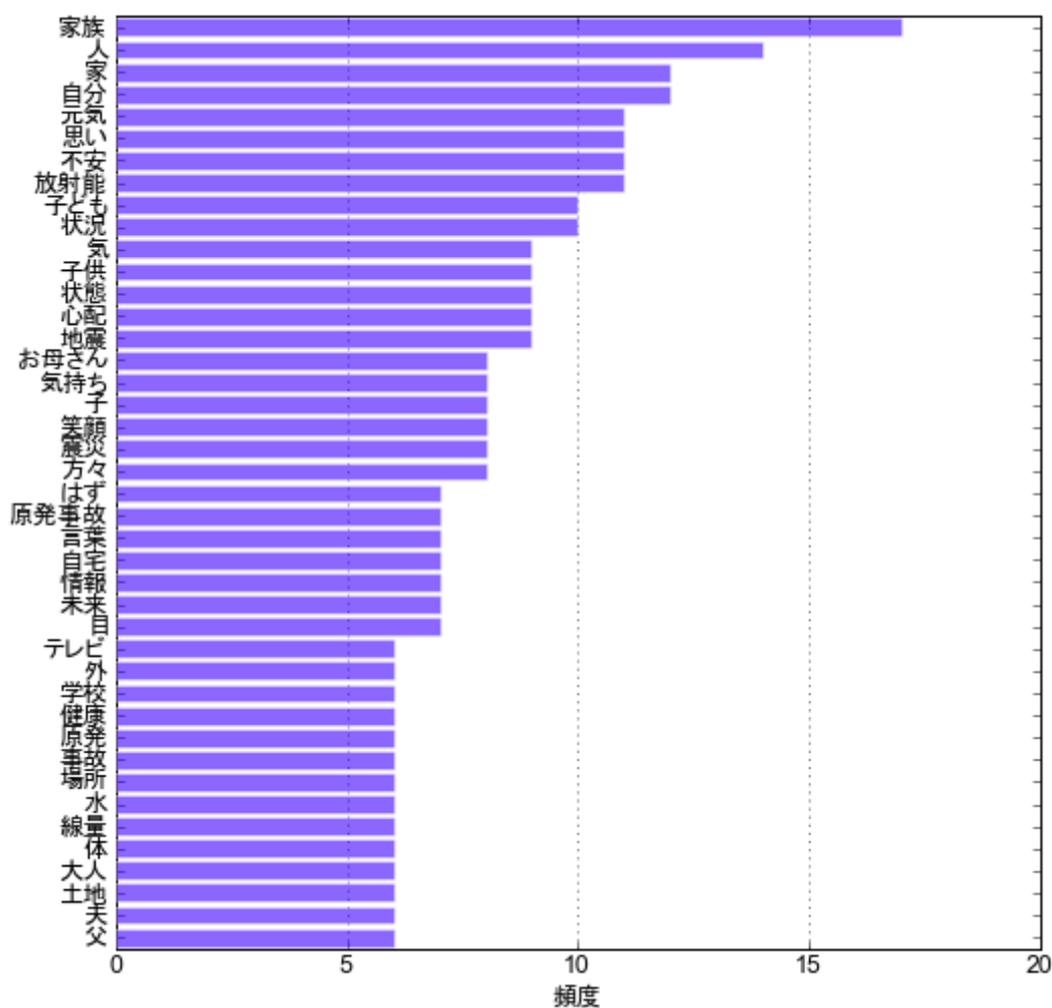


図2 手紙の単語頻度分析（名詞）

2-3 形容詞

図3は手紙を単語頻度分析し、形容詞上位39の単語を横棒グラフで表したものである。この分析を行うことで、手紙の中ではどの形容詞が多く用いられているかを明らかにし、避難者の考えを汲み取る。図3を見ると、形容詞は「良い」という単語が最も多く17回使われていた。次点で「元気」、「不安」という単語が11個と2番目に多く使われていることがわかった。また、「大切」が10回、「心配」、「多い」、「大好き」が9回と同頻度扱われていることも分かった。

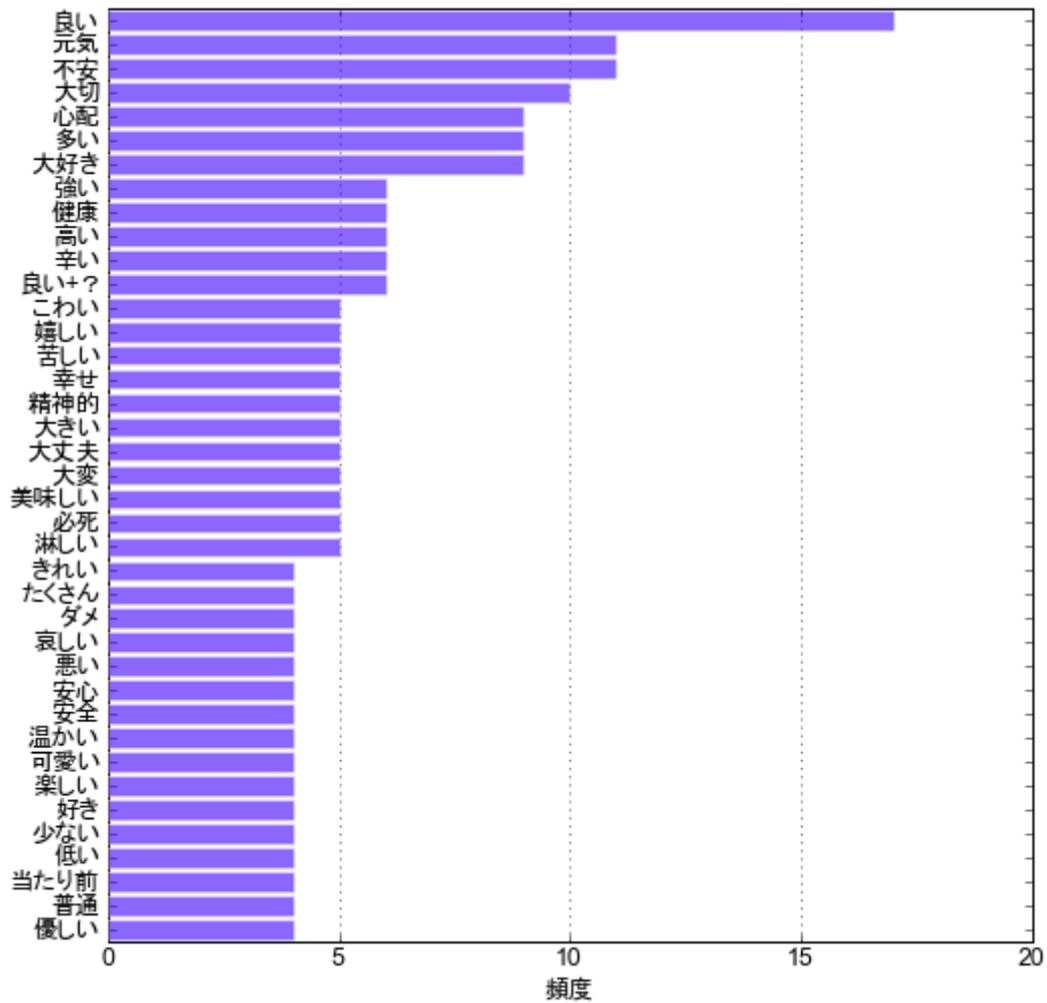


図3 手紙の単語頻度分析（形容詞）

2-4 動詞

図4は手紙を単語頻度分析し、動詞上位37の単語を横棒グラフで表したものである。この分析を行うことで、手紙の中ではどの動詞が多く用いられているかを明らかにし、避難者の考えを汲み取る。図4を見ると、動詞は「言う」、「思う」という単語が最も多く18回使われていた。次点で「いる」という単語が15個と2番目に多く使われていることがわかった。「生活」、「避難」が14回、「一緒」が12回、また、「いう」、「考える」、「持つ」、「心」、「生きる」が11回と同頻度扱われていることも分かった。

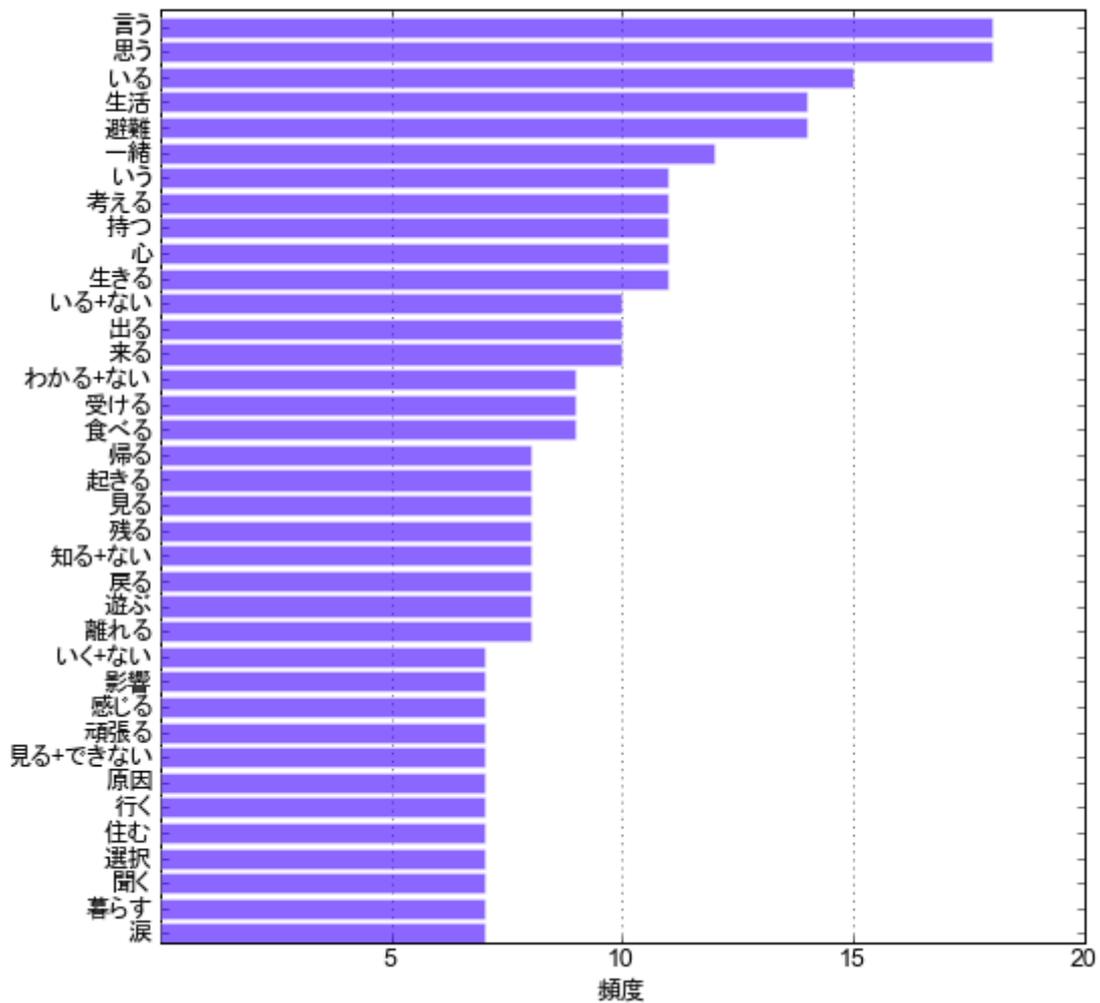


図4 手紙の単語頻度分析（動詞）

2-5 願望・要望の表現

図5は手紙を単語頻度分析し、願望・要望上位12の単語を横棒グラフで表したものである。この分析を行うことで、手紙の中ではどの願望・要望が多く用いられているかを明らかにし、避難者の考えを汲み取る。図5を見ると、願望・要望は「守る+したい」という単語が最も多く9回使われていた。次点で「帰る+したい」という単語が7個と2番目に多く使われていることがわかった。また、「暮らす+したい」、「戻る+したい」が4回、「いる+したい」、「知る+したい」、「伝える+したい」、「返す+したい」が3回と同頻度扱われていることも分かった。

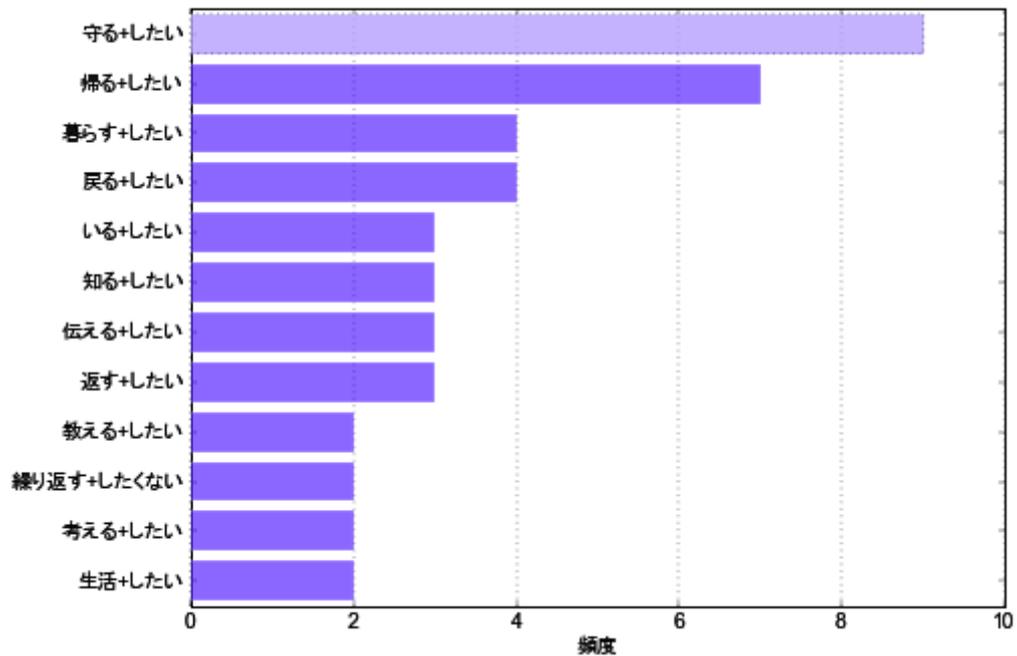


図5 手紙の単語頻度分析（願望・要望の表現）

3. 係り受け頻度分析

図6は使われた単語の中で、どの単語との係り受けが多いのかを係り受け頻度分析を行って横棒グラフにして表したものである。グラフの横軸の数値は、係り受け関係にある単語の出現項数（頻度）を表している。この分析を行うことで、特定の単語との繋がりを明らかにし、どのようなことが手紙の中で語られているかを見出す。図6を見ると、「人—いる」が最も多く、6通であった。次点で「お父さん—いる+ない」が5通と係り受け頻度が多かった。また、「福島」という単語が複数回係り元単語として機能しており、「福島—いる」と「福島—帰る」、「福島—帰る+したい」、「福島—戻る」が同数の3通あり、全部で4通りの係り元単語として扱われていることがわかった。

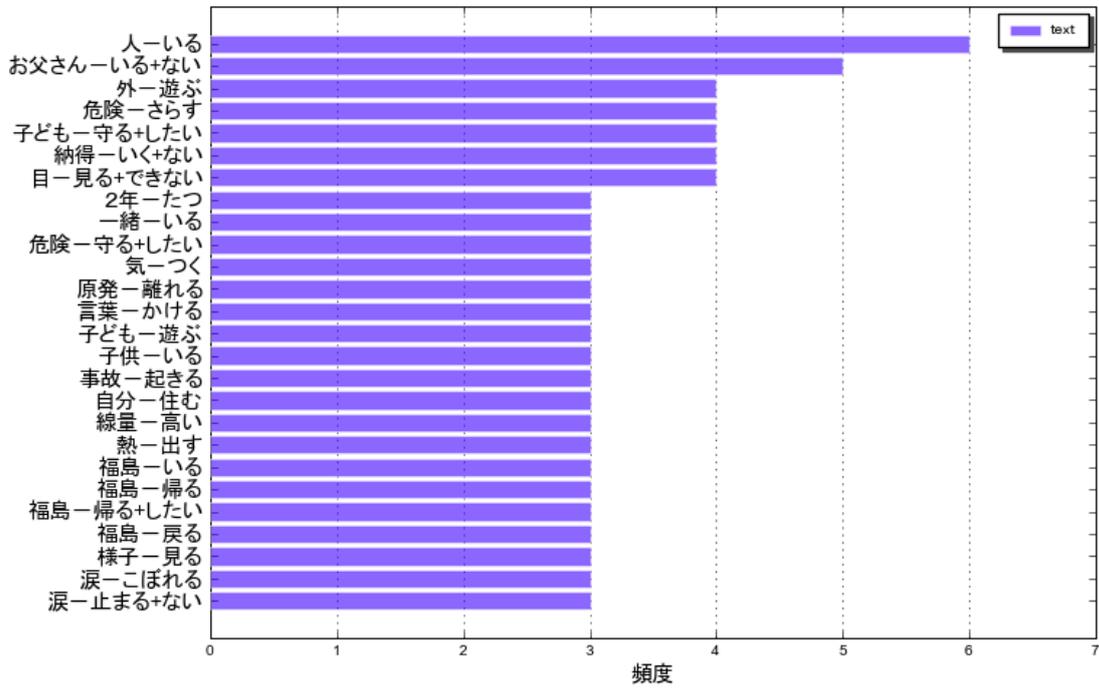


図6 手紙の係り受け頻度分析

4. ビジュアル集計

図7は原発避難者の手紙26通を男女別にし、グラフに表したものである。男性は5通、女性は21通となったが、男性の5通はいずれも「Cさん」という一人の男性から送られており、女性も「うたはさん」から3通、「Dさん」から2通と男女で合計19名から手紙を送られていた。

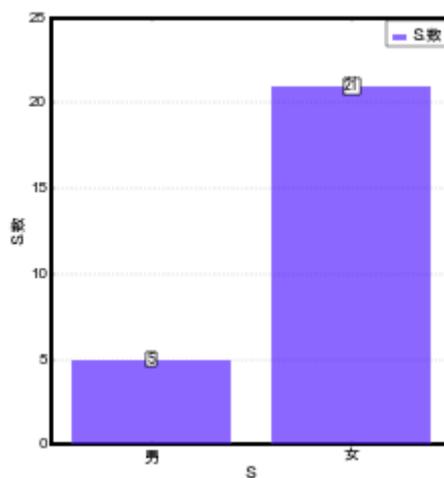


図7 手紙のビジュアル集計

5. 特徴語抽出

手紙の頻度上位 20 位を対象とした特徴語抽出を行なった結果、男女ともに「子供」やそれに同義の言葉について頻繁に使用されていることが明らかとなった。また、頻度上位の単語を男女別に見ると、女性は全体的に状況を想定させる単語が目立つが、男性は全体的に人に使われる単語が多いことが分かった。

	S-女	指標値 S-女	S-男	指標値 S-男
1	避難	10.132	子どもたち	25.366
2	福島	9.243	父	25.136
3	生活	8.980	体	15.760
4	思う	8.354	愛	13.029
5	放射能	8.059	幼い	13.029
6	子供	7.829	弟	12.798
7	人	7.401	心	12.075
8	不安	6.678	時間	11.417
9	母	5.526	笑顔	11.417
10	危険	5.066	大好き	11.417
11	子供たち	5.066	気	11.186
12	住む	5.066	淋しい	10.956
13	原発	4.605	食べる	10.726
14	地震	4.375	見る	9.575
15	戻る	4.375	可能	8.686
16	震災	3.914	自主避難	8.686
17	生きる	3.914	由希	8.686
18	考える	3.684	写真	8.455
19	自宅	3.684	顔	8.225
20	知る+ない	3.684	姿	8.225

表2 手紙の特徴語抽出

6. ことばネットワーク

複数のことばからなる意味的なかたまりを分析することで、手紙の中でどのようなことが話題になっているかを明らかにした。図8より、手紙では「福島」「人」「いる」「子ども」「危険」の5つに関する話題が多く語られていたことが明らかになった。「子ども」と「子供」は同義であると考え、繋がっているのは「いる」「守る+したい」「遊ぶ」の3つになる。また、「福島」の周りには「帰る」「帰る+したい」「戻る」といった同じような意味の言葉が繋がっており、手紙を書いた避難者の気持ちが推測できる。

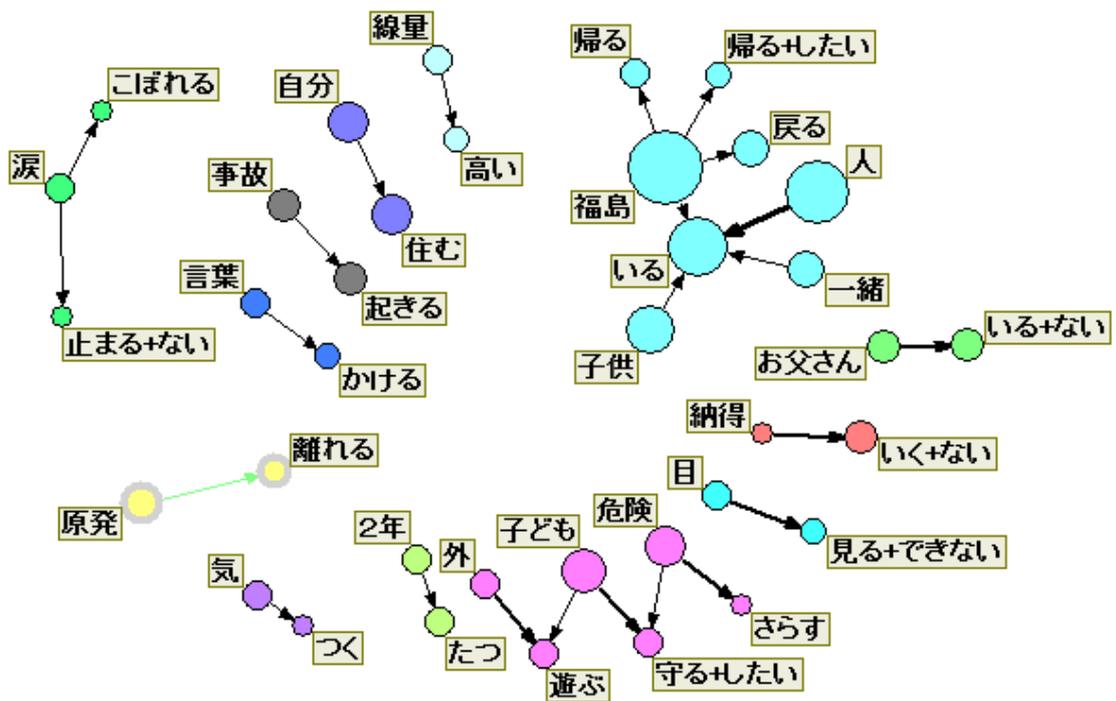


図8 手紙のこぼネットワーク

7. 評判抽出

「影響」と「人」は、ポジティブな単語として頻繁に使用されていた。また、「生活」と「思い」は、ネガティブな単語として頻繁に使用されていた。

手紙において、出現回数の多い上位21位の単語は図9の通りである。最もポジティブに使用されている単語は「影響」と「人」であり、全7通で使用されているうち、どちらもポジティブな単語として使用されていたのは6通、ネガティブな単語として使用されていたのは1通である。また、一番使用されている単語は「思い」で13通使用されているうち、ポジティブな単語として使用されていたのは3通、ネガティブな単語として使用されていたのは10通である。

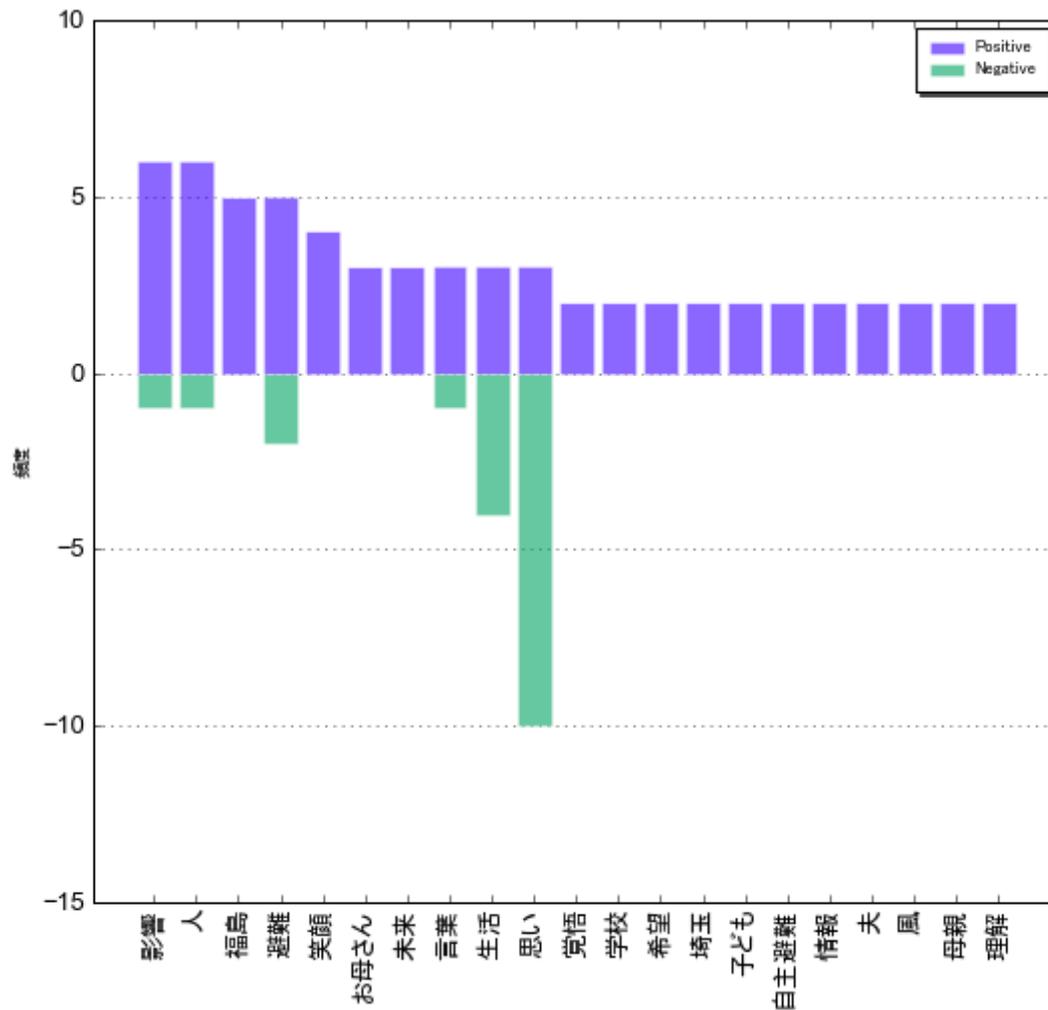


図9 手紙の評判抽出

考察

本研究では、主に単語の使用頻度や言葉の繋がりについてテキストマイニングにより量的に分析を行い、原発避難者の手紙を通して避難者の願いや家族への思い等の心情を考察することを目的とした。

この分析を行った結果、単語頻度分析で全体的に多様された単語は「言う」「思う」「家族」「人」「良い」だった。この中でも特に顕著だったのは「言う」「思う」であり、特に「思う」は特徴語抽出においても手紙の頻度上位に入っていることが見て分かる。これは「今、あなたに伝えたい」という手紙の募集テーマが起因している。思いを伝えることがこの手紙本来の本分なのだから、これは必然と言える。また、願望や要望の表現では「守る+したい」の

形が多く見られた。これは主に放射能汚染による健康被害から乳幼児等の子どもや妊産婦を守りたいという意味合いで使われているのであろう。さらに名詞の単語頻度分析では「家族」が最も頻度が高いことから、家族全体に対しても「守る+したい」が用いられている。

評判抽出を見ると「影響」や「人」といった単語がポジティブに使われている単語として目立つが、それよりも一番頻繁に使用されている単語が「思い」だとわかる。全13通のうちポジティブに使用されているのが3通、ネガティブに使われているのが10通と他を圧倒する多さだが、これには当時の福島の状況や政府の対応の甘さから感じた避難者の悲痛さが感じ取れる。またそれ以外にも先の見えぬ現状の生活を思えば、ネガティブに考えてしまうことは想像に難くない。

次に「福島」という単語に注目してみると、ことばネットワークでは一番話題として使われている単語なのが一目瞭然となっており、特徴語抽出では女性における手紙の頻度が上位二番目にあることがわかった。このことから分かった当時私たちに手紙で伝えたかったことは、自分たちの置かれている状況や苦悩だけでなく、自分たちが暮らしていた福島の状態や環境、またはそれに伴う福島に対しての不安だと考えられる。特徴語抽出では「不安」が女性における手紙の頻度の上位八番目にあることがその裏付けになるだろう。

さらに係り受け頻度分析とことばネットワークを見ると、「福島」という単語が複数回係り元単語として機能しており、「福島—いる」と「福島—帰る」、「福島—帰る+したい」、「福島—戻る」の全部で4通りの係り元単語として扱われていることがわかる。これは、戻りたいという苦悩の部分にあたる、故郷である福島に対しての各々の正直な気持ちであると考えられる。しかし、地震や津波で何もかも破壊され、福島第一原子力発電所の事故により放射能被害までが明らかとなった福島に果たして本当に戻ってもいいのか、福島にこのまま居続けていいのかの葛藤も大きい。帰りたいけど帰れない人がいる、これが現実である。

「今、あなたに伝えたい」が手紙のテーマのため、大多数の手紙で自分の状況や置かれている環境等が綴られていると考えていたがそれだけではなかった。係り受け頻度分析を見ると「人—いる」が6通と最も多く、ことばネットワークの図を見ても分かるように単語としての話題もかなり目立つ結果となった。これは自分の事だけでなく、家族や他人について思いを馳せる人が多かった結果と言える。評判抽出を見てみると「人」は全7通で使用されているうち、6通でポジティブな単語として使われていることがわかる。場所を失い、時間が経過しても人々は思い合い、こうして繋がることができることを、このサイトは多様なケースから表現豊かに示している。

今回の研究では、実際に原発被害によって避難した人たちから送られてきた手紙を分析した。しかし男性からの手紙は5通あるとはいえ、その実一人の男性から送られているため内容に広がり弱いので解釈には注意が必要である。また、あくまで乳幼児や妊産婦を含む家族と、避難中の乳幼児や妊産婦を含む家族のみを対象として募集した手紙の公開がこの分析対象のサイトの目的である。したがって原発被害から避難された全ての人に当てはめるには限界があるかもしれない。このように今回当てはまらなかった人々を分析するこ

とが出来なかった限界があるとはいえ、原発避難者の手紙という貴重な手紙を分析することにより、当時の避難した人々の願いや思いが明らかになったことが本研究の成果である。

謝辞

学生研究奨励賞の原稿作成にあたり、「Text Mining Studio バージョン 6.1.2」を使用させて頂きました NTT データ数理システム様に感謝いたします。また、本論文を作成するにあたり、指導教官の伊藤武彦教授から丁寧かつ熱心なご指導を賜りましたことに感謝いたします。

引用文献

福島乳幼児妊産婦ニーズ対応プロジェクト(FnnnP) お手紙プロジェクト
<http://tegamifukushima.blog.fc2.com/blog-category-4.html>