

硫黄島戦争体験記のテキストマイニング

佐野誠(和光大学)

問題

太平洋戦争の末期に行われた硫黄島の戦いは1945年の2月19日から1945年3月26日まで東京小笠原諸島で行われた日米間の戦いである。この戦いで日本側は守備兵力20933名のうち96%の20129名が戦死あるいは行方不明となった戦いであった。現在も遺骨が回収されたのは2008年時点でわずか8638人である。この戦いの中で捕虜となったのが3月末までに200名、終戦までで1023名であった。また、最後の生存者として、終戦から4年後に2名が投降したのは有名な話である。その戦闘の壮絶さは硫黄島での戦争体験者の手記や経験者からインタビューなどを通じて書かれた書籍などを通して知ることができる。太平洋戦争から71年を経た今、当時の硫黄島の激戦を知ることは戦争という体験から離れて育った自分達の世代にも戦争の悲惨さを知ることや、当時の日本の状況を知ることにも繋がる。しかし、当時硫黄島の戦いに参加した人本人の手記は少なく、次第に戦争のことも風化しつつある。硫黄島の戦争体験者達は当時どういった心情で硫黄島での戦争状況下で生き残り、何を見たのか。

本研究では硫黄島の戦いを体験した3名の手記をテキストマイニングし、3人の心情や行動の関連性を探ろうと考えている。

目的

硫黄島での戦争経験者3名の手記から見える当時の硫黄島の状況と3人の心情、また日本とアメリカの動きを検討する。3人がどう感じて何を見て硫黄島の戦いを生き延びたのか、また3人の共通の心情を理解、分析することで当時の戦争状況下について考察する。

方法

1. 分析対象

秋草鶴次（2011）『硫黄島を生き延びて』 清流出版、川相昌一（2007）『硫黄島体験記』 光人社、石井周治（1982）『硫黄島に生きる』 国書刊行会を対象とした。秋草氏の内容は主に、全体の3分の2が硫黄島での体験で残りは米軍捕虜になってからの体験であった。部隊は通信隊であった。川相氏の内容は全体の3分の2ほどが硫黄島での戦闘で、残りがやはり米軍捕虜になってからの体験であった。また、部隊は秋草氏と同じく通信隊であった。最後の石井氏は全体の3分の1ほどが硫黄島での体験で、残りが米軍捕虜になってからの体験であった。また、石井氏は師団直轄の野戦病院での衛生兵であった。

今回この3冊を選んだ理由は個人的に硫黄島の戦いに興味があったこと、前述した通り貴重な記録であると同時に、硫黄島での体験と捕虜になってからの体験が含まれていたからであり、また死と隣合わせの戦場でどういったことを考え感じているのかがよく記されていると思ったからである。

2. 分析手順

本書のPDFファイルをPCソフト「読取革命 ver.15」で文章ファイルに変換、タブ区切りテキストにしてExcelファイルにしたものを「Text Mining Studio ver.5.2」で分析した。

はじめに、対象の本を裁断し、スキャナーでPCに読み込んでPDFファイルにする。さらに、そのファイルを「読取革命 ver.15」で読み込んで文章ファイルに変換した。次に、その文章ファイルを「Word2013」に読み込み、目次など不要な部分を取り除き、誤字や脱字、文字化け、乱丁の部分修正した。そして、文章を各著者ごとに分け、それぞれの間タブを入力してタブ区切りテキストにした。それらの文章をコピーして、「Excel2013」に1行空けて2行目から貼り付けのオプションの「貼り付け先の書式を合わせる」を指定、ペーストした。そして、空けた1行目のAには「行ID」、Bには「著者」、Cには「出版年月」、Dには「タイトル」、Eには「章」、Fには「本文」と入力した。3人の著書の「まえがき」と「あとがき」には章タイトルが無かったため、その2つタイトルは省いた。

「Excel97-2003ブック」の形式で保存した前述のファイルを「Text Mining Studio ver.5.2」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、対応バブル分析の順に行った。その際、単語頻度分析では上位20を抽出し、名詞別、動詞別、形容詞別に分けたものも抽出した。係り受け頻度分析では属性を「著者」と選択し、上位20語を抽出した。対応バブル分析において、動作を「属性とことばの関係を図示する」、属性を「著者」と設定し、上位20を抽出した。これら以外は「オリジナル設定」の「品詞設定」からそれぞれ名詞、動詞、形容詞ごとにまとめ、抽出した。

今回対象にした3冊は戦後から70年以上が経ち、戦争が風化してしまっている現代において激戦の地であった硫黄島での3人の体験を記している貴重な記録である。

結果

1. 基本情報

表1は3冊の基本情報である。ここでは総行数、平均行数、総文数、平均文長、延べ単語数、単語種別数である。まず、総行数は分析対象の3人の著書の項数を表しており、67項であった。次に、1項あたりの文字数を表す平均行長(文字数)は2614文字であった。この著書の総文数は10985文、その平均文長(文字数)は15.9であった。内容語の延べ単語数は72728個、単語種別数は15426個だった。

	項目	値
1	総行数	67
2	平均行長(文字数)	2614.4
3	総文数	10985
4	平均文長(文字数)	15.9
5	延べ単語数	72728
6	単語種別数	15426

表1 3冊合計の基本情報

2. 単語頻度分析

図1は本書を単語頻度分析し、上位20の単語を横棒グラフで表したものである。図2は名詞別にしたもので、図3は動詞、図4は形容詞で絞り分析したものだ。この分析を行うことで、著書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、著者の心情を汲み取る。図1を見ると「思う」という単語が最も多く次点で「見る」という単語が多く、その次に「いう」が続く形になった。

図2の名詞別に分けたものを見ると、3人中で共通して「我々」という単語が最も多く361個用いられた。次点で「ここ」という単語が254個と2番目に多く使われていた。また、「それ」が252個、「人」が234個使われていることも分かった。「壕」という単語は6番目に多く使われていて、191個、その次に硫黄島という単語が多く使われていた。図3は動詞を単語頻度分析したものだが、上位3つの単語は図1のものと変わらなかった。図4は形容詞別のもので、「良い」が最も多く次に「無い」が多く、また、「多い」がその次に多く使われていた。

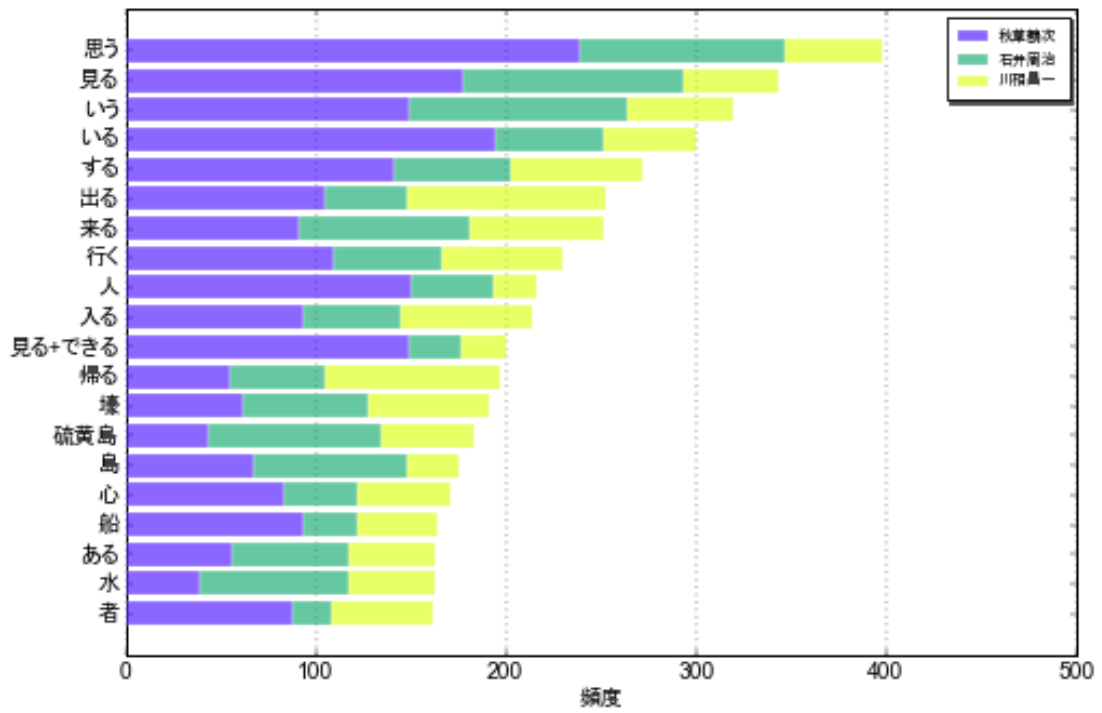


図1 単語頻度分析(著者別)

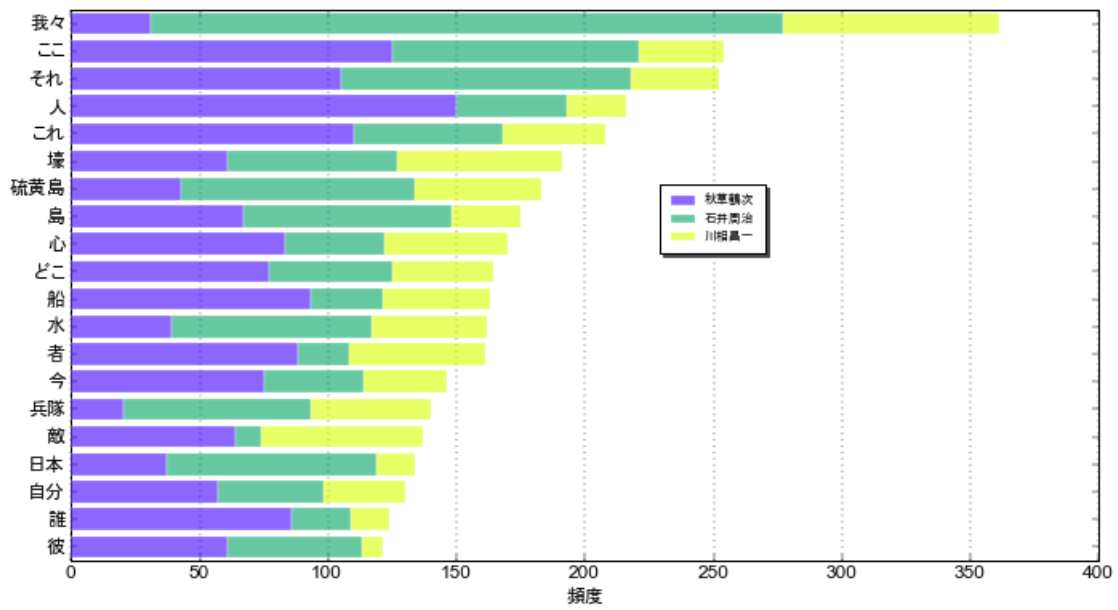


図2 単語頻度分析(名詞・著者別)

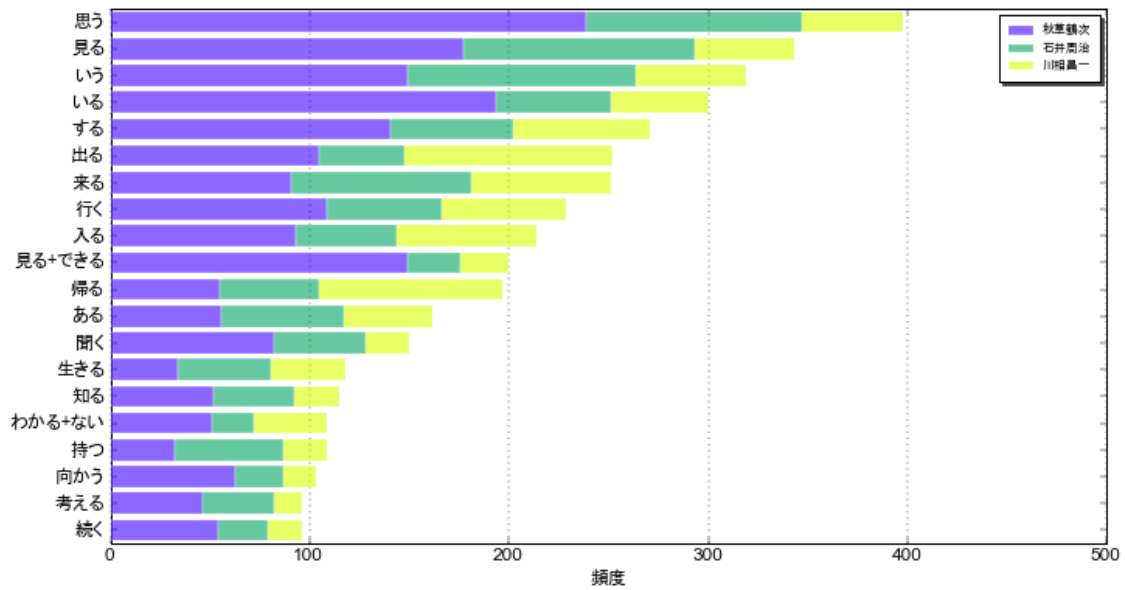


図3 単語頻度分析(動詞・著者別)

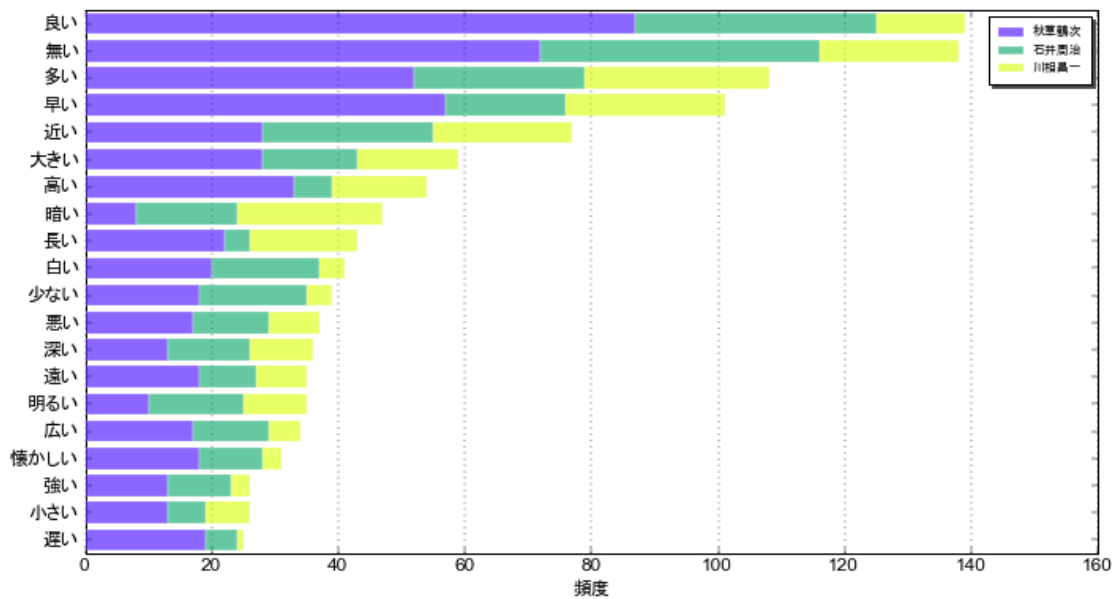


図4 単語頻度分析(形容詞・著者別)

3. 係り受け頻度分析

図5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、どのようなことが著書の中で語られているかを見出す。図5を見ると、「者—いる」が最も多く17項で、次点で「音—する」が多く、14項であった。また、「声—聞く」

と「音—聞く」に対して「聞く—できる」も係り受け頻度が高いことも特筆される。さらには「壕—出る」や「敵—攻撃」、「兵隊—いる」などの現代では使われないような単語は三人とも使っていることが分かる。

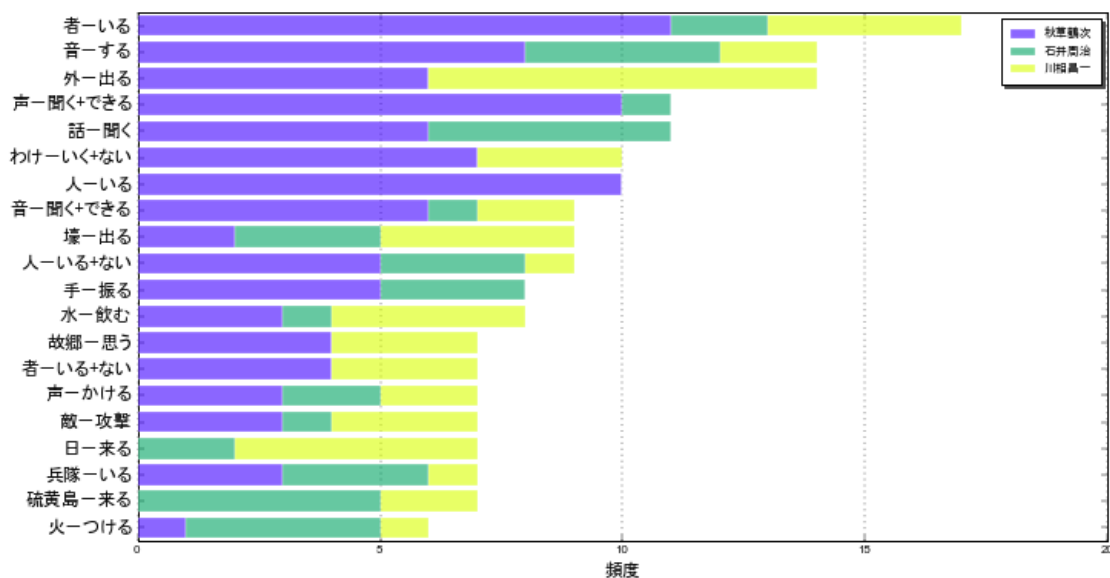


図 5 係り受け頻度分析

4. 対応バブル分析

対応バブル分析では、著者と単語の関係性を分析した。3人の著者の間の言葉の関係性をバブルの大きさで表す。対応バブル分析でも単語頻度分析と同じように単語別、名詞別、動詞別、形容詞別で分析を行った。図 6 を見ると、3人が共通して使用している単語に「水」という単語と「無い」という単語があり、水や物資がどの部隊も不足していたことがわかる。また、図 7 の名詞別を見ると「今」が共通して使われていることが分かる。3人が「今」を使用している部分は、「今あの人はどこにいるだろうか」のような表現に使われていることが共通してあり、それは図 7 の「今」に「どこ」が少し重なっていることから確認が取れる。次に、図 8 を見ていくと動詞の多くは川相さんがよく使用していたことが分かる。最後に図 9 の形容詞別を見ると「深い」が一番共通して使われており、これは特に「壕が深い」のように使用されている。「壕」は名詞別でも石井氏、川相氏が多く使用しており、戦時中、栗林中将が突撃を禁止し、壕での持久戦に持ち込むように出した指示が兵士に行き渡っていたことが分かる。また秋草氏はアメリカ兵によってガソリン入りの海水を投げられた壕の中で水に浮いているところを気づいた犬とアメリカ兵に発見される形で生存している。

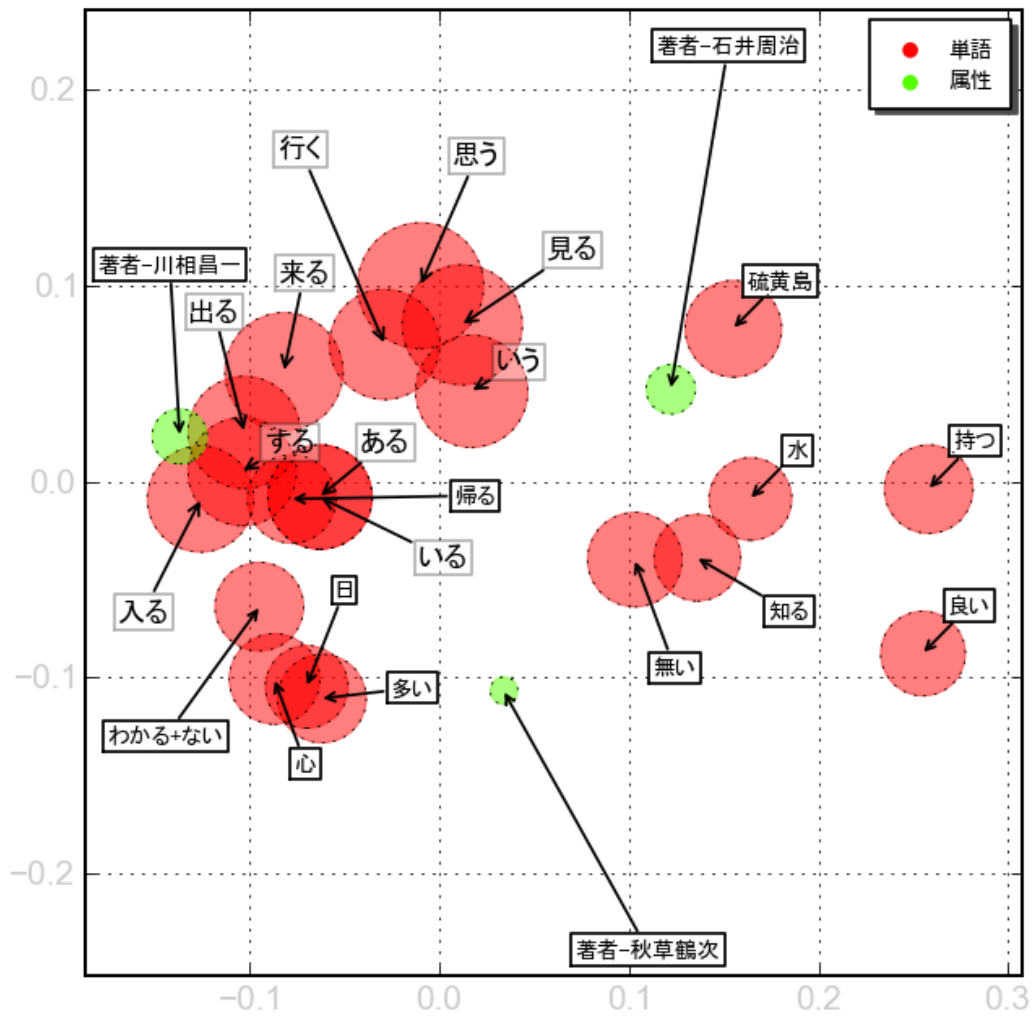


図 6 対応バブル分析(名詞・形容詞・動詞×著者)

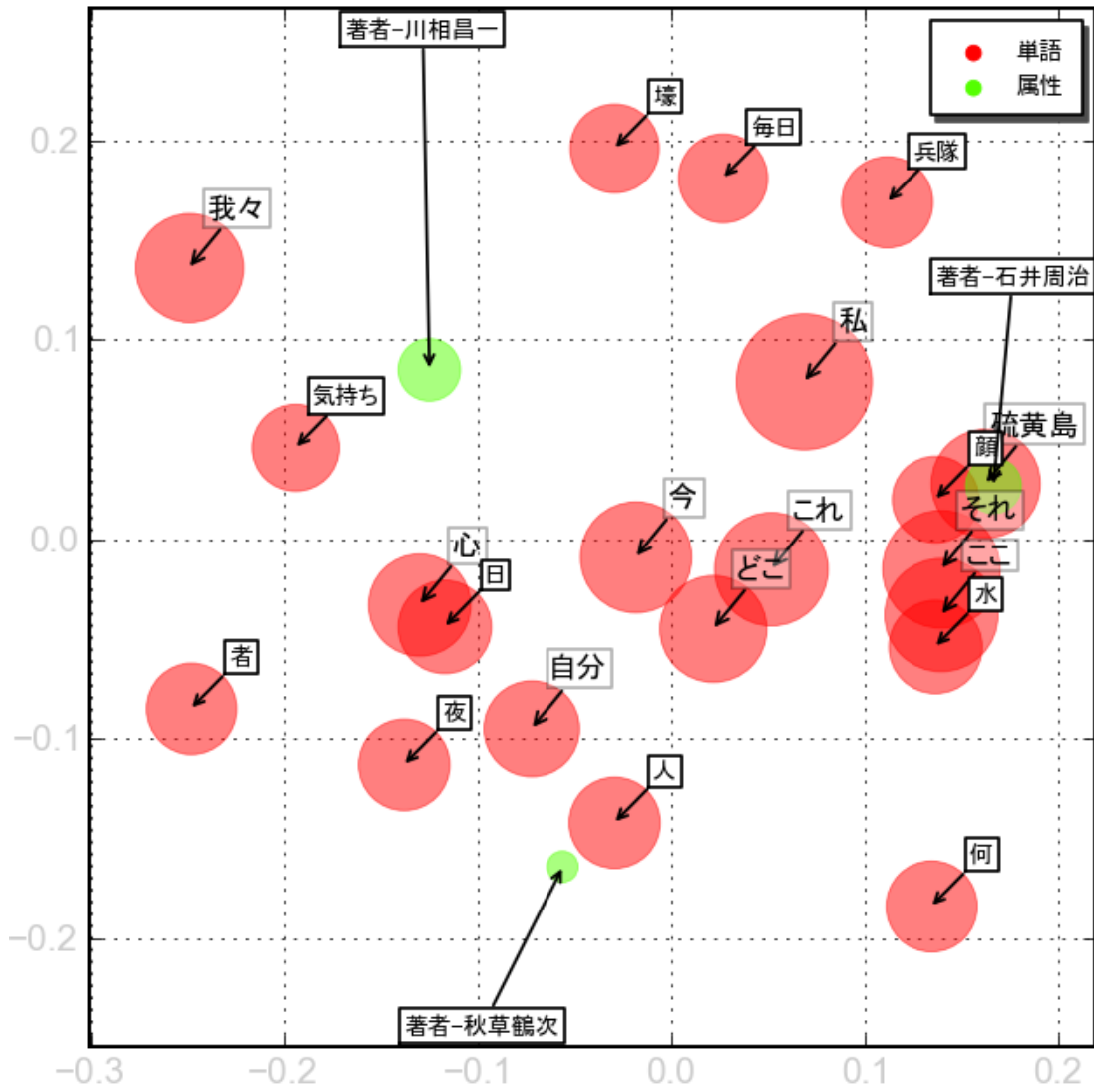


図7 対応バブル分析(名詞×著者)

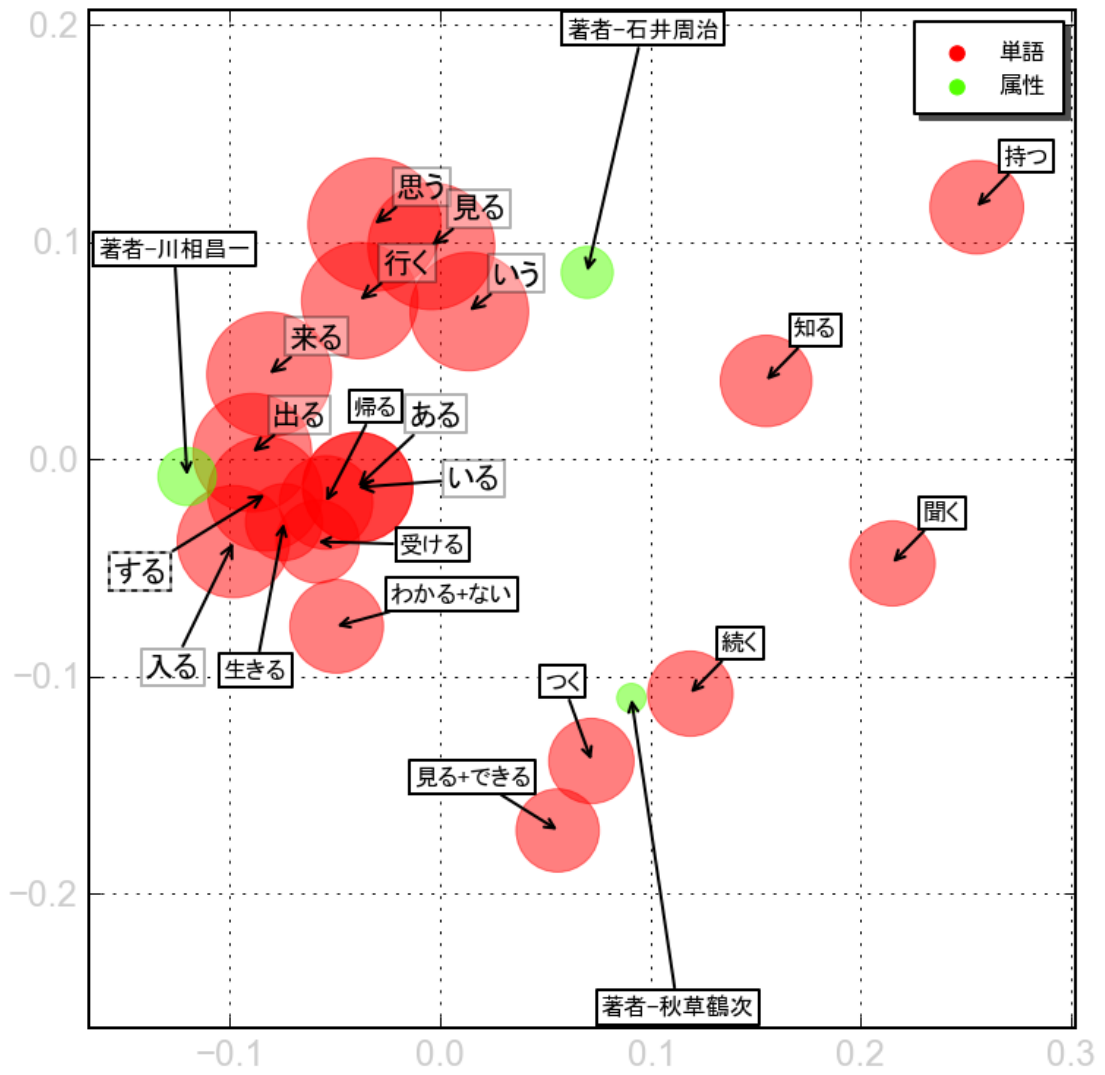


図 8 対応バブル分析(動詞×著者)

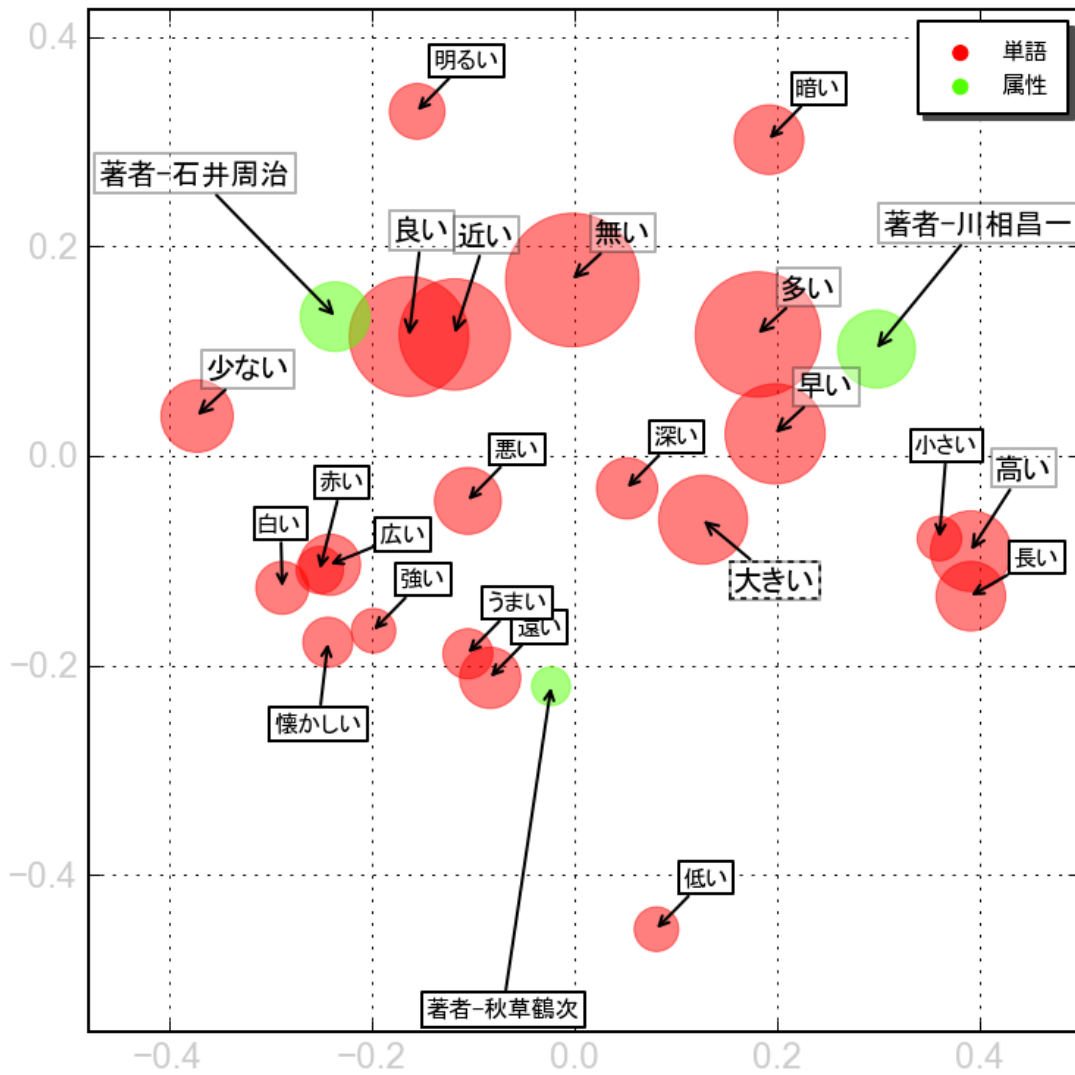


図9 対応バブル(形容詞×著者)

考察

1. 硫黄島の戦いを体験した3人の特徴

今回対象にした3冊は硫黄島の戦いから生存した3人の貴重な記録である。その3人の単語頻度解析で特に注目できるのは「壕」、「水」である。硫黄島の戦いで日本軍守備隊の最高指揮官である栗林中将はいわゆる玉砕覚悟の「バンザイ突撃」を禁止し、壕の中での持久戦にもちこんだ。そのため、「壕」という単語を原文参照すると、硫黄島での体験の多くは壕の中で起きていることが明らかになった。

そして単語頻度分析でもこの「壕」は出てきていて、当時の日本軍の伝達が行き渡っていたことが分かる。「壕」という単語は名詞別の分析だと6番目に多いこともこれを確信付けている。

次に注目できる「水」という単語は係り受け頻度分析でも「水一飲む」が上位20位に入っており、対応バブル分析でも「水」は3人とも使用していることも明らかになった。今回行った2種類の分析全ての上位20位に入ってきた単語で特に特徴的であったものは「水」しかなかった。この3人は全員が雨水を飲んで生き残っていることから頻度が限定されている表現であっても頻度表に上がってきているといえる。

「水」という単語が使用されている理由は、硫黄島は文字の如く硫黄が吹き出す島であり壕内の温度は50℃を超え、水も湧かないため水が不足していた。秋草氏の『硫黄島を生き延びて』(P.130)には『ある時、足の早い雨がきた。土砂降りだった。まずは口を開けて飲んだ。雨雲を丸ごと呑み込む思いで大口を開けて雲を睨んだ』と綴られており、水不足が非常に深刻であったことが分かる。

次に本研究の限界点を検討する。本研究では3人の日本人の硫黄島での戦争体験記に絞り分析を行ったが、時系列の統一を優先したため書籍自体の文量に差ができてしまったことがあげられる。しかし、3人の文を比較することによって当時の日本軍の末端まで行き渡る伝達や深刻な水不足について知れたことは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 ver.5.2」を使用させて頂きました。また、本論文を作成するにあたり、指導教員の伊藤武彦教授及び伊藤ゼミ生の関川巧真、早川尚貴から長時間に渡り、熱心にご指導頂いたことに感謝いたします。

文献

- 秋草鶴次 (2011) 『硫黄島を生き延びて』 清流出版
- 川相昌一 (2007) 『硫黄島体験記』 光人社
- 石井周治 (1982) 『硫黄島に生きる』 国書刊行会