

異文化間コンフリクトの解決方法について

学生レポートの Text Mining Studio による分析

松上伸丈（和光大学大学院）

問題

日常生活にコンフリクトすなわち対立や葛藤は、つきものである。コンフリクトの解決のしかたによって、それは、問題のポジティブな解決やよりよき人間関係につながる。しかし、解決方法を失敗すると、課題達成の失敗や人間関係の悪化などを生じさせる。コンフリクト解決のカギを握る要素の一つとして、コンフリクトの当事者間の文化の違いがある。ここで文化とは、多義的であり、表 A のような要素が考えられる。

表 A. RESPECTFUL カウンセリングからみた文化の要因 (Lewis 他、井上監訳 2006)

R : 宗教的・霊的同一性	religious/spiritual identity
E : 経済的階層的背景	economic class background
S : 性同一性	sexual identity
P : 心理的成熟	level of psychological maturity
E : 民族性・人種的同一性	ethnic/racial identity
C : 年齢・発達段階に応じた課題	chronological/developmental challenges
T : トラウマ（心理的外傷）と幸福感への脅威	various forms of trauma and other threats to one's sense of well-being
F : 家族の背景と家族史	family background and history
U : 独特な身体的特徴	unique physical characteristics
L : 居住地域と言語の違い	location of residence and language differences

目的

本研究では、異文化間コンフリクトの解決方法について、大学生を対象に自由記述を用いたデータに基づいて、その構造の特徴をテキストマイニングの手法を用いて明らかにすることを目的とする。

方法

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。サンプル構成は、男性 6 人 (18%)、女性 27 人 (82%) で、学年は全員が 3 年次生であった。レポートの課題は、「異文化間コンフリクトの解決方法について」とし、自由記述させた。実施は 2006 年 7 月である。

結果

MSエクセルを使いテキストの基本情報をみると、33サンプル、データの平均文字数(Len関数)は 925.5 文字(SD=346.2、最大=1845、最小=282)と、1 サンプルあたりの文字数が多めである。

Text Mining Studio ver2.1 を使い分析を行なった結果は以下の通りである。

前処理を「分かち書きと係り受けの自動連結」で行なった。テキスト情報の基本情報をみると、表 1-1 より、総行数は、MS エクセルと同じ結果となった。平均行長では MS エクセルに比べ小さい値を返された、理由は不明である。総文数は、632 文で、1 サンプル平均 19.2 文であった。述べ単語数は、6,118 単語で、1 サンプルあたり 185.4 単語、1 文あたり 9.7 単語となった。

テキスト情報の品詞別出現回数では、表 1-2 より、名詞が 4,048 回で 66%を占めている。次いで動詞の 1,171 回(19%)、副詞 312 回(5%)となっている。

表1-1. 基本情報:テキスト基本情報

項目	値
総行数	33
平均行長(文字数)	487.2
総文数	632
平均文長(文字数)	25.4
述べ単語数	6118
単語種別数	2030

表1-2. 基本情報:品詞別出現回数(上位)

品詞	出現回数
名詞	4048
動詞	1171
副詞	312
連体詞	210
接続詞	167
形容詞	163
感動詞	27
記号	17
複合助詞	1
連体助詞	1
助動詞	1

単語頻度解析を合計と性別でみる(表 2)。性別は女性が多数のため、男女比が同一であった場合を仮定したウエイトを掛けたもの(=全数÷2×性別サンプル数)をMSエクセルにて再計算させ「女性比率」「男性比率」を作成した。単語出現頻度では「コンフリクト」という課題語単語が最も多い。ついで「日本」これも「異文化」という課題語単語の一つではないかと考えられる単語となっている。前記のように課題語に関する単語が多く見られる。また、男女別を比率に置き換えた値から男女別をみると。女性で相対的に多く出現している単語は「日本人」「気持ち」など、また、男性では「文化」「人」「日本」などとなっている。表 2 の男女比率の値には差が 10 人以上の項目にマーキングを入れている。

表2. 単語頻度解析:男女別(人)

単語	品詞	頻度	女性	男性	女性比率	男性比率
コンフリクト	名詞	76	65	11	39.7	30.3
日本	名詞	68	49	19	29.9	52.3
人	名詞	60	42	18	25.7	49.5
自分	名詞	57	46	11	28.1	30.3
お互い	名詞	50	43	7	26.3	19.3
文化	名詞	44	22	22	13.4	60.5
考える	動詞	42	32	10	19.6	27.5
相手	名詞	38	28	10	17.1	27.5
日本人	名詞	35	34	1	20.8	2.8
国	名詞	34	28	6	17.1	16.5
韓国	名詞	34	22	12	13.4	33.0
気持ち	名詞	34	33	1	20.2	2.8
異文化間	名詞	31	24	7	14.7	19.3
考え方	名詞	27	21	6	12.8	16.5
夫	名詞	26	24	2	14.7	5.5
トランセンド法	名詞	24	20	4	12.2	11.0
解決	名詞	23	20	3	12.2	8.3
考え	名詞	21	17	4	10.4	11.0
話す	動詞	20	19	1	11.6	2.8
理解	名詞	20	13	7	7.9	19.3

※比率のマーキングは仮定ウェイトを掛けた値の差が10人以上

係り受け頻度解析を品詞フィルタ=イメージで出力し合計でみる(表3-1)。このデータの持つキーワードの組み合わせが挙がってくるが、単語解析頻度上位単語は挙がらなかった。また、最大頻度も2件と少数であった。

表3-1. 係り受け頻度解析(品詞フィルタ=イメージ)

係り元単語	係り元品詞	係り先単語	係り先品詞	頻度
理解	名詞	必要	名詞	2
げっぷ	名詞	おいしい	形容詞	2
職種	名詞	ダメ+?	名詞	2
お互い	名詞	必要	名詞	2
理解	名詞	大切	名詞	2
雰囲気	名詞	よい	形容詞	2
行儀	名詞	悪い	形容詞	2

フィルタ条件を変更、すなわち入力品詞種類数、より緩やかな条件で再実行すると、表 3-2 と表 3-3 の結果が得られた。フィルタ行動とフィルタ話題は同じ結果となった。「コンフリクト」と「解決」の組み合わせが 10 件で原文に戻して 3 件の一致がみられた。

表3-2. 係り受け頻度解析(品詞フィルタ=行動)

係り元単語	係り元品詞	係り先単語	係り先品詞	頻度
コンフリクト	名詞	解決	名詞	10
制服	名詞	着る	動詞	6
相手	名詞	理解	名詞	6
コンフリクト	名詞	生じる	動詞	5
線	名詞	引く	動詞	5
トランセンド法	名詞	解決	名詞	5
人	名詞	和解	名詞	4
人	名詞	いる	動詞	4
コンフリクト	名詞	起こる	動詞	4
靖国神社	名詞	参拝	名詞	4
話	名詞	聞く	動詞	4

表3-3. 係り受け頻度解析(品詞フィルタ=話題一般)

係り元単語	係り元品詞	係り先単語	係り先品詞	頻度
コンフリクト	名詞	解決	名詞	10
相手	名詞	理解	名詞	6
制服	名詞	着る	動詞	6
トランセンド法	名詞	解決	名詞	5
線	名詞	引く	動詞	5
コンフリクト	名詞	生じる	動詞	5
話	名詞	聞く	動詞	4
人	名詞	和解	名詞	4
靖国神社	名詞	参拝	名詞	4
コンフリクト	名詞	起こる	動詞	4
人	名詞	いる	動詞	4

注目分析

注目分析とは、注目したあることばについて分析する。

注目語の設定を課題であり最多頻出後の「コンフリクト」とし共起関係(一行の中に両方の単語が使われている組み合わせ)上位をみる。パラメータの設定をデフォルトで行ったところ 100 以上のノード(矢印)が出たため、ノード数を共起単語ネットワーク図としてみやすい 20~30 程度になるように調整をおこなった。共起抽出設定では最低信頼度 80、共起ルールを 20 回以上、詳細設定では注目語を含む係り受け表現を 10 回、単語の最大表示数を 50 と(図 1-1)とすると、3 単語以上の多重共起のみられる「対立点」が課題となる単語と考えられる。前記のパラメータに抽出設定を最低信頼度 60 に変更する(図 1-2)と、多重共起語が「対立点」以外に「人」「解決」「気持ち」「聞く」「相手」となった。

多重共起(3 つ以上)単語は、注目単語と共起の高い他の単語とも、共起している単語であるため、文全体の課題を浮き彫りにすると考えられる。

図 1-1

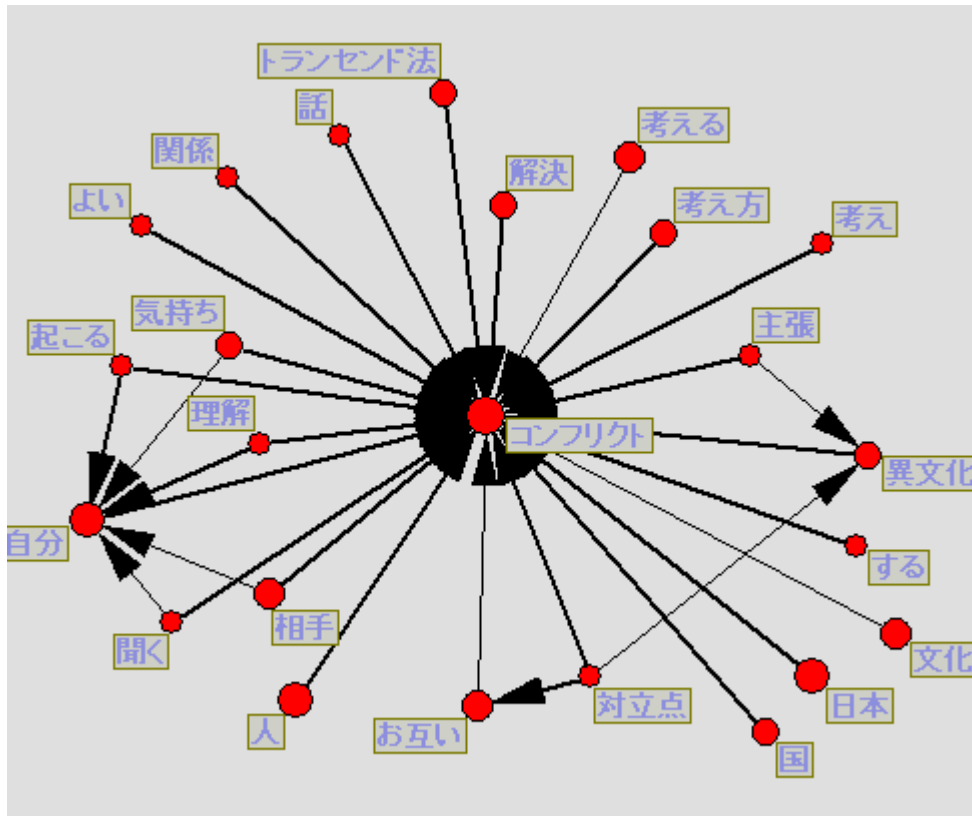
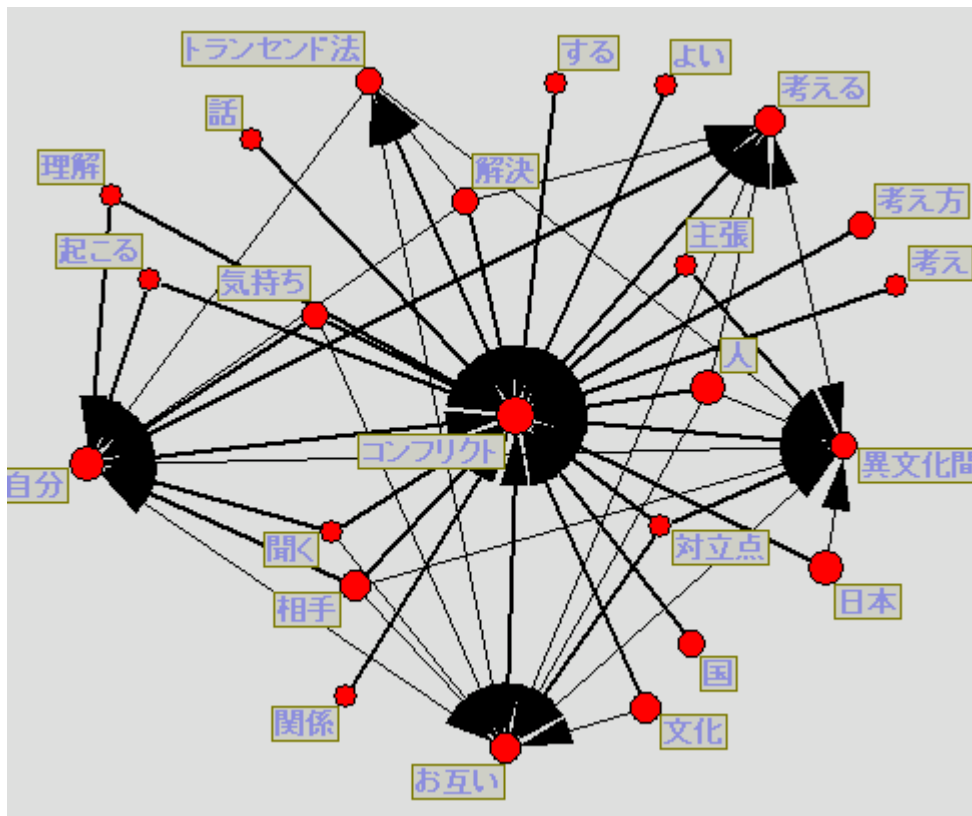


図 1-2



特徴分析

特徴分析とは、データに付随する属性毎に、特徴的に出現する単語及び係り受け表現を抽出する。

性別の特徴語(表 4)は、パラメータは抽出指標「補完類似度」を使用した。男女で扱っている内容の違いが一目でわかる。女性では対象が「夫」「子供」など親族、男性では「太田」「キム」といった友人となっている。共通項目として課題の「異文化」の一方を「日本」としてみていることが伺える。

属性別のサンプル数の違いにより最大値が大きく異なっていると考えられる。

表4. 特徴語抽出:男女別

A3-女性	指標値	A3-男性	指標値
日本人	15.2	文化	32.5
気持ち	14.7	選手	25.8
コンフリクト	11.0	太田	25.8
話し合う	8.6	キム	19.8
夫	8.1	BSE	17.8
お互い	7.8	サポーター	17.8
話す	7.6	人	14.5
げっぷ	7.6	人種差別	13.9
意見	7.6	試合	13.4
A君	6.6	日本	13.0
言葉	6.6	韓国	12.7
子供	6.6	自分たち	12.4
中国	6.6	コミュニケーションスタイル	11.9
一緒	6.1	ハン	11.9
見る	6.1	フォロー・サポート	11.9
子	6.1	思う	11.3
島	6.1	協調的	9.9
父	6.1	主張的	9.9
使う	5.6	カレー	7.9
反対	5.5	感情	7.9

特徴表現抽出

特徴表現抽出とは、データに付随する属性毎に、特徴的に出現する係り受け表現を抽出する。

性別に、パラメータを抽出する表現品詞の設定は「話題一般」、抽出指標「補完類似度」を使用した(表 5)。女性は、特徴語で3位であった「コンフリクト」が、男性では1位の「文化」が特徴表現において上位にきている。

表5. 特徴表現一覧:男女別

A3-女性	指標値	A3-男性	指標値
コンフリクト-解決	5.0	協調的-文化	8.0
制服-着る	3.0	主張的-文化	8.0
コンフリクト-生じる	2.5	相手-理解	4.5
線-引く	2.5	カレー-好む	4.0
コンフリクト-起こる	2.0	キム-フォロー・サポート	4.0
感謝-言葉	2.0	家-ホームステイ	4.0
人-いる	2.0	過ち-繰り返す+ない	4.0
靖国神社-参拝	2.0	叶える+ない-悲しみ	4.0
良い-環境	2.0	感情-理解	4.0
話-聞く	2.0	韓国-統一	4.0

評判抽出

評判抽出とは、単語に対して、好意的な表現・非好意的な表現それぞれで語られた回数をカウントし、それをもとに好評語・不評語のランキングを作成する。

使用したデータが「異文化間コンフリクトについて」という抽象的な課題に対する回答だったため、同単語に対し、肯定的使われ方、否定的な使われ方の両極は少なかった。その中で両方に出現した単語は「気持ち」「国」の2語が双方での出現となっている(表6)。「国」は文化単位の一つと考えると、課題単語と捕らえられる。「気持ち」は注目分析や特徴語分析の女性で上位に挙げた単語でもあるため、この課題を象徴する単語と考えられる。

表6. 評判抽出(人)

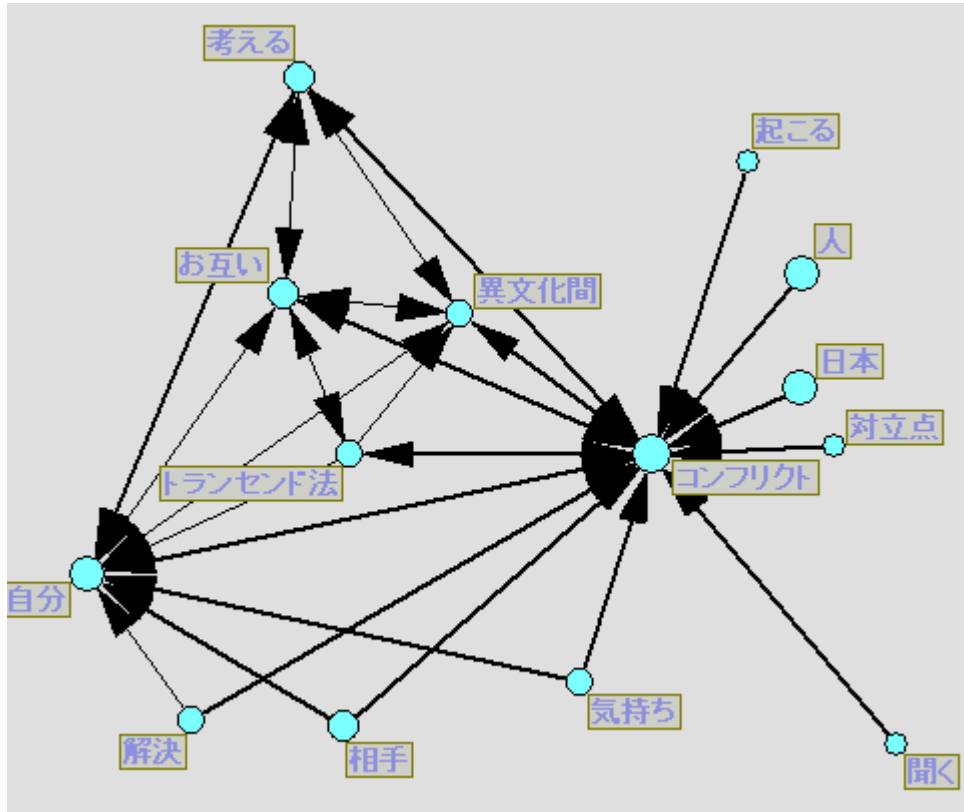
好評語ランキング				不評語ランキング			
単語	品詞	Positive	Negative	単語	品詞	Positive	Negative
関係	名詞	7	0	人	名詞	0	-4
お互い	名詞	6	0	思い	名詞	0	-3
気持ち	名詞	6	-3	話	名詞	0	-3
環境	名詞	5	0	気持ち	名詞	6	-3
理解	名詞	5	-1	顔	名詞	0	-2
仲	名詞	4	0	行儀	名詞	0	-2
プレー	名詞	3	0	自己主張	名詞	0	-2
考え方	名詞	3	0	声	名詞	0	-2
場所	名詞	3	0	着こなし	名詞	0	-2
本	名詞	3	0	国	名詞	2	-2
げっぷ	名詞	3	-1				
フォロー	名詞	2	0				
ライフイベント	名詞	2	0				
意味合い	名詞	2	0				
価値観	名詞	2	0				
家族	名詞	2	0				
自然	名詞	2	0				
重視	名詞	2	0				
生活	名詞	2	0				
大切	名詞	2	0				
方法	名詞	2	0				
解決	名詞	2	-1				
送金	名詞	2	-1				
雰囲気	名詞	2	-1				
国	名詞	2	-2				

ことばネットワーク

ことばネットワークとは、単語間及び単語と属性の関連をネットワーク図で表す。

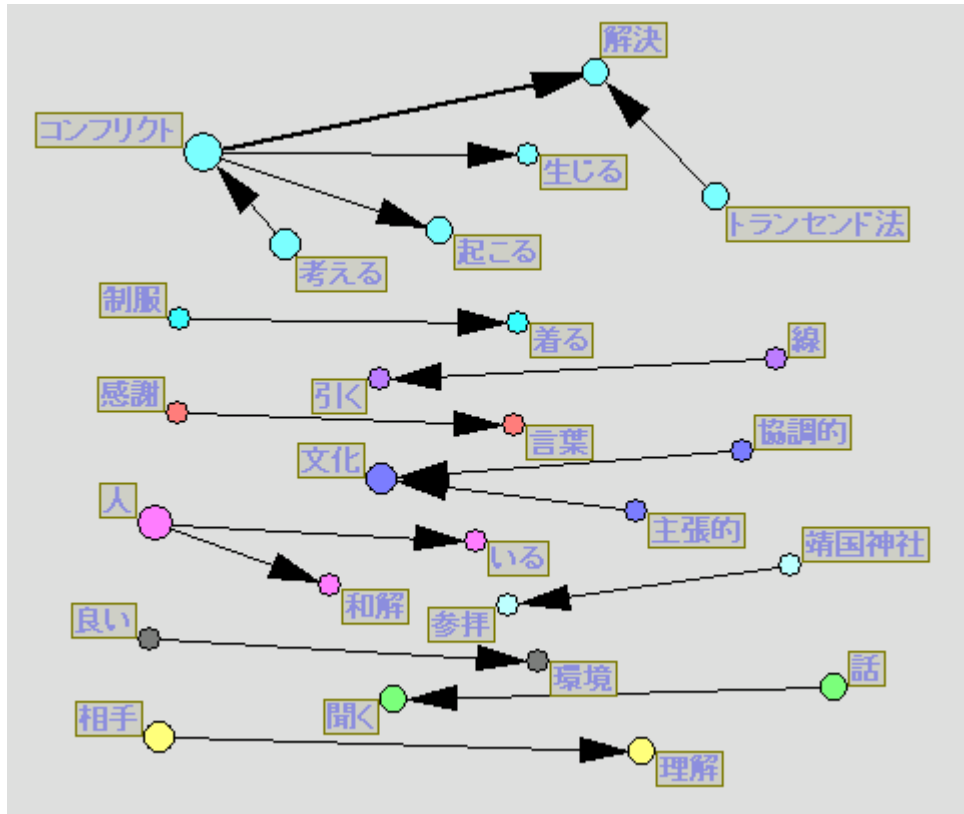
抽出する関係を共起とし、パラメータは抽出指標を最低信頼度 60、共起ルールを 12 回で実行した(図 2-1)。3 つ以上の単語と共起している単語は「コンフリクト」「自分」「考える」「異文化間」「トランセンド法」「お互い」の 6 項目、2 つの単語と共起している単語は「気持ち」「相手」「解決」の 3 項目となった。課題単語以外をみると「自分」「考える」「トランセンド法」「お互い」「気持ち」「相手」となる。

図 2-1. ことばネットワーク図(共起関係を抽出する)



抽出する係り受け品詞設定は、話題一般、係り受け頻度が5回以上を選択(図2-2)。データ全体で係り受けが低いため、広めに設定をおこなった。図は係り受け頻度が上位にあるものが上に来るように配置した。係り受け頻度解析でつながりの薄さはイメージされていた通りで、大きな塊のあるネットワーク図にはならず、3単語以上の図は3つにとどまった。2つ以上の単語に係っている単語は「コンフリクト」「人」、また受けている単語は「解決」と「文化」となっている。ここから課題語をはずすと、「人」が重要な単語と考えられる。

図2-2. ことばネットワーク図(係り受け関係を抽出する)

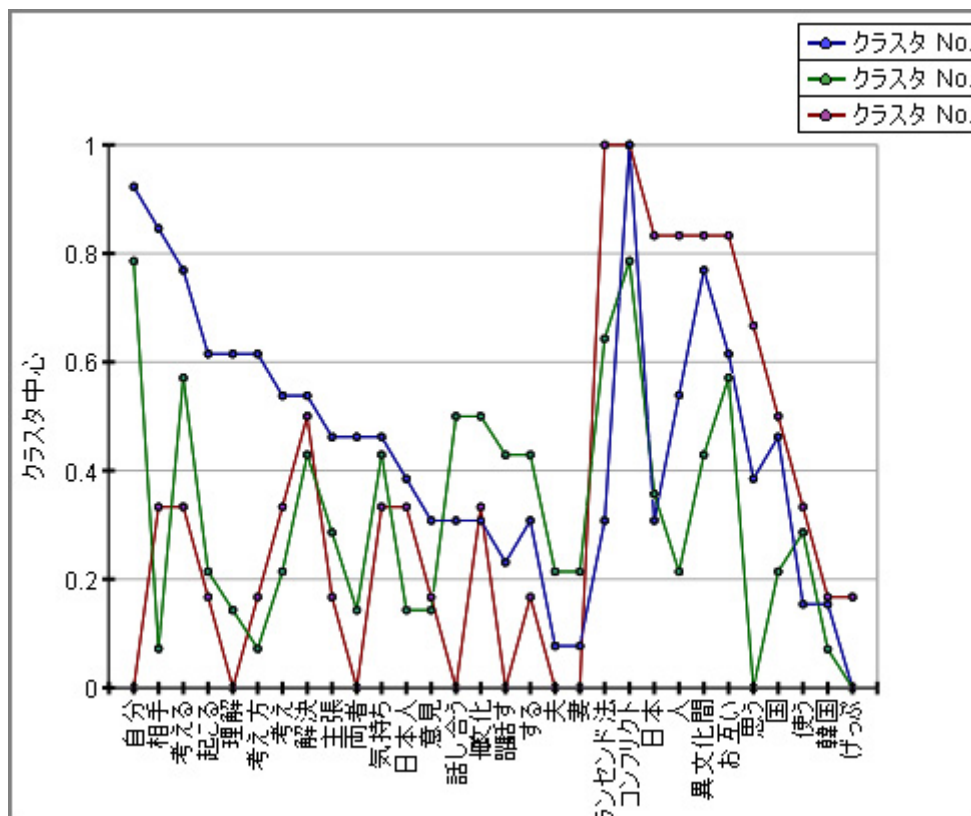


文章分類

文章分類とは、使われている単語に応じて、テキストの行もしくは文章をクラスタに分類し、クラスタ毎の属性情報や特徴単語を表示することである。

文章分類は、テキストを分類する指標を頻度の上位 30 単語を用い、3 クラスタとした(図 3)。1 クラスタの特徴単語は「相手」「起こる」「理解」「考え方」「両者」など、2 クラスタでは「話し合う」「文化」「話す」「する」など、3 クラスタは「日本」「人」「思う」などとなっている。

図 3. 全クラスタ情報



考察

今回使用したデータは、課題に対し学生がそれぞれ場面の設定をし、意見を書かせたものであったため、1 サンプル内の共起が高く、サンプル間での共起が低いものであった。注目分析では「対立点」「人」「解決」「気持ち」「聞く」「相手」、ことばネットワーク図では「自分」「考える」「トランセンド法」「お互い」「気持ち」「相手」など重要と考えられる単語として挙げられた。これらの結果から「気持ち」「相手」の入った原文に注目しサンプル全体の傾向を検討することが重要ではないかと考える。

Text Mining Studio ver2.1 の利点を生かして、男女別をみると、女性で頻出単語は「日本人」「気持ち」など、また、男性では「文化」「人」「日本」となっていることが明らかになった。今後データを増やしたり、コンフリクトのテーマによってグループ化したりすることにより、コンフリクト解決をとらえる際のジェンダーの差や、テーマごとのコンフリクト表現の違いがより明確になると期待される。

今回の分析をおこなって

最も時間を割いた事項は原文の整理であった。全文の句点、句読点の入り方を目視による確認である。その後はパラメータの試行錯誤に手間取ったもののスムーズな進行が見られた。

今後の課題としては、今回は原文が長く、単語量と意味が多岐に渡ったため辞書の整理過程をはずした、例えば「する」と「行う」を一つにまとめるべきかどうかの作業はできなかった。今後は 1 つの課題について類似した内容が多数書かれたデータを元に辞書整理に力を入れ、データ全体が訴えている内容の探索を模索してみたいと考える。また 192 人の本分析データがあるので、内容にしたがって被験者をグループ化し、Text Mining Studio ver2.1 の利点を生かして分析することも一つのやり方であろう。

謝辞

本研究を遂行するにあたり和光大学大学院の伊藤武彦教授のご指導を受けました。記して感謝します。また、データエントリーにおいて献身的な働きをしていただいた守下理氏、久木田隼氏に感謝します。

【参考文献】

- 井上孝代(編) (2005). コンフリクト転換のカウンセリング 川島書店
Lewis, J. et al. (2003). Community counseling. 井上孝代監訳 伊藤武彦・石原静子訳
(2006) コミュニティ・カウンセリング 川島書店
数理システム (2006) Text Mining Studio ver2.1 操作マニュアル 数理システム