

<研究ノート>

KHコーダーを用いた テキストマイニングの覚書

— 大学生の結婚観の分析から —

中西啓喜

キーワード：KHコーダー，テキストマイニング，
量的研究と質的研究，
ロマンティックラブ・イデオロギー

1. テキストデータ分析に客観性をもたせる

社会調査における観察データは、量的（定量的）研究（Quantitative Research）と質的（定性的）研究（Qualitative Research）に大別される。佐藤郁哉（2006, p. 76）は、それぞれの調査法的な特徴を以下のように整理している。

- 量的調査：統計データの分析やサーベイ調査の結果をもとにした社会調査のように、数値データを中心にして分析を進め、その結果については、主にグラフや数値表あるいは数式などで表現する調査法
- 質的調査：主にインフォーマル・インタビューや参与観察あるいは文書資料や歴史資料の検討などを通して、文字テキストや文章が中心と

なっているデータを集め、その結果の報告に際しては、数値による記述や統計的な分析というよりは、日常言語に近い言葉による記述と分析を中心とする調査法

これらは決して対立的な関係ではなく、人間や社会を深く理解するためには、両方の立場について理解しておく必要がある。

しかし、インタビューやテキストから得られるデータを分析した研究は、「主観的」だという批判が向けられることも少なくない。質的データは、量的データと比較すると規則的ではない。また、研究論文は紙幅が限られている場合も多く、膨大なインタビューデータなどをすべて掲載することはほぼ不可能である。論文内で提示されるデータが、研究者の都合の良い箇所を抜き出しているかもしれないという指摘を免れないこともある。そのため、テキストデータをある程度は規則的に分析することも求められる。

こうした動機から質的に収集されたテキストデータを計量的に分析するのが内容分析(Content Analysis)ないしテキストマイニングという手法である。これらの手法により、分析ソフトウェアを用いてテキストデータを探索し、統計的な意味のあるパターンを抽出することが可能となる(樋口2020)。

本稿では、KHコーダーというソフトウェアを用いて、大学生の結婚観についてのテキストマイニングを行う。KHコーダーは決してマイナーなソフトウェアではない。年々利用される研究事例が蓄積されるばかりか、フリーソフトにも関わらず非常に精巧かつ初心者フレンドリーに作成されている。また、日本語のテキストデータの分析を得意とすることも非常に大きなメリットである。例えば、高史明(2015)がインターネット上に溢れる在日コリアンへの偏見に関するテキストデータの分析からレイシズムの研究を行うなど、日本の社会学研究への貢献も蓄積されている。本稿は、KHコーダーの操作方法のリサーチノート作成と実際のデータ分析を行いながら、テキス

トマイニングへの理解を深めることを目的としたものである¹⁾。

なお、本稿の構成は以下のようなものである。第2章では、少子化社会における家族形成についての分析課題を設定し、第3章では分析データを詳述する。第4章では、それぞれの節で抽出語リスト、共起ネットワーク、対応分析の順で紹介し、それぞれの目でKHコーダーの操作方法と分析結果を示す。最後に、第5章では知見をまとめつつ、KHコーダーを大学教育で活用することの可能性について言及する。

1. テキストデータ分析に客観性をもたせる
2. 少子化社会における結婚
3. データと方法
 3. 1. 使用データ
 3. 2. テキストマイニングとKHコーダーの基本的な操作法
4. テキストマイニング
 4. 1. テキストマイニング①：抽出語リストと前後文の検討
 4. 1. 1. KHコーダーにおいて抽出語リストを出力する
 4. 1. 2. 恋愛結婚と婚活サービスのイメージの比較
 4. 2. テキストマイニング②：共起ネットワーク
 4. 2. 1. KHコーダーにおいて共起ネットワークを出力する
 4. 2. 2. 恋愛結婚と婚活サービスの共起ネットワークを比較する
 4. 3. テキストマイニング③：対応分析
 4. 3. 1. KHコーダーにおいて対応分析を実行する
 4. 3. 2. 恋愛結婚と婚活サービスの対応分析の結果を比較する
5. まとめ

1) KHコーダーに関するホームページ (<https://kxhcodcr.net/> (2021年10月1日取得)) および樋口 (2020)などを参照にした。なお、KHコーダーの開発に込められた意図は、樋口 (2004, 2011, 2020)などに詳しい。

2. 少子化社会における結婚

日本では長らく少子化が社会問題として取りざたされているが、改善の兆しは一向に見られない。図1はわが国の出生数の推移である²⁾。団塊ジュニア世代のピークが1973年(約203万人)で、最後に200万人を記録したのが翌年の1974年である。その後、出生数は下がり続け、2016年には100万人を割り、今後も低下する見込みである(国立社会保障・人口問題研究所「日本の将来推計人口」による)。

ただし、データを見る限り子どもを産まない夫婦が大幅に増えたとはいいいにくい。表1は、国立社会保障・人口問題研究所(2017)が調査した夫婦の出生子ども数分布の推移である³⁾。これを見ると、1977年から2015年にかけて、子どものいない夫婦が3%から6.2%へと変化している。この数値の変化を「微増」と読むか「倍増」と読むかは読者に任せるが、いずれにしても子どもを産まない夫婦は1割に満たないということである。

また、完結出生児数(夫婦が最終的に出生する子ども数の平均値)は2.19人から1.94人へと微減している。感覚的にいえば、一夫婦が生む子どもの人数が、1977年から2002年頃までは「2人~3人」だったのが、2005年以降では「1人~2人」へと変化したことになる。確かに少子化の傾向を読み取れるが、「結婚したら子どもを産む」というルーティンはある程度維持されているといえよう。

少子化問題において、むしろ問題視すべきは晩婚化と結婚件数の減少である。図2は、内閣府(2017)の『平成29年版 少子化社会対策白書』に見

2) 図1は、厚生労働省政策統括官付参事官付人口動態・保健社会統計室「人口動態統計」より筆者作成 (<https://www.mhlw.go.jp/stf/wp/hakusyo/kousei/19/backdata/01-01-01-07.html>)より、2021年10月1日データ取得。

3) 表1は、国立社会保障・人口問題研究所(2017) http://www.ipss.go.jp/ps-do/ukou/j/doukou15/gaiyou15html/NFS15G_html07.htmlより2021年10月1日取得。

られる未婚率の推移である⁴⁾。データの的には男性よりも女性の方が顕著である。1985年頃までは、女性の30代の未婚率は10%程度であったが、2015年には3割程度が30代でも未婚である。初婚行動が出生率に与える影響が大きいのは、岩澤（2008）など学術的にも支持されているところである。

結婚が減っている原因は諸説ありだが、図3の恋愛結婚・見合い結婚構成の推移を見ると、「恋愛でないと結婚しない」という傾向が読み取れる。ただし、表2の「調査別にみた、夫妻が出会ったきっかけの構成」も同時に確認してみよう⁵⁾。すると、「職場や仕事で」は1982年から2015年にかけて約3割程度を推移している一方で、「友人・兄弟姉妹を通じて」が1982年から2015年にかけて10ポイント程度（20.5%から30.9%へ）増加している。結婚に繋がる出会いにおける他者からの紹介が、見合いサービスから友人・兄弟姉妹へと変化しているように思われる。

このような現状を踏まえ、家族社会学における結婚格差への問題関心は、収入による格差（三輪 2019）、学歴同類婚による階層結合（白波瀬 1999、打越 2016）、容姿ないし美醜による格差（小林・谷本 2016）などを扱ってきた。つまり、結婚への障害は、まず資金の不足（当面のお金を手元にあるかだけでなく、安定した十分な収入を含む）があり、学歴や年齢などの属性的な条件、ルックスが自分好みであるかどうか加味される。

また、「若者が恋愛に興味関心を持たなくなった」という“草食化”言説は、図3やいくつかの文献（荒川 2017、谷本・渡邊 2019）を見る限りは支持されておらず、恋愛を通じて結婚したいというロマンティックラブ/ロマンティックマリッジ・イデオロギーはいまだに強固である。要するに、自身の恋愛相手として“適当な異性”とめぐり合い、恋愛できないことが結婚件

4) 図2は、内閣府（2017）（<https://www8.cao.go.jp/shoushi/shoushika/whitepaper/measures/w-2017/29pdfhonpen/pdf/s1-2.pdf>）より2021年10月1日取得。

5) 図3、表2共に国立社会保障・人口問題研究所（2017）http://www.ipss.go.jp/ps-doukou/j/doukou15/gaiyou15html/NFS15G_html06.htmlより2021年10月1日取得。

数減少の背景だというシンプルな話である。

それでは、現代の若者は、恋愛結婚とその対となる「婚活サービス」の利用それぞれに対してどのような意見を持っているのだろうか。本稿では、大学生を対象とした結婚観についてのテキストデータを分析することで、結婚件数の減少について検討してみたい。

図1. 出生数の推移

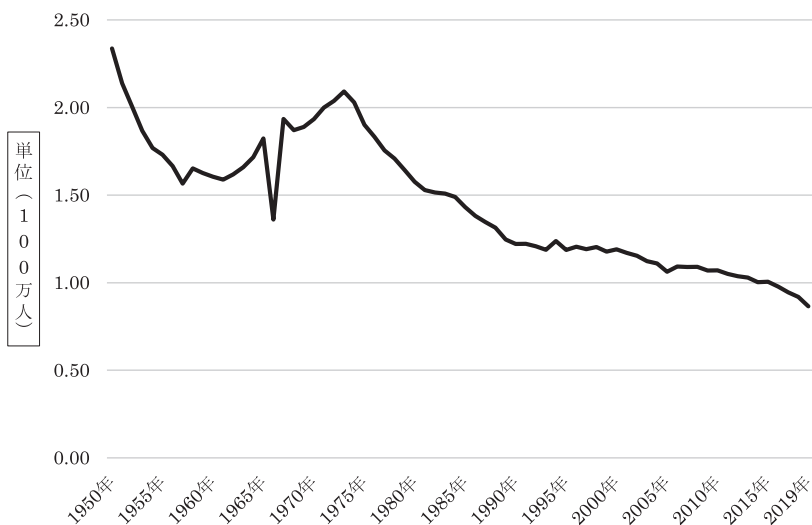
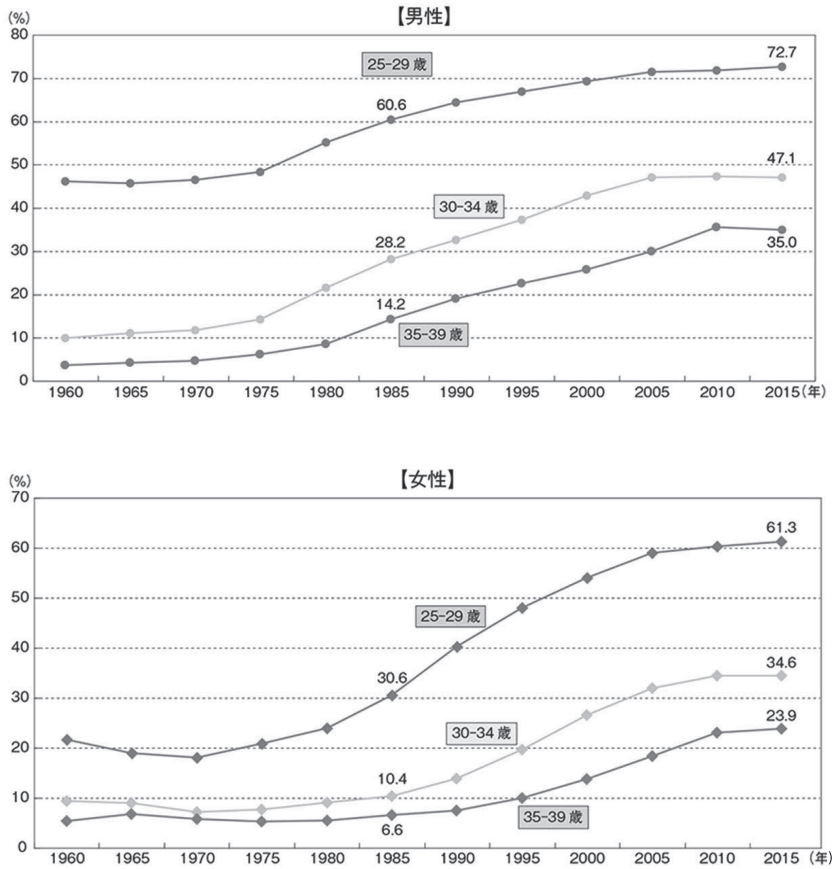


表1. 調査別にみた、夫婦の出生子ども数分布の推移 (結婚持続期間 15~19年)

調査(調査年次)	総数 (客体数)	0人	1人	2人	3人	4人以上	完結出生児数
第7回調査(1977年)	100.0 % (1.427)	3.0 %	11.0	57.0	23.8	5.1	2.19 人
第8回調査(1982年)	100.0 (1.429)	3.1	9.1	55.4	27.4	5.0	2.23
第9回調査(1987年)	100.0 (1.755)	2.7	9.6	57.8	25.9	3.9	2.19
第10回調査(1992年)	100.0 (1.849)	3.1	9.3	56.4	26.5	4.8	2.21
第11回調査(1997年)	100.0 (1.334)	3.7	9.8	53.6	27.9	5.0	2.21
第12回調査(2002年)	100.0 (1.257)	3.4	8.9	53.2	30.2	4.2	2.23
第13回調査(2005年)	100.0 (1.078)	5.6	11.7	56.0	22.4	4.3	2.09
第14回調査(2010年)	100.0 (1.385)	6.4	15.9	56.2	19.4	2.2	1.96
第15回調査(2015年)	100.0 (1.233)	6.2	18.6	54.1	17.8	3.3	1.94

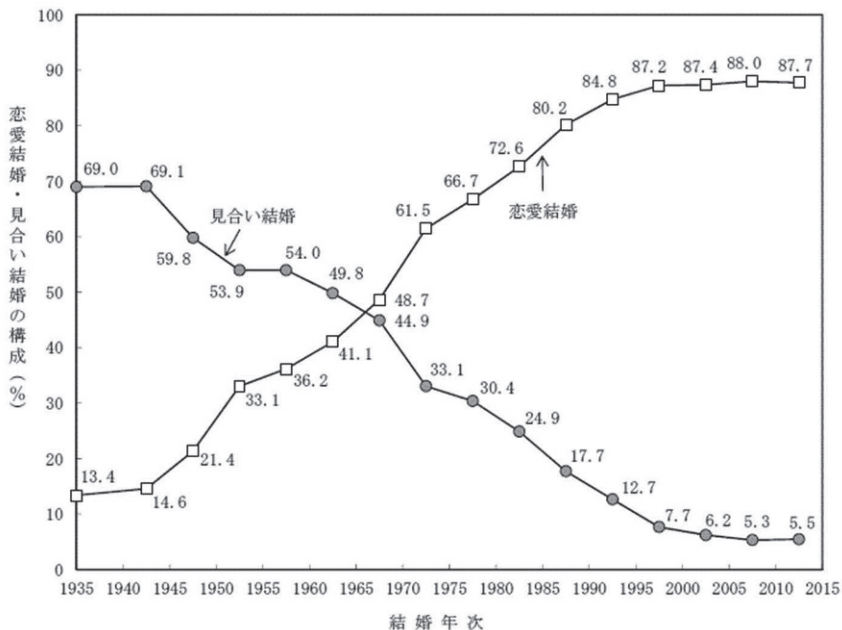
出所: 国立社会保障・人口問題研究所(2017)

図2. 年齢（5歳階級）別未婚率の推移



出所:内閣府(2017)『平成29年版 少子化社会対策白書』

図3. 結婚年次別にみた、恋愛結婚・見合い結婚構成の推移



出所：国立社会保障・人口問題研究所(2017)

表2. 調査別にみた、夫妻が出会ったきっかけの構成

調査 (調査年次)	総 数(客体数)	恋愛結婚							見合い 結婚	その他	不詳
		職場や 仕事で	友人・ 兄弟姉妹 を通じて	学校で	街なかや 旅先で	サークル・ クラブ 習い事で	アルバイト で	幼なじみ ・隣人			
第8回調査 (1982年)	100.0 % (1,295)	25.3 %	20.5	6.1	8.2	5.8	...	2.2	29.4	0.3	2.2
第9回調査 (1987年)	100.0 (1,421)	31.5	22.4	7.0	6.3	5.3	...	1.5	23.3	1.9	0.8
第10回調査 (1992年)	100.0 (1,525)	35.0	22.3	7.7	6.2	5.5	4.2	1.8	15.2	1.6	0.3
第11回調査 (1997年)	100.0 (1,304)	33.5	27.0	10.4	5.2	4.8	4.7	1.5	9.7	1.9	1.2
第12回調査 (2002年)	100.0 (1,488)	32.9	29.2	9.3	5.4	5.1	4.8	1.1	6.9	3.0	2.2
第13回調査 (2005年)	100.0 (1,076)	29.9	30.9	11.1	4.5	5.2	4.3	1.0	6.4	4.5	2.3
第14回調査 (2010年)	100.0 (1,136)	29.3	29.7	11.9	5.1	5.5	4.2	2.4	5.2	4.8	2.0
第15回調査 (2015年)	100.0 (901)	28.1	30.9	11.7	5.7	4.9	3.7	1.7	6.5	5.0	2.0

出所：国立社会保障・人口問題研究所(2017)

3. データと方法

3.1. 使用データ

本稿で用いるのは、関東A大学の社会調査科目の授業内において2021年6月に実施した調査から得られたデータである。アンケートはグーグルフォームで作成し、履修学生がSNS等を通じて協力依頼をした。その結果、165人からの協力が得られ、これを本レポートの分析ケースとする。なお、自由記述については複数人の無回答がある。使用するデータの基本的な属性構成は表3にまとめた。140人（約85%）が結婚したいと考えていることがわかる。

分析のメインに使用する恋愛結婚と婚活サービスについてのテキストデータは、それぞれ「あなたは恋愛結婚についてどのようなイメージを持っていますか。あなたの恋愛結婚へのイメージを自由に書いてください」と「あなたは婚活サービスの利用についてどのようなイメージを持っていますか。あなたの婚活サービス利用へのイメージを自由に書いてください」というワーディングでデータを収集した。

むろん、ひとくちに「婚活サービス」といっても無料で利用できるアプリのようなものから高額な利用料を要する仲人を介した「お見合い」に近いものなど多岐にわたる。しかし本稿では、大学生の大まかなイメージを把握するため、こうした弁別は行わずにデータ収集を行った。

表3. 使用するデータの基本的な属性構成

		性別		全体 N = 165
		男子 N = 88	女子 N = 77	
学年	1年生	16	14	30
	2年生	23	26	49
	3年生	38	27	65
	4年生	11	10	21
結婚願望	したい	77	63	140
	したくない	11	14	25

3.2. テキストマイニングとKHコーダーの基本的な操作法

本稿で用いる方法は、テキストマイニングである。テキストマイニングとは、文書（テキスト）データから有益と思われる情報を構造化してとらえる分析手法である。そのために、KHコーダーという分析ソフトウェアを用いるのだが、ここでは基本的な操作方法を記載しておきたい。

KHコーダーのダウンロード等々は済ませたうえで、図4-1のようにエクセルファイルで関係形式テーブルを作成する。1行目は変数名として認識される。また、分析したいテキストデータはA列に配置する必要がある。量的データに慣れていると、ついA列に「調査ID」などを配置してしまうことがあるのだが、A列に分析したいテキストデータを配置しておくことでKHコーダーのデフォルトで分析を開始してくれる。

次にKHコーダーへ分析データを読み込む。「プロジェクト」→「新規プロジェクト」から図4-2のウィンドウを呼び出し、「参照」をクリックし、ディレクトリから分析したいエクセルファイルを読み込む。続いて、図4-3のように「前処理」のタブから「前処理の実行」を行う。「この処理には時間がかかる場合があります。続行してよろしいですか？」と表示されるが気にせずに「OK」をクリックする。これで分析の前準備が完了する⁶⁾。

- 6) テキストマイニングにおいて、一般的によく使用する単語（特に動詞）や句読点は分析データから除去しておく必要がある。例えば、本稿データの場合、「思う」や「。」がひとつのカテゴリとして頻出語リストや共起ネットワークに表示される。しかし、これらは分析カテゴリとしてほとんど意味がない。そこで、「前処理」→「語の取捨選択」をクリックし、図10のウィンドウを呼び出す。分析から排除する語を「使用しない語の設定」に直接入力し、

図10. 「語の取捨選択」の操作ウィンドウ



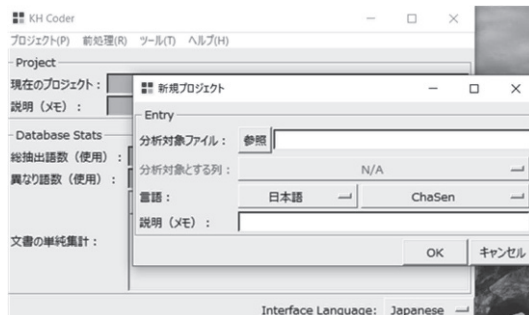
(*) 「強制抽出する語」の指定は、再度前処理を行うまで反映されません。

次章からは、KHコーダーを用いたテキストマイニングの基本的な流れをおさえつつ、分析結果を読み解いていこう。

図 4-1. KHコーダーに読み込ませるエクセルのデータ形式

	A	B	C	D
1	恋愛結婚	性別	結婚願望	結婚願望+性別
2	良い	女子	結婚したい	結婚したい+女子
3		女子	結婚したい	結婚したい+女子
4	大変そう	女子	結婚したくない	結婚したくない+女子
5	理想	男子	結婚したい	結婚したい+男子
6	後期近代特有の結婚	男子	結婚したくない	結婚したくない+男子
7	失敗する人が多い	女子	結婚したい	結婚したい+女子
8	美しいもの	男子	結婚したい	結婚したい+男子
9	大変なことが多い	女子	結婚したい	結婚したい+女子
10	まず癒しで、辛いこ	男子	結婚したい	結婚したい+男子
11	人権が尊重されてい	男子	結婚したくない	結婚したくない+男子
12	好きという気持ちだ	女子	結婚したい	結婚したい+女子
13		女子	結婚したい	結婚したい+女子
14	熱が覚めた時に大変	男子	結婚したい	結婚したい+男子
15	若い年代の人に多い	女子	結婚したい	結婚したい+女子
16	幸せ	女子	結婚したい	結婚したい+女子
17	最近の時代において	男子	結婚したい	結婚したい+男子
18	よく分からない	女子	結婚したくない	結婚したくない+女子
19		女子	結婚したい	結婚したい+女子
20	幸せそう	女子	結婚したい	結婚したい+女子
21	自然な流れだと思う	男子	結婚したい	結婚したい+男子
22	肯定的	女子	結婚したい	結婚したい+女子
23	幸せだけリスクが	女子	結婚したい	結婚したい+女子

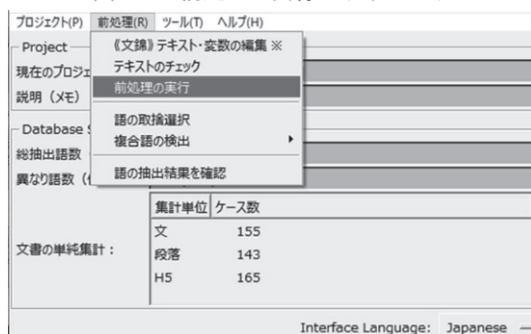
図 4-2. KHコーダーへのデータの読み込みウィンドウ



分析を実行する。量的研究における欠損値の指定のような作業である。

また本稿の場合では、「婚」と「活」が分かれて抽出されてしまったため、「強制抽出する語の指定」の箇所に入力した。こうしておけば、目的に応じた分析が可能となる。そして、このように語の指定を行った上で「前処理の実行」を再度行うことを忘れないよう。

図4-3. 前処理の実行のウィンドウ



4. テキストマイニング

4.1. テキストマイニング①：抽出語リストと前後文の検討

4.1.1. KHコーダーにおいて抽出語リストを出力する

テキストマイニングでは、最初に抽出語リストを出力させる。これは、量的分析における記述統計量を検討することに相当するだろう。図5-1のように、「ツール」→「抽出語リスト」と操作すると図5-2のようなウィンドウが出力される。

図5-2は、大学生の恋愛結婚イメージについての自由記述の抽出語リス

図5-1. 抽出語リストの出力ウィンドウ

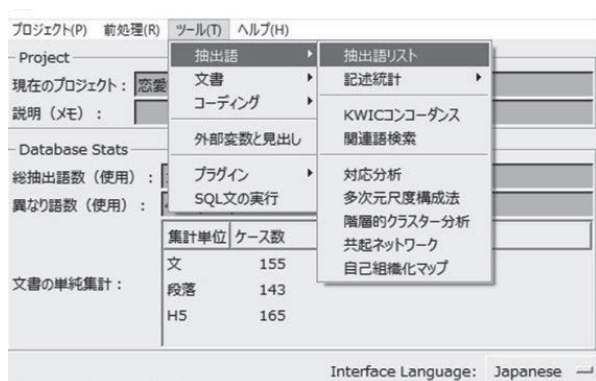


図5-2. 抽出語リストの出力ウィンドウ

1	結婚	サ変名詞	38	
2	幸せ	形容動詞	16	
3	恋愛	サ変名詞	16	
4	理想	名詞	14	
5	人	名詞C	13	
6	イメージ	サ変名詞	11	
7	良い	形容詞	8	
8	関係	サ変名詞	7	
9	自分	名詞	7	
10	大変	形容動詞	7	
11	好き	形容動詞	6	
12	自由	形容動詞	6	
13	素敵	形容動詞	6	
14	多い	形容詞	6	
15	形	名詞C	5	
16	相手	名詞	5	

図5-3. KWICコンコーダンスの出力ウィンドウ

■ KWICコンコーダンス

Search Entry
抽出語: 品詞: 活用形: 追加条件

ソート1: ソート2: ソート3: (前後 24 語を表示)

Result

近代的特有の結婚形態が失敗する人が多い。決め手が分からないからタイミングが合えばなんとなく結婚している	結婚	形態が失敗する人が多い。決め手が分からないからタイミングが合えばなんとなく結婚している	結婚
3歳になるイメージ きっと辛いこともあるけど、それ以上に楽しかったり安心地味かつ汚くない	結婚	はできないかなあと思う人の人権が尊重されていて民主主義的<h5>----cell---	
と思う人権が尊重されていて民主主義的の好きという気持ちだけで	結婚	ずるとあとから嫌な部分がたくさん見えてきそう。<h5>熱が覚めた	
まそう。<h5>熱が覚めた時に大変さなどは思いいつ、恋愛	結婚	には異敵なイメージはあります。<h5>若い年代の人に多い<h5>幸せ<h5>----cell---	
、仕事、家事の時間等含めても気楽な大変さであるかという点	結婚	は相手との価値観が異なると上手くいかないと思うので、大変であるというイメージがります	
ます。<h5>顔のネジがぶんとでる人たちがするもの<h5>恋愛	結婚	できる幸せ。<h5>顔のネジがぶんとでる人たちがするもの<h5>	
幸せ。<h5>顔のネジがぶんとでる人たちがするもの<h5>恋愛	結婚	を経た夫婦が良好な2者間の関係を築ける可能性は非常に低いと考えて	
から情性で関係が続いていった夫婦が多いのではないだろうか。こうした自身の恋愛	結婚	のイメージは、自分の最も身近な夫婦であるところの両親の関係を良好にはなく、また	
とは重ならないところを感じている。<h5>多様性の理想的な	結婚	<h5>現代の風潮において、必ずしも必要なものではない。<h5>持続が困難<h5>----cell---	
現象はそんなに甘くない。<h5>離婚率が高い<h5>若い恋愛感情を持った状態で	結婚	に陥るのば永遠に一緒にいたいという人間の心理を含めて<h5>自然なことだ	
に一緒にいたいという人間の心理を含めて<h5>自然なことだと思ってくれど、恋愛で	結婚	が同じものか?と書ければ、別物だと答えるような気がします。恋愛は何事	
と答えるような気がします。恋愛は何事も無難な状態で乗り越える印象があるけれど、	結婚	は多角的な面から将来を見極める必要を感じます。結婚したことないからわかんない	
で乗り越える印象があるけれど、結婚は多角的な面から将来を見極める必要を感じます。	結婚	したことないからわかんないけど、<h5>思いと思う<h5>最近のもの<h5>	
解消されてしまう場合もある。<h5>出来たら幸せだと思います。<h5>恋愛	結婚	と書いても男女とも打算なしという可能性とて感じているので特に自由という感じはない	
も男女とも打算なしという可能性としてので特に自由という感じもないが、政府	結婚	よは選択が広いと思うのでいい人間がするものから結婚したい	
が、政府結婚は選択がないと思うのでいい人間がするものから	結婚	したいと思える人に出会えて結婚できるなら幸せだけどどこまで思える人が本当にいるのか疑問	
と思うのでいい人間がするものから結婚したいと思える人に出会えて	結婚	できるなら幸せだけどどこまで思える人が本当にいるのか疑問。子供が生まれたら海<h5>----cell---	

トである。自由記述の回答には「結婚」という単語が38個、「幸せ」という単語が16個、「恋愛」という単語が16個登場していることが把握できる。

こうして分析データにおいて使用された単語を抽出した後、図5-3の「KWICコンコーダンス」を出力する。KWICコンコーダンスは、抽出語がどのような文章中に共起しているのかを把握するためのものである。例えば、図5-2中の「結婚」という単語箇所をクリックすると、図5-3のようなウィンドウが出力される。すると、「結婚」という単語がどのような文脈において用いられているのかを把握できる。

4. 1. 2. 恋愛結婚と婚活サービスのイメージの比較

表4-1は、恋愛結婚と婚活サービスそれぞれの頻出語リストを比較しやすいように並べたものである。この頻出語リストを見るだけではそれほど多くのことはわかりそうではないが、それでもいくつか特徴的なことも見られる。例えば、恋愛結婚には、「幸せ」や「理想」という言葉が散見されるが、婚活サービスでは「使う」、「サービス」、「利用」、「効率」などが多くロマンティックラブの様相は伺えない。

表4-1を見ると「イメージ」という単語が、恋愛結婚と婚活サービスの両方で頻出語として挙がっているので、それぞれについてKWICコンコーダンスから前後の文脈を確認しよう。すると、やはり恋愛結婚にはロマンティックラブの様相が見られ、婚活サービスは手段としての側面が強調されているように思われる。

表4-1. 恋愛結婚と婚活サービスの自由記述の抽出語リスト

恋愛結婚		婚活サービス	
結婚	38	人	36
幸せ	16	出合い	23
恋愛	16	使う	21
理想	14	結婚	20
人	13	イメージ	18
イメージ	11	サービス	16
良い	8	利用	15
関係	7	自分	12
自分	7	効率	11
大変	6	合う	11
好き	6	良い	11
自由	6	相手	9
素敵	6	婚活	7
多い	6	見つける	5
形	5	現代	5
相手	5	手段	5
楽しい	4	怖い	5
信頼	4	恋愛	5
お互い	3	悪い	4
一緒	3	出会う	4
一般	3	出会える	4
印象	3	少し	4
感情	3	場	4
現代	3	ツール	3
考える	3	印象	3
高い	3	機会	3
最も	3	見つかる	3
最近	3	個人	3

表 4-2. 恋愛結婚における「イメーজ」の前後文

多い。◇あまり癒して、辛いことは半分が嬉しいことは2倍、3倍になる	イメージ	もっと辛いこともあるけど、それ以上に楽しかったり居心地良かったりしない結婚はでもないの
<h5>◇熱が覚めた時に大変そうだしとは思いつつ、恋愛結婚には素敵な	イメージ	◇若い年代の人に多い◇幸せ◇最近の時代において
◇結婚は相手との価値観が異なると思うので、大変であるという	イメージ	◇しがらみ◇1番良い形、好きになった
で関係が続くといった夫婦が多いのではないだろうか。こうした自身の恋愛結婚の	イメージ	は、自分の最も身近な夫婿であるところの両親の関わりが良好ではなく、また世の中を
けど。◇いいかと思う◇最近のもの◇いつか終わる◇自由な	イメージ	親に束縛されないイメージ◇離婚率が高い◇自由というようにイメージ
◇最近のもの◇いつか終わる◇自由なイメージ	イメージ	◇離婚率が高い◇自由があるが、その分簡単に
イメージ	イメージ	親に束縛されないイメージ◇自由というようにイメージがあるが、その分簡単に解消されてしまう場合もある。◇出来たら幸せだ
かが、最も重要だと考える◇◇難しいもの◇仲睦まじく暮らしている	イメージ	◇いいいのである◇自分は好きになった人と真剣に交際して結婚
ている感じ◇憧れ◇恋愛と結婚は必ずしも一致しないと思うし、大変な	イメージ	◇あるのが、恋愛結婚をする人は尊敬するし、憧れもある◇いいよなあ
な人と家族になる◇運命の人とは必然的に結ばれる◇ポジティブな	イメージ	◇◇良き◇最高◇大変そう◇幸せそう
◇良いことだと思う◇新たな家族を受け入れる、人生で重大なイベントという	イメージ	◇不自由◇とても良いと思う◇感情の起伏が激しい◇いいい

表 4-3. 婚活サービスにおける「イメージ」の前後文

手段◇登録上のステータスとは違う人が現れそう。◇◇特に悪い	イメージ	などは持っています。恋愛や結婚の形も多種多様かと思うので。◇
忙しい現代人にとって必須サービスだと思ふ。◇否定的◇合理的◇悪い	イメージ	がある◇少子化中だからとっても良いこと◇今のネット社会ではうまぐ
いるもの◇婚活サービスに書かれている情報が本当とは限らないので、少し怖い	イメージ	があります。◇最終手段◇出会いの場としては悪いものでは
人達が集まる効率的な場所◇ありだと思ふけどプロがある人が使いそうな	イメージ	◇使う人は自由に使えばいいと思うが、私は利用しないと思う。< h5>
と少し疑っている。◇どうして結婚したい人や一度離婚した人が使っている	イメージ	◇現代では自分の属するコミュニティの外にパートナーを狩りたいという人や、そもそも出会いが
便明だも思います。◇出会える人の性格が似通っていいそう◇◇あまり、よい	イメージ	はもっていない。危険がありそうなので、自分としてはあまり使いたいとは思わない。
的に出会いの場が少なくなっている現代においては効率的が良いと思う。利用者への	イメージ	があまり良くないから、非難してしまうくらいのもう少しオープンになれば良いと思う
、非難してしまいう人もいと思うので多少オープンになれば良いと思う◇◇あまりいい	イメージ	がない◇◇ちょっと危ない人がいる◇あまりいい印象はない◇◇あまりいい
ないでし。◇◇ちょっと危ない人がいる◇◇あまりいい印象はない◇◇あまりいい	イメージ	がない◇◇自分の理想とする出会い方ではない◇特になし◇個人
が自由に感ずるサービスとして好意的に思っている。◇◇こで知り合ってもなんだかんだ上手いく	イメージ	◇マイナスなイメージ◇若い人は使わなそう◇悪いイメージはないが、
もなんだかんだ上手いくイメージ◇マイナスなイメージ◇若い人は使わなそう◇悪い	イメージ	◇若い人は使わなそう◇悪いイメージはないが、自分は使おうとは思
繋がることに自然だし不安がある。◇◇あまり使いたくはない。集つてくる人が使い	イメージ	はないが、自分は使おうとは思わない。◇◇めつてしまいいいと思
ーcellー< h5> 全然いいと思う◇◇あまり使いたくはないと思われない◇◇あまりいい	イメージ	◇◇あまり利用したくない◇全然いいと思う◇◇ちょっと怖いけど
◇◇あまりお互いを知らない◇誰かとの出会いに期待する◇ネガティブな	イメージ	◇◇出会いがない人が使う◇出会いの場として、うまぐ利用され
利用さればいい機会になると思う◇◇いいと思う◇◇突如して相手を探す	イメージ	◇◇結婚に必死な人◇◇ヤリモク多数◇余
と思う◇少し不安がある◇余裕がない人、もしくはは避んでいる人が使う	イメージ	◇微妙な人とも思われない◇条件に合った相手を見つけれられるため効率が

4. 2. テキストマイニング②：共起ネットワーク

4. 2. 1. KHコーダーにおいて共起ネットワークを出力する

テキストマイニングの醍醐味のひとつが共起ネットワークである。共起ネットワークは、それぞれの抽出語の距離が近いか遠いかを計算し図化されたものである。共起ネットワークでは、抽出語は「node」と呼び円で表現され、距離を「edge」と呼び線で表現される。共起ネットワークでは、nodeはedgeによって結ばれている。

共起関係のedgeには強い弱いがある。テキスト中に共通に登場する抽出語どうしはedgeが強くなり、重要なのは、図中の実際の距離ではなく、線の濃さである。こうしたedgeの強度を判定するのがJaccard係数である。

図6-1がKHコーダーで共起ネットワークを出力させるための操作方法である。メニューバーから「ツール」→「抽出語」→「共起ネットワーク」を選択する。すると図6-2のようなウィンドウが開く。デフォルトはソフト開発者が最適だと考える分析が設定されているので、そのまま「OK」をクリックしても問題ない⁷⁾。

なお、共起の強さを測るJaccard係数は、デフォルトでは0.2で表示される。また、「Jaccard」のタブをクリックすれば、Cosine係数とEuclid距離という別の測定も選択できる。図6-2中の「強い共起関係ほど濃い線に」や「係数を表示」にチェックを入れることで、出力される共起ネットワークを調整できる。

図 11. 共起ネットワークの調整ウィンドウ

- 7) 共起ネットワークは、デフォルトではカラーのバブルプロットで出力される。そのため、度数が大きい頻出語ほど大きな円で表示され、それぞれの円が色違いで弁別可能である。しかし、バブルの大きさに関心を置かず、論文掲載に際してグレースケールで表示することが求められる場合もある。その時には、共起ネットワークの出力画面下にある「調整」をクリックし図11のようなウィンドウを呼び出し、「バブルプロット」のチェックを外し、「グレースケールで表現」にチェックを入れる。

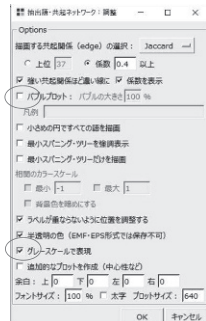


図 6-1. 共起ネットワークの出力ウィンドウ

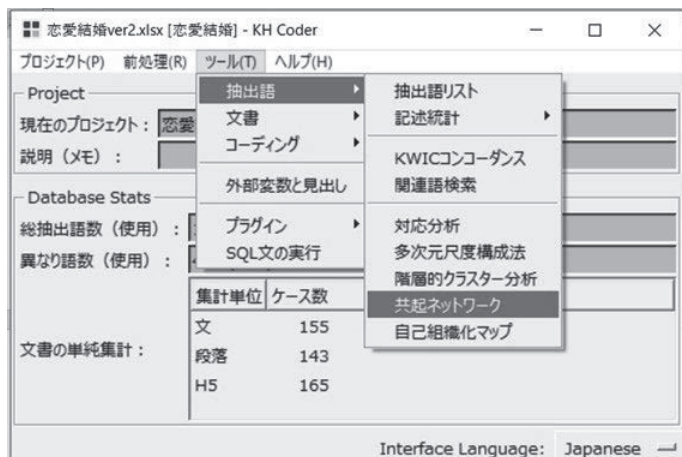
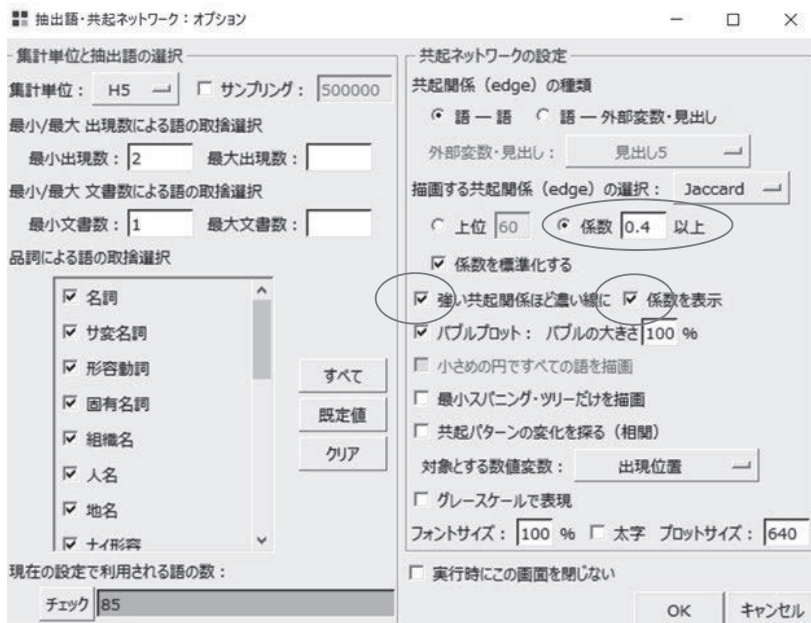


図 6-2. 共起ネットワークのオプション操作ウィンドウ



4. 2. 2. 恋愛結婚と婚活サービスの共起ネットワークを比較する

図7-1と図7-2が、恋愛結婚と婚活サービスに対するイメージの共起ネットワークによる分析結果である。なお、ここでの分析結果は、Jaccard係数が0.4以上を条件として設定した。Jaccard係数は単語が集合する類似度を数値化したものであり、この係数が大きいほどテキストの集合の類似度が高い。KHコーダーのデフォルト設定は0.2であるが、本データのサンプルサイズが大きくないことなどもあってか、それほど明確な共起ネットワークを見いだせなかった。そのため、Jaccard係数0.4以上を条件として図7-1と図7-2を提示している。

この共起ネットワークを見ると、表4-2と表4-3で確認した頻出語がそれぞれどのように組み合わせられているのかが把握できよう。恋愛結婚について繋がる言葉は、「夫婦—良好—続ける—子供—関係」、「家庭、築ける」「互いに、一緒」のような非常にロマンティックなものである。

一方で、婚活サービスの方は、「最終—手段」、「個人—自由」、「少し—不安」のような言葉が繋がっている。さらに、婚活サービスにおいてKWICコンコーダンスで「個人」と「自由」の前後の文脈を確認すると以下のものであった。

- ◆ どんどん広まってほしい◇効率良い◇新しい形の一つ◇個人情報の流出が怖い◇真剣な人同士で結婚しやすいのでいいと思う。
- ◆ 恋愛観を押し付け合ってしまうなどの弊害もあるのではないかとも思う。個人の自由だと思いますが、私は使いたいとは思いません。
- ◆ 自分の理想とする出会い方ではない◇個人が自由に恋愛するサービスとして好意的に思っている。◇そこで知り合ってもなんだかんだ上手くいく。
- ◆ 多様性◇結婚できない人が使うイメージ◇使う人は自由に使えばいいと思うが、私は利用しないと思う。

このように共起ネットワークからは、恋愛結婚へのロマンティックなイメージと婚活サービスへの「最終手段」、「少し不安」、「個人の自由（だが使いたくない）」のようなイメージが伺える。

図 7-1. 恋愛結婚の共起ネットワーク

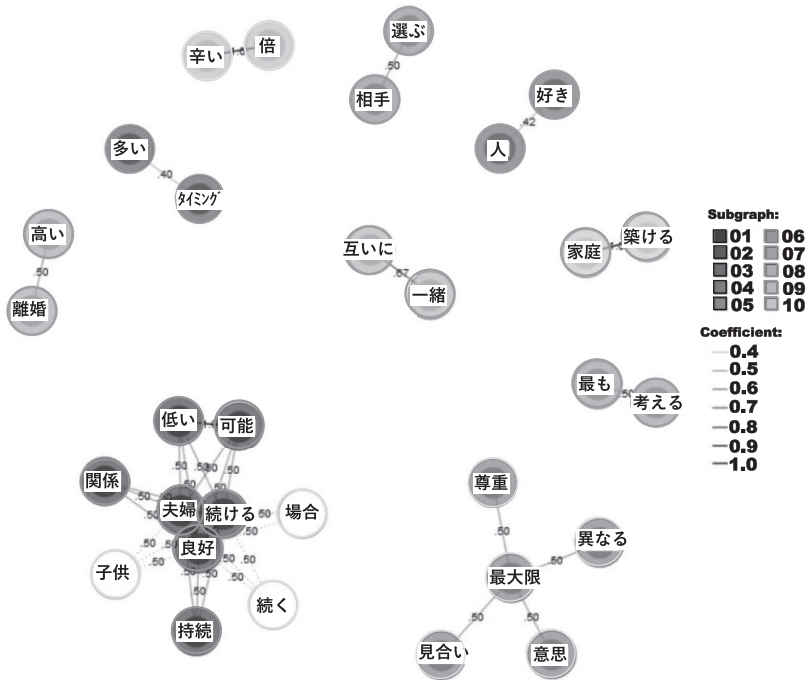
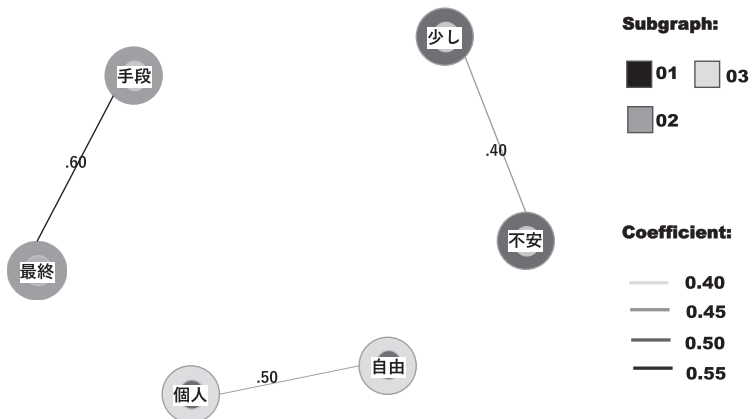


図 7-2. 婚活サービスの共起ネットワーク



4. 3. テキストマイニング③：対応分析

4. 3. 1. KHコーダーにおいて対応分析を実行する

対応分析は、抽出語が外部変数（性別など）によってどのように特徴的に表れるのかを視覚的に把握することができる。図9-1と図9-2を眺め、イメージを掴んでくれても良いだろう。図中の縦軸と横軸それぞれの0(0, 0)の交点(=原点)には特徴のない言葉が集まる。反対に、図中の外側には特徴的な言葉が配置されることになる。

図8-1がKHコーダーで共起ネットワークを出力させるための操作方法である。メニューバーから「ツール」→「抽出語」→「対応分析」を選択する。図8-2のようなウィンドウが開いたら、「抽出語×外部変数」の箇所に自分の関心のある変数を選択する。言うまでもなく、自身の関心のある外部変数は、データを読み込む際にエクセルファイル上で作成しておく必要がある。そして、「OK」をクリックすれば結果が出力される。

図8-1. 対応分析の出力ウィンドウ

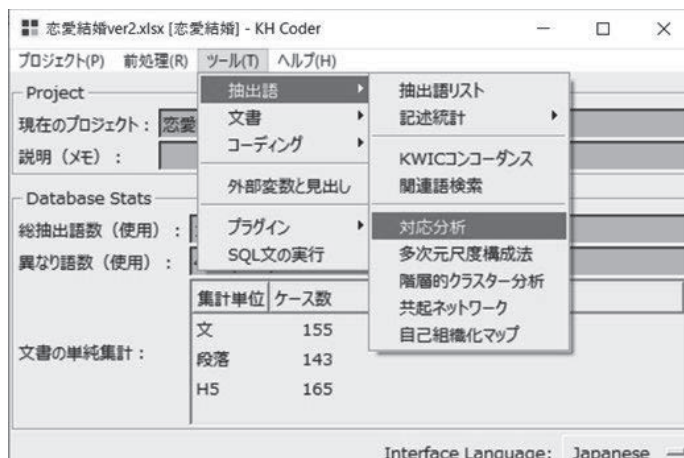
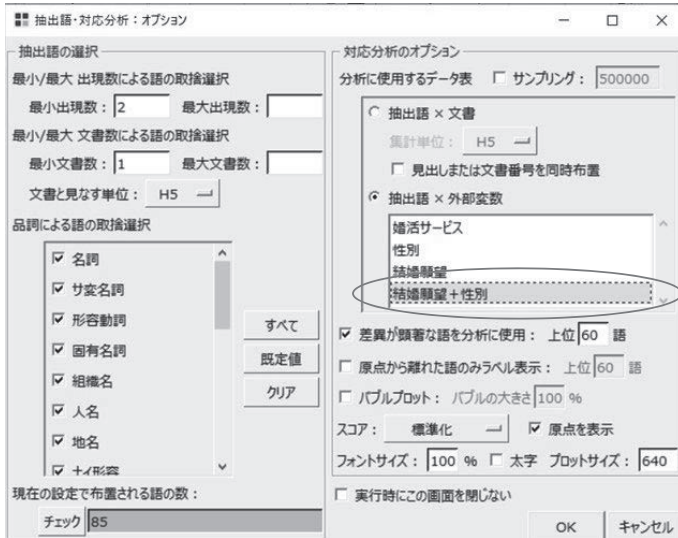


図 8-2. 対応分析のオプション操作ウィンドウ



4. 3. 2. 恋愛結婚と婚活サービスの対応分析の結果を比較する

それでは、結婚希望か否かと性別で4つにカテゴライズした変数を用いて対応分析の結果を見ていこう。

図 9-1 は、恋愛結婚についての対応分析の結果である。男女ともに、「結婚したくない」と回答していると近くに言葉が集まっていない。結婚を希望していない場合には、恋愛結婚へのイメージをそれほど明確に持っていないことが伺える。

一方で、結婚を希望している場合には、男女それぞれで次のように言葉が集まっている。それぞれにポジティブな言葉が並ぶが、男子には「辛い」や「責任」という言葉も見られ、「所帯を背負う」ことへの責任感が伺える。

- ✓ 結婚したい男子 = 良い, 楽しい, 憧れる/難しい, 辛い, 責任
- ✓ 結婚したい女子 = 愛, 素敵, 大事, 好き, 幸せ

次に、図9-2に示した婚活サービスのイメージに対する対応分析の結果を見ていこう。まず、結婚を希望していない男子以外のカテゴリーは比較的近い位置に配置されていることがわかる。それ以外のカテゴリーでは、「最終手段、少し不安、個人の自由（だが使いたくない）」といったややネガティブな言葉が散見される。大まかにまとめると以下ようになる。

- ✓ 結婚したい女子 = 少し不安
- ✓ 結婚したくない男子 = 個人の自由

こうしたネガティブと思われる言葉がどのような文脈で用いられているのかについて、KWICコンコーダンスから「悪い」と「不安」の前後の文脈を確認しよう。

- ◆ 特に悪いイメージなどは持っていません。恋愛や結婚の形も多種多様かと思うので。
- ◆ 否定的◇合理的◇悪いイメージがある◇少子化中だからとっても良いこと
- ◆ 少し怖いイメージがあります。◇最終手段◇出会いの場としては悪いものではない。どんな人がいるのか分からないから怖い。
- ◆ マイナスなイメージ◇若い人は使わなそう◇悪いイメージはないが、自分は使おうとは思わない。
- ◆ デジタルを使用した画期的なサービスだとは思いますが使おうとは思わない。◇少し不安◇自分と合う相手が見つかるなら利用してもいいのではないか。
- ◆ 出来れば使いたくない。不特定多数と繋がることに漠然とした不安がある。◇あまり使いたくない。焦っている人が使うイメージ。
- ◆ あまりいいと思わない◇コロナを機にみなが主体的になると思う◇少し

不安がある◇余裕がない人、もしくは遊んでいる人が使うイメージ◇微妙次に、婚活サービスへの漠然としたイメージではなく、自分事としてはどのように考えているのかを把握するために「自分」という言葉をKWICコーダンスで見よう。

- ◆ どんな人がいるのか分からないから怖い◇合理的◇◇自分は将来的に結婚を望んでいないため、使用したいとは思わない。
- ◆ 自分以外の第三者が使うことに対しては何も思わないが、どれくらいの割合で婚活サービスを通して出会ったがあるなら供給されて然るべしだし、他人の需要に関してとやかく言うつもりはない。ただ自分はあまり興味を持たない。◇20代後半になってから使うもの◇
- ◆ 注意する必要があると感じる◇最近のもの◇良い出会いの見つけ場所◇自分に合った人が効率的に見つけられて良いと思う。
- ◆ ◇ありだと思う◇あまり良い出会いがなさそう◇◇自分に合った人を探してくれる。
- ◆ どうしても結婚したい人や一度離婚した人が使っているイメージ◇現代では自分の属するコミュニティの外にパートナーを持ちたいという人や、そもそも出会いがない人も多いので非常に似通ってそう
- ◆ ◇あまり、よいイメージはもっていない。危険がありそうなので、自分としてはあまり使いたいとは思わない。◇趣味の合う人どうしをマッチングさせる。
- ◆ ◇あまり好ましくない◇自然でない◇あまり利用したくない◇自分に合う人が見付きやすいと思うので良いサービスだと思う。
- ◆ 画期的なサービスだとは思いますが使おうとは思わない。◇少し不安◇自分と合う相手が見つかるなら利用してもいいのではないかな。
- ◆ ◇やろうとは思わないが、否定もしない。◇自分が結婚したい年齢で結婚できなかった人が、婚活サービスを利用しよう。

- ◆ ちょっと危ない人がある◇あまりいい印象はない◇あまりいいイメージがない◇自分の理想とする出会い方ではない◇特になし◇個人が自由に恋愛するサービス
- ◆ ◇マイナスなイメージ◇若い人は使わなそう◇悪いイメージはないが、自分は使おうとは思わない。◇◇めっちゃいいと思います！

このようにKWICコンコダンスも合わせて確認すると、「悪いイメージはない」とはしつつも利用には不安を持ち、自分自身では利用に積極的ではないことが伺える。また、手段的側面が強調され、婚活サービスを通じた結婚をロマンティックだとは感じていなそうではある。

図 9-1. 恋愛結婚の対応分析

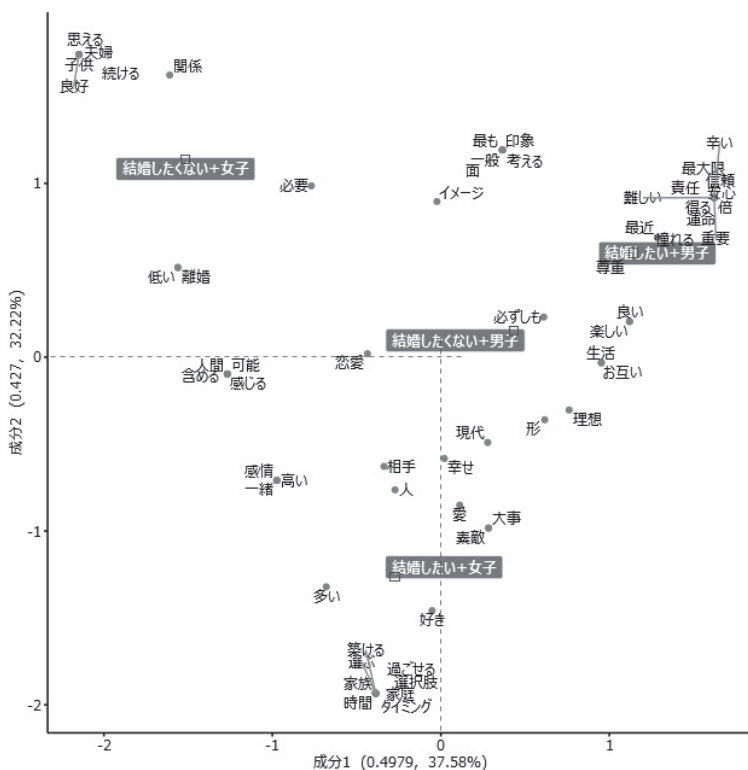
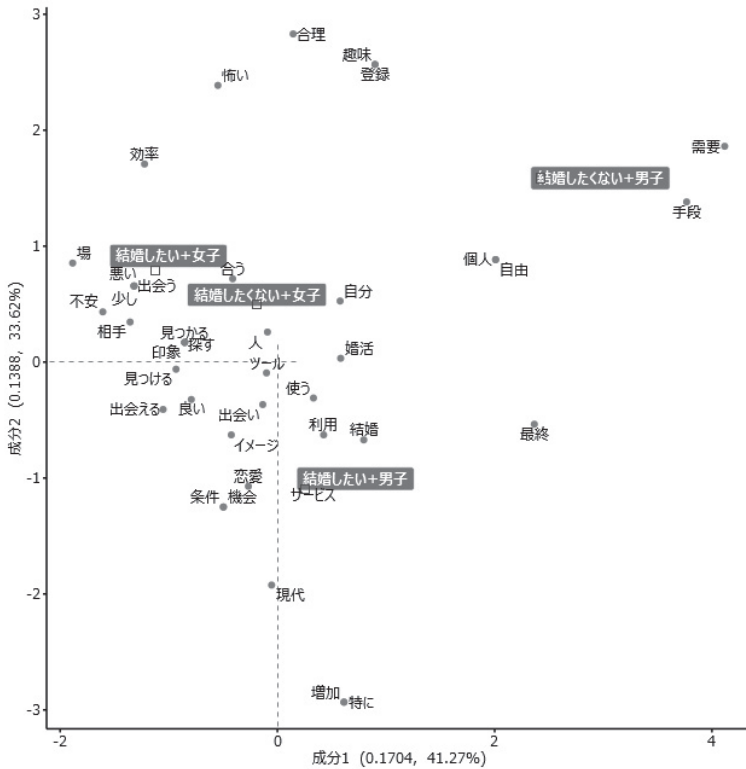


図9-2. 婚活サービスの対応分析



5. まとめ

ここまで大学生の恋愛結婚と婚活サービスのイメージについてのテキストマイニングを行ってきた。まず量的な側面を再確認しておく、調査対象者のうち140人(約85%)が結婚したいと考えているため、結婚希望者が多数派なことが把握できる。そして、テキストマイニングの結果、大学生の多くは恋愛結婚がしたいのであって、婚活サービスを使うのは「手段」でしかないととらえていることが伺える。結婚したい男子には、「サービス」、「利用」、「機会」などの手段的側面の言葉が並ぶが、結婚したい女子では、「悪

い]、「出会う」、「不安」、「探す」など男子に比べややネガティブイメージが伺える。つまり、大学生の結婚観におけるロマンティックラブ・イデオロギーが再確認できたことになる。

改めて、本稿のデータでも結婚希望と子どもが欲しいかどうかの関連を確認しておこう。表5は、それらの変数をかけ合わせたクロス集計表である。これを見ると、結婚を希望する140人のうち90%が子どもを欲しいと考えており、やはり結婚と子どもはセットとして家族形成をとらえていることがわかる。

少子化を社会問題として考え、その解決を目指すためのインプリケーションを述べるとすれば、「恋愛結婚を促す」ということになろうか。あるいは、見合い結婚や「歳の差婚」を再び増やすべきという主張になろうか。いずれのインプリケーションであっても、個人の価値や選択の問題であるため、それほどシンプルな解決策はなさそうである。

表5. 結婚希望と子どもが欲しいかどうかのクロス集計表

	子どもが ほしい	子どもは ほしくない	合計	N
結婚したい	90.0%	10.0%	100.0%	140
結婚したくない	16.0%	84.0%	100.0%	25
全体	78.8%	21.2%	100.0%	165

カイ二乗値=69.5, 自由度=1, $p < .001$

最後に、大学での社会調査教育におけるKHコーダー利用の可能性をフリーソフトであることの利点に着眼して記述しておこう。

社会調査科目において、最も大きな課題のひとつに分析ソフトの利用料金がある。例えば、量的調査法の授業であるが、IBM社のSPSSは大学の社会調査科目でも頻繁に使われるソフトウェアだが非常に高価である⁸⁾。テキス

8) JASPなどのフリーソフトの開発も進んでいるが、これについては耳塚・中西(2021)のコラム「データ分析と統計ソフト」も合わせて参照されたい。

トマイニングの分析ソフトでは、Text Mining StudioやMAXQDAが有名だが、やはりそれなりに高価である。むしろ有料ソフトにはそれに見合ったサービスを受けられるというメリットがあるのだが、大学で教育利用する数十台のPCへ導入するには金銭的なハードルは高く、全ての学生に購入させることも難しい⁹⁾。その点では、KHコーダーはフリーソフトのため教育利用に適している。

さらに、KHコーダーはインストールせずに実行ファイル(.exe)で操作することも大学教育に利用しやすい点である。しばしば大学のPCでフリーソフトを利用する場合には、ウィルス対策の観点からインストールを避けたいことも多いが、KHコーダーであればその障壁は低い。

〈文献〉

- 荒川和久, 2017, 『超ソロ社会—「独身大国・日本」の衝撃』PHP新書。
- 樋口耕一, 2004, 「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』19(1), pp.101-115。
- 樋口耕一, 2011, 「計量テキスト分析の提案と必要なソフトウェアの開発」『ソシオロジ』55(3), pp.102-108。
- 樋口耕一, 2020, 『社会調査のための計量テキスト分析—内容分析の継承と発展を目指して』ナカニシヤ出版; 第2版。
- 岩澤美帆, 2008, 「初婚・離婚の動向と出生率への影響」『人口問題研究』Vol. 64(4), pp.19-34。
- 小林盾・谷本奈穂, 2016, 「容姿と社会的不平等—キャリア形成, 家族形成, 心理にどう影響するのか」『成蹊大学文学部紀要』第51号, pp.99-113。
- 高史明, 2015, 『レイシズムを解剖する—在日コリアンへの偏見とインターネット』勁草書房。
- 国立社会保障・人口問題研究所, 2017, 『現代日本の結婚と出産—第15回出生動向基

9) なお、近年では有料の分析ソフトもサブスクリプション・アプリやアカデミックバックを併用などすれば金額的に利用しやすくなっている。こうした動向もキャッチアップする必要はあるだろう。

- 本調査（独身者調査ならびに夫婦調査）報告書】。
- 内閣府，2017，『平成29年版 少子化社会対策白書』。
- 耳塚寛明監修・中西啓喜編，2021，『教育を読み解くデータサイエンス—データ収集と分析の論理』 ミネルヴァ書房。
- 三輪哲，2019，「結婚を阻む「壁」の在り処—結婚意識と配偶者選択」石田浩監修・佐藤博樹・石田浩編『格差の連鎖と若者2 出会いと結婚』勁草書房，pp.15-43。
- 谷本奈穂・渡邊大輔，2019，「ロマンティックラブ・イデオロギーとロマンティックマリッジ・イデオロギー—変容と誕生」小林盾・川端健嗣編『変貌する恋愛と結婚—データで読む平成』新曜社，pp.48-70。
- 佐藤郁哉，2006，『フィールドワーカー書を持って街へ出よう 増訂版』新曜社。
- 白波瀬佐和子，1999，「階級・階層，結婚とジェンダー—結婚に至る階層結合パターン」『理論と方法』Vol.14(1)，pp.5-18。
- 打越文弥，2016，「学歴同類婚の世代間連鎖とその趨勢—大規模調査データの統合による計量分析」『家族社会学研究』Vol.28(2)，pp.136-147。

〈謝辞〉

調査の設計，実施，分析にご協力いただきました南口拓巳氏（早稲田大学大学院）に厚く感謝いたします。