

## 報告

# 学生が作成したキャリアプランのテキストマイニング

生田和重

徳島文理大学 一般総合科目

要約：これまでの指導経験をもとに学習支援 Moodle サイトを構築し、それを活用して学生主体の授業を展開している。その一つである「情報コミュニケーション C」では、新聞投稿文やキャリアプランなどの文章を作成し、学生各々が楽しみながら日本語文章力を高めている。本報では、このキャリアプランを学生指導で活用することを目的として、学習支援 Moodle サイトにアップロードされた文書ファイルをテキストマイニング技術で解析した。その結果、出現する語の共起ネットワークや対応分析などの図は、学生の考えを把握するための一助となることが分かった。今後は、このような視覚資料を蓄積するとともに、学生指導で活用したい。

(キーワード：高等教育、キャリアプラン、学生指導、テキストマイニング、Moodle)

## Text Mining Analysis of Career Planning Documents Created by My Students

Kazushige IKUTA

Liberal Arts, Tokushima Bunri University

Abstract: A learning support website was constructed using Moodle in 2012 and has been used in my classroom. The students in my classes can experience student-based and cooperative activities on the Moodle website. In a class called "Information Communication C," the students can improve their writing proficiency in Japanese by enjoying making letters to a newspaper and career planning documents. This paper provides the results of the text mining analysis of career planning documents created by my students. The results indicate that the graphs of word co-occurrence networks and correspondence analyses contribute to comprehending what my students are thinking. I want to accumulate the visual information and exploit it in mentoring.

(Key words: higher education, career planning, mentoring, text mining, moodle)

### 1. はじめに

前報<sup>1)</sup>で紹介したように、Moodle<sup>2)</sup>を活用した学習支援サイトを構築し、学生自身が主体的に学べる環境を提供している。2013年度では、この学習支援 Moodle サイトを活用して、「情報コミュニケーション C」や「教育方法・技術論」などの授業を実施した。その結果、学生は課題をアップロードしたり、評価を確認したりできる Moodle の機能に好印象を持っていることが分かった。また、フォーラムへの投稿を積極的に参照しつつディスカッションを行なうことにより、学生同士のやりとりを活性化することができた。これにより、中央教育審議会の答申<sup>3)</sup>に則した教育改善への第一歩を踏み出せたと考える。

学習支援 Moodle サイトを構築したサーバー内には、学生各々が作成したレポート (Word 文書ファイル) が保存されている。その都度、このレポートに対する評価やフィードバックを返しているため、それを参考にして学生は学習を進めること

ができる。これによって、レポートの役割を一応は果たしたと考えられる。しかし、それだけではもったいない。レポートに込められた学生各々の思いを把握して、個別指導に活かすことができれば、さらに教育効果は上がるだろう。

そこで本報では、学生がマインドマップ<sup>4)</sup>を活用して作ったキャリアプランをテキストマイニング技術で解析し、その結果を学生指導で活かす術を考える。

### 2. 解析方法

筆者が担当する「情報コミュニケーション C」では、学生各々のキャリアプランを Word で作成して Moodle サイトにアップロードしてもらっている。この演習では、3年後、5年後、そして10年後の自分の状況を表したマインドマップを描画し、その内容を文章で説明する。今回は、この説明文 (28件) を対象として解析を行なう。

解析の手順は以下の通りである。まず、Word

の文章校正機能を活用して、説明文に入力ミスや不自然な表現が無いかどうかを確認する。もし該当箇所があれば、内容を変えることなく修正する。その後、各説明文を個別にテキストファイル形式で保存する。つぎに、各テキストファイルを KH Coder<sup>5)</sup> にデータとして読み込んでテキストマイニングを行なう。この KH Coder では、データから語を抽出する際に形態素解析システム「茶釜」<sup>6)</sup> を用いる。また、対応分析<sup>7)</sup> や共起ネットワーク分析<sup>8)</sup> を行う際には、統計計算とグラフィックスの環境を提供する R<sup>9)</sup> を活用する。なお共起ネットワーク分析では、語と語との共起関係の強弱を Jaccard 係数<sup>10)</sup> の大小により判断している。さらに今回は、共起ネットワーク分析と対応分析に加えて、Word Cloud を描く。Word Cloud とは、出現頻度が高い順に単語を選び出し、その頻度に応じた大きさで図示する手法である。この手法は KH Coder の機能として提供されていないので、出現頻度のデータを R に読み込んで、Jockers<sup>11)</sup> を参考にして作図する。

### 3. 解析結果

学生が作成したキャリアプランを対象にテキストマイニングを行なった結果を紹介する。まず、28 件全てに対して共起ネットワークを作成した。その際には、図 1 に示すように、抽出語の最小出現数を「2」、文書とみなす単位を「文」と設定し、全品詞を選択した。また図 2 に示すように、Jaccard 係数を「0.1」に設定した。

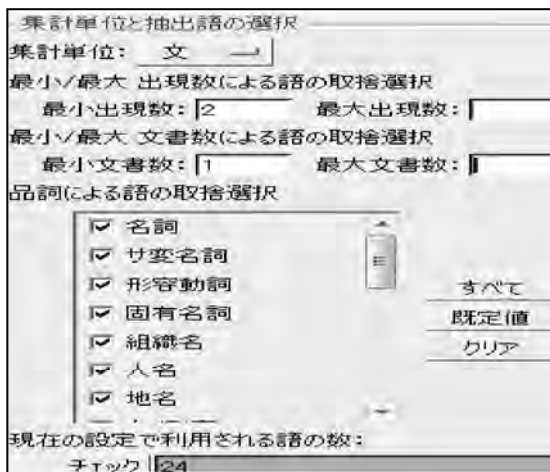


図 1 解析の基本条件

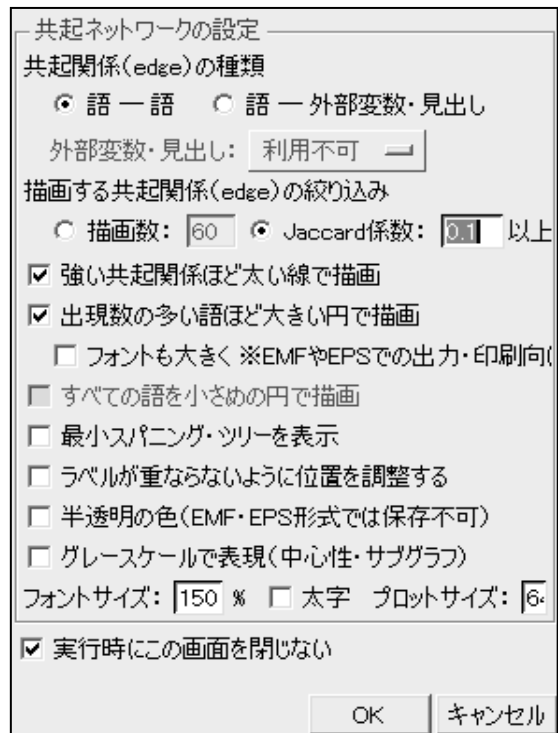


図 2 共起ネットワーク分析の設定

このような条件で求めた共起ネットワークを図 3 から図 7 にまとめて示した。Case 番号の右に示した 3 つの数値は、原文の文字数、抽出語数、そして Jaccard 係数である。今回対象としたキャリアプランの作成では、文章量の制限を設けていなかったため、文字数は 446 字から 1165 字まで様々である。それに対応して、抽出語数は 13 から 41 までばらついている。

Case 19 や Case 24 のように抽出語数が極端に少ない場合には、原文の主旨を推測するために想像力を豊かに働かせる必要がある。そのため、図を見る人によって解釈がばらつくと考えられる。なお Case 24 で、最小出現数を 1 に設定すると、抽出語数は 65 まで増える。これでは共起ネットワークの表示が煩雑になるので、「品詞による語の取捨選択 (図 1 参照)」で「名詞」と「動詞」のみを指定すると、抽出語数は 36 となる。その場合の共起ネットワーク (図 8) は、図 6 の Case 24 と比較して内容を推測しやすい。ただ、文章量に合わせて品詞を選択するという煩わしい作業は、できれば省略したい。そこで、抽出語数が極端に少なくなるように、文章量の制限を設けるのが合理的であると考えられる。

括弧内: 原文の文字数, 抽出語数, Jaccard係数

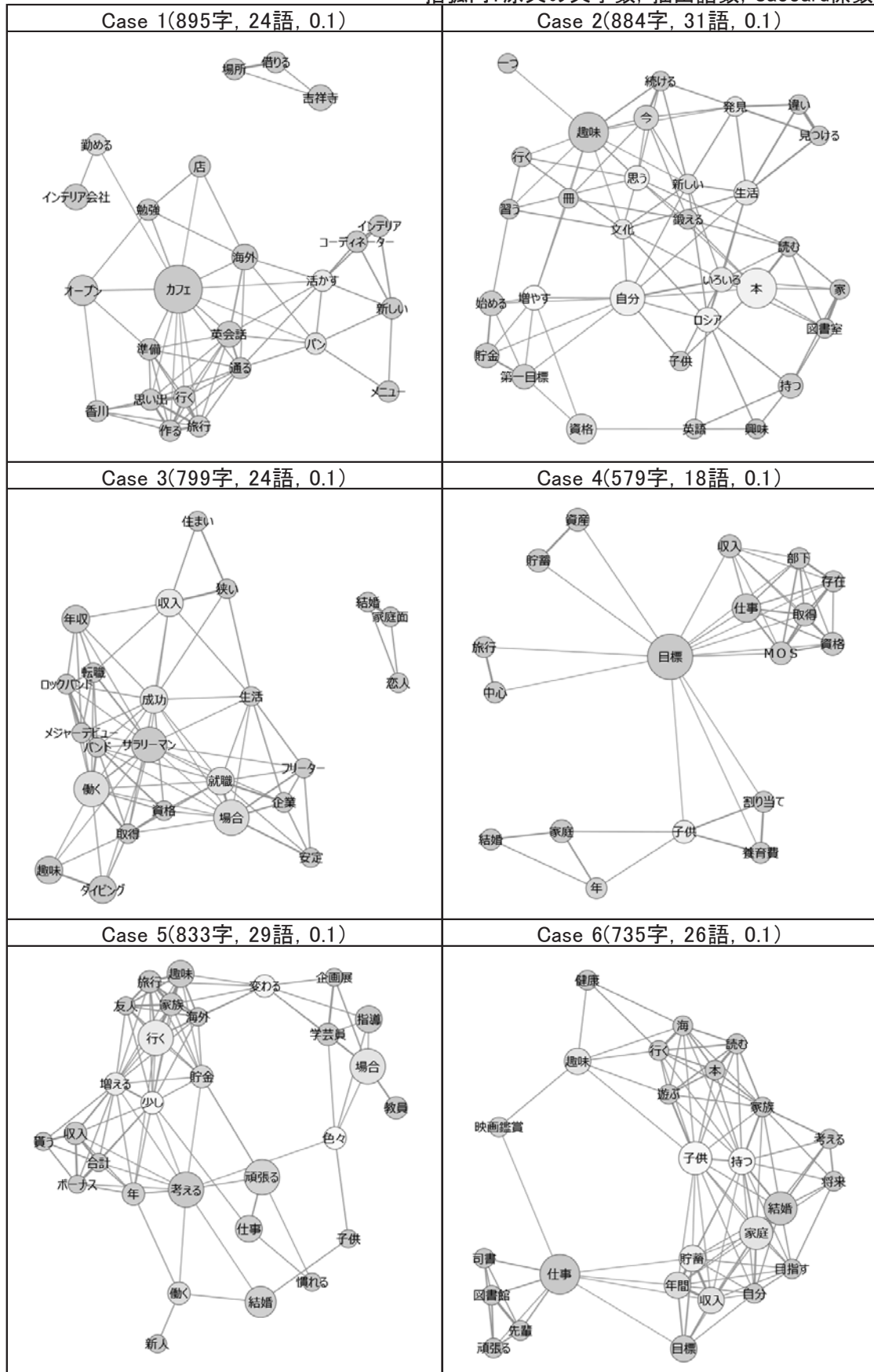


図 3 共起ネットワーク (Case 1~Case 6)

括弧内: 原文の文字数, 抽出語数, Jaccard係数

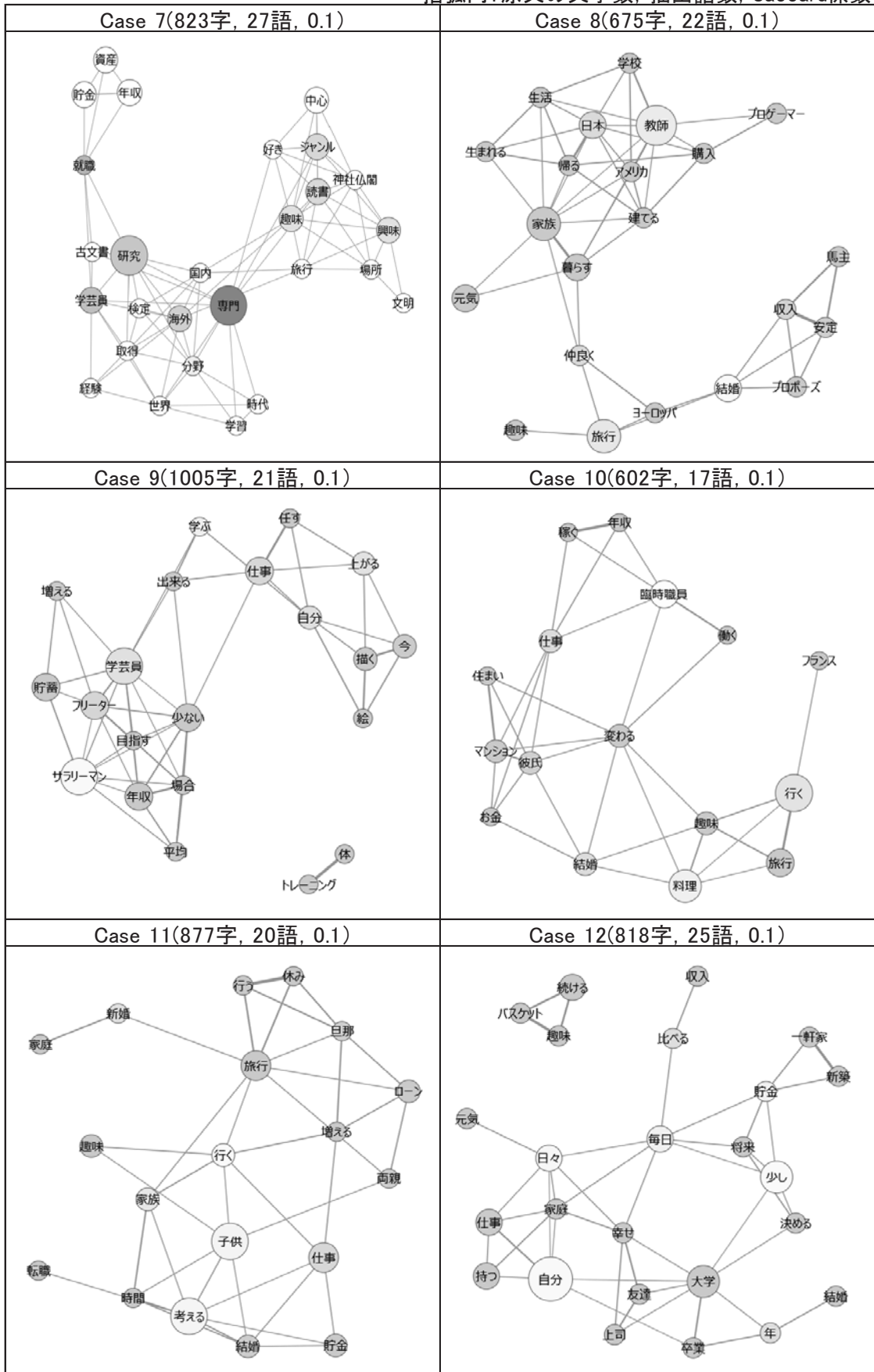


図 4 共起ネットワーク (Case 7~Case 12)



括弧内: 原文の文字数, 抽出語数, Jaccard係数

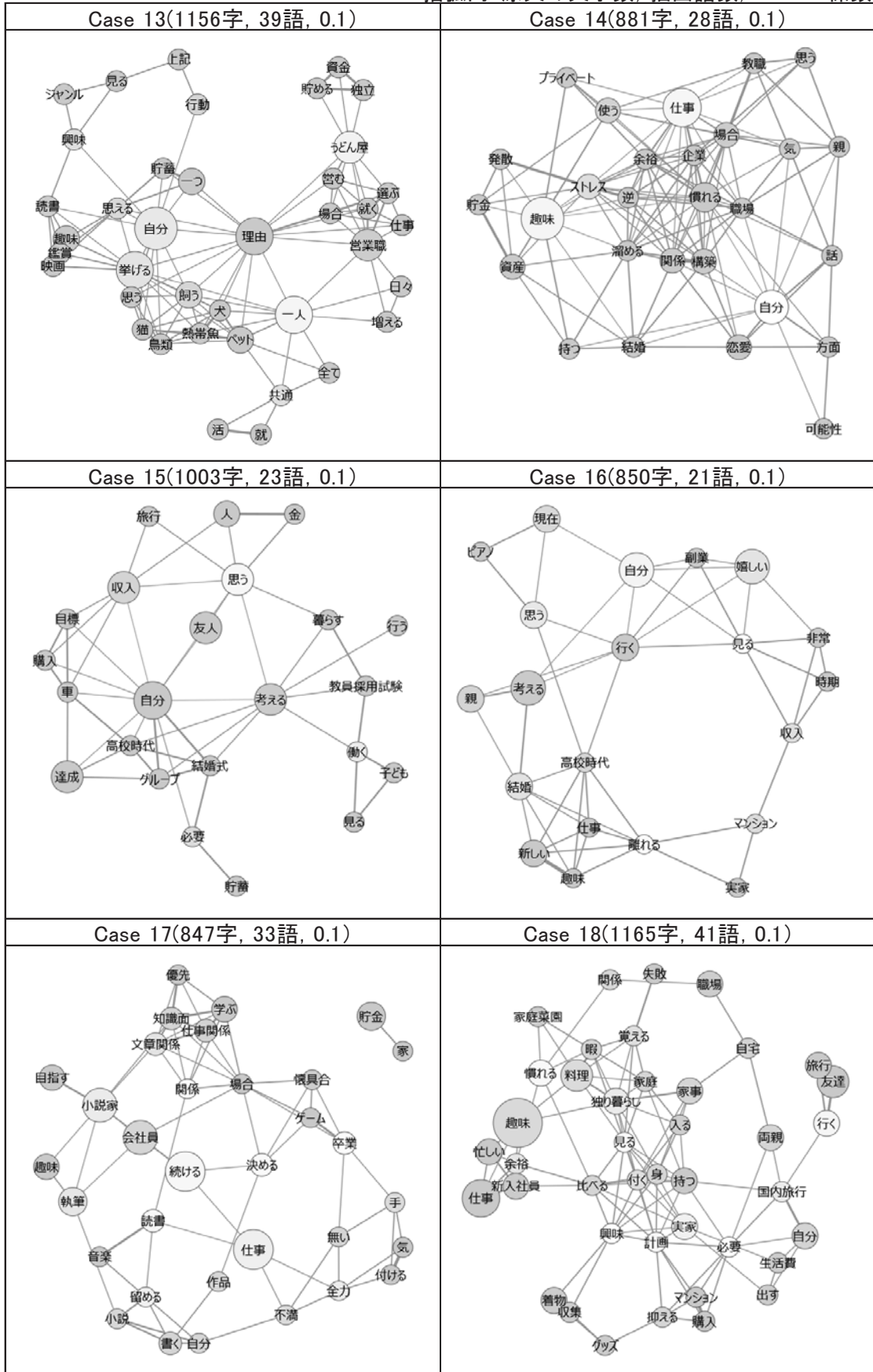


図 5 共起ネットワーク (Case 13~Case 18)

括弧内: 原文の文字数, 抽出語数, Jaccard係数

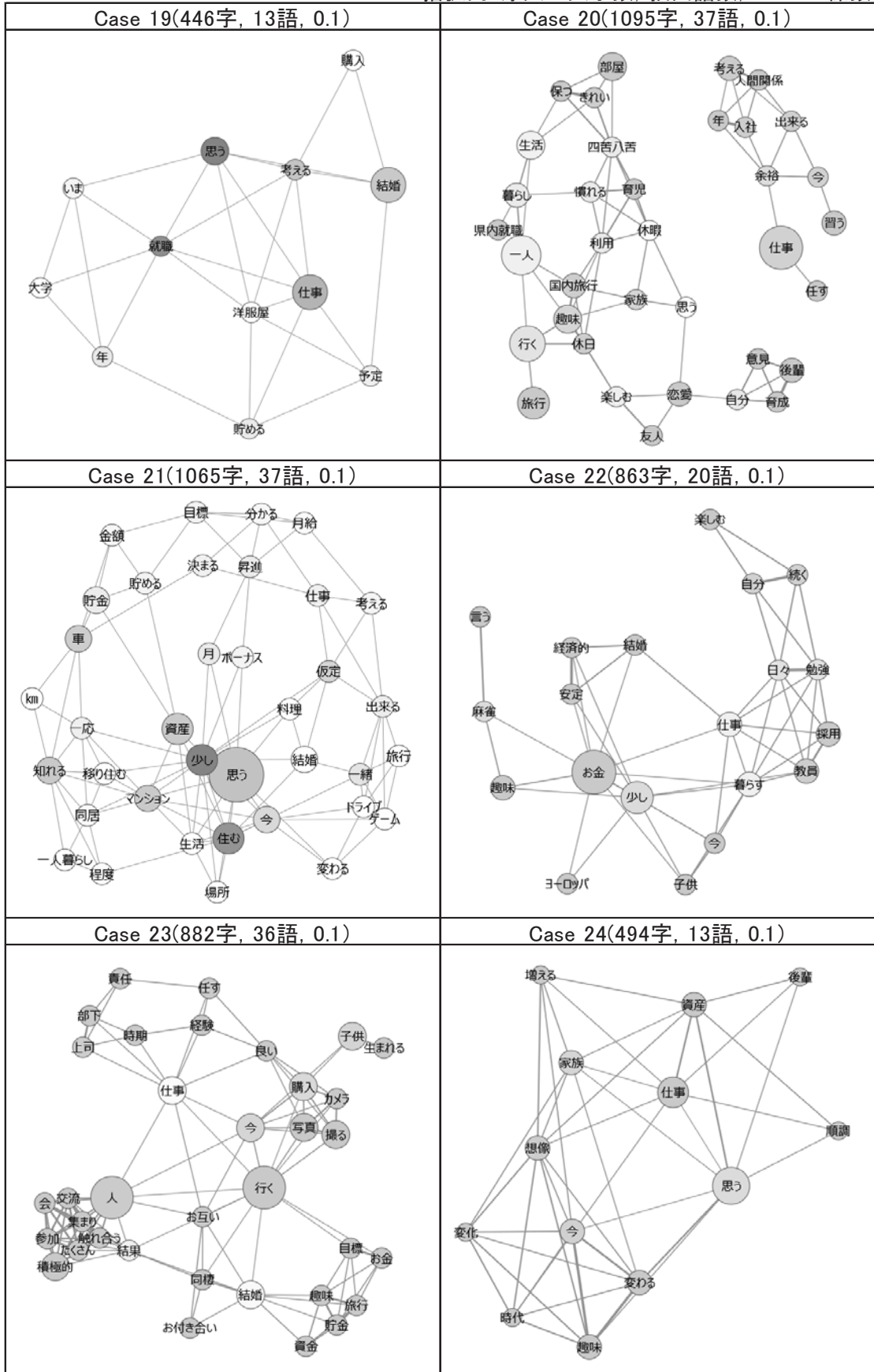


図 6 共起ネットワーク (Case 19~Case 24)

括弧内: 原文の文字数, 抽出語数, Jaccard係数

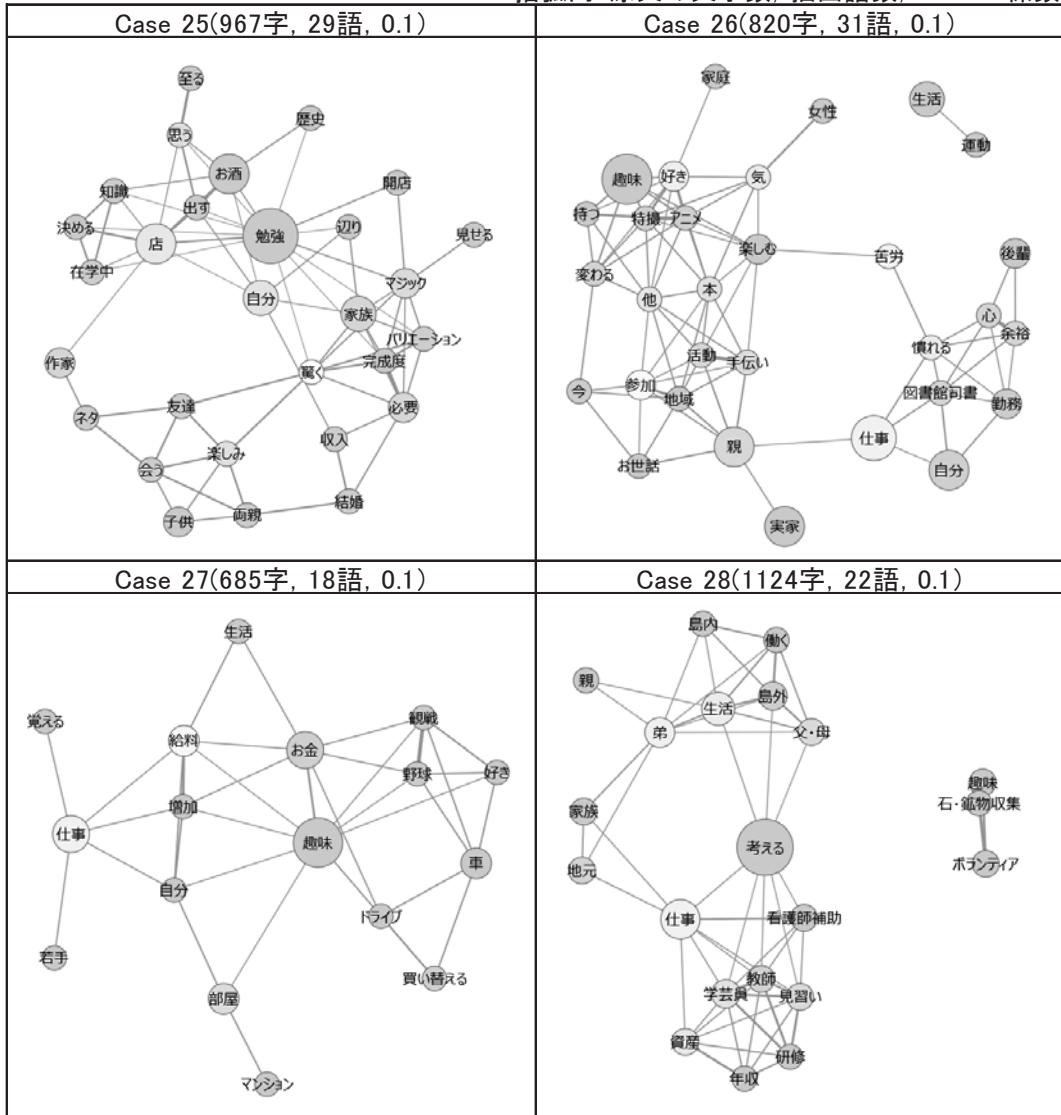


図 7 共起ネットワーク (Case 25~Case 28)

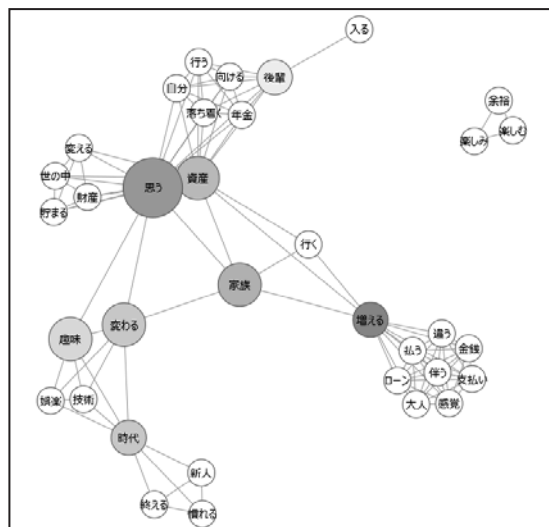


図 8 共起ネットワーク (Case 24, 抽出語数 : 36 語)

図 4 の Case 10 と図 7 の Case 27 を見ると分かるように、文字数が 600 字台で抽出語数が 20 語弱でも、趣味や仕事に対する学生の思いを推し量ることはできそうである。しかし、図 3 の Case 1 や Case 5 のように、文字数が 800 字以上で抽出語数が 20 語を超えている場合と比較すると、読み取れる情報は格段に少ない。また、図 5 の Case 13 や Case 18、そして図 6 の Case 20 や Case 21 のように、文字数が 1000 字以上で抽出語数が 40 語に近い場合には、複数の抽出語が複雑に絡み合う。このような場合にも、図を見る人によって解釈がばらつくと考えられる。したがって、今回の解析結果をもとに文章量の制限を設けるとすれば、800 字から 1000 字程度に設定するのが妥当であろう。

以下では、800 字から 1000 字までの文字数範囲に含まれる Case 1 (図 3) を例にとって、原文の内容をどの程度表せているかを検討する。その際に、対応分析と Word Cloud についても合わせて言及したい。なお対応分析では、図 1 の基本条件の他に、図 9 のようにオプションを設定した。また、Word Cloud では、2 回以上出現した語を、その頻度に応じた大きさで表示した。

Case 1 の原文を図 10 に示す。図中の枠線で囲った部分は、この文章のキーセンテンスを表している。また図 11 と図 12 は、この原文に対する対応分析の結果と Word Cloud である。

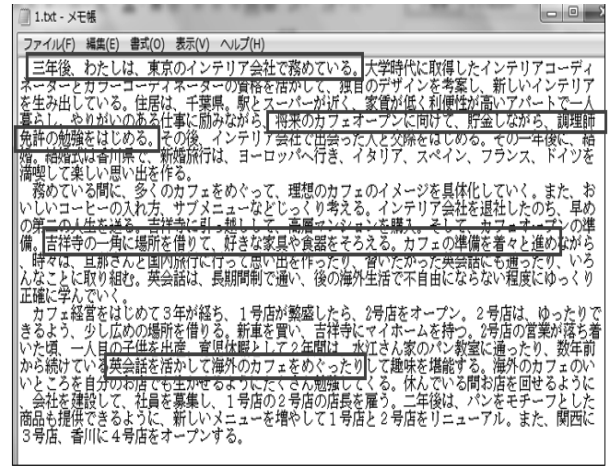


図 10 Case 1 の原文

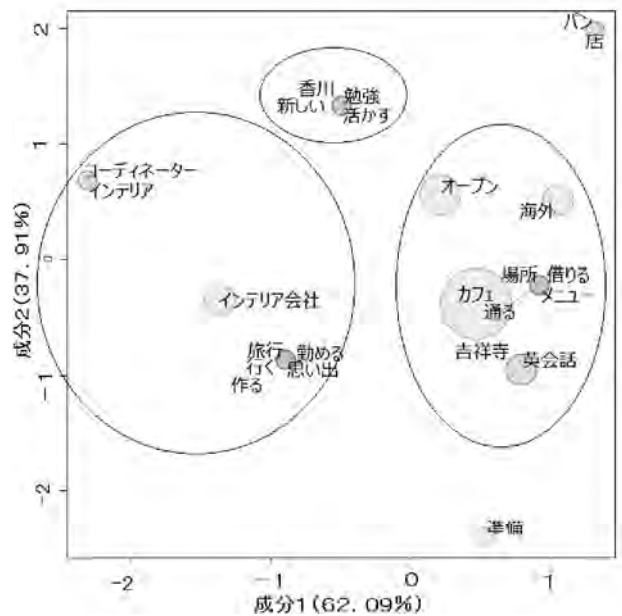


図 11 Case 1 の対応分析の結果

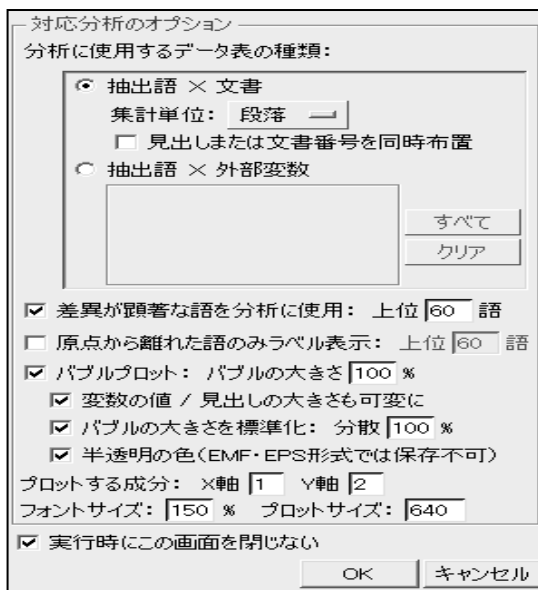


図 9 対応分析のオプション

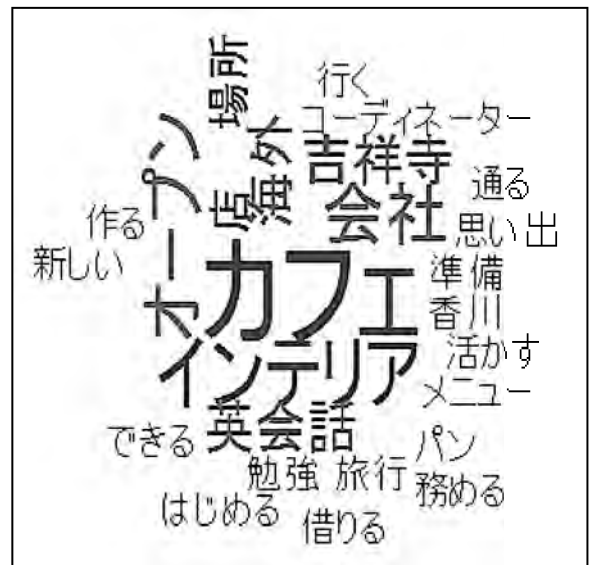


図 12 Case 1 の Word Cloud



図 3 に示した Case 1 の共起ネットワークを見ると、「カフェ」を囲む円が最も大きく、出現頻度が最も高いことが分かる。また、この語を中心に他の語群が関連付けられていることも読み取れる。その主たる語群として、「オープンー準備」、「インテリア会社ー勤める」、「コーディネーター活かす」、「英会話ー海外ー勉強ー店」、「旅行ー行くー思い出ー作る」を挙げることができる。さらに、「吉祥寺ー場所ー借りる」という語群が衛星のように配置されている。以上の情報があれば、図 10 のキーセンテンスは容易に推測できるとともに、頭の中で一つのストーリーを構築できそうである。

図 11 の対応分析の結果を見ると、円で囲った三つの語群が目につく。この情報からも、インテリア会社に勤めてコーディネータの能力を身につけ、それを活かしてカフェをオープンさせたいという要旨を推測できるであろう。ただ、共起ネットワークと異なり、対応分析では関係が深い語群を線で結んでいないため、見る人による推測結果のばらつきが大きくなるかもしれない。図 12 に示した Word Cloud についても同じことが言える。すなわち、出現頻度に応じた大きさで単語がランダムに表示されているため、ストーリーを組み立てるのは至難の業である。ただ、出現頻度が特に高い「カフェ」、「オープン」、「インテリア」という語が目飛び込むので、「インテリアを工夫したカフェをオープンしたい」ことは直感的に分かるであろう。以上のことを勘案すると、対応分析の結果と Word Cloud は各々単独で使用するのではなく、共起ネットワークを補足する視覚資料として活用すべきであろう。

最後に、共起関係の強弱を表す Jaccard 係数について言及する。今回使用した KH Coder では、Jaccard 係数の設定値よりも小さい共起関係は線で結ばない。また、線で結ばれていない単独の抽出語は表示されない。すなわち、共起ネットワークの表示内容が Jaccard 係数の設定値によって変化する。表示の変化を確認するために、Case 1 を対象にして、Jaccard 係数を 0.2、0.3、0.4 と変えて解析を行なったところ、図 13 に示す共起ネットワークが得られた。なお、Jaccard 係数の取り得る範囲は、0 から 1 までである。

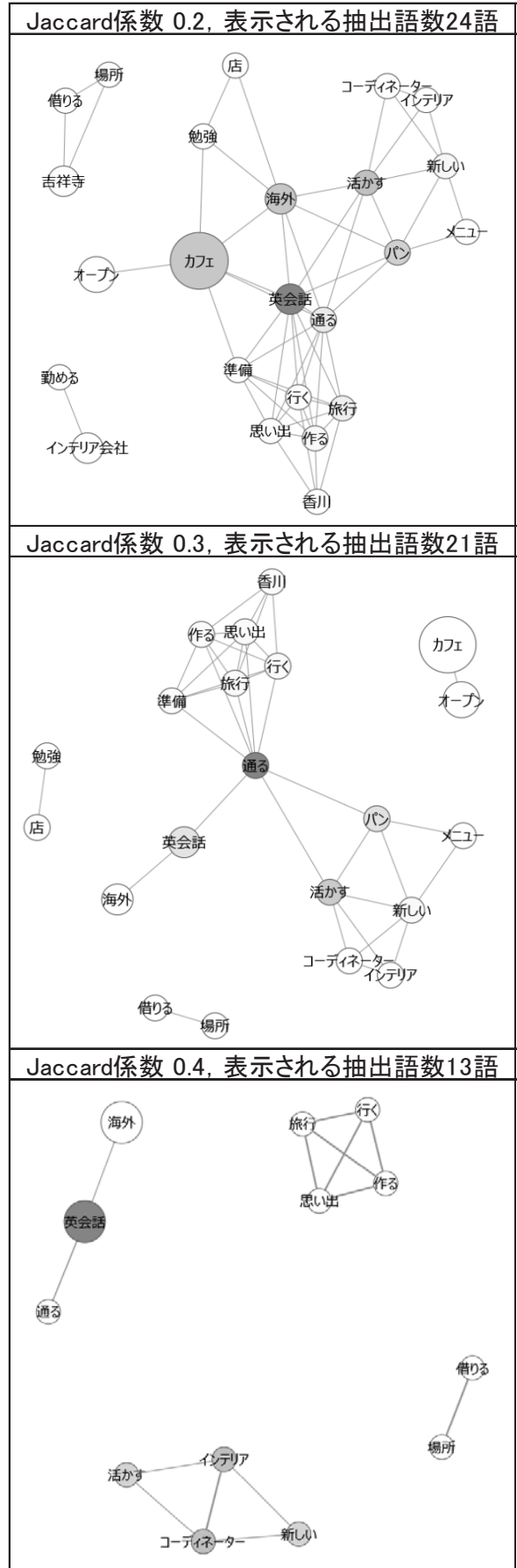


図 13 共起ネットワーク (Case 1)

まず、図 3 の Case 1 と図 13 の最上段、すなわち Jaccard 係数の設定値が 0.1 の場合と 0.2 の場合とを比較する。両者ともに表示される抽出語数は 24 語である。また、0.1 から 0.2 へと共起判定の閾値を高めることにより、「勤めるーカフェ」や「オープンー勉強」など数か所の共起関係の線が消える。Jaccard 係数の設定値をさらに大きくすると、共起関係の線が消えて単独となった抽出語は削除される。図 13 の中段に示したように、Jaccard 係数の設定値が 0.3 の場合には、「インテリア会社」、「勤める」、「吉祥寺」という三つの抽出語が表示から消える。さらに、「カフェ」と他の語群とを結ぶ重要な線も切れる。図 13 の下段に示したように、Jaccard 係数の設定値が 0.4 の場合には、表示される抽出語が 13 語まで減り、視覚資料としての意味が極端に薄れる。以上の結果から、Jaccard 係数の設定値として 0.1 を採用する。

#### 4. 学生指導での活用法

本学では、2013 年度よりポートフォリオ学習支援統合システム（図 14）を導入している。新入生には、「入学時の状況」や「将来像」などを入力してもらっている。この入力データを対象に前述のテキストマイニングを実施し、1 年次のデータとして担任やチュータに活用してもらおうことができると考える。



図 14 ポートフォリオ学習支援統合システム

また、2 年次の必修科目の授業時間外学習として、「マインドマップを活用したキャリアプランの作成」を課してもよいであろう。キャリアプランは、学生の成長によって変わると考えられるので、3 年次や 4 年次でもデータを適宜に採取して、それを対象にテキストマイニングを行なうこともできる。これらの解析結果を学内システムの面接記録（図 15）に添付ファイルとして登録しておけば、担任やチュータが学生指導やキャリアサポートで活用できる。



図 15 学内システムの面接記録

#### 5. おわりに

学生指導で活用することを目的に、学生が作成した文書をテキストマイニング技術で解析した。その結果、原文の主旨を損なわない範囲でイメージ化するための解析条件や文章量を把握することができた。

今回の解析をしていて、Cyndi Lauper（歌手）の「True Colors」という曲を思い出した。その歌詞は、勇気を出して「true color（本当の自分）」を表現しようを訴えている。しかし、「true color」を出すためには、外部からのエネルギーが必要であろう。私たち教師は、学生が独自の色を表現できるように応援していかなくてはならない。そのためには、学生の気持ちや経験談を聞く機会を積極的に設ける必要がある。その際に、学生が作成し

た文書の解析結果は視覚資料として役立ってくれると確信している。

今後は、このような視覚資料を蓄積するとともに、学生指導で活用したい。

#### 参考文献

- 1) 生田和重： Moodle を活用した授業実践と学習活動の分析，大学教育研究ジャーナル，第 11 号，pp.125-135，2014
- 2) Rice, W., 2011, *Moodle 2.0 E-Learning Course Development*, Birmingham, Packt Publishing
- 3) 中央教育審議会，新たな未来を築くための大学教育の質的転換に向けて， [http://www.mext.go.jp/b\\_menu/shingi/chukyo/chukyo0/toushin/1325047.htm](http://www.mext.go.jp/b_menu/shingi/chukyo/chukyo0/toushin/1325047.htm)，2012.8.28（参照日：2013.2.14）
- 4) 片岡俊行：マインドマップ練習帳．秀和システム，東京，2007
- 5) 樋口耕一： 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，京都，2014
- 6) 松本裕治ほか： 形態素解析システム『茶釜』 version 2.3.1 使用説明書．NAIST Technical Report, 奈良先端科学技術大学院大学松本研究室, 2003
- 7) 金明哲： R によるデータサイエンス，pp.87-91，森北出版，東京，2009
- 8) Danowski, J. A.: Network analysis of message content, *Progress in communication sciences*, vol. 12, pp. 198-221, 1993
- 9) R Core Team, R: A language and environment for statistical computing, R Foundation for Statistical Computing, Vienna, Austria, <http://cran.r-project.org/doc/manuals/r-release/fullrefman.pdf>, 2014
- 10) Konchady, M., 2006, *Text Mining Application Programming*, pp.269-271, Boston, Charles River Media
- 11) Jockers, M. L., 2014, *Text Analysis with R for Students of Literature*, pp.150-151, Switzerland, Springer International Publishing