

研究ノート

計量テキスト分析ツール KH Coder による分析結果の再現性 —— KH Coder による再現と R・SPSS による再分析や検算

樋口 耕一ⁱ

信頼できるデータ分析ツールであるためには、そのツールで分析した結果を何度でも再現できなくてはならない。また、そのツールによる計算結果を、別の分析ソフトウェアによって検算できるとなお良いだろう。こうした観点から、本稿では筆者が開発している計量テキスト分析ツール KH Coder について、分析結果を再現するために記録しておくべきオプション項目をまず確認する。次に分析結果を容易に再現できる「R Source」形式ファイルを保存・活用する方法を紹介する。その上で R や SPSS といった統計ソフトウェアを用いて、再分析や検算を行う方法を示す。KH Coder は分析の過程で、テキストデータを「文書×抽出語」表に変換して、この表に対して統計的な分析を行なう。この「文書×抽出語」表を取り出して、R や SPSS といった統計ソフトウェアに入力すれば、自在に再分析・検算を行える。KH Coder はただ信頼して使うしかないというツールではなく、必用に応じて、ユーザーが分析の過程を詳しく追跡・確認するための仕組みを備えている。

キーワード：計量テキスト分析，KH Coder，再現性，検算，R，SPSS

1 目的

1.1 KH Coder とは

「計量テキスト分析」とは、アンケート自由記述やインタビュー記録のようなテキスト型データを統計的に分析するための方法として、筆者らが提案してきたものである（川端・樋口 2003; 川端 2021; 樋口 2004, 2020）。この計量テキスト分析を実現するためのソフトウェア・ツールとして、筆者は KH Coder を開発し、2001年に無償公開を始めた。この KH Coder をお使いになる際にはぜひ押さえておいていただきたいポイントを樋口・中村・周（2022）にまとめている。また、開発の背景にある考え方や応用

事例、それぞれの機能については樋口（2020）が詳しい。KH Coder は思いのほか多くの皆さまにご利用いただき、これを用いた研究は現在、書籍・論文・学会発表をあわせて約7,000件を数えている¹⁾。

2023年12月から KH Coder は、（株）SCREEN アドバンスドシステムソリューションズ社の製品「KH Coder オフィシャルパッケージ」として販売されている。20年ほどのあいだ、開発・メンテナンス・サポートなどのすべてを筆者が1人で担うことで、KH Coder の無償公開を実現してきた。しかし、KH Coder の多機能化が進むとともにメンテナンスにはより細かな注意が必要になり、普及が進むとともにお問い合わせをいただく回数も増えた。正直なところ1人ですべてを行なう体制は限界で、継続できなかった。こうした状況から、産学連携によって KH Coder を企業様の製品とすることで、開発・メンテ

i 立命館大学産業社会学部教授

ナンス・サポートなどに企業様のご協力を得る運びとなった。この製品化とともに無償公開は終了したが、アカデミック割引や格安の「ゼミ&研究室パック」のほか、無料の授業用ライセンスが準備されている。

こうした製品化の判断には、筆者に健康不安が生じ、従来の体制ではKH Coderの持続可能性について確信を持てなくなったことも関係している。個人が開発・公開しているソフトウェアは、様々な事情からその個人が開発を続けられなくなった場合、以降のメンテナンスが行なわれない。不具合や脆弱性が見つかって修正されないし、Windowsのアップデートによって動作に問題が生じて改善されない。そうしたソフトウェアを使い続けることは難しいだろう。それに対して企業の製品となれば、一定程度の売上げを計上しているかぎり、当該企業によるメンテナンスを期待できる。製品版を購入することで支えてくださる方々がいらっしゃるあいだは、KH Coderも継続を許されるだろう²⁾。

1.2 分析結果の再現と検算

応用研究でのご利用に加えて、近年ではKH Coderの利用法をご検討・ご紹介いただくケースも散見される。そのなかには書籍で取り上げていただいたものもあれば(牛澤・和泉 2024; 牛澤 2021; 有馬 2021; 内田・大賀・中藤編 2021; 左古編 2021)、論文の形で公表されたものもある(たとえば最近では大多和ら 2023; 中西 2022; 吉澤・奥・堀 2021; 西村・清水 2021; 林2023など)。そうした中には、KH Coderによる計算の結果について、自分自身でも別途に計算して合致するかどうかを確かめようとする試みが見られる(林 2023)。すなわち検算してみようという試みである³⁾。

確かに、KH Coderを信頼できる分析ツールとして利用するためには、次のような意味で、KH Coderによる分析結果が再現可能でなければならない。第一に、分析者自身が同じ分析結果を何度でもKH Coderを用いて再現できることが必要である。第二

に、KH Coderによる分析結果をほかのソフトウェアによって再現できれば、KH Coderによる計算が正しいことを確認できる。すなわち検算できる。よって本稿では、これら2つの形でKH Coderによる分析結果を再現する方法について述べる。

第一の点、すなわちKH Coderによる分析結果の再現は当然可能なことのように思われるかもしれない。しかしマウス操作によって、条件を少しずつ変えながら分析を繰り返し行なっていると、設定した条件を混同してしまうこともある。そうなると、分析結果を自分で再現することが難しくなってしまうので、その対策を記載しておくことには意味があるだろう。また第二の点、検算を行なえることは、分析ツールの信頼性を確認するために重要である。それに加えて検算が行なえるということは、分析を自分自身で再現できるということであり、再現できるならばその過程でカスタマイズを加えることも当然可能であろう。すなわち検算だけでなく、自分の研究データや目的にあわせて、分析法をカスタマイズできるようになる点で、その手順には意義があるだろう。

2 分析結果に影響する 2つのオプション群の記録

2.1 統計分析用データ「文書×抽出語」表の作成

KH Coderの分析画面には多くのオプション(選択項目)があるが、それらは「文書×抽出語」表を作成するためのものと、統計分析・可視化手法の詳細を決めるものとの2種類に大別される。分析結果を後から再現できるようにするには、分析結果とあわせて、この2種類のオプションについて、分析時にどのような選択を行なったかを記録しておく必要がある。以下ではまず「文書×抽出語」表を作成するためのオプション群について述べる。

KH Coderは分析の過程で、テキストデータをもとに、統計分析が可能な形式のデータ「文書×抽出語」表を作成する。「文書×抽出語」表とは表1(b)

表1 統計分析用データの作成

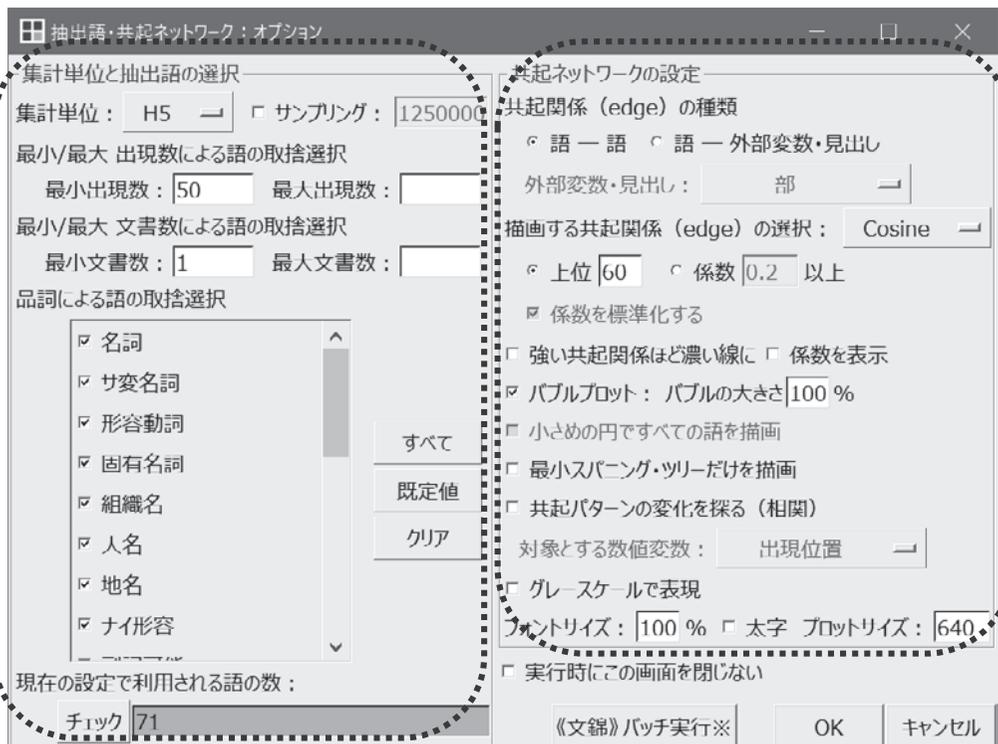
(a)もとのテキスト		(b)「文書×抽出語」表			
		データ	多い	読む	難しい
文書1	データが多い。とても多い。	1	2	0	0
文書2	多すぎて読むのは難しい。	0	1	1	1
文書3	それでもデータを読むべきだ。	1	0	1	0

のように、「データ」「多い」などの抽出語が、各文書に何回出現していたのかをまとめた表である⁴⁾。「文書×抽出語」表を統計ソフトウェアに入力すれば、様々な統計分析を行なえる。たとえば、よく一緒に使われることが多かった語、すなわち共起することが多かった語を線で結んだ共起ネットワークを作成できる。そのほか KH Coder には、クラスター分析、多次元尺度構成法、自己組織化マップ、トピックモデル (LDA)、対応分析などの分析機能を備えている。

図1は KH Coder で共起ネットワークを作成する

際のオプション画面である。図1の左半分 (a) の部分で、「文書×抽出語」表をどのように作成するかという設定を行なう。まず「集計単位」を選ぶことで、1つの「文書」と見なす単位を設定する。Excel ファイルのセル (h5) を1つの文書と見なすか、改行で区切られた段落を1つの文書と見なすか、「。」で区切られた個々の文を文書とみなすかといった選択を行なえる。その上で、「文書×抽出語」表にどの抽出語を含めるかを、出現回数や品詞といった基準で設定できる。

図1のような分析のオプション画面以外にも、「文



(a) 「文書×抽出語」表の設定

(b) 統計分析の設定

図1 共起ネットワークのオプション画面

書×抽出語」表に影響を及ぼすオプションがある。1つは「前処理」メニューの「語の取捨選択」画面での設定である。たとえば「使用しない語」に指定した語は、「文書×抽出語」表でも使用されない。また「強制抽出」を行なうと、語の抽出結果が変化するので、「文書×抽出語」表も変化する。

もう1つは、「プロジェクト」メニューの「設定」画面にある、「基本形が同じ語は、品詞名が異なっても同じ語と見なす」オプションである。このオプションはデフォルトで無効になっているので、たとえば形容動詞の「不思議」と名詞の「不思議」が別の語として認識されているかもしれない。このオプションを有効にして前処理を再度行なうと、「不思議」という語は1種類に統合される。

最後の1つは、文章中から語を取り出すための形態素解析ツールとしてChaSenを使うのかMeCabを使うのかという選択である。この選択は新規プロジェクトの作成時に行なえる。プロジェクト作成後は「プロジェクト」メニューで「開く」を選択し、開いた画面で当該プロジェクトを選択して「編集」ボタンをクリックすると、設定を確認または変更できる。

これらのオプションを記録しておけば、KH Coderで分析を再現するための第一歩として、統計分析に用いた「文書×抽出語」表を再現できる。KH Coder上での操作に関しては、上記のオプション設定を再現すれば、同一の「文書×抽出語」表が再現されるはずである。ただし、もしも自分でChaSenやMeCabの辞書に変更を加えている場合は、それらの辞書を保存しておく必要がある。

2.2 統計分析・可視化手法の詳細

同一の「文書×抽出語」表が再現されれば、あとは同じ方法で統計分析・可視化を行なうことで、同一の結果が再現される。統計分析・可視化手法の詳細を決めるためのオプションは、図1の右半分(b)の部分に配置されている。それに加えて、分析結果の画面で「調整」ボタンをクリックすると開く「調

整」画面でも、統計分析・可視化手法の詳細を設定できる。

統計分析・可視化手法の詳細については、分析方法にもよるのだが、図1に示す画面の右半分(b)でのみ設定できるオプションもあれば、「調整」画面でのみ設定できるオプションもある。したがって、「調整」を行なった場合は、図1での設定内容に加えて、「調整」画面での設定内容も記録しておく必要がある。これら2つの画面のスクリーンショットを保存しておくのが簡便であろう。変更箇所が少ないようなら、デフォルトから変更した箇所のみを記録しておくという方法もある。ただ、この方法ではバージョンアップによって万一デフォルトが変更された場合には再現が難しくなるだろう。

なお「調整」画面で変更できるのは統計分析・可視化手法のオプションのみである。分析の入力データとなる「文書×抽出語」表を作成するオプションは変更できない。したがって「文書×抽出語」表から修正したい場合には、分析結果を出す前の、図1の画面に戻る必要がある。

3. Rを用いた分析結果の再現および再分析

上記2つのオプション群を記録しておくのが確実であるが、補助的な方法として分析結果を「R Source」形式で保存しておくことも役立つだろう。「R Source」形式のファイルには、「文書×抽出語」表のデータと、その表をもとに統計分析を行なうRコマンドとが保存されている。(1)このファイルを統計ソフトウェアRで実行するとKH Coderで行なった分析の結果がそのまま再現される。また(2)「文書×抽出語」表を取り出すこともできるし、(3)分析時に選択した統計分析・可視化手法の詳細オプションをRコマンドから読み取ることができる。ここでは(1)と(2)の手順について述べる⁵⁾。

まず、分析結果を「R Source」形式で保存するには、分析結果の画面で「保存」ボタンをクリックする。そして、保存画面でファイル形式として「R

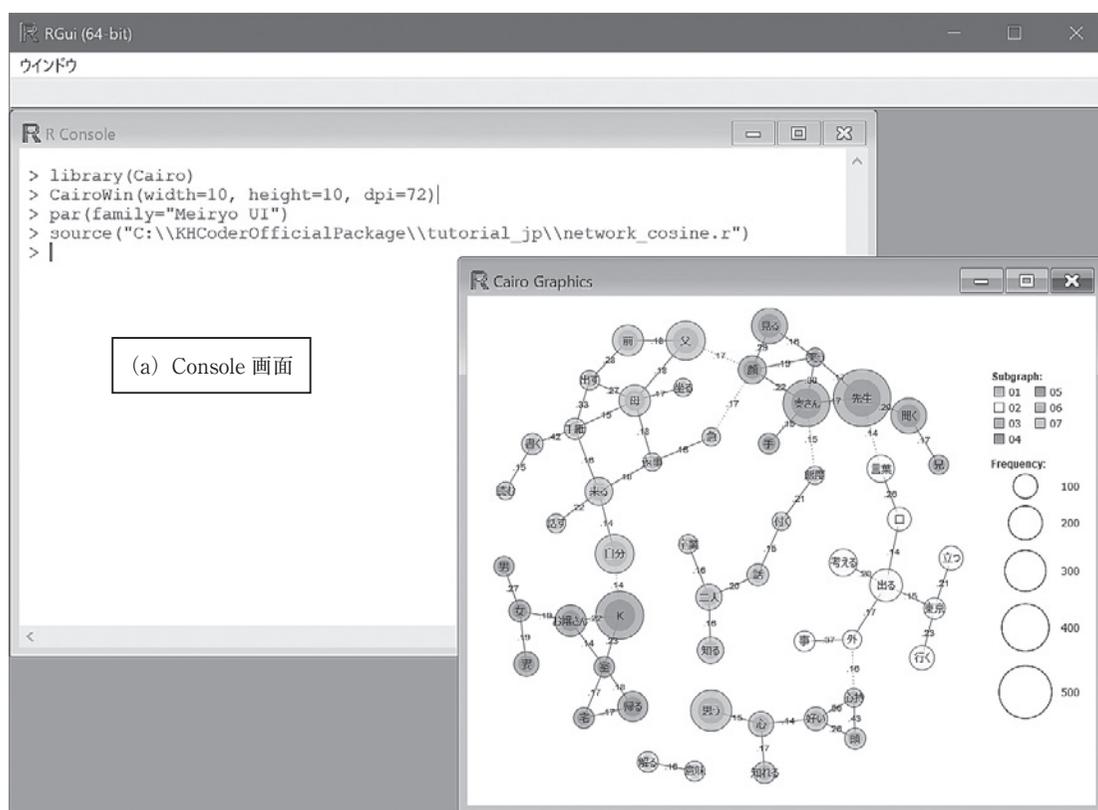


図 2 R による分析結果の再現

Source」を選択すれば、「文書×抽出語」表と分析用 R コマンドを記録した「R Source」ファイルが保存される。

次に (1) 保存した「R Source」形式のファイルを R で実行し、分析結果を再現する手順は、以下の通りである。KH Coder をインストールしたフォルダを開き、「kh_coder.exe」と同じ場所にある「Rgui64.bat」をダブルクリックすれば、KH Coder に同梱の R が起動する。あとは図 2 (a) の「Console」画面上に、保存した「R Source」形式のファイルをドラッグ&ドロップするだけで良い。これによって「R Source」形式のファイルに記載されたコマンドが R によって実行され、KH Coder 上に表示されたのと同じ分析結果が、R 画面に表示される (図 2)。

(2) 「文書×抽出語」表を取り出すには、R コマンドの編集・追記が必要である。「R Source」形式の

ファイルは、R コマンドが記載されたテキストファイルである。よって「メモ帳」のようなテキストエディタで開けば編集できる。「R Source」形式のファイルを「メモ帳」で開き、「編集」メニューの「検索」から、「DATA」という文字列を検索すると、「#END: DATA」という行が見つかるだろう。この行の直後に、「文書×抽出語」表を取り出すためのコマンドを書き加えれば良い。

```
write.table(
  t(d),
  "C:/KHCoderOfficialPackage/matrix.txt",
  sep = "¥t",
  row.names = FALSE,
  fileEncoding = "UTF-8"
)
```

以上のコマンドを書き加えてから「R Source」形

式のファイルを上書き保存し、図2の(a)に示すRの「Console」画面上にドラッグ&ドロップすればよい。そうすると、コマンドが実行されて、KH Coderをインストールしたフォルダ、すなわち kh_coder.exeと同じ場所に「matrix.txt」というファイル名でタブ区切り形式の「文書×抽出語」表が保存される。このファイルをダブルクリックすると「メモ帳」などのテキストエディタ上に「文書×抽出語」表が表示される。すべて選択してからコピーし、Excelに貼り付けると確認しやすいだろう⁶⁾。

以上に、Rで分析結果を再現する方法や、Rコマンドを書き加えることで「文書×抽出語」表を取り出す方法を記述した。「文書×抽出語」表には、当然、分析用に選択された語が列挙されている。これを見れば、図1における(a)の部分で、分析時にどのような選択を行なったかを確認するための有力な手がかりとなるだろう。また、統計分析・可視化手法の詳細オプションについてもRコマンドの形式で記録されている。ここから、論文に掲載するような分析結果については「R Source」形式でも保存しておくと、年数が経ってからも確認がしやすくなって便利だろう。なお、ここではデータをファイルに書き出すためのRコマンドを示したが、分析部分のRコマンドを編集すれば、分析や可視化の方法をカスタマイズしての再分析も自在に行なえる。

4. SPSS を用いた検算

KH Coderによる計算の結果について、ほかのソフトウェアでも計算して検算したい場合はどうすればよいだろうか。たとえば図3は、KH Coderを用いて、Cosine 係数によって語と語の共起の強さを測定し、共起ネットワークを作成したものの一部である。図1の画面で「係数を表示」オプションを有効にすることで、Cosine 係数をネットワーク上に表示している。図3に表示されているCosine 係数を、SPSSを用いて検算する手順を以下に示す。

手順全体としては、(1) KH Coder から「文書×

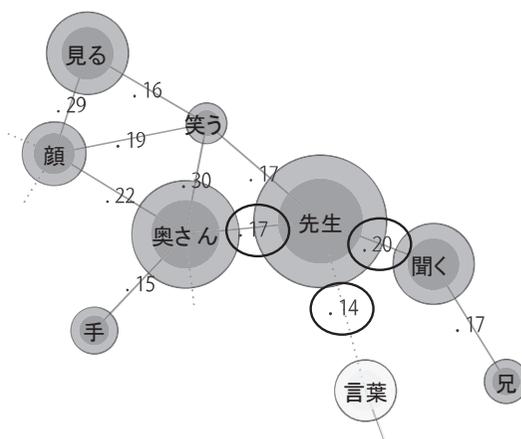


図3 KH Coder の計算した Cosine 係数

抽出語」表を取り出して、(2) その表を SPSS で開き、(3) SPSS 上で Cosine 係数を計算するという流れになる。

(1) KH Coder から「文書×抽出語」表を取り出すには、KH Coder のメニューから「ツール」「エクスポート」「文書×抽出語」表」「SPSS ファイル」を順に選択する。そして、KH Coder で分析を行なったときと同じオプションを選択して「OK」をクリックすれば、「文書×抽出語」表が SPSS シンタックス (*.sps ファイル) として保存される。

(2) 保存した SPSS シンタックス・ファイルをダブルクリックすると、SPSS が開くだろう。SPSS のメニューから「実行」「すべて」を選択すると、「文書×抽出語」表が SPSS 上に読み込まれる。語の出現数以外の追加的な変数も読み込まれるが、それらの詳細については樋口 (2020) の A.3.5 節を参照されたい。

(3) 以下の SPSS シンタックスを実行すれば、SPSS 上で Cosine 係数を計算できる。図4に示す計算結果を見ると、図3と数値が一致していることを確認できるだろう。

```
FORMATS w0 to w70 (f8.2).
DO REPEAT x = w0 to w70.
  COMPUTE x = x / length_w * 1000.
END REPEAT.
```

EXECUTE.

PROXIMITIES w0 to w70

/VIEW=VARIABLE

/MEASURE=COSINE

/STANDARDIZE= NONE.

なお今回用いたデータでは、「文書×抽出語」表の作成時に71種類の語が使用されている。71語なのでシNTAXスでは「w0」から「w70」までという指定になっている。出力された語の数に合わせて「w70」となっている部分を修正する必要がある。たとえば80語が出力された場合は、「w70」となっている3箇所を「w79」に修正する必要がある。

おわりに

以上に、KH Coder による分析結果を再現するために記録しておく必要があるオプションについて述べた上で、記録の補助として「R Source」形式ファイルを保存しておくこと役立つことを示した。また R や SPSS で再分析・検算を行なう手順についても触れた。

これらの方法を使えば、KH Coder を用いて分析結果を再現できるだけでなく、再分析の過程で、統計手法のカスタマイズを行なうこともできる。また

ピボットテーブル 近接行列

ファイル(E) 編集(E) 表示(V) 挿入(I) ピボット(P) 形式(O) ヘルプ(H)

SansSerif 12 A↑ A↓ B I U A. | | | | | | | | | |

近接行列

値のベクトルのコサイン

	w0 先生	w1 奥さん	w2 自分	w3 お嬢さん	w4 言葉	w5 手紙	w6 叔父	w7 人間	w8 様子	w9 心持
w0 先生	1.000	.169	.031	.000	.142	.072	.007	.034	.056	.029
w1 奥さん	.169	1.000	.086	.135	.075	.001	.006	.007	.073	.016
w2 自分	.031	.086	1.000	.078	.070	.031	.046	.035	.035	.026
w3 お嬢さん	.000	.135	.078	1.000	.012	.003	.010	.021	.041	.030
w4 言葉	.142	.075	.070	.012	1.000	.005	.023	.028	.041	.007
w5 手紙	.072	.001	.031	.003	.005	1.000	.010	.004	.009	.001
w6 叔父	.007	.006	.046	.010	.023	.010	1.000	.005	.090	.004
w7 人間	.034	.007	.035	.021	.028	.004	.005	1.000	.004	.007
w8 様子	.056	.073	.035	.041	.041	.009	.090	.004	1.000	.003
w9 心持	.029	.016	.026	.030	.007	.001	.004	.007	.003	1.000
w10 態度	.066	.150	.088	.076	.081	.010	.056	.022	.056	.009
w11 話	.060	.042	.047	.068	.024	.037	.034	.017	.022	.007
w12 意味	.037	.050	.043	.019	.055	.012	.006	.078	.062	.002
w13 病氣	.058	.021	.046	.002	.068	.060	.001	.002	.009	.007
w14 卒業	.043	.016	.045	.014	.006	.002	.004	.029	.004	.000
w15 返事	.111	.011	.047	.025	.046	.077	.000	.003	.061	.001
w16 急	.041	.122	.012	.023	.031	.003	.019	.012	.041	.003
w17 東京	.019	.005	.028	.000	.009	.065	.079	.021	.129	.009
w18 問題	.078	.051	.021	.031	.004	.002	.020	.002	.004	.002
w19 前	.109	.092	.117	.078	.060	.048	.019	.047	.027	.013
w20 今	.026	.031	.058	.060	.025	.019	.027	.045	.029	.002
w21 K	.001	.081	.143	.219	.089	.024	.002	.026	.059	.021
w22 二人	.074	.048	.031	.084	.008	.004	.005	.005	.021	.021
w23 一人	.076	.009	.036	.032	.015	.001	.057	.066	.022	.011
w24 思う	.099	.054	.112	.085	.025	.061	.034	.061	.028	.016
w25 見る	.108	.119	.066	.104	.025	.014	.033	.011	.078	.008
w26 聞く	.201	.092	.025	.056	.087	.001	.032	.020	.031	.005

図4 SPSSによるCosine係数の計算結果

R や SPSS といった統計ソフトウェアを使った検算によって、KH Coder の計算結果を確かめられる。林 (2023) のように実際に検算を試みる例というのは、それほど多くないかもしれない。しかし、もし必用になった場合には検算が可能であることと、検算の詳しい方法を開示しておくことには、KH Coder の信頼性を保持する上で一定の意義があると考えられる。以上のように、KH Coder はただ信頼して使うしかないというツールではなく、必用に応じて、ユーザーが分析の過程を詳しく追跡・確認するための仕組みを備えたツールである。

注

- 1) KH Coder を用いた研究で、筆者が把握しているものを <https://khcoder.net/bib.html> にリストアップしている。現時点での掲載数は7,020件である。
- 2) 「KH Coder オフィシャルパッケージ」製品のうち、Q & A サービスの付属しない Base エディションのアカデミック価格は現在 ¥24,750- である。KH Coder の製品化をお願いした (株) SCREEN アドバンストシステムソリューションズ社には、筆者の勤務校である立命館大学との間に技術指導契約を結んでいただき、コミュニケーションをとっている。また以前より、KH Coder に機能を追加する有料プラグイン「文錦」シリーズが同社より販売されている。「分からない」のように否定されている語を、「分かる (否定)」のように区別して抽出できる「文錦 否定表現チェッカー for KH Coder」。また、「子供」「子ども」のように似通った語を統一する「文錦 表記ゆれ&同義語エディタ for KH Coder」など、7種類の機能追加プラグインが販売中である。
- 3) この試みにおいて、KH Coder が算出した Jaccard 係数と Simpson 係数については検算が成功しているが、コサイン係数については成功していない。
- 4) 「文書×抽出語」表の作成プロセスを追跡・確認するには、入力した文章がどのように語に分解されて、どの語が分析対象となっているかを調べればよいだろう。このためには KH Coder のメニューで「前処理」「語の抽出結果を確認」を選択し、確認したいフレーズまたは文を入力する。「検索」ボタンをクリックして、表示された検索結果をダブルクリックすれば、どこで語が区切られているかに加えて各語の品詞名も確認できる。品詞名が「その他」となっている語は、デフォルトの設定では「文書×抽出語」表から除外されている。すなわち、分析から除外されている。どのような文章の中にもあらわれる一般的で分析に使いにくい語が「その他」に分類されている。さらに図 1 (a) の「品詞による語の取捨選択」箇所でも、デフォルトで選択されておらず、分析から除外される品詞として「名詞 B」「動詞 B」などがある。「その他」「名詞 B」のような KH Coder の品詞名について詳しくは樋口 (2020) の A.2.2 節を参照されたい。
- 5) R コマンドから選択したオプション内容を読み取るには、R のコマンドをある程度まで理解する必要がある。R の基礎については舟尾 (2016) が詳しい。さらに共起ネットワークや対応分析といった分析方法ごとに、KH Coder が利用しているライブラリのマニュアルを参照するとよいだろう。R コマンドを自分で読解せずに、選択したオプションを調べるには、(株) SCREEN アドバンストシステムソリューションズが提供しているコンサルティングサービスを利用できるだろう。
- 6) もし中国語データを扱っている場合は、ここで示した R コマンドの冒頭に「Sys.setlocale (category="LC_ALL", locale="Chinese")」という行を追加する必要があるだろう。韓国語データの場合は Chinese の部分を Korean に変更すればよいし、ほかの言語の場合も同様である。

文献

- 有馬明恵, 2021, 『内容分析の方法 第2版』ナカニシヤ出版。
- 舟尾暢男, 2016, 『The R Tips 第3版—データ解析環境 R の基本技・グラフィックス活用集』オーム社。
- 林直樹, 2023, 「テキストマイニングで作る人物関連図—河上肇『貧乏物語』を例に」『尾道市立大学経済情報論集』23(1): 83-100。
- 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): 101-115.
- 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 第2版』ナカニシヤ出版.
- 樋口耕一・中村康則・周景龍, 2022, 『動かして学ぶ! はじめてのテキストマイニング—フリー・ソフトウェアを用いた自由記述の計量テキスト分析』ナカニシヤ出版.
- 川端亮・樋口耕一, 2003, 「インターネットに対する人々の意識—自由回答の分析から」『大阪大学大学院人間科学研究科紀要』29: 163-181.
- 川端亮, 2021, 「ライフヒストリーにおける言葉のあいまいさの活用」内田論・大賀哲・中藤哲也編『知を再構築する—異分野融合研究のためのテキストマイニング』ひつじ書房, 217-232.
- 中西啓喜, 2022, 「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」『桃山学院大学社会学論集』55(2): 175-202.
- 西村奏咲・清水忠, 2021, 「テキストマイニングを用いたアンケート解析」『薬学教育』5: 2020-009.
- 大多和直樹・小川清華・神谷潤・権野禎・高橋沙綾, 2023, 「KH Coder をどう効果的に利用するか—実習からみえてきた初学者が共有しておくべきこと」『人間発達研究』お茶の水女子大学, 37: 1-22.
- 左古輝人編, 2021, 『テキスト計量の最前線—データ時代の社会知を拓く』ひつじ書房.
- 内田論・大賀哲・中藤哲也編, 2021, 『知を再構築する—異分野融合研究のためのテキストマイニング』ひつじ書房.
- 牛澤賢二, 2021, 『やってみようテキストマイニング—自由回答アンケートの分析に挑戦! 増訂版』朝倉書店.
- 牛澤賢二・和泉茂一, 2024, 『やってみようアンケートデータ分析—選択式回答のテキストマイニング流分析』朝倉書店.
- 吉澤弥生・奥彩子・堀新, 2021, 「デジタル人文学の研究と教育に関する基礎的研究」『共立女子大学・共立女子短期大学総合文化研究所紀要』27: 73-86.

Research Note

How to Reproduce the Analysis Results of the Quantitative Text Analysis Tool KH Coder: Reproduction and Verification Using KH Coder, R, and SPSS

HIGUCHI Koichiⁱ

Abstract : A reliable data analysis tool must be able to reproduce the results of its analysis. It is even better if the tool's calculation results can be verified by other analysis software. From this perspective, this paper focuses on KH Coder, a quantitative text analysis tool developed by the author, and first lists the selectable items to be recorded in order to reproduce the analysis results. It also introduces how to save and utilize "R Source" format files that can easily reproduce the analysis results. KH Coder converts the text data into a document-term matrix and performs statistical analysis on this matrix during the analysis process. After extracting this matrix and inputting it into statistical software such as R or SPSS, analysis results can be freely reproduced and verified. KH Coder is not a tool that users simply have to trust. It provides valuable functionalities enabling users to track and review the analysis process in detail.

Keywords : quantitative text analysis, KH Coder, reproducibility, recalculation with R and SPSS

i Professor, College of Social Sciences, Ritsumeikan University