

Word Miner[®] 事例集
導入編

テキスト・マイニング研究会

ご注意

- ・事前の許可なく、本書のいかなる部分も無断転載、複製、複写を禁じます。
- ・本書の内容に関しては、将来予告なしに変更することがあります。
- ・本書の内容について、万全を期して作成致しましたが、万一ご不審な点や誤り、記載漏れなどお気付きのことがありましたらご連絡下さい。
- ・運用した結果の影響について上項にかかわらず、一切の責任を負いかねますのでご了承下さい。
- ・WordMiner®は日本電子計算株式会社の登録商標です。

WordMiner® 事例集 ~導入編~ 目次

はじめに

事例集の目的	1
事例集の構成	1
画面名称	2
事例集で行う分析の流れ	4

1. 準備

1-1 サンプルデータの準備	6
1-2 WordMiner の起動	8
1-3 プロジェクトの作成	10

2. 変数を生成する（分析を始めるまでの前処理）

2-1 構成要素変数の生成（分かち書きツール WinAiBASE を起動させる）	14
2-2 構成要素変数の生成（分かち書きを行う）	16
2-3 質的変数の生成	20
2-4 構成要素の編集	22

3. 分析を行い、結果を観察する [構成要素×質的変数]

3-1 多次元データ解析を用いて 構成要素と質的変数「性年齢区分」の関連性を観察する	26
— 多次元データ解析 —	
3-2 構成要素の特徴や傾向を図に表わし観察する	30
— 布置図 —	
3-3 カテゴリーごとに特徴のある構成要素の一覧を観察する	34
— 頻度による有意性テスト —	

4. 分析を行い、結果を観察する [構成要素×サンプル]

4-1 多次元データ解析を用いて構成要素とサンプルの関連性を観察する	38
— 多次元データ解析 —	
4-2 サンプルの類型化を行い、回答群の観察をする	42
— クラスター化 —	

5. 分析結果の観察（構成要素の使われ方）

5-1 特徴的な構成要素の前後の構成要素を表示し、 その構成要素がどのように用いられているか観察する	46
— コンコーダンス —	

6. 変数を編集する

- 6-1 構成要素の編集（置換）.....50
- 6-2 構成要素の編集（削除）.....56
- 6-3 構成要素の編集（実行）.....60

7. 分析を行い、結果を観察する [構成要素×質的変数]

- 7-1 多次元データ解析を用いて
 - 構成要素と質的変数「性年齢区分」の関連性を観察する.....62
 - 多次元データ解析 —
- 7-2 構成要素の特徴や傾向を図に表わし観察する.....64
 - 布置図 —
- 7-3 カテゴリーごとに特徴のある構成要素の一覧を観察する.....68
 - 頻度による有意性テスト —

8. 分析を行い、結果を観察する [構成要素×サンプル]

- 8-1 多次元データ解析を用いて構成要素とサンプルの関連性を観察する.....72
 - 多次元データ解析 —
- 8-2 サンプルの類型化を行い、回答群の観察をする.....74
 - クラスター化 —

付録

- 1 データビューア.....80
- 2 構成要素の検索.....82
- 3 頻度グラフ.....84
- 4 エクスポート.....86
- 5 分析に使用するデータ.....90
- 6 キーワードの抽出.....94

WordMiner®事例集

— 導入編 —

■ 事例集の目的

この事例集は WordMiner®の導入編としてテキスト型データの解析を体験します。添付されているサンプルデータ【好きな料理】を用いて、WordMiner®による分析の基本的な流れを把握します。

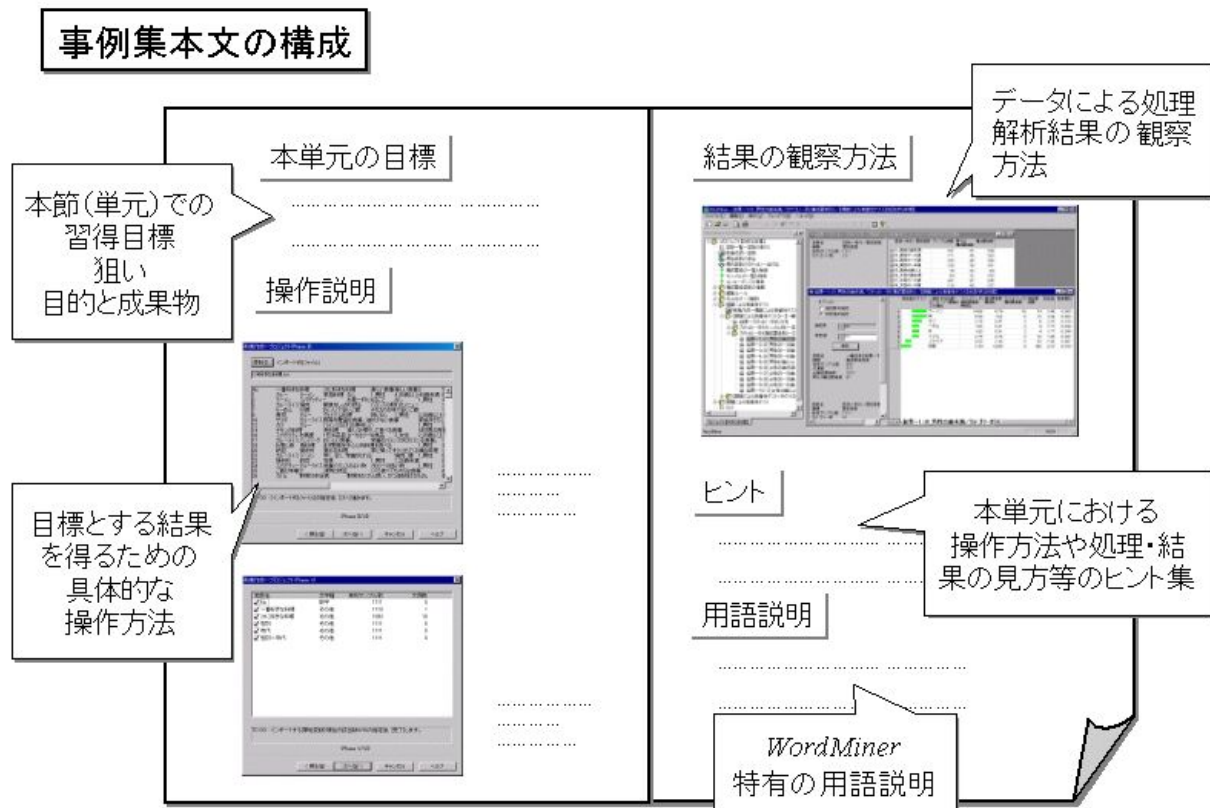
ここではテキスト型データから「キーワード」を抽出し、分析手法である「同時配置図」と「頻度による有意性検定」、「クラスター化」、「コンコーダンス」を用いて、性別・年代等の属性との関連性を分析し観察します。

解析に用いるデータの説明は[付録5 分析に用いるデータ]を参照してください。

この事例集は WordMiner® version 1.01d を基に記載されています。お使いのバージョンと画面レイアウト、処理結果等が異なる場合があります。

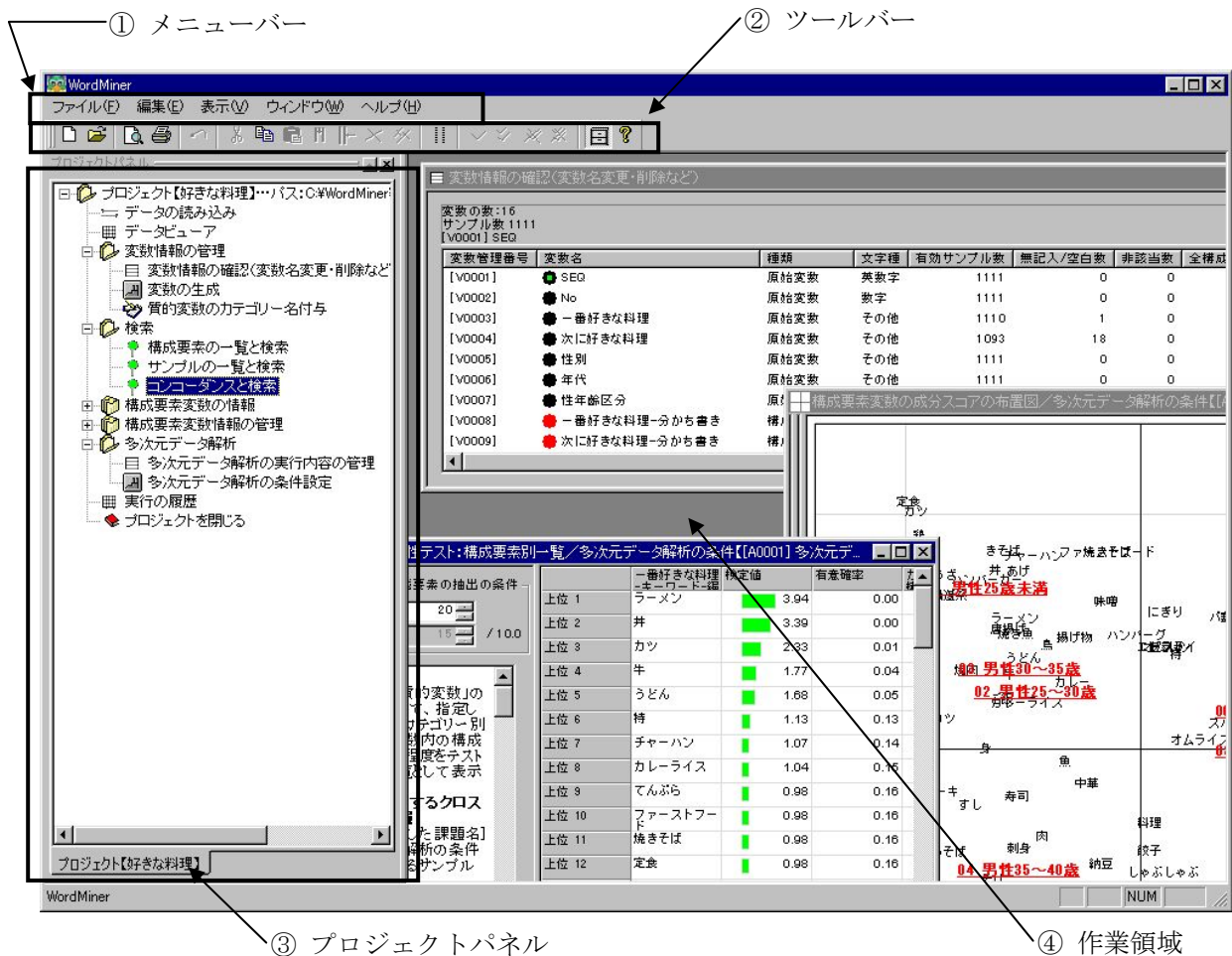
■ 事例集の構成

次の図のように、項目ごとに[本単元の目標]→[操作説明]→[結果の観察]→[ヒント]→[用語説明]の順に説明し WordMiner®の使い方や解析、出力結果の解釈や見方を学びます。



画面名称

WordMiner®を起動し、分析を行うと以下の画面が表示されます。
この事例集では画面内の各部について、以下の名称で説明します。



① メニューバー

プロジェクトの新規作成や画面の表示部分の指定ができます。

② ツールバー

プロジェクトの新規作成や印刷などを行うことができます。

③ プロジェクトパネル

起動した画面の左端にある縦長の枠で、上部に「プロジェクトパネル」と書かれてある部分のことをいいます。WordMiner®の基本機能がここに示されています。

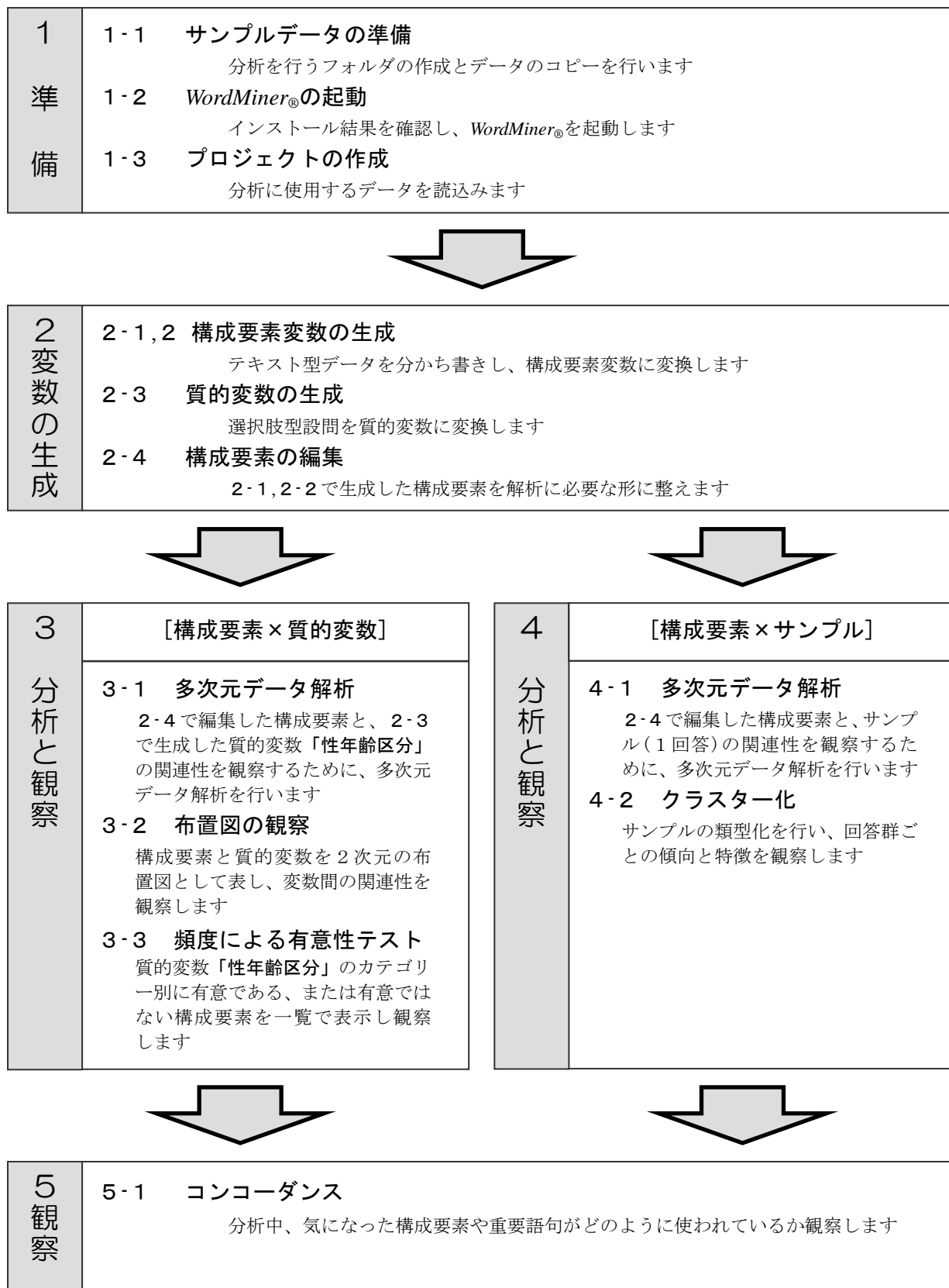
④ 作業領域

処理結果が表示される領域です。

※ プロジェクト

1つの解析の単位をまとめたものをいいます。

■ 事例集で行う分析の流れ



6 変数の編集	6-1 構成要素の編集（置換） 同義語の構成要素を置換編集する辞書を作成します
	6-2 構成要素の編集（削除） 解析に必要でないと思われる単語を削除編集する辞書を作成します
	6-3 構成要素の編集（実行） 6-1, 6-2 で作成した辞書を使用して構成要素を解析に必要な形に整えます



7 分析と観察	[構成要素×質的変数]
	7-1 多次元データ解析 [8. 変数の編集]で編集した構成要素と、2-3で生成した質的変数「性年齢区分」の関連性を観察するために、多次元データ解析を行います
	7-2 布置図の観察 構成要素と質的変数を2次元の布置図として表し、変数間の関連性を観察します
	7-3 頻度による有意性テスト 質的変数「性年齢区分」のカテゴリ一別に有意である、または有意ではない構成要素を一覧で表示し観察します
8 分析と観察	[構成要素×サンプル]
	8-1 多次元データ解析 [8. 変数の編集]で編集した構成要素と、サンプル(1回答)の関連性を観察するために、多次元データ解析を行います
	8-1 クラスター化 サンプルの類型化を行い、回答群ごとの傾向と特徴を観察します

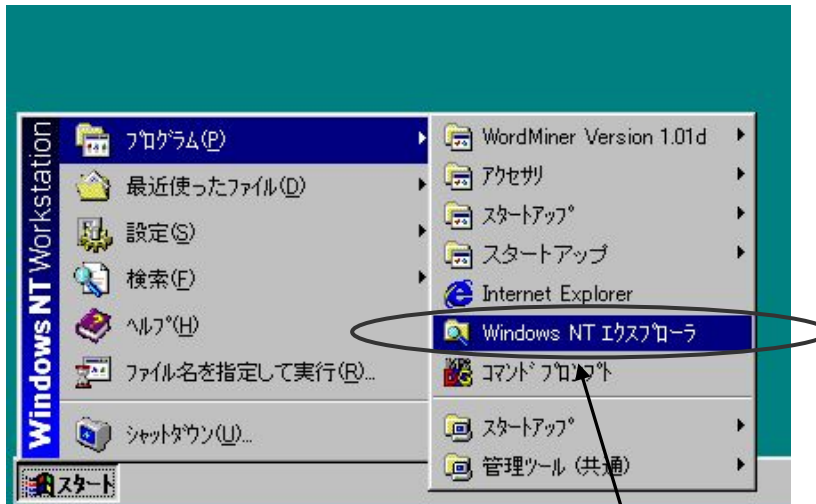
付 録	1 データビューアによる確認 作成した変数の内容を確認するときに用います
	2 構成要素の検索 作成した構成要素変数の一覧を観察するときに用います
	3 構成要素の頻度グラフの観察 構成要素の頻度分布や異なり構成要素数等を観察するときに用います
	4 エクスポート 画面上に表示された結果をファイルに出力します
	5 分析に使用するデータ この導入編で使用するデータの説明です
	6 キーワード抽出 WinAiBASE(分かち書きツール)のオプション設定の説明です

1-1 サンプルデータの準備

WordMiner®で分析を行う前に、処理結果を保存するフォルダの作成と、フロッピーディスクから分析に用いるサンプルデータ【好きな料理.txt】のコピーを行います。

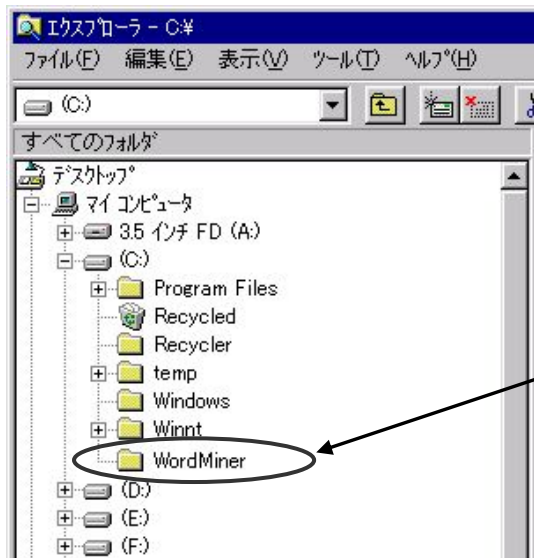
操 作 説 明

(1) エクスプローラの起動



① [スタート]－[プログラム(P)]－[エクスプローラ]をクリックすると(2)の画面が表示されます。

(2) フォルダの作成



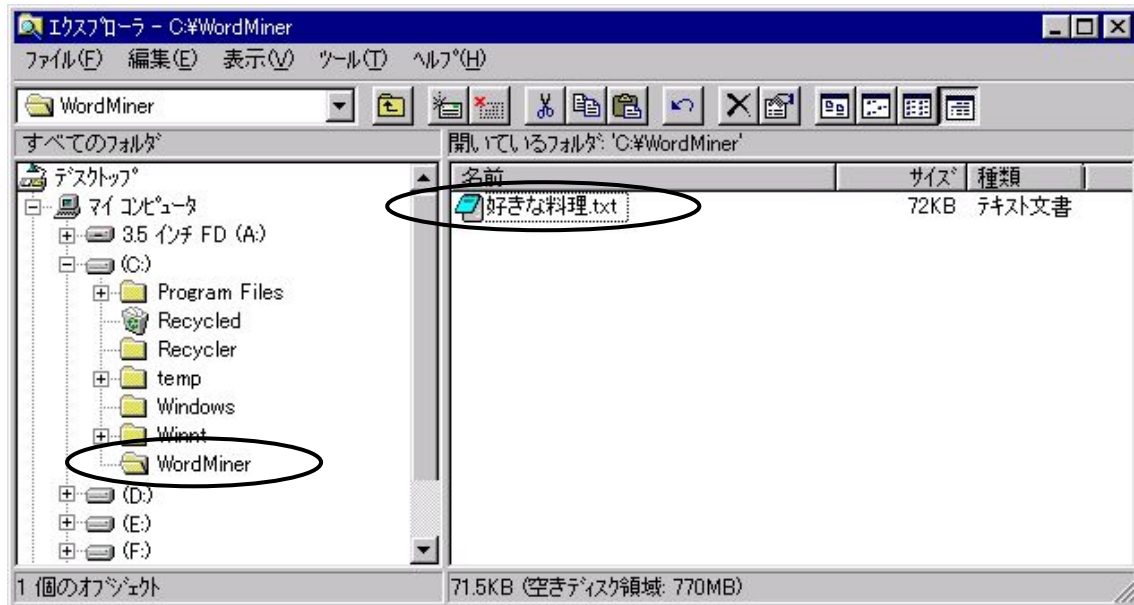
② 「C:」ドライブに「WordMiner」というフォルダを新規作成します。

③ フロッピーディスクに入っている【好きな料理.txt】を作成したフォルダの中にコピーします。

結果の観察

操作を終了すると以下のようなフォルダ構成になります。

処理結果を保存するフォルダ「WordMiner」が【C:¥ WordMiner】にあり、作成したフォルダの中にサンプルデータ【好きな料理.txt】があることを確認してください。



1-2 WordMiner®の起動

テキスト型データ解析ソフト WordMiner®が正常にインストールされていることを確認し、WordMiner®を起動させます。

操 作 説 明

(1) WordMiner®の起動方法 (その1)



- ① [スタート]－[プログラム(P)]－[WordMiner version 1.01d]－[WordMiner version 1.01d]をクリックすると、WordMiner®が起動されます。

(2) WordMiner®の起動方法 (その2)

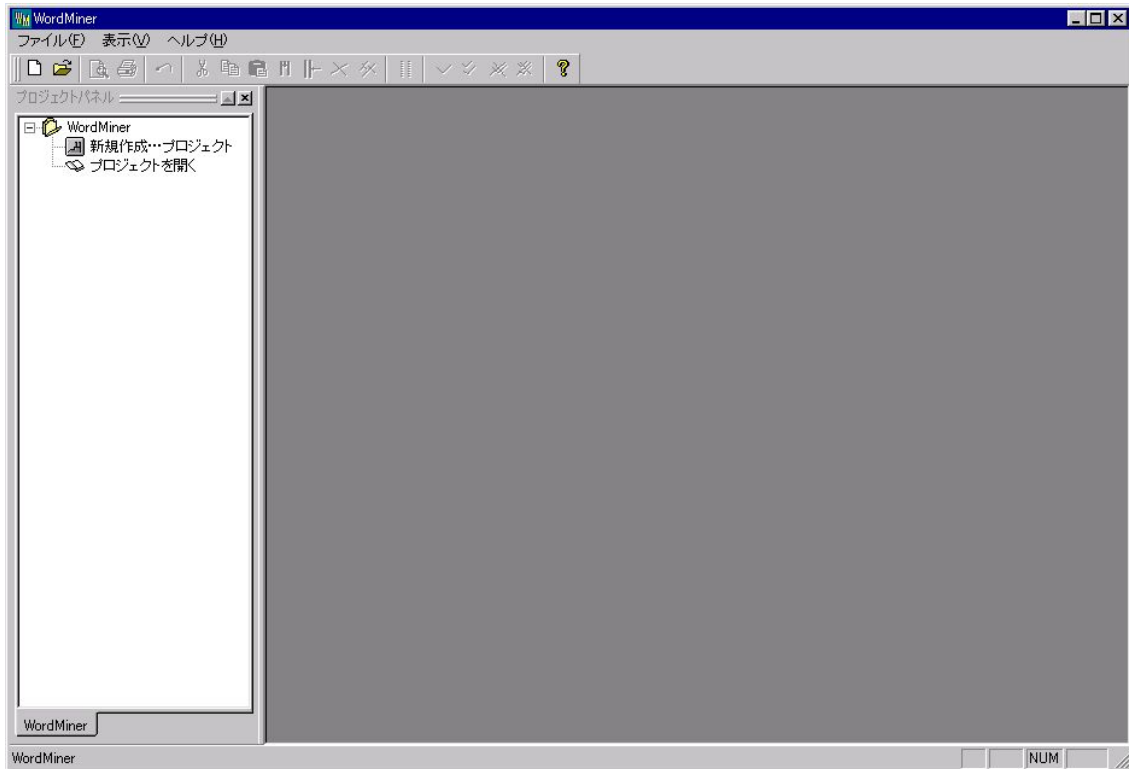


- ② インストール後、デスクトップに作成されたアイコンをダブルクリックすると、WordMiner®が起動されます。

結果の観察

WordMiner®起動直後の画面

以上の操作を行うと下の図のように WordMiner®が起動されます。



ヒント

1. WordMiner®のインストールについて

(1) のショートカット、または (2) のアイコンが表示されない場合、WordMiner®が正しくインストールされていない可能性があります。もう一度 WordMiner®をインストールし直すことをお勧めします。(詳細は「WordMiner®インストールガイド」をご覧ください)

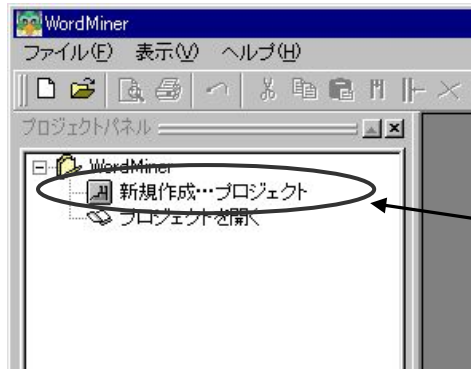
1-3 プロジェクトの作成

WordMiner®を起動し、最初に「プロジェクト」を作成します。

サンプルデータ【好きな料理.txt】を WordMiner®に読み込ませ、使用するデータが正常に読み込まれていることを[変数情報の確認]画面で確認します。

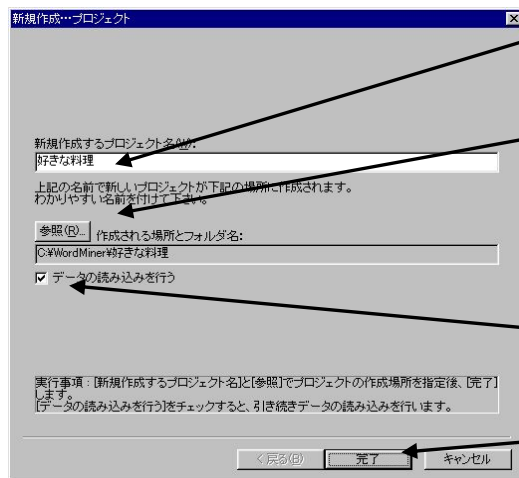
操 作 説 明

(1) プロジェクトを作成する



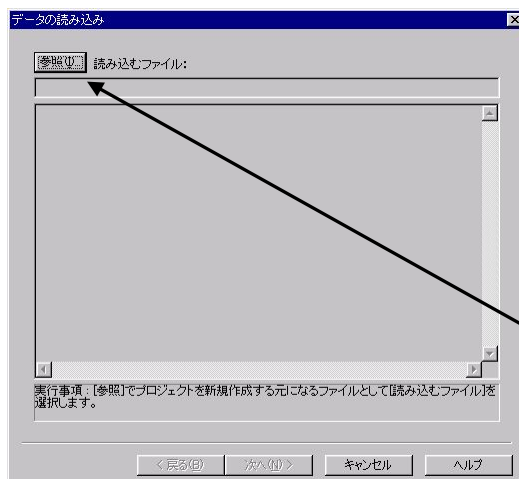
- ① WordMiner®を起動し、「プロジェクトパネル」にある【新規作成...プロジェクト】をダブルクリックすると(2)の画面が表示されます。

(2) プロジェクトの作成場所とプロジェクト名を設定する



- ② 【新規作成するプロジェクト名(W)】に「好きな料理」と入力します。
- ③ 【参照(R)...】をクリックし、【フォルダの参照】から「C:¥WordMiner¥」を指定し【OK】をクリックすると【作成される場所とフォルダ名】に、【C:¥WordMiner¥好きな料理】と表示されます。
- ④ 【データの読み込みを行う】をチェックします。
- ⑤ 【完了】をクリックすると、次の(3)の画面が表示されます。

(3) 解析に用いるファイルを指定する



- ⑥ 【参照(I)...】をクリックして解析に用いるファイル【C:¥WordMiner¥好きな料理.txt】を選びます。

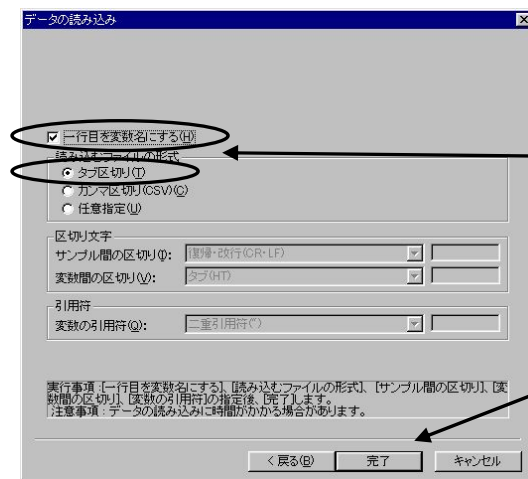
(4) 読み込んだファイルの確認をする



⑦ 読み込んだファイルが正しいか確認します。

⑧ [次へ(N)]をクリックすると次の(5)の画面が表示されます。

(5) 読み込むファイルの形式を指定する



⑨ ここで解析に使用するファイル【好きな料理.txt】は、各行がタブ区切りで、一行目に変数名が付与されていますので、[一行目を変数名にする(H)]と[タブ区切り(T)]をチェックします。

⑩ [完了]をクリックすると、データの読み込みが始まり、「作業領域」に結果の[変数情報の確認]画面が表示されます。

操作説明

プロジェクト作成直後に「変数情報の確認」画面が表示されます。

正しくサンプルデータが読み込まれると下の図のように表示されます。

変数管理番号	変数名	種類	文字種	有効サンプル数	無記入/空白数	非該当数	全構成要素数	全構成要素数(編集前)	異なり構成要素数	カテゴリー数
[V0001]	SEQ	原始変数	英数字	1111	0	0	-	-	-	-
[V0002]	No	原始変数	数字	1111	0	0	-	-	-	-
[V0003]	一番好きな料理	原始変数	その他	1110	1	0	-	-	-	-
[V0004]	次に好きな料理	原始変数	その他	1093	18	0	-	-	-	-
[V0005]	性別	原始変数	その他	1111	0	0	-	-	-	-
[V0006]	年代	原始変数	その他	1111	0	0	-	-	-	-
[V0007]	性年齢区分	原始変数	その他	1111	0	0	-	-	-	-

ヒント

1. プロジェクトの作成場所と名前について

インストール直後の初期状態ではプロジェクトの作成場所は **WordMiner®** をインストールしたフォルダになり、「プロジェクト」という名前が自動付与されます。プロジェクトは分かりやすい作業場所に置き、分かりやすい名前にした方が後の分析に便利です。

2. 「変数情報の確認」画面について

この画面には、変数名^{*3}、構成要素^{*4}の数などデータの情報が表示されます。
列ごとの説明は用語説明を参照してください。

3. データビューアによる確認

データビューアを用いて、読み込んだデータの確認をします。使用方法是（付録1 データビューア(P80)）を参照してください。

4. ファイルから読み込んだ変数について

読み込んだファイルの変数が「変数情報の確認」に正しく表示されているか確認してください。読み込みが正常ならば、サンプルデータの一行目(表題・タイトル)が変数名になっています。

今回は **WordMiner®** が自動付与した連番である変数「SEQ」と、ファイルから読み込んだ「No」「一番好きな料理」「次に好きな料理」「性別」「年齢」「性年齢区分」の計7つの変数が読み込まれ、表示されます。この時点では読み込んだ変数の「種類」は全て「原始変数」になっています。次の単元で分析を行う形式ごとに種類を変換します（「2-1. 2-2. 構成要素変数の生成」「2-3. 質的変数の生成」にて説明）。

5. 「変数情報の確認」画面を閉じてしまった場合

「変数の情報確認」画面右上の[×] をクリックして画面を閉じてしまった場合、「プロジェクトパネル」にある「変数情報の管理」-「変数情報の確認(変数名変更・削除など)」をダブルクリックすると「変数情報の管理」画面が表示されます。

用語説明

1. サンプル

1 回答（サンプルデータの 1 行）のことをいいます。

2. 変数

性別、年齢、構成要素等、サンプルデータの 1 列のことをいいます。

3. 変数名

変数についている表題、またはタイトルのことをいいます。

4. 構成要素

WordMiner®では、もとのテキストデータを半角空白で分割した単位のことをいいます。

空白で区切られた文字列には意味を持つものもありますし、単なる文字・数値の集まりであることもあります。

5. 種類

変数の種類を示します。種類は「原始変数」「構成要素変数」「質的変数」の 3 種類があります。

6. 原始変数

WordMiner®が入力ファイルを読み込んだときに標準認識する変数の種類です。

7. 構成要素変数

構成要素からなる変数のことをいいます。

8. 質的変数

選択肢型設問の回答データや属性データのことをいいます。

9. 文字種

1 変数の内容がすべて数値で入力されているならば「**数字**」、数値に英字が混在している場合は「**英数字**」と表示されます。文字列のみ、または英数字と文字列が混在している場合は「**その他**」と表示されます。

10. 有効サンプル数

その変数に、1 文字以上の文字・数値が入力されている回答数のことです。

11. 無記入/空白数

文字や数値を一切含まない空白の回答数（無記入/空白数が 18 ということは、空白回答が 18 回答あるということ）を示しています。

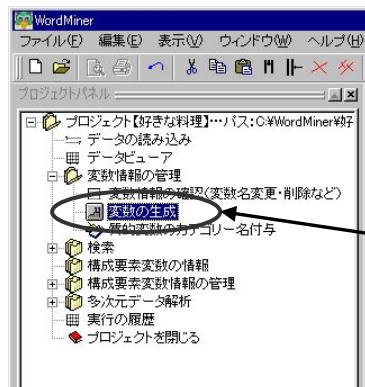
「有効サンプル数 + 無記入/空白数 = 総データ数」となります。

2-1 構成要素変数の生成(分かち書きツール WinAiBASE を起動する)

原始変数を構成要素変数に変換する操作を行います。テキスト型データに相当する原始変数を分かち書き^{*1}し、構成要素変数に変換します。ここでは分かち書き処理ツール (WinAiBASE) を起動させて指定したデータが正しく読み込まれているか確認します。

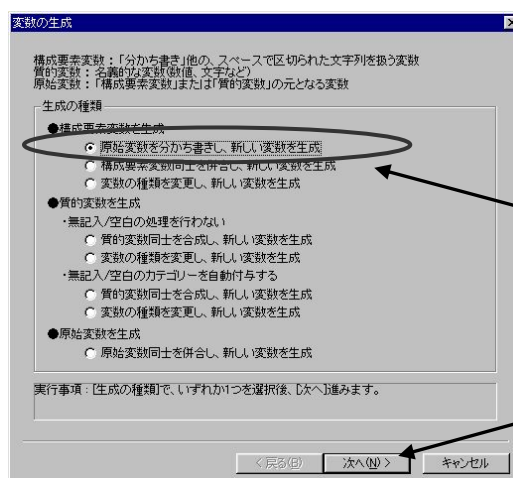
操作説明

(1) 構成要素変数を生成する



① 「プロジェクトパネル」にある「変数情報の管理」－「変数の生成」をダブルクリックすると次の(2)の画面が表示されます。

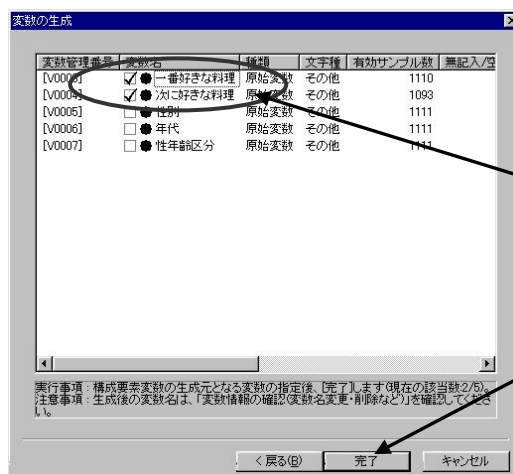
(2) 分かち書きを指定する



② 構成要素変数を生成し、分かち書きを行いますので、[原始変数を分かち書きし、新しい変数を生成]をチェックします。

③ [次へ(N)]をクリックすると、次の(3)の画面が表示されます。

(3) 分かち書きを行い構成要素変数に変換する原始変数を指定する



④ 分かち書きを行う変数名をチェックします。今回は、「一番好きな料理」と「次に好きな料理」をチェックします。

⑤ [完了]をクリックすると分かち書きツールが起動します。指定した変数が表示されていることを確認してください。

結果の観察

分かち書きツール(WinAiBASE)の起動後の画面。

指定したデータが WinAiBASE に正しく読み込まれていると、下の画面が表示されます。



The screenshot shows the WinAiBASE2.0 application window. It has a menu bar with options like 'ファイル(F)', '編集(E)', '表示(V)', '実行(X)', 'オプション(O)', 'ウィンドウ(W)', and 'ヘルプ(H)'. Below the menu is a toolbar with icons. The main area displays a table titled '原文' (Original Text). The table has four columns: 'No.', 'SEQ', '一番好きな料理' (Favorite Dish), and '次に好きな料理' (Second Favorite Dish). The data is as follows:

No.	SEQ	一番好きな料理	次に好きな料理
1	[00000001]	カレー	ラーメン
2	[00000002]	ラーメン	スパゲッティ
3	[00000003]	カレーライス	焼肉
4	[00000004]	ラーメン	炒飯
5	[00000005]	寿司	カレー
6	[00000006]	サラダ	カレーライス
7	[00000007]	カツ	カレー
8	[00000008]	チキンの料理	魚料理
9	[00000009]	スパゲッティ	お蕎麦
10	[00000010]	カレーライス	ハンバーグ
11	[00000011]	お刺し身	鳥料理
		納豆	焼き肉

At the bottom of the window, there is a status bar that says 'ヘルプを表示するには [F1] を押してください。' (Press [F1] to display help.) and a page indicator '1/1111行、1/2列' (1/1111 rows, 1/2 columns) along with a 'NUM' button.

ヒント

1. 分かち書きデータ

WordMiner®では半角空白で区切られている文字列を構成要素として認識しますので、半角空白で区切られているデータが準備されているときは、そのデータを分かち書き処理せずに使用することもできます。その場合は(2)の画面で「**変数の種類を変更し、新しい変数を生成する**」をチェックして操作を行うと、分かち書きツールを使わずに構成要素変数に変換することができ、以後の分析を行うことが可能です。

2. 分かち書きツールに読み込んだデータ

(3)で指定した変数に対応するデータが正しく読み込まれているかを確認してください。

分かち書きツールを起動しただけでは分かち書き処理は実行されません。次の単元(「2-2. 構成要素変数の生成(分かち書きを行う)」)の手順を行ってください。

用語説明

1. 分かち書き

日本語は単語間に空白がありません(膠着(こうちゃく)言語といいます)。分かち書き処理を行うことで文章をある単位に区切り、構成要素として扱います。

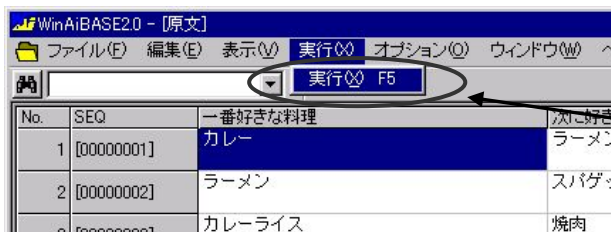
2-2 構成要素変数の生成（分かち書きを行う）

読み込んだ原始変数（テキスト型データ）に分かち書き処理を適用し、構成要素変数を生成します。構成要素変数が正常に生成されたかを、[変数情報の確認]画面で確認します。

この事例集では WordMiner version 1.01d を基に記載されています。お使いのバージョンと処理結果が異なる場合があります。この事例集と処理結果を同一にするには「付録6 キーワード抽出(P94)」を参照してオプションを変更してください。

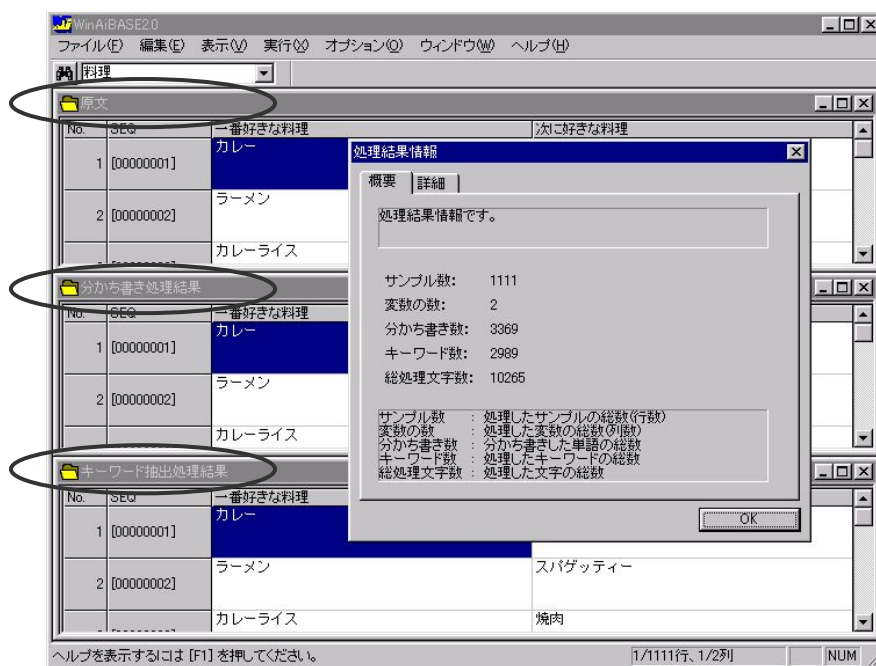
操作説明

(1) 分かち書きを実行する



① 「メニューバー」の[実行(X)]-[実行(X)]をクリックすると分かち書きを実行します。処理が終了しますと、次の(2)の画面が表示されます。

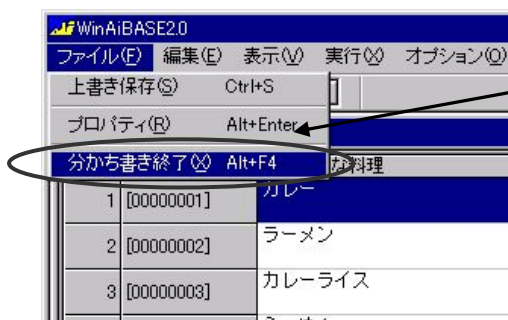
(2) 分かち書き処理実行後



② 上から「原文」「分かち書き」「キーワード」の順に表示されます。

③ [処理結果情報]に処理結果が表示されます

(3) 分かち書きツールを終了する



④ 「メニューバー」の[ファイル(F)]-[分かち書き終了(X)]をクリックするとデータが保存され、分かち書きツールが終了します。[変数情報の確認]画面の末尾に生成した変数が追加されます。生成した変数名が表示されていることを確認してください。

結果の観察

構成要素変数生成後（分かち書き終了後）に[変数情報の確認]画面が表示されます。
正しく構成要素変数が生成されると下の図のように表示されます。

変数情報の確認(変数名変更・削除など)										
変数の数:15 サンプル数 1111 [V0007] 性年齢区分										
変数管理番号	変数名	種類	文字種	有効サンプル数	無記入/空白数	非該当数	全構成要素数	全構成要素数(編集前)	異なり構成要素数	カテゴリー数
[V0001]	SEQ	原始変数	英数字	1111	0	0	-	-	-	-
[V0002]	No	原始変数	数字	1111	0	0	-	-	-	-
[V0003]	一番好きな料理	原始変数	その他	1110	1	0	-	-	-	-
[V0004]	次に好きな料理	原始変数	その他	1093	18	0	-	-	-	-
[V0005]	性別	原始変数	その他	1111	0	0	-	-	-	-
[V0006]	年代	原始変数	その他	1111	0	0	-	-	-	-
[V0007]	性年齢区分	原始変数	その他	1111	0	0	-	-	-	-
[V0008]	一番好きな料理-分かち書き	構成要素変数	その他	1110	1	0	2049	2049	373	-
[V0009]	次に好きな料理-分かち書き	構成要素変数	その他	1093	18	0	1866	1866	384	-
[V0010]	一番好きな料理-キーワード	構成要素変数	その他	1103	8	0	1541	1541	278	-
[V0011]	次に好きな料理-キーワード	構成要素変数	その他	1080	31	0	1448	1448	308	-

ヒント

1. 分かち書き処理結果の編集

分かち書き処理した結果は期待通りに区分されているとは限りません。日本語の文章を構成要素に切断することは一意とならないからです。そのため編集機能を用意してあります。

分かち書き結果を編集する場合、対象のセルをダブルクリックして内容を編集してください。または「メニューバー」の[編集(E)]-[置換(R)]等を使用して構成要素の編集を行ってください。

2. 分かち書き処理結果情報

分かち書き処理の実行後、画面中央に[処理結果情報]が表示されます。[概要]は分かち書き処理結果の全体が出力され、[詳細](右図)には変数ごとの処理結果が出力されます。[詳細]にある[出力(O)]をクリックすることで処理結果の数値をファイルに出力することができます。

処理結果情報			
概要 詳細			
処理結果情報です。			
	一番好きな料理	次に好きな料理	
原文文字数	5429	4836	
分かち書き数	1757	1612	
キーワード数	1541	1448	
出力(O)			
原文文字数 : 各列毎の原文文字数合計 分かち書き数 : 各列毎の分かち書き単語数合計 キーワード数 : 各列毎のキーワード数合計			
OK			

3. 生成した変数(構成要素変数)

分かち書き処理を行った変数ごとに「分かち書き」と「キーワード」が生成されます。[変数情報の確認]画面に表示されますので、確認してください。

ここで新たに生成された4つの変数の種類は「構成要素変数」になっています。

2. 変数を生成する

4. キーワード抽出

「キーワード」は、分かち書き処理結果から不要語・重複語を削除したものです。つまり、

- ・不要語となる記号、助詞等を削除します
- ・1回答に同じキーワードが複数存在した場合、そのキーワードは1つしか抽出されません

例)

原文 : カレーが好き。インドで食べたカレーが本当に美味しかった。

分かち書き : カレー□が□好き□。□インド□で□食べた□カレー□が□本当□に□美味しかった□。

処理結果
キーワード抽出 : カレー□インド□本当
処理結果

(※ □ は半角空白)

用語説明

1. 全構成要素数

ある変数内に含まれるすべての構成要素数のことです。

2. 異なり構成要素数

構成要素の種類数をいいます。

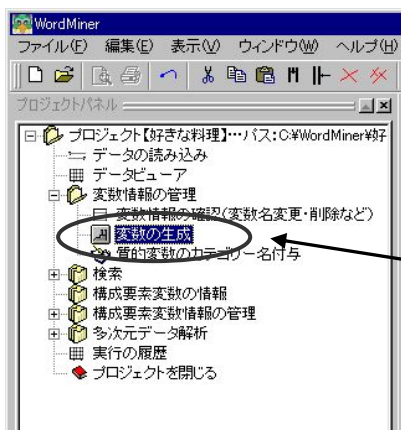
例えば、「カレー」という同一の構成要素が3回、「インド」が2回出現しても、これを2と数えます。

2-3 質的変数の生成

原始変数を質的変数に変換する操作を行い、選択肢型設問や属性項目を質的変数に変換します。質的変数が正常に生成されていることを[変数情報の確認]画面で確認します。

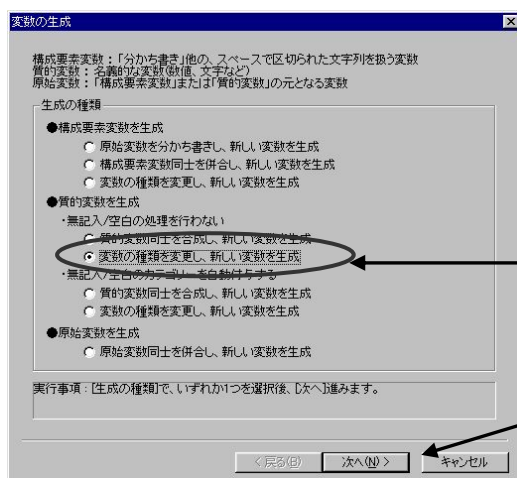
操作説明

(1) 質的変数を生成する



① 「プロジェクトパネル」にある
[変数情報の管理]－[変数の生成]をダブルクリックすると次の(2)の画面が表示されます。

(2) 質的変数を指定する



② 質的変数を生成しますので、[変数の種類を変更し、新しい変数を生成]をチェックします。

③ [次へ(N)]をクリックすると、次の(3)の画面が表示されます。

(3) 質的変数に変換する変数を指定する



④ 質的変数に変換する変数をチェックします。今回は、「性別」と「年代」、「性年齢区分」をチェックします。

⑤ [完了]をクリックすると[変数情報の確認]画面の末尾に生成した変数が追加されます。生成した変数名が表示されていることを確認してください。

結果の観察

質的変数生成後に[変数情報の確認]画面が表示されます。

正しく質的変数が生成されると下の図のように表示されます。

変数情報の確認(変数名変更・削除など)

変数の数:15
サンプル数 1111
[V0007] 性年齢区分

変数管理番号	変数名	種類	文字種	有効サンプル数	無記入/空白数	非該当数	全構成要素数	全構成要素数(編集前)	異なり構成要素数	カテゴリー数
[V0001]	SEQ	原始変数	英数字	1111	0	0	-	-	-	-
[V0002]	No	原始変数	数字	1111	0	0	-	-	-	-
[V0003]	一番好きな料理	原始変数	その他	1110	1	0	-	-	-	-
[V0004]	次に好きな料理	原始変数	その他	1093	18	0	-	-	-	-
[V0005]	性別	原始変数	その他	1111	0	0	-	-	-	-
[V0006]	年代	原始変数	その他	1111	0	0	-	-	-	-
[V0007]	性年齢区分	原始変数	その他	1111	0	0	-	-	-	-
[V0008]	一番好きな料理-分かち書き	構成要素変数	その他	1110	1	0	2049	2049	373	-
[V0009]	次に好きな料理-分かち書き	構成要素変数	その他	1093	18	0	1866	1866	384	-
[V0010]	一番好きな料理-キーワード	構成要素変数	その他	1103	8	0	1541	1541	278	-
[V0011]	次に好きな料理-キーワード	構成要素変数	その他	1080	31	0	1448	1448	308	-
[V0012]	性別-質的変数	質的変数	その他	1111	0	0	-	-	-	2
[V0013]	年代-質的変数	質的変数	その他	1111	0	0	-	-	-	5
[V0014]	性年齢区分-質的変数	質的変数	その他	1111	0	0	-	-	-	10

ヒント

1. 質的変数に変換する対象となる変数

構成要素変数（つまりテキスト型データ）との関連性の有無を知りたい原始変数を質的変数として生成します。例えば、ある構成要素変数と性年齢区分の関係、満足度との関係というように考えてみます。

2. 生成した変数(質的変数)

質的変数に変換した変数が[変数情報の確認]画面に表示されますので、確認してください。
ここで新たに生成された3つの変数の種類は「質的変数」になっています。

用語説明

1. カテゴリー数

ある質的変数のカテゴリー数を示します。

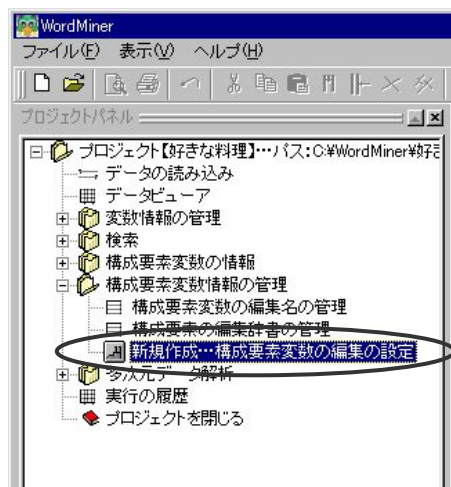
この例で、性別は男性・女性ですので、数値（カテゴリー数）は2となります。

2-4 構成要素の編集

構成要素を編集するルールを作成します。ここでは解析に用いる構成要素を再編集して、分析処理に適した形に整える操作を行い、結果を[変数情報の確認]画面で確認します。解析に用いる構成要素の編集方法はいろいろありますが、今回の解析では閾値(しきいち)*1のみを用いて構成要素を編集します。

操作説明

(1) 編集ルールを作成する



① 「プロジェクトパネル」にある「構成要素変数情報の管理」-「新規作成...構成要素変数の編集の設定」をダブルクリックすると、次の(2)の画面が表示されます。

(2) 編集する構成要素変数を指定する



② ルールを適用する変数名を指定します。今回の処理は、[編集する構成要素変数名(N)]に「[V0010]一番好きな料理-キーワード」を指定します。

③ [次へ(N)]をクリックすると、次の(3)の画面が表示されます。

(3) 作成する編集ルールの名前を設定する

④ ここでは生成する[編集ルール]の名前を指定します。[新規作成する構成要素変数の編集名(N)]は標準で自動付与されている名称「[V0010]一番好きな料理ーキーワードの編集」のまま、[完了]をクリックすると、次の(4)の画面が表示されます。

(4) 編集ルールを作成し、実行する

⑤ [新規作成する構成要素編集名]は標準で自動付与されている名称「一番好きな料理ーキーワードー編集」のままにします。

⑥ [閾値による構成要素の抽出]に「2」を入力します。

⑦ 他の条件は設定せずに、[実行]をクリックすると、[変数情報の確認]画面の末尾に作成した変数が追加されます。作成した変数が表示されていることを確認してください。

結果の観察

編集ルール（閾値）適用後に[変数情報の確認]画面が表示されます。

正しく構成要素変数が生成されると下の図のように表示されます。

変数情報の確認(変数名変更・削除など)										
変数の数:15 サンプル数:1111 [V0007] 性年齢区分										
変数管理番号	変数名	種類	文字種	有効サンプル数	無記入/空白数	非該当数	全構成要素数	全構成要素数(編集前)	異なり構成要素数	カテゴリ数
[V0001]	● SEQ	原始変数	英数字	1111	0	0	-	-	-	-
[V0002]	● No	原始変数	数字	1111	0	0	-	-	-	-
[V0003]	● 一番好きな料理	原始変数	その他	1110	1	0	-	-	-	-
[V0004]	● 次に好きな料理	原始変数	その他	1093	18	0	-	-	-	-
[V0005]	● 性別	原始変数	その他	1111	0	0	-	-	-	-
[V0006]	● 年代	原始変数	その他	1111	0	0	-	-	-	-
[V0007]	● 性年齢区分	原始変数	その他	1111	0	0	-	-	-	-
[V0008]	● 一番好きな料理-分かち書き	構成要素変数	その他	1110	1	0	2049	2049	373	-
[V0009]	● 次に好きな料理-分かち書き	構成要素変数	その他	1093	18	0	1866	1866	384	-
[V0010]	● 一番好きな料理-キーワード	構成要素変数	その他	1103	8	0	1541	1541	278	-
[V0011]	● 次に好きな料理-キーワード	構成要素変数	その他	1080	31	0	1448	1448	308	-
[V0012]	● 性別-質的変数	質的変数	その他	1111	0	0	-	-	-	2
[V0013]	● 年代-質的変数	質的変数	その他	1111	0	0	-	-	-	5
[V0014]	● 性年齢区分-質的変数	質的変数	その他	1111	0	0	-	-	-	10
[V0015]	● 一番好きな料理-キーワード-編集	構成要素変数	その他	1030	81	0	1366	1463	103	-

ヒント

1. 構成要素と閾値

閾値を変えたときの異なり構成要素率の推移を探索することは、構成要素の示す特徴を知る指標として考えられます。

構成要素変数の情報を表示（付録3 P84）することで、異なり構成要素率の推移を観察することができます。

2. 閾値以外の編集ルール

編集ルールは「閾値」以外に「削除」や「置換」があります。「6. 変数を編集する」(P60)で説明します。

3. 閾値を指定する目安

「プロジェクトパネル」にある[構成要素変数の情報]—[【[V0010]一番好きな料理—キーワード】]に出力された「その頻度以上の異なり構成要素率」を観察すると、上から「18.0%」「7.5%」「5.1%」「4.5%」「4.2%」となっています。グラフを見て数値の減少が急に小さくなったあたりを目安として閾値を設定します。

4. 名前の付け方のコツ

ある程度解析を進めると、変数名が次第に長くなります。これが長くなると変数名が画面内に入りきらず分かりにくくなります。「キーワード」ならば「KWD」とするなど「短く、分かりやすい」変数名に書き換えることをお勧めします。

5. 生成した変数（構成要素変数）について

編集ルール（閾値）を用いて生成した構成要素変数が[変数情報の確認]画面に表示されますので確認してください。

用語説明

1. 閾値（しきいち）

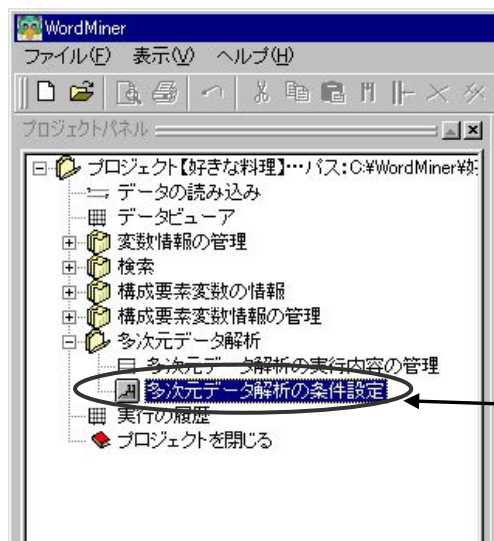
指定した数値未満の出現頻度の構成要素を切り捨て、それ以上の構成要素を残す、頻度による取捨選択の基準となる値のことです。例えば、閾値 2 以上ということは、出現頻度が 2 回以上の構成要素のみを解析に用いるということです。

3-1 多次元データ解析を用いて構成要素と質的変数[性年齢区分]の関連性を観察する

ある構成要素と質的変数の関連性を観察するために、多次元データ解析*1の機能を用いて分析を行います。

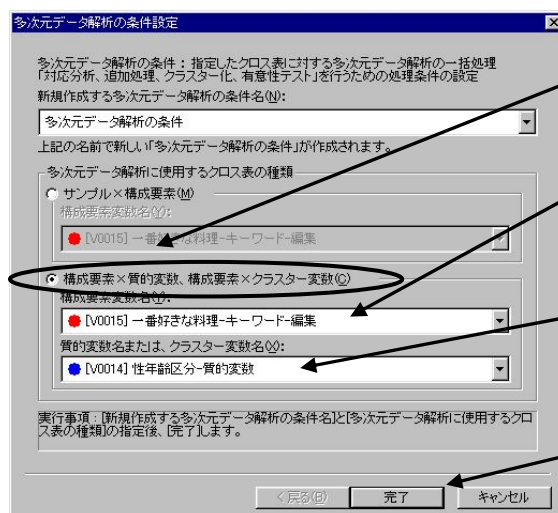
操 作 説 明

(1) 多次元データ解析を行う



① 「プロジェクトパネル」にある [多次元データ解析] - [多次元データ解析の条件設定] をダブルクリックすると、次の (2) の画面が表示されます。

(2) 多次元データ解析に用いる構成要素変数と質的変数を指定する



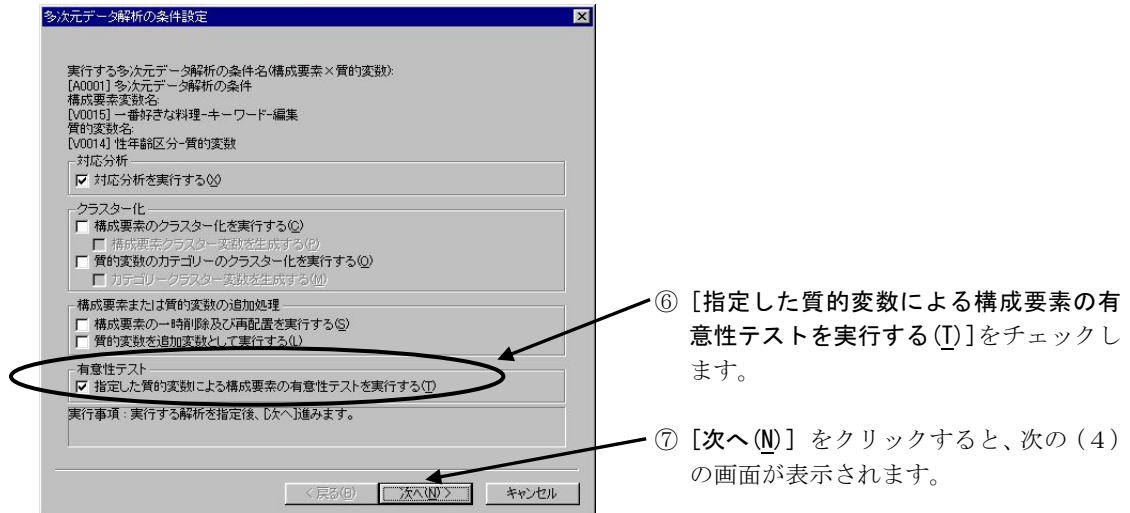
② [構成要素×質的変数、構成要素×クラスター変数(C)] をチェックします。

③ [構成要素変数名(Y)]に[[V0015]一番好きな料理-キーワード-編集]を指定します。

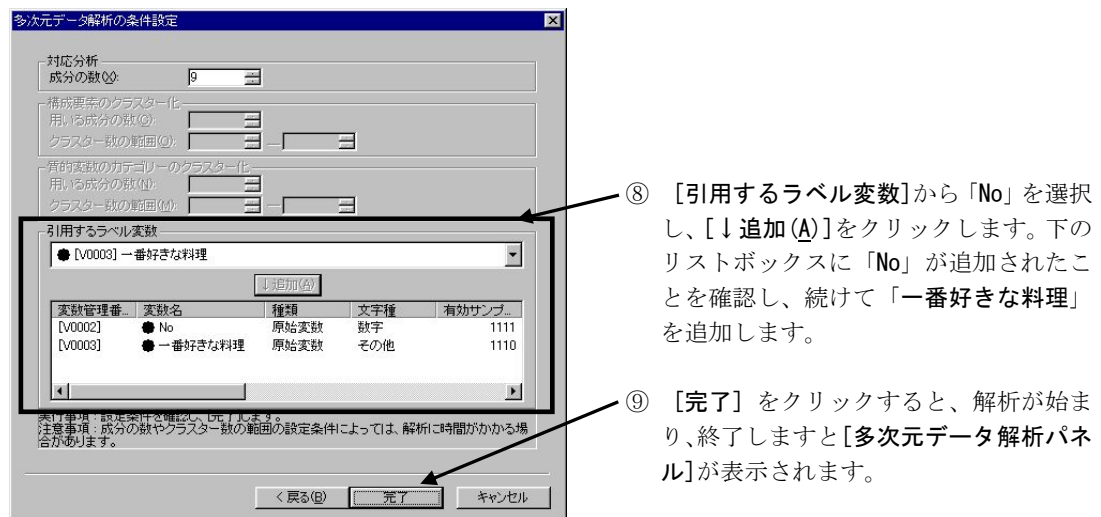
④ [質的変数名または、クラスター変数名(X)]に[[V0014]性年齢区分-質的変数]を指定します。

⑤ [完了] をクリックすると、次の (3) の画面が表示されます。

(3) 実行する多次元データ解析の処理条件を指定する

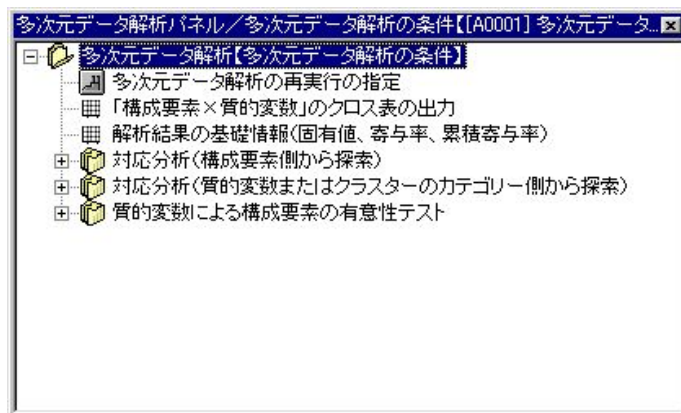


(4) 作成する対応分析の名前を指定する



結果の観察

以上の操作を行いますと、下の図の[多次元データ解析パネル]が表示されます。



ヒント

1. 多次元データ解析とは

構成要素と質的変数、または構成要素とサンプルとの生成クロス表の対応分析*2を行う機能です。

つまり、構成要素と質的変数、構成要素とサンプルの関連を探索するときに用いる機能です。

2. 「構成要素」×「性年齢区分」の解析

抽出した構成要素が「性年齢区分」とどのような関連があるのか、その程度はどのくらいか、といったことを分析します。

3. [多次元データ解析パネル]を閉じてしまった場合

[多次元データ解析パネル]右上の[×]をクリックして画面を閉じてしまった場合、「プロジェクトパネル」にある[多次元データ解析]－[多次元データ解析の実行内容の管理]をダブルクリックすると[多次元データ解析の実行内容の管理]画面が表示されます。

[多次元データ解析の条件名]や[種類]、[行の変数(表側)]、[列の変数(表頭)]を確認して解析の条件名をクリックすると[多次元データ解析パネル]が画面に再表示されます。

用語説明

1. 多次元データ解析

多くの場合、データは多次元データであり、様々な要素、要因が複雑に絡みあっています。複雑なデータ構造の中から意味的特徴を探索的に引き出す方法論の総称を多次元データ解析といいます。

2. 対応分析

多次元データ解析に含まれる分析技法の1つで所与のデータ表（ここでは「構成要素」×「性年齢区分」のクロス表）の分析を行う方法をいいます。なお対応分析とは **Correspondence Analysis** の和訳です。

3. はずれ値

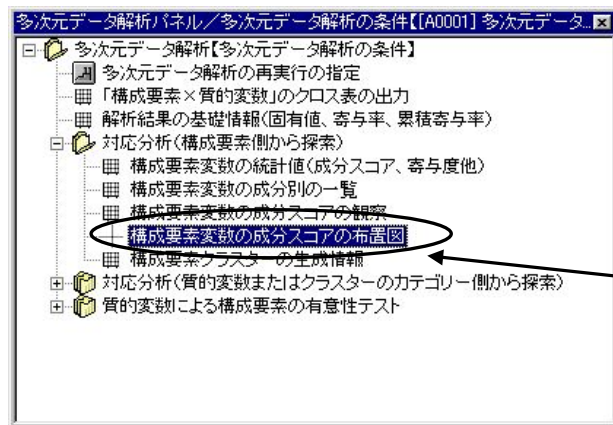
ある変数に含まれる情報を表す分布の姿を観察したとき、多数のデータの傾向からはずれて特異な位置にあるデータののことをいいます。

3-2 構成要素の特徴や傾向を図に表わし観察する

布置図を表示することによって、構成要素と質的変数の関連性の有無の程度を観察することができます。

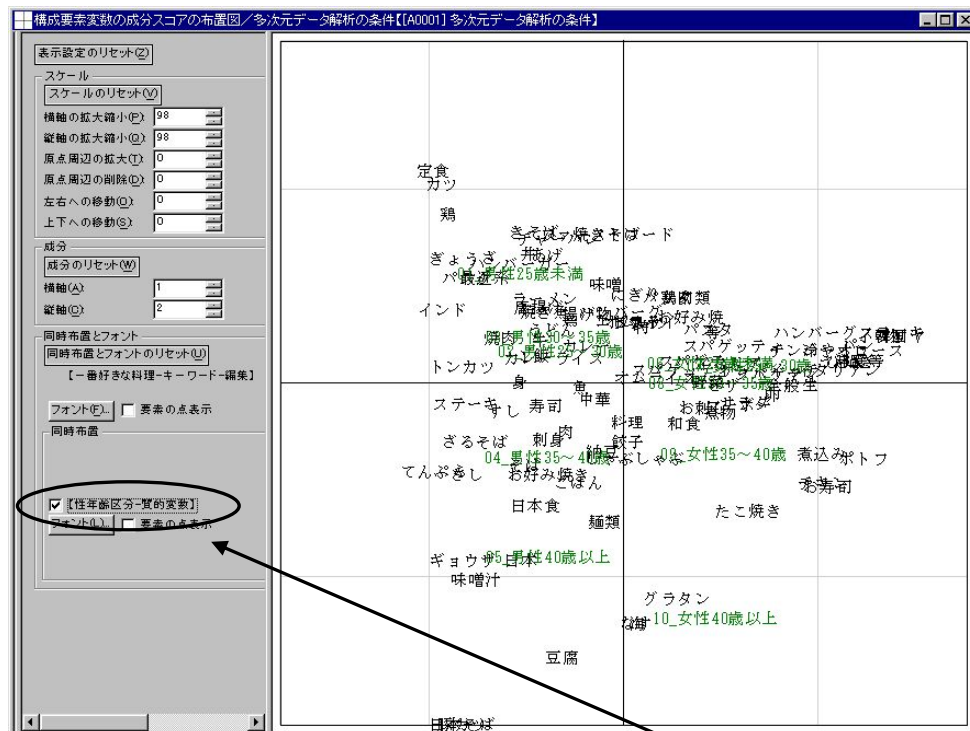
操作説明

(1) 布置図を観察する



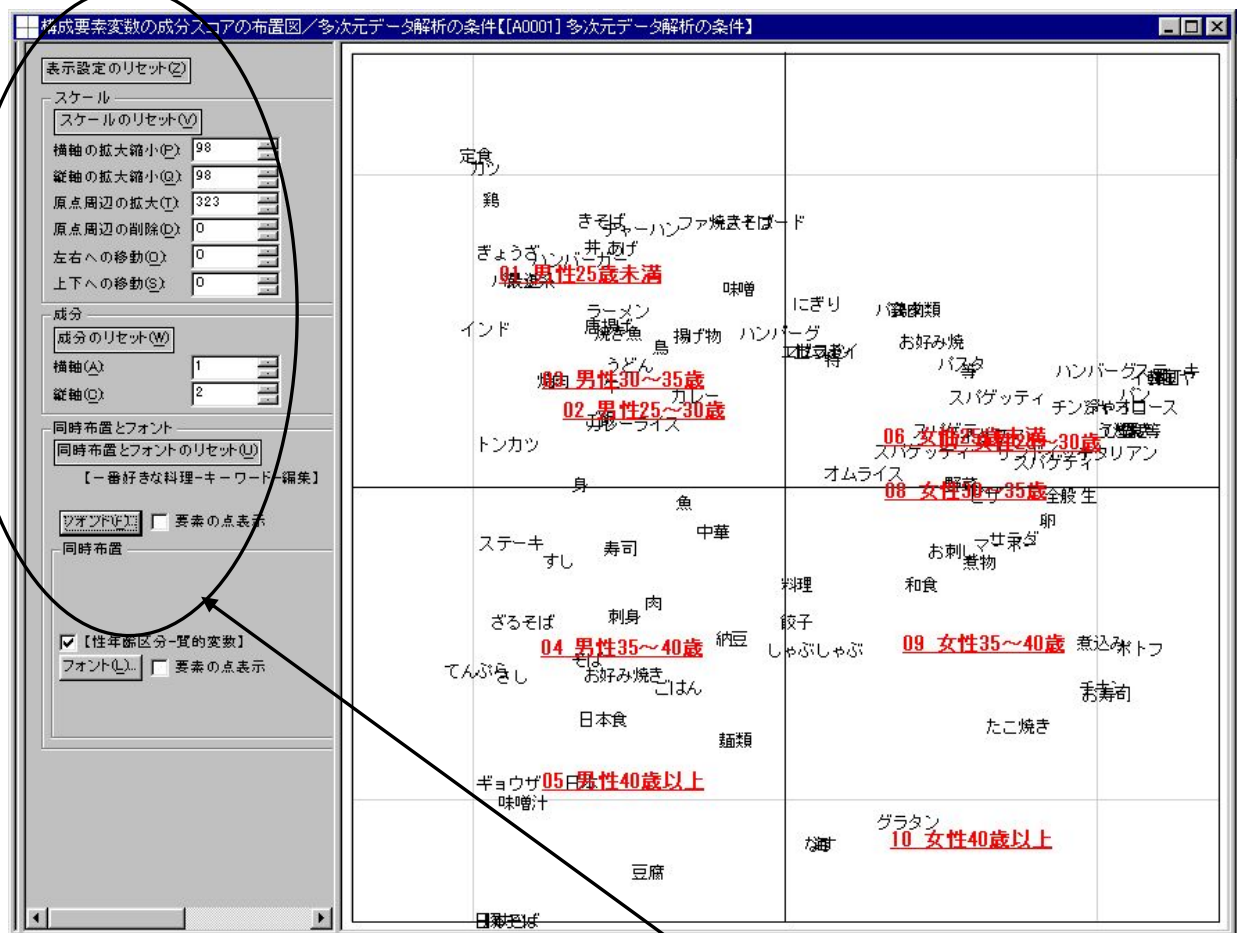
① [対応分析(構成要素側から探索)]－
[構成要素変数の成分スコアの布置図]
をクリックすると、(2)の画面が表示
されます。

(2) 同時布置図を表示する



② [性年齢区分－質的変数]をチ
ェックすると左図のように
「性年齢区分」が同時布置さ
れます。

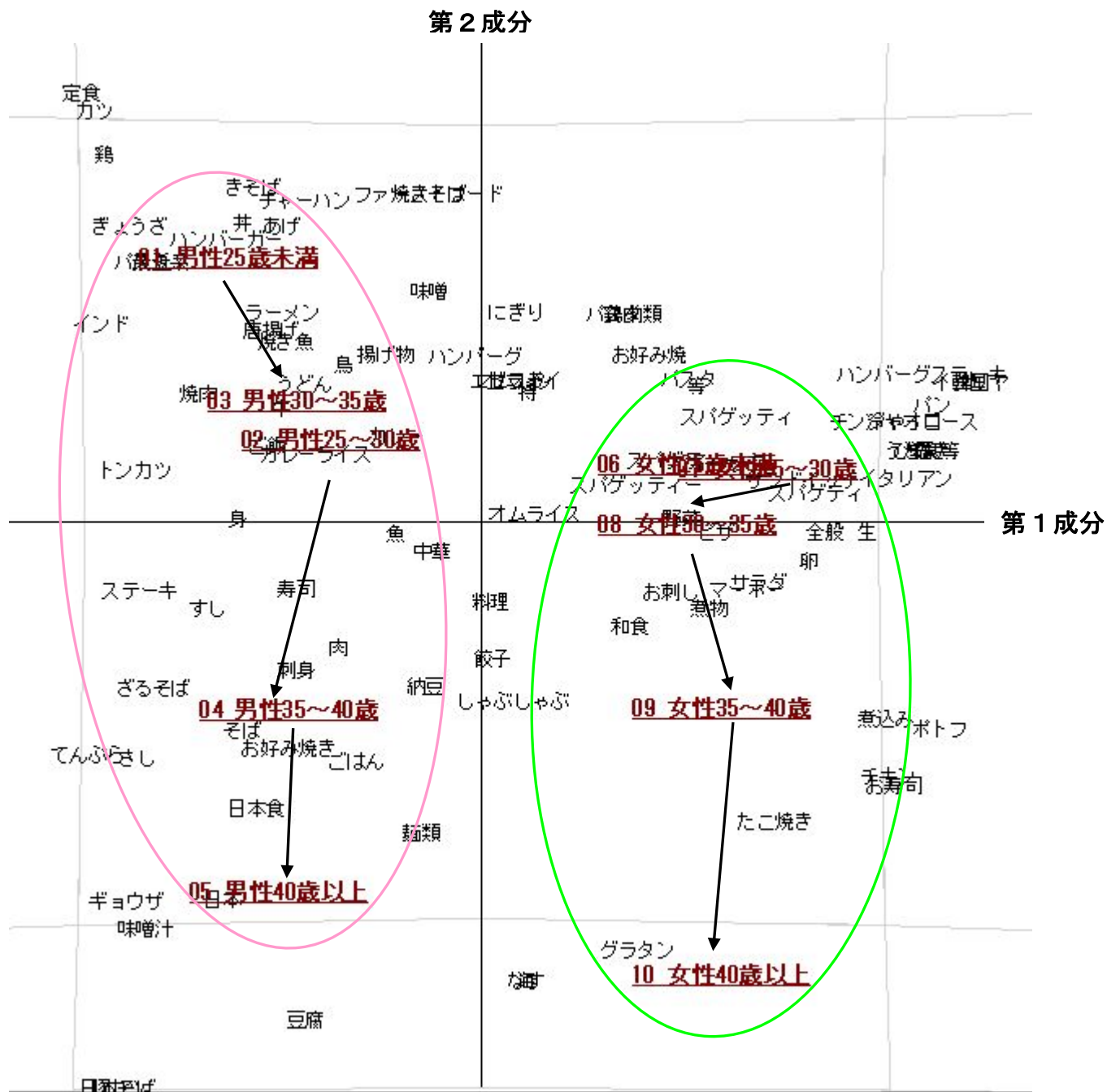
- (3) 構成要素変数「一番好きな料理ーキーワード編集」と質的変数「性年齢区分」との関係を見やすくする



- ③ 左のフレームにある[スケール]、[フォント]、[横・縦軸の拡大縮小]、[原点周辺の拡大・削除]等の数値や条件を変えることによって布置図を見やすくします。

結果の観察

多次元データ解析による分析と結果の解釈



ここでは、使用している全構成要素数 1,366 語、異なり構成要素数 103 語（異なり構成要素率 7.5%）、閾値 2 語以上、「性年齢区分」が 10 区分となっています。

特徴として質的変数の布置を観察すると、

1. 性別の差があることが見られる
2. 男女とも加齢（図の上から下）に伴って好きな料理が移り変わることが見られる

次の単元で出力される「頻度による有意テスト」と共に、性年齢区分カテゴリーの有意に働く構成要素を観察します。

ヒント

1. 構成要素スコアの観察

〔対応分析（構成要素側から探査）〕－〔構成要素別統計値（構成要素のスコア、寄与度他）〕等を表示することで、ある注目した成分にどのような構成要素が寄与しているかを数値で詳細に観察することができます。

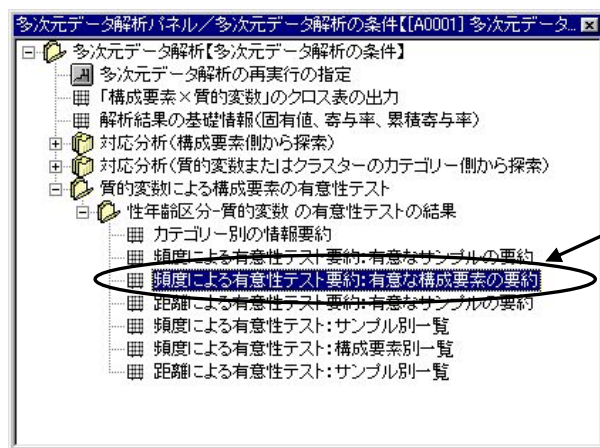
3-3 カテゴリーごとに特徴のある

構成要素の一覧を観察する

構成要素の出現頻度による有意性テスト*1の結果を観察します。「性年齢区分」のカテゴリーごとに特徴のある構成要素の一覧を観察します。

操作説明

(1) 頻度による有意性テストの結果を観察する



① [質的変数による構成要素の有意性テスト] - [性年齢区分 - 質的変数の有意性テストの結果] - [頻度による有意性テスト要約: 有意な構成要素の要約]をダブルクリックすると、次の(2)の画面が表示されます。

(2) 結果の一覧を観察する



② [上位・下位の構成要素の抽出の条件] - [検定値]を選択し、エディットボックスに「1」を入力します。

③ 解釈に必要な情報が表示されます。

結果の観察

ここで[検定値]を1にし、性別・年代のカテゴリーごとに有意に働く構成要素と、逆にそのカテゴリーに寄与しない構成要素を表にしてみました。

表1. 「性年齢区分」の構成要素による有意性テスト

性年齢区分		男性 25 歳未満	男性 25～30 歳未満	男性 30～35 歳未満	男性 35～40 歳未満	男性 40 歳以上
カテゴリー内 サンプル数		101	171	183	125	79
カテゴリーに寄与する構成要素	1	ラーメン	鳥	トンカツ	すし	肉
	2	井	そば	鶏	中華	豆腐
	3	カツ	ステーキ	魚	料理	料理
	4	牛	ラーメン	ぎょうざ	麺類	日本
	5	うどん	焼肉	味噌	お好み焼き	そば
	6	特	日本食	カレーライス	肉	日本食
	7	チャーハン	カレーライス	チャーハン	寿司	納豆
	8	カレーライス	焼き魚	きそば	カレー	和食
	9	てんぷら	刺身	日本	焼肉	餃子
	10	ファーストフード	イタリア	焼き魚	豆腐	豚カツ
	11	焼きそば	あげ	カレー	さし	日本そば
	12	定食	中華	ラーメン	日本	ギョウザ
	13	唐揚げ	唐揚げ	寿司	煮物	味噌汁
	14	インド	カツ	ハンバーグ	ざるそば	てんぷら
	15	焼肉	ご飯	ハンバーガー	しゃぶしゃぶ	なす
カテゴリーに寄与しない構成要素	15	刺身	寿司	特	イタリアン	焼肉
	14	お寿司	パスタ	全般	ハンバーグ	すし
	13	すし	納豆	和食	チャーハン	サラダ
	12	サラダ	スパゲッティ	麺類	全般	中華
	11	日本	野菜	料理	ピザ	煮物
	10	中華	等	中華	お寿司	スパゲティー
	9	スパゲティー	お好み焼き	お寿司	スパゲッティー	イタリア
	8	煮物	料理	豆腐	井	イタリアン
	7	和食	和食	イタリア	特	焼き魚
	6	パスタ	麺類	サラダ	野菜	スパゲティ
	5	イタリアン	お寿司	井	スパゲティ	ハンバーグ
	4	スパゲティ	豆腐	スパゲティ	スパゲッティ	ラーメン
	3	魚	スパゲティー	煮物	和食	スパゲッティ
	2	イタリア	餃子	イタリアン	イタリア	カレー
	1	料理	イタリアン	スパゲッティ	パスタ	パスタ

性年齢区分		女性 25 歳未満	女性 25～30 歳未満	女性 30～35 歳未満	女性 35～40 歳未満	女性 40 歳以上
カテゴリー内 サンプル数		94	207	110	35	6
カテゴリーに寄与する構成要素	1	スパゲッティー	パスタ	全般	麺類	お寿司
	2	スパゲッティ	イタリアン	スパゲティ	スパゲティ	豆腐
	3	サンドイッチ	お寿司	生	サラダ	日本
	4	スパゲティー	イタリア	和食	チキン	野菜
	5	中華	韓国	サラダ	ポトフ	スパゲッティ
	6	お好み焼き	物	納豆	煮込み	和食
	7	イタリアン	スパゲッティ	煮物	お刺し	
	8	しゃぶしゃぶ	イタリア	うどん等	マーボー	
	9	にぎり	等	ひじき	和食	
	10	ハンバーグステーキ	パン	鍋	サンドイッチ	
	11	煮込み	野菜	洋食	全般	
	12	ピザ	和食	なす	イタリア	
	13	うどん	たこ焼き	海	お寿司	
	14	パスタ	煮物	エビフライ	唐揚げ	
	15	ごはん	餃子	オムレツ	スパゲッティー	
カテゴリーに寄与しない構成要素	15	焼肉	納豆	チャーハン		
	14	身	そば	お好み焼き		
	13	すし	お好み焼き	カレー		
	12	日本	刺身	唐揚げ		
	11	そば	うどん	カレーライス		
	10	煮物	麺類	スパゲッティー		
	9	肉	焼き魚	豆腐		
	8	ラーメン	身	焼肉		
	7	餃子	寿司	すし	寿司	
	6	カレー	日本	日本	トンカツ	
	5	麺類	ラーメン	そば	魚	
	4	焼き魚	肉	ハンバーグ	パスタ	
	3	トンカツ	中華	肉	料理	
	2	和食	トンカツ	トンカツ	ラーメン	
	1	魚	カレーライス	ラーメン	中華	

3. 分析を行い、結果を観察する [構成要素×質的変数]

性年齢区分のカテゴリーに有意に働く構成要素と、逆にそのカテゴリーにあまり寄与しない構成要素をそれぞれ 15 語、要約しました (表 1)。あるカテゴリーに寄与する構成要素の程度を測る検定値、構成要素数等の数値は省略してあります。数値の詳細は[多次元データ解析パネル]－[質的変数による構成要素の有意性テスト]－[性年齢区分－質的変数の有意性テストの結果]－[頻度による有意性テスト:構成要素別一覧]を参照してください。

寄与する構成要素とあわせて寄与しない構成要素を観察することで、性年齢区分の特徴をよりよく観察することができます。ある構成要素に焦点を合わせ観察すると性年齢区分の差異を見つけることができます。例えば「ラーメン」は男性の若年層 (25 歳未満、25～30 歳未満) では上位に位置しますが、年代が上がるごとに下がっていることが見られます。さらに男性 40 歳以上では「ラーメン」は寄与しない構成要素になっています。女性を見ると「ラーメン」はほとんどの年代で寄与しないということを見ることができます。女性の場合「スパゲティ」が年代を問わず、カテゴリーに有意な単語として現れているのが見え、逆に男性は女性より好きな料理として「スパゲティ」をあまり記述しなかった、ということを表から読み取ることができます。

女性 40 歳以上の寄与しない構成要素が出力されていませんが、このカテゴリーにあまり寄与しない構成要素がなかったことを示します。

ヒント

1. 有意性テストの出力結果

ここでは構成要素別の有意性を出力していますが、WordMiner®はサンプル別の有意性も出力しています。サンプル別の有意性テストの出力結果は P27 (4) の[引用するラベル変数]で指定した変数の内容を出力します。

2. 距離による有意性テストの出力

あるカテゴリーに所属する構成要素が、その重心 (つまり平均的な傾向) からどの程度離れているか、あるいは近いかを測ります。構成要素やサンプルに付与された数値が小さいほど、まとまりがよいことになります。

3. 構成要素の表示内容の切り替え

インストール直後の初期状態では最初に表示される構成要素の数は検定値の値を基準にして表示されます。[上位・下位の構成要素の抽出の条件]－[抽出数]を指定することで入力した数の構成要素を表示することができます。

用語説明

1. 頻度による有意性テスト

ある構成要素が、当該カテゴリー内における出現頻度と全構成要素のそのカテゴリー内での出現頻度とを比較し、有意か否かをテストし、その結果を構成要素別に有意性テストの結果として表示します。さらに、構成要素の検定値を用いて、その構成要素を含む回答文（すなわち、分析に用いたサンプル内の構成要素群）の有意性テストを行い、その回答文を有意順に表示し、解釈の手がかりを与えます。

構成要素やサンプルに付与された数値が大きいほど、その構成要素やサンプルが所属するカテゴリー内で有意であることを表しています。

2. カテゴリー内構成要素数構成比

「当該カテゴリー内構成要素数 ÷ カテゴリー内全構成要素数 × 100」（%）を表示しています。

3. 構成要素数構成比

「当該構成要素数 ÷ 全構成要素数 × 100」（%）を表示しています。

4. カテゴリー内構成要素数

そのカテゴリー内に出現する当該構成要素の頻度数です。

5. 構成要素数

当該構成要素の出現頻度です。

6. 検定値

各構成要素について、その出現頻度に基づく有意性テストの結果を示す検定値です。

7. 有意確率

検定値に対応する有意確率を示しています。

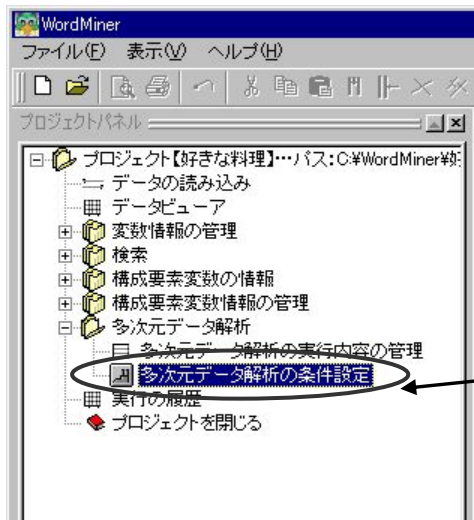
4-1 多次元データ解析を用いて

構成要素とサンプルの関連性を観察する

ある構成要素とサンプルとの関連性を観察するために、多次元データ解析の機能を用いて分析を行います。

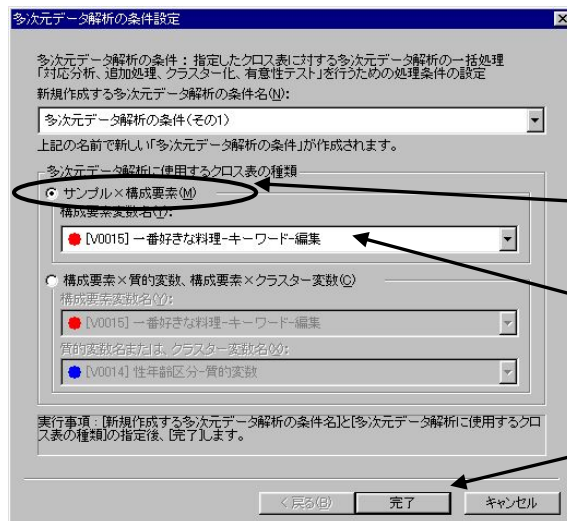
操 作 説 明

(1) 多次元データ解析を行う



① 「プロジェクトパネル」にある[多次元データ解析]－[多次元データ解析の条件設定]をダブルクリックすると、次の(2)の画面が表示されます。

(2) 多次元データ解析に用いる構成要素変数を指定する



② 「サンプル×構成要素(M)」をチェックします。

③ [構成要素変数名(Y)]に[[V0015]一番好きな料理-キーワード-編集]を指定します。

④ [完了]をクリックすると、次の(3)の画面が表示されます。

(3) 実行する多次元データ解析を指定する

実行する多次元データ解析の条件名(サンプル×構成要素):
[A0002] 多次元データ解析の条件(その1)
構成要素変数名:
[v0015] → 番号好きな料理-キーワード 編集
対応分析およびクラスター化
☒ サンプルのクラスター化を実行する(S)
☒ サンプルクラスター変数を生成する(M)
☒ サンプルクラスター変数による構成要素の有意性テストを実行する(P)
☐ 構成要素のクラスター化を実行する(C)
☐ 構成要素クラスター変数を生成する(Q)
変数の追加処理
☐ 構成要素変数を追加変数として実行する(A)
☐ 質的変数を追加変数として実行する(L)
実行事項: 実行する解析を指定後、次へ進みます。
注意事項: 生成後のクラスター変数名は、「変数情報の確認(変数名変更・削除など)」を確認して下さい。

戻る(B) 次へ(N) > キャンセル

- ⑤ [サンプルのクラスター化を実行する(S)]と[サンプルクラスター変数を作成する(M)]、「サンプルクラスター変数による構成要素の有意性テストを実行する(P)」の3つにチェックを入れます。

- ⑥ [次へ(N)] をクリックすると、次の
(4) の画面が表示されます

(4) 作成する対応分析の名前を指定する

多次元データ解析の条件設定

対応分析
成分の数(M): 15

サンプルのクラスター化
用いる成分の数(S): 15
クラスター数の範囲(U): 20 - 20

構成要素のクラスター化
用いる成分の数(Q):
クラスター数の範囲(Q):

引用するサンプル表数
● [V0003] - 一番好きな料理

追加(A)

変数管理番号	変数名	種類	文字種	有効サンプル数
[V0002]	● No	原始変数	数字	1111
[V0003]	● 一番好きな料理	原始変数	その他	1110

実行事項 設定条件を適用し、解析を行います。
注意事項 成分の数やクラスター数の範囲の設定条件によっては、解析に時間がかかる場合があります。

< 戻る(B) 完了 キャンセル

- ⑦ [クラスター数の範囲(P)]の両方のエディットボックスに「20」を入力します。(20 群に分けます)

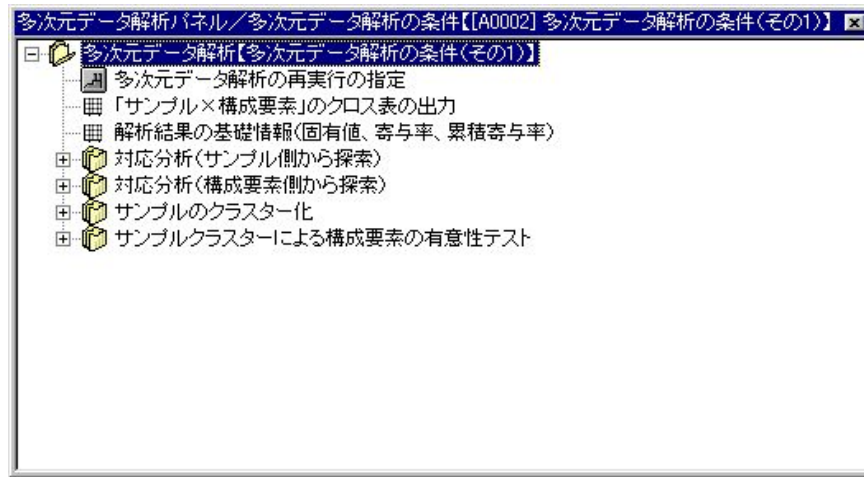
- ⑧ [引用するラベル変数]から「No」を選択し、[↓追加(A)]をクリックします。下のリストボックスに「No」が追加されたことを確認し、続けて「一番好きな料理」を追加します。

- ⑨ [完了] をクリックすると、解析が始まり、終了しますと[多次元データ解析パネル]が表示されます。

4. 分析を行い、結果を観察する [構成要素×サンプル]

結果の観察

以上の操作を行いますと、下の図のように[多次元データ解析パネル]が別ウインドウで表示されます。



ヒント

1. 「クラスター数の範囲(P)」(画面(4)－⑦)

全サンプルを分類したいグループ数、すなわちクラスター数を指定します。

この例では「20」と指定することで全サンプルを 20 群のクラスターに分類します。なお、クラスター数の範囲を指定することで、何通りものクラスター化の結果が得られます。

例えば、左のエディットボックスに「10」、右のエディットボックスに「15」と指定したときは 10 群から 15 群まで 6 通りのクラスター化の結果が得られます。

1. 「構成要素」×「サンプル」の解析

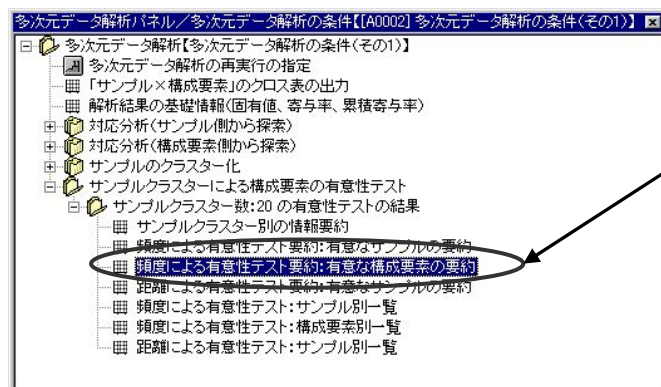
抽出した構成要素がサンプルとどのような関連があるのか、その程度はどのくらいか、といったことを分析します。選んだ構成要素とサンプル(回答文)の関係を多次元的に探査します。また、構成要素、サンプルそれぞれのクラスター化を行うことができます。

4-2 サンプルの類型化を行い、回答群の観察をする

クラスター化（自動分類）を行った結果の観察を行います。まず、サンプルの類型化（クラスター生成）を行い、どのような回答群があるのかを観察します。まず初めに各クラスターに特徴的な構成要素の観察をします。

操作説明

(1) 各クラスターに有意な単語の一覧を表示する



① [サンプルクラスターによる構成要素の有意性テスト] - [サンプルクラスター数: 20 の有意性テストの結果] - [頻度による有意性テスト要約: 有意な構成要素の要約] をダブルクリックすると、(2) の画面が表示されます。

(2) 結果の一覧を観察する



② [上位・下位の構成要素の抽出の条件] - [検定値] にチェックを入れエディットボックスに「1」を入力します。

③ 解釈に必要な情報が表示されます。

結果の観察

[上位・下位の構成要素の抽出の条件]－[検定値]に「1」を入力し、クラスターごとに有意に働く構成要素と、逆にそのクラスターに寄与しない構成要素 15 語を表にしてみました。構成要素が 15 語に満たないクラスターは有意検定の基準に満たなかったということです。

表 2. クラスターごとの構成要素による有意性テスト

クラスター	クラスター1	クラスター2	クラスター3	クラスター4	クラスター5	クラスター6	クラスター7	クラスター8	クラスター9	クラスター10
カテゴリ内サンプル数	49	14	21	6	10	13	6	8	2	16
カテゴリに寄与する構成要素	1 カレー 2 3 4 5 6 7 8 9 10 11 12 13 14 15	スパゲティ	トンカツ等	オムライス 焼きそば	日本食 たこ焼き	焼肉 ハンバーグステーキ さし	ハンバーガー	お好み焼き	ぎょうざ	唐揚げ あげ 鶏 鳥 揚げ物
カテゴリに寄与しない構成要素	15 トンカツ 14 寿司 13 スパゲティ 12 ハンバーグ 11 魚 10 イタリア 9 スパゲッティ 8 中華 7 和食 6 カレーライス 5 ラーメン 4 肉 3 パスタ 2 1 料理									カレー 和食 カレーライス ラーメン 肉 パスタ 料理

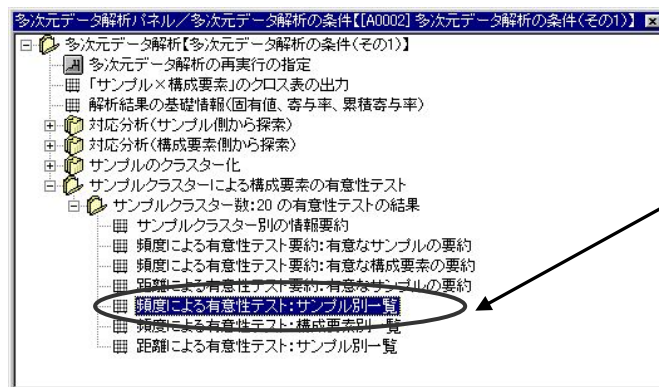
クラスター	クラスター11	クラスター12	クラスター13	クラスター14	クラスター15	クラスター16	クラスター17	クラスター18	クラスター19	クラスター20
カテゴリ内サンプル数	2	24	50	674	6	10	8	27	11	73
カテゴリに寄与する構成要素	1 ギョウザ 2 3 4 5 6 7 8 9 10 11 12 13 14 15	スパゲティ物	寿司 餃子 納豆 ご飯 ごはん にぎり 味噌汁 なす 生 豆腐	料理 肉 パスタ ラーメン 和食 中華 スパゲッティ イタリア 魚 焼き魚 イタリアン 野菜 そば 煮物 日本	きそば ファーストフード	冷やっこ お好み焼 ざるそば オムレツ 日本そば	サンドイッチ エビフライ	ハンバーグ 煮込み	スパゲッティ	カレーライス しゃぶしゃぶ 最近
カテゴリに寄与しない構成要素	15 14 13 12 11 10 9 8 7 6 5 4 3 2 1		トンカツ スパゲティ ハンバーグ 魚 イタリア スパゲッティ 中華 カレー 和食 カレーライス ラーメン 肉 パスタ 料理	トンカツ 納豆 日本食 お好み焼き 唐揚げ 焼肉 スパゲッティ 餃子 スパゲティ ラーメン ハンバーグ カレー 料理				スパゲッティ 中華 カレー 和食 カレーライス ラーメン 肉 パスタ 料理		麺類 寿司 トンカツ スパゲティ ハンバーグ 魚 イタリア スパゲッティ 中華 カレー 和食 ラーメン パスタ 料理

4. 分析を行い、結果を観察する [構成要素×サンプル]

[表2. クラスターごとの構成要素による有意性テスト]を観察しながら、各クラスター内のサンプルがどのように分類されているか観察します。一例として、クラスター10に含まれる16名、クラスター16に含まれる10名のサンプルを取り上げ表にします。

操作説明

(1) 各クラスターに有意なサンプルの一覧を表示する



① [サンプルクラスターによる構成要素の有意性テスト]—[サンプルクラスター数:20の有意性テストの結果]—[頻度による有意性テスト: サンプル別一覧]をダブルクリックすると、(2)の画面が表示されます。

(2) 結果の一覧を観察する



② [上位・下位の構成要素の抽出の条件]—[抽出数]に「300」を入力します。

③ ウィンドウ下のタブをクリックすることで各クラスターの内容を観察することができます。

結果の観察

クラスター10 とクラスター16 のサンプルを全抽出し、どのような回答があるのか観察します（表3）。

表3. クラスター10、16 のサンプル

クラスター10			クラスター16		
番号	検定値	本文	番号	検定値	本文
1	8.88	唐揚げ	1	4.94	冷やっこ
2	8.88	唐揚げ	2	4.94	冷やっこ
3	8.88	唐揚げ	3	4.94	冷やっこ
4	8.88	唐揚げ	4	3.85	ざるそば
5	8.88	唐揚げ	5	3.85	お好み焼
6	7.07	鶏の唐揚げ	6	3.85	日本そば
7	7.07	鳥の唐揚げ	7	3.85	日本そば（温かいものとざるそば）
8	7.07	鳥の唐揚げ	8	3.85	お好み焼
9	7.07	鶏の唐揚げ	9	1.93	かた焼きオムレツ
10	5.99	からあげ	10	1.28	オムレツ、カレーライス
11	5.62	鶏のからあげ			
12	5.62	鳥のからあげ			
13	5.26	鳥のから揚げ			
14	3.94	揚げ物（鶏の唐揚げやヒレカツなど）			
15	2.99	鳥肉のからあげ			
16	2.99	若鳥のからあげ			

以上の操作で、クラスター10 は「唐揚げ」が一番好きであると回答したグループであり、クラスター16 は「冷やっこ」「そば」が好きであると回答したグループであるということが分かります。

他に、クラスター14 は 674 名の群になっていますが、WordMiner®はクラスター14 のみを取り出して、もう一度さらにクラスター化することができます。また、ここではクラスター数を 20 と指定しましたが、ユーザの指定によりクラスター数を増減することができます（P39-（4））。

ヒント

1. 検定値について

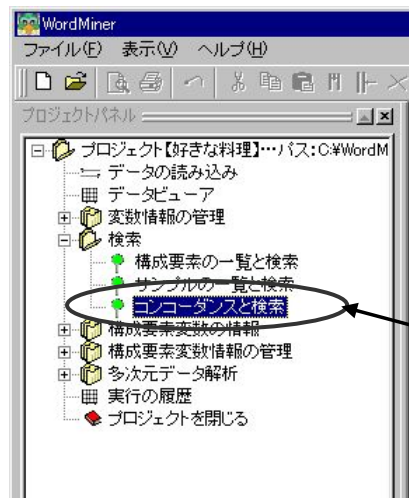
この表の検定値は、各クラスター内の構成要素による有意性テストにより得られた検定値から計算されたクラスター基準値です。この数値が大きいほどそのクラスターに有意な構成要素が多く使われているということになります。

5-1 特徴的な構成要素の前後の構成要素を表示し、その構成要素がどのように用いられているか観察する

結果の観察・分析中に気になった構成要素の文章中での使われ方を探索します。コンコーダンス*1機能を用いて検索対象の構成要素がどのように使われているか観察します。

操 作 説 明

（１）コンコーダンスを行う



- ① 「プロジェクトパネル」にある [検索]－[コンコーダンスと検索]をダブルクリックすると（２）の画面が表示されます。

（２）特徴的な構成要素を検索する



- ② [検索する変数(V)]－[一番好きな料理]を指定します。
- ③ 今までの処理結果のなかで気になった構成要素を [検索文字列(C)] に入力します。今回の例では[味噌]を入力します。
- ④ [検索(X)] をクリックします。

結果の観察

文章中に[味噌]をもつサンプルが[味噌]を中央に配置して一覧表示されます。



ヒント

1. コンコーダンスの使用法について

コンコーダンスは重要な語句や意味不明の語があるときなど、この機能を用いてデータを探索します。

用語説明

1. コンコーダンス（用語検索）

用語が回答文（テキスト文）の中で、どのように用いられているかを知るための機能です。指定をした用語が含まれる文を検索し、その指定語を中央に配置し、文章（構成要素変数）を表示します。

■ ここまでのまとめ

ここまでの分析・観察で *WordMiner*®によるテキスト型データ分析の一連の流れを説明しましたが、テキスト型データを詳細に分析したとは言いきれません。

表をみても分かる通り「スパゲティ」「お寿司」等にはいくつかの表記が見られます。「スパゲティー」「スパゲッティ」、「すし」「寿司」等です。このような表記のゆれを1つの構成要素に置換する機能を *WordMiner*®は持っています。

ここまでの説明では省略しましたが[編集ルール]を用いることによって、上記のような表記のゆれや、同意語、同義語等を1つの単語として扱うことが可能です。

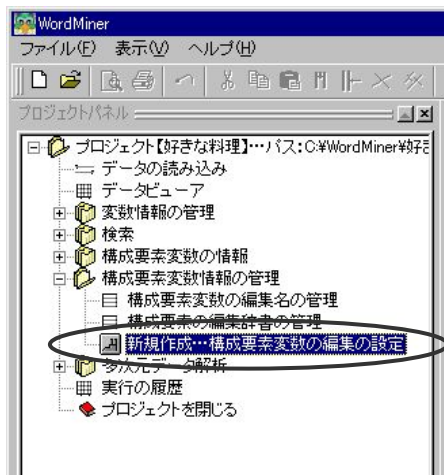
次の単元から上記の処理を行い、テキスト型データを分析します。

6-1 構成要素の編集(置換)

構成要素を編集する置換ルールを作成します。ここでは置換辞書ウインドウに構成要素を入力し、同義語を置換する構成要素を登録します。

操 作 説 明

(1) 編集ルールを作成する



① 「プロジェクトパネル」にある「構成要素変数情報の管理」－「新規作成...構成要素変数の編集の設定」をダブルクリックすると、次の(2)の画面が表示されます。

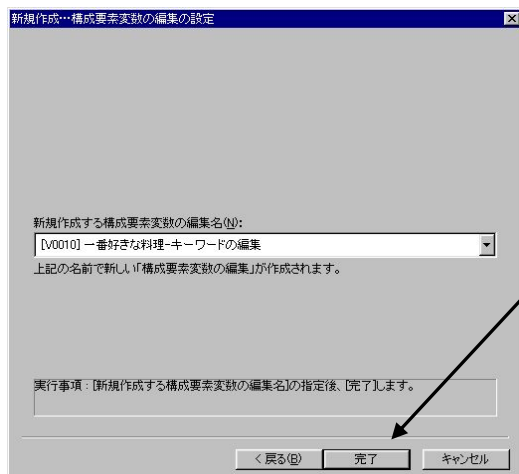
(2) 編集する構成要素変数を指定する



② ルールを適用する変数名を指定します。今回の処理は、[編集する構成要素変数名(N)]に「[V0010]一番好きな料理－キーワード」を指定します。

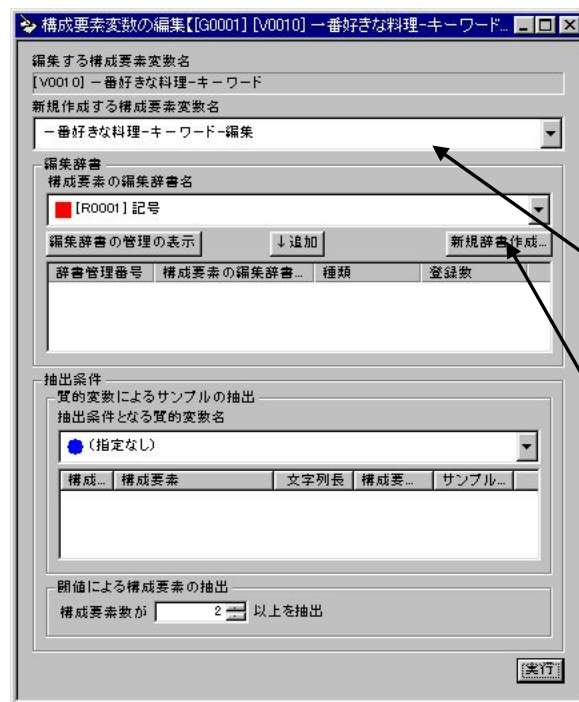
③ [次へ(N)] をクリックすると、次の(3)の画面が表示されます。

(3) 作成する構成要素変数の編集名を指定する



④ ここでは作成する[編集ルール]の名前を指定します。[新規作成する構成要素変数の編集名(N)]は標準で自動付与されている名称「[V0010]一番好きな料理－キーワードの編集」のまま、[完了]をクリックすると、次の(4)の画面が表示されます。

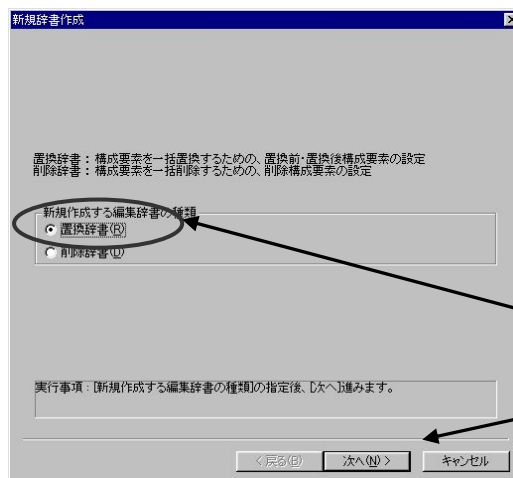
(4) 新規辞書編集ルール（置換）を作成する



⑤ [新規作成する構成要素編集名]は標準で自動付与されている名称「一番好きな料理-キーワード-編集」にします。

⑥ [新規辞書作成...] をクリックすると、次の(5)の画面が表示されます。

(5) 作成する辞書を選択する



⑦ [置換辞書(R)] をチェックします。

⑧ [次へ(N)] をクリックすると、次の(6)の画面が表示されます

(6) 作成する辞書の名前を指定する



⑨ ここでは作成する[編集辞書]の名前を指定します。[新規作成する編集辞書名(N)]は標準で自動付与されている名称「置換辞書」のまま、[完了]をクリックすると、次の(7)の画面が表示されます。

6. 変数を編集する

(7) 置換する構成要素を入力し登録します

置換辞書【R0008】置換辞書」／構成要素変数の編集【G0001】 [V00...

置換後 納豆

置換前 なっとう

追加(A)↓

置換後	置換前
-----	-----

⑩ 例として[置換後]のエディットボックスに「納豆」、[置換前]のエディットボックスに「なっとう」を入力します。

⑪ [追加(A)↓] をクリックすると、次の画面のようになります。

結果の観察

置換辞書へ編集する構成要素が追加されると下の図のようになります。

「なっとう」は「納豆」に置換されます。

置換辞書【R0008】置換辞書」／構成要素変数の編集【G0001】 [V00...

置換後 納豆

置換前 なっとう

追加(A)↓

置換後	置換前
納豆	なっとう

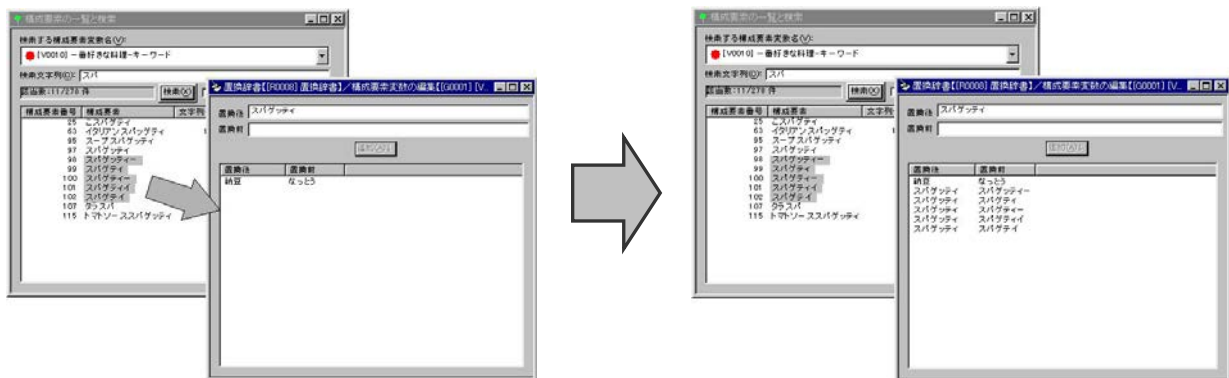
ヒント

1. 置換辞書登録について

【構成要素の一覧と検索】ウインドウからドラッグ&ドロップで登録が可能です。

例として、「プロジェクトパネル」の【検索】－【構成要素の一覧と検索】を開き、【検索(X)】ボタンをクリックして構成要素を表示します。

「スパゲッティ」を置換後とし、【置換後】エディットボックスに入力します。表記のゆれと思われる「スパゲティー」「スパゲテイ」などを選択し、ドラッグ&ドロップで置換辞書に登録します。



2. 誤入力について

間違えて登録してしまった場合は、リストから削除する構成要素を選択し、キーボードの Delete ボタンを押すことで置換辞書から除くことができます。

6. 変数を編集する

3. この事例集との処理結果をあわせるためには

以降の事例と内容と整合性をあわせるには以下の 64 語の登録を行ってください。

	置換後	置換前
1	うどん	うどん等
2	お好み焼き	お好み焼
3	お刺身	おさし
4	お刺身	お刺し
5	お刺身	刺身
6	お刺身	刺し身
7	お刺身	さし
8	お刺身	刺し
9	お寿司	寿司
10	お寿司	すし
11	お寿司	お鮨
12	お肉	肉
13	お肉	肉系
14	お肉	肉類
15	ご飯	ごはん
16	ご飯	ライス
17	ご飯	御飯
18	そば	ソバ
19	そば	日本そば
20	そば	ざるそば
21	たまご	卵
22	たまご焼き	卵焼き
23	てんぷら	天ぷ
24	めん類	麺類
25	イタリア	イタリア
26	イタリアン	イタリア
27	カレーライス	カレー
28	サンドイッチ	サンドウィッチ
29	スパゲッティ	スパゲッティー
30	スパゲッティ	スパゲティ
31	スパゲッティ	スパゲティー
32	スパゲッティ	スパゲティイ
33	スパゲッティ	スパゲテイ
34	スパゲッティ	こスパゲティ
35	チャーハン	炒飯
36	トンカツ	豚カツ
37	ナス	なす
38	ハンバーグ	ハンバ
39	パスタ	パスタ系
40	パスタ	パスタ類
41	パスタ	バスタ
42	ピザ	p i z z a

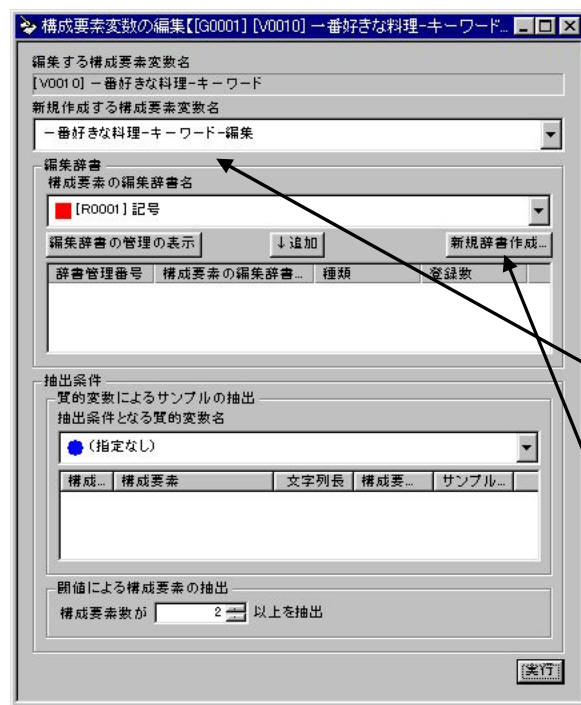
	置換後	置換前
43	フライ	フライ物
44	魚	お魚
45	鶏	鳥
46	鶏	チキン
47	鶏	チキン系
48	焼きそば	きそば
49	焼き魚	焼魚
50	中華	中華風
51	鳥肉	鶏肉
52	漬け物	お漬物
53	唐揚げ	あげ
54	豆腐	お豆腐
55	鍋	鍋物
56	納豆	なっとう
57	野菜	野菜ごちやまぜ
58	野菜	サラダ
59	冷やっこ	冷や奴
60	和食	和食系
61	和食	日本食
62	和食	日本
63	餃子	ぎょうざ
64	餃子	ギョウザ

6-2 構成要素の編集(削除)

構成要素を編集する削除ルールを作成します。ここでは削除辞書ウインドウに構成要素を入力し、解析から不必要と思われる構成要素を取り除く構成要素を登録します。

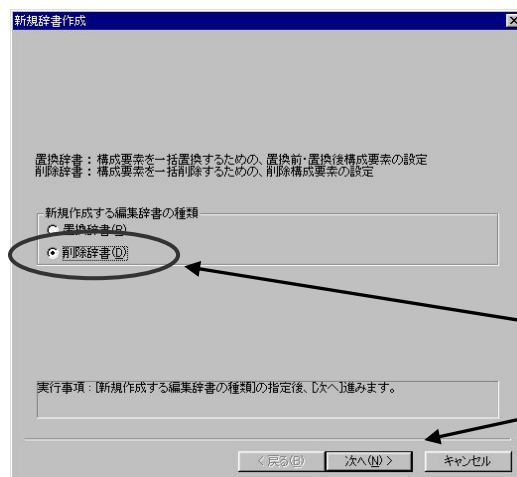
操 作 説 明

(1) 新規辞書編集ルール(削除)を作成する



- ① 置換辞書を作成したときに用いた左図のウインドウを開きます。(「プロジェクトパネル」にある[構成要素変数情報の管理]から[構成要素変数の編集名の管理]を開き[[V0010] 一番好きな料理-キーワードの編集]をクリックすることで開くことができます)
- ② [新規作成する構成要素編集名]は標準で自動付与されている名称「一番好きな料理-キーワード-編集」にします。
- ③ [新規辞書作成...] をクリックすると、次の(2)の画面が表示されます。

(2) 作成する辞書を選択する



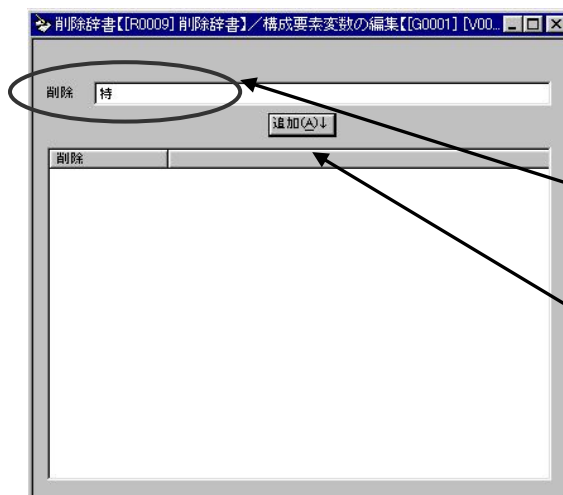
- ④ [削除辞書(D)] をチェックします。
- ⑤ [次へ(N)] をクリックすると、次の(3)の画面が表示されます

(3) 作成する辞書の名前を指定する



- ⑥ ここでは作成する[編集辞書]の名前を指定します。[新規作成する編集辞書名(N)]は標準で自動付与されている名称「削除辞書」のまま、[完了]をクリックすると、次の(4)の画面が表示されます。

(4) 置換する構成要素を入力し登録します



- ⑦ 例として、[削除]のエディットボックスに「特」を入力します。
- ⑧ [追加(A)↓] をクリックすると、次の画面のようになります。

結果の観察

削除辞書へ解析から取り除く構成要素が追加されると下の図のようになります。
「特」という構成要素は解析から取り除かれます。



ヒント

1. 削除辞書登録について

置換辞書登録と同様に[構成要素の一覧と検索]ウインドウからドラッグ&ドロップで登録が可能です。

2. 誤入力について

間違えて登録してしまった場合は、リストから削除する構成要素を選択し、キーボードの Delete ボタンを押すことで削除辞書から除くことができます。

3. この事例集との処理結果をあわせるためには

以降の事例と内容と整合性をあわせるには以下の 7 語の削除登録を行ってください。

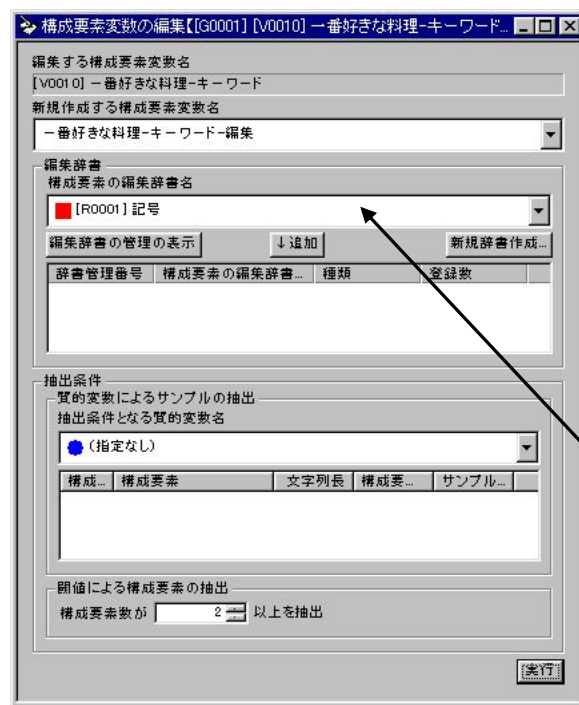
	削除
1	特
2	物
3	全般
4	等
5	身
6	最近
7	料理

6-3 構成要素の編集(実行)

前の単元で作成した置換辞書、削除辞書を用いて、構成要素を編集実行します。使用するデータが正常に生成されていることを[変数情報の確認]画面で確認します。

操 作 説 明

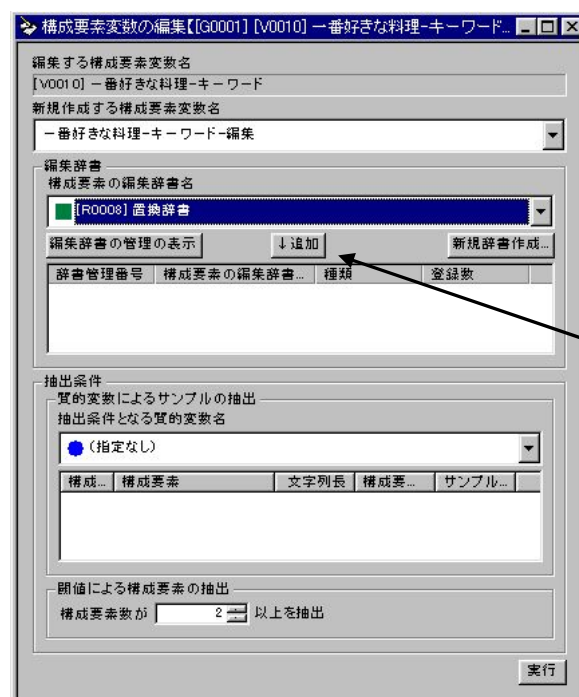
(1) 編集ルールを選択する



- ① 置換辞書、削除辞書を作成したときに用いた左図のウィンドウを開きます。
(「プロジェクトパネル」にある[構成要素変数情報の管理]から[構成要素変数の編集名の管理]を開き[[V0010]一番好きな料理-キーワードの編集]をクリックすることで開くことができます)

- ② [構成要素の編集辞書名]コンボボックスをクリックし[[R0008]置換辞書]を選択しますと、次の(2)の画面のようになります。

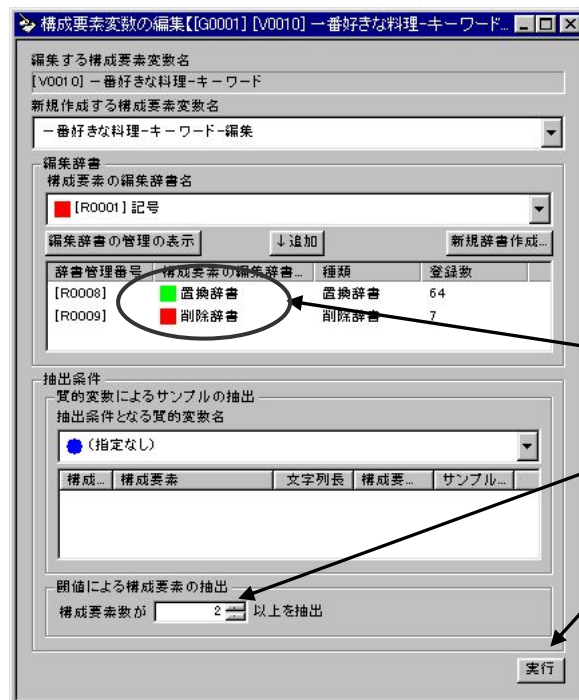
(2) 使用する辞書を追加する



- ③ [↓追加] をクリックすると、画面中央のリストボックスに[[R0008]置換辞書]が追加されます。

- ④ 続いて[構成要素の編集辞書名]コンボボックスから[[R0009]削除辞書]を選択し、[↓追加]をクリックしますと、次の(3)の画面のようになります。

(3) 編集ルールを設定し実行する



⑤ 作成した[置換辞書]と[削除辞書]が表示されていることを確認します。

⑥ [閾値による構成要素の抽出]に「2」を入力します。

⑦ [実行]をクリックすると、[変数情報の確認]画面の末尾に作成した変数が追加されます。作成した変数が表示されていることを確認してください。

結果の観察

構成要素の編集実行後に [変数情報の確認] 画面が表示されます。

正しく構成要素変数が生成されると下の図のように表示されます。

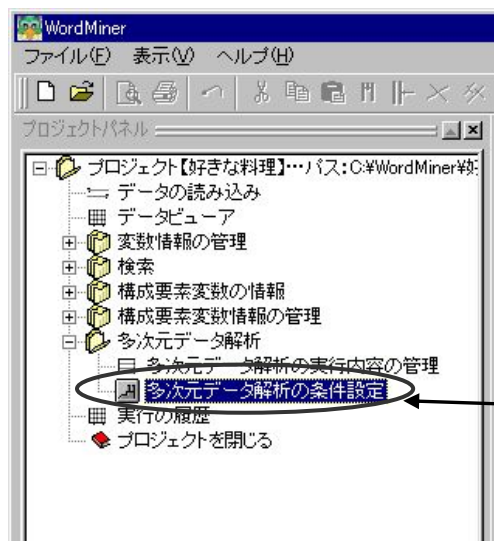
変数情報の確認(変数名変更・削除など)									
変数の数:15 サンプル数:1111 [v0001] SEQ									
変数管理番号	変数名	種類	文字種	有効サンプル数	無記入/空白数	非該当数	全構成要素数	全構成要素数(編集前)	異なり構成要素数
[v0001]	SEQ	原始変数	英数字	1111	0	0	-	-	-
[v0002]	No	原始変数	数字	1111	0	0	-	-	-
[v0003]	一番好きな料理	原始変数	その他	1110	1	0	-	-	-
[v0004]	次に好きな料理	原始変数	その他	1093	18	0	-	-	-
[v0005]	性別	原始変数	その他	1111	0	0	-	-	-
[v0006]	年代	原始変数	その他	1111	0	0	-	-	-
[v0007]	性年齢区分	原始変数	その他	1111	0	0	-	-	-
[v0008]	一番好きな料理-分かち書き	構成要素変数	その他	1110	1	0	2049	2049	373
[v0009]	次に好きな料理-分かち書き	構成要素変数	その他	1093	18	0	1866	1866	384
[v0010]	一番好きな料理-キーワード	構成要素変数	その他	1103	8	0	1541	1541	278
[v0011]	次に好きな料理-キーワード	構成要素変数	その他	1080	31	0	1448	1448	308
[v0012]	性別-質的変数	質的変数	その他	1111	0	0	-	-	-
[v0013]	年代-質的変数	質的変数	その他	1111	0	0	-	-	-
[v0014]	性年齢区分-質的変数	質的変数	その他	1111	0	0	-	-	-
[v0015]	一番好きな料理-キーワード-編集	構成要素変数	その他	1030	81	0	1166	1451	71

7-1 多次元データ解析を用いて構成要素と 質的変数[性年齢区分]の関連性を観察する

ある構成要素と質的変数の関連性を観察するために、多次元データ解析のいろいろな機能を用いて分析を行います。

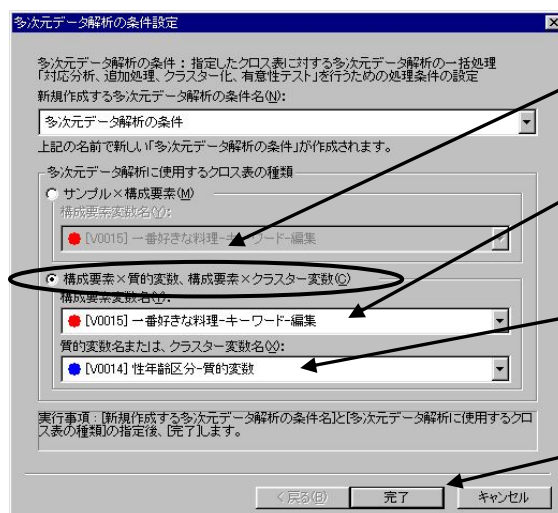
操 作 説 明

(1) 多次元データ解析を行う



① 「プロジェクトパネル」にある[多次元データ解析]―[多次元データ解析の条件設定]をダブルクリックすると、次の(2)の画面が表示されます。

(2) 多次元データ解析に用いる構成要素変数と質的変数を指定する



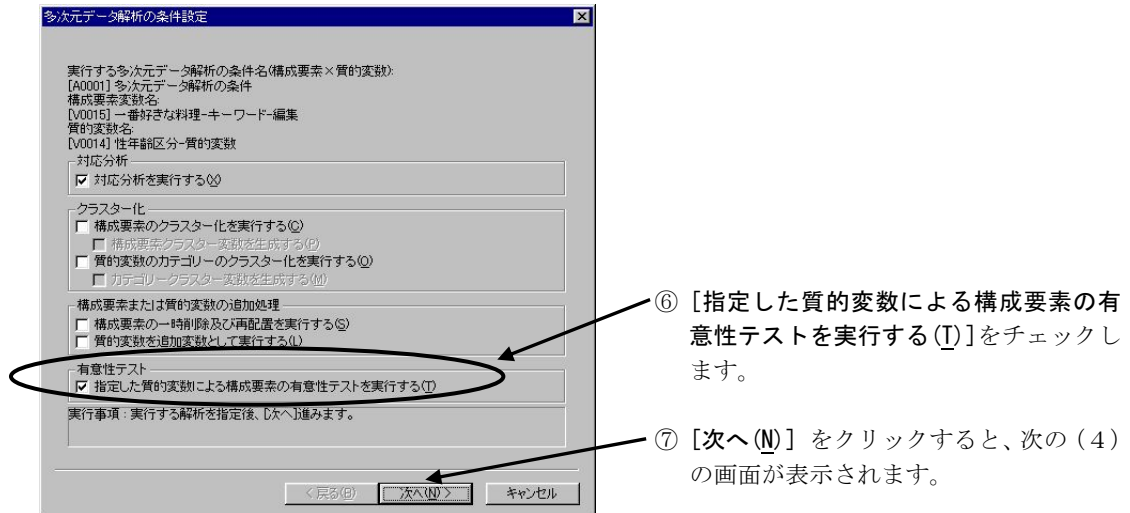
② [構成要素×質的変数、構成要素×クラスター変数(C)] をチェックします。

③ [構成要素変数名(Y)]に[[V0015]一番好きな料理―キーワード編集]を指定します。

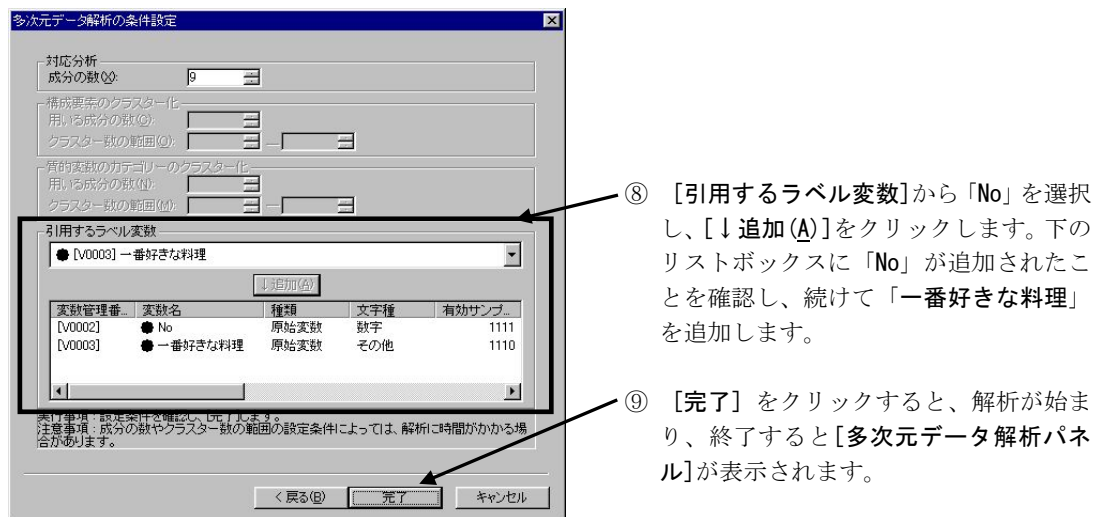
④ [質的変数名または、クラスター変数名(X)]に[[V0014]性年齢区分―質的変数]を指定します。

⑤ [完了] をクリックすると、次の(3)の画面が表示されます。

(3) 実行する多次元データ解析の処理条件を指定する

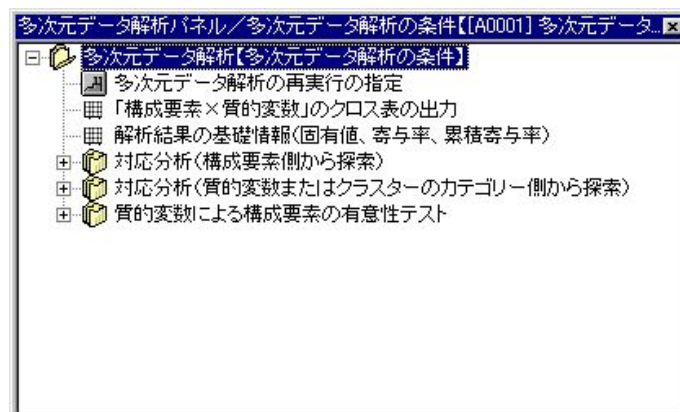


(4) 作成する対応分析の名前を指定する



結果の観察

以上の操作を行いますと、下の図の[多次元データ解析パネル]が表示されます。

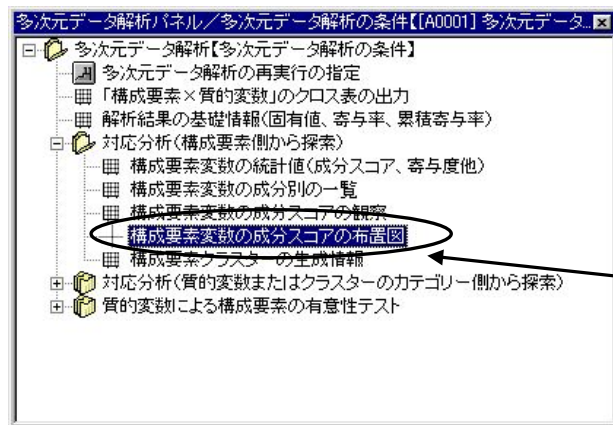


7-2 構成要素の特徴や傾向を図に表わし観察する

布置図を表示することによって、構成要素と質的変数の関連性の有無の程度を観察することができます。

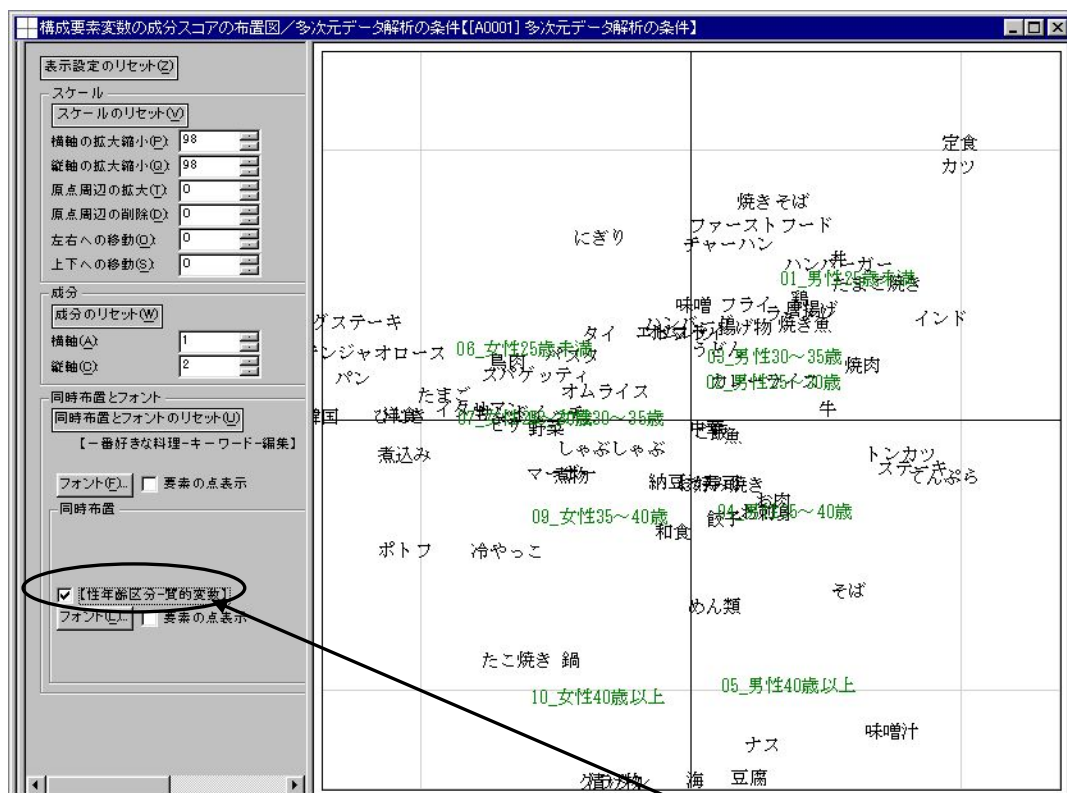
操 作 説 明

(1) 布置図を観察する



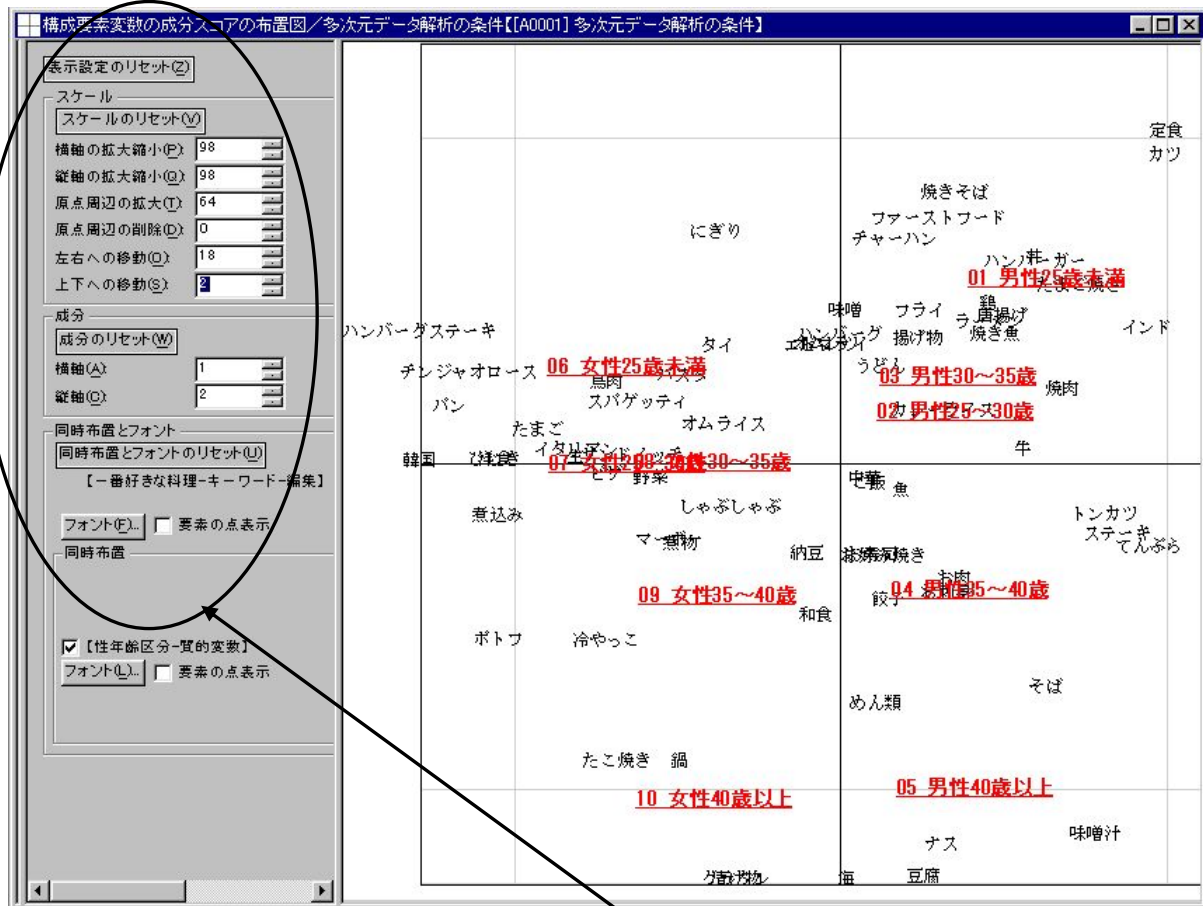
① [対応分析(構成要素側から探索)]－
[構成要素変数の成分スコアの布置図]
をクリックすると、(2)の画面が表示
されます。

(2) 同時布置図を表示する



② [性年齢区分－質的変数]をチ
ェックすると左図のように
「性年齢区分」が同時布置さ
れます。

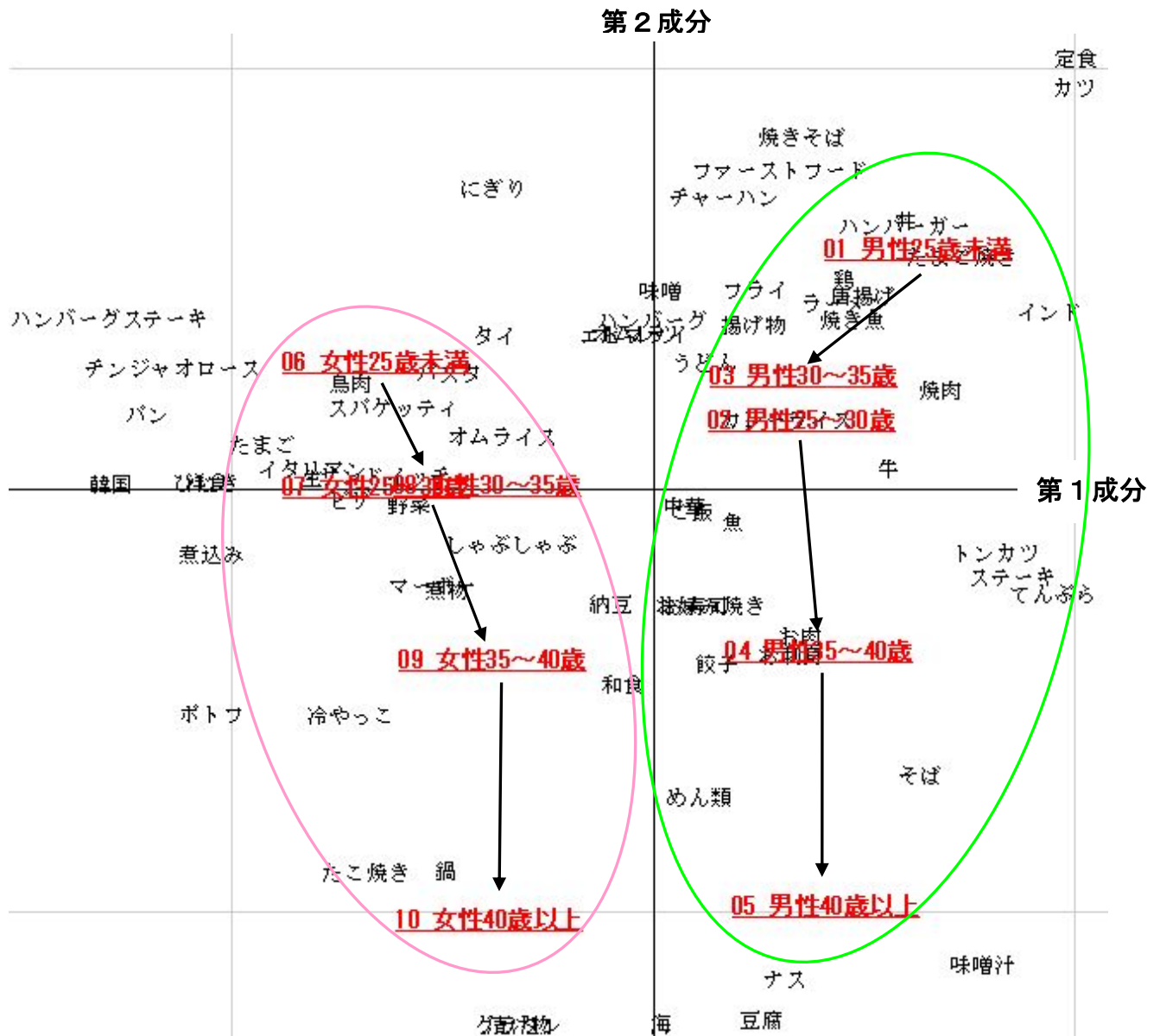
(3) 構成要素変数「一番好きな料理－キーワード－編集」と質的変数「性年齢区分」との関係を見やすくする



- ③ 左のフレームにある[スケール]、[フォント]、[横・縦軸の拡大縮小]、[原点周辺の拡大・削除]等の数値や条件を変えることによって布置図を見やすくします。

結果の観察

多次元データ解析による分析と結果の解釈



ここでは、使用している全構成要素数 1,116 語、異なり構成要素数 71 語（異なり構成要素率 6.4%）、閾値 2 語以上、「性年齢区分」が 10 区分となっています。（編集前：全構成要素数 1,366 語、異なり構成要素数 103 語、異なり構成要素率 7.5%）

特徴として質的変数の布置を観察すると、

1. 性別の差があることが見られる
2. 女性の若年層は「スパゲッティ」「韓国」「イタリアン」等の「洋食系」が多く見られる
3. 男性の若年層には「ハンバーグ」「カレーライス」「焼き魚」等の「食堂・定食系」が多く見られる
4. 男女とも加齢（図の上から下）に伴って好きな料理が移り変わることが見られる

次の単元で出力される「頻度による有意テスト」と共に、性年齢区分カテゴリーの有意に働く構成要素を観察します。

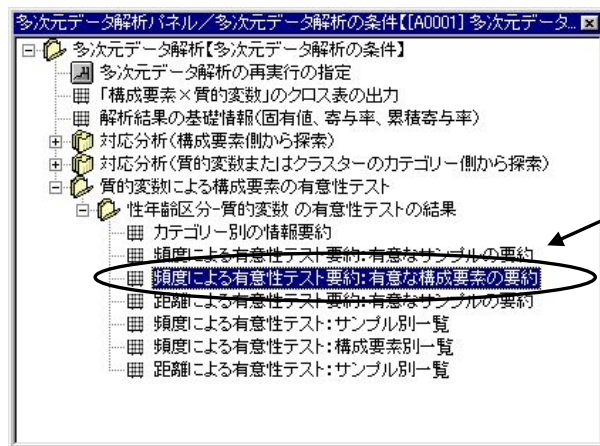
7-3 カテゴリーごとに特徴のある

構成要素の一覧を観察する

構成要素の出現頻度による有意性テスト*1の結果を観察します。「性年齢区分」のカテゴリーごとに特徴のある構成要素の一覧を観察します。

操作説明

(1) 頻度による有意性テストの結果を観察する



① [質的変数による構成要素の有意性テスト] - [性年齢区分 - 質的変数の有意性テストの結果] - [頻度による有意性テスト要約: 有意な構成要素の要約]をダブルクリックすると、次の(2)の画面が表示されます。

(2) 結果の一覧を観察する

上位・下位の構成要素の抽出の条件

抽出数: 15 / 100

検定値: 1 / 10

●表示内容
「構成要素×質的変数」のクロス表の質的変数の各カテゴリーについての構成要素の出現頻度による有意性テストの結果に基づき、有意となった構成要素を質的変数のカテゴリー別に要約一覧表示する。

●解析対象とするクロス表の基礎情報
[ユーザが指定した課題名] 多次元データ解析の条件 [解析対象とするサンプル数] 1030
[解析対象から除外されたサンプル数] 81
[解析対象とする構成要素変数名(表側)] 一番好きな料理-キーワード編集
[解析対象とする異なり構成要素数] 71
[解析対象とする総構成要素数] 1166
[解析対象とする質的変数名(表頭)] 性年齢区分-質的変数
[解析対象とするカテゴリー数] 10

●「頻度による有意性テスト要約: 有意な構成要素の要約」の表示内容

	01_男性25歳未満 サンプル数: 93 異なり構成要素数: 33	02_男性25～30歳 サンプル数: 161 異なり構成要素数: 34	03_男性30～35歳 サンプル数: 170 異なり構成要素数: 45	04_男性35～40歳 サンプル数: 118 異なり構成要素数: 31	05_男性40歳以上 サンプル数: 70 異なり構成要素数: 33	06_女性 サンプル数: 103 異なり構成要素数: 33
上位 1	ラーメン	ステーキ	トンカツ	中華	お肉	スパゲッ
上位 2	丼	そば	魚	お寿司	和食	たまご
上位 3	カツ	唐揚げ	餃子	お肉	豆腐	サンドイ
上位 4	牛	ラーメン	カレーライス	そば	そば	中華
上位 5	うどん	焼肉	フライ	めん類	餃子	納豆
上位 6	焼きそば	カレーライス	味噌	お好み焼き	トンカツ	しゃぶし
上位 7	唐揚げ	焼き魚	焼き魚	カレーライス	納豆	にぎり
上位 8	ファーストフード	中華	ハンバーグ	焼肉	味噌汁	ハンバー
上位 9	定食	鶏	チャーハン	豆腐	グラタン	煮込み
上位 10	チャーハン	カツ	ハンバーガー	お刺身	海	イタリア
上位 11	カレーライス	ハンバーガー	焼きそば	煮物	漬け物	お好み焼
上位 12	てんぷら	たまご焼き	鶏	しゃぶしゃぶ	めん類	ピザ
上位 13	インド	味噌汁	ラーメン	インド	ご飯	野菜
上位 14	鶏	丼	たまご焼き	ナス	ナス	ご飯
上位 15	焼肉	お肉	定食	魚	てんぷら	チンジャ
上位 16	ステーキ	お刺身	にぎり	オムライス	たこ焼き	うどん
上位 17	ハンバーガー	マーボー	エビフライ		鍋	パスタ
上位 18	お肉	揚げ物	オムレツ		牛	冷やっこ
上位 19		ハンバーグ	タイ		冷やっこ	オムライ
上位 20			トマト		ステーキ	パン
上位 21			お刺身		お刺身	ハンバー
上位 22			揚げ物			焼きそば
上位 23			てんぷら			
上位 24			インド			

② [上位・下位の構成要素の抽出の条件] - [検定値]を選択し、エディットボックスに「1」を入力します。

③ 解釈に必要な情報が表示されます。

結果の観察

ここで[検定値]を1にし、性別・年代のカテゴリーごとに有意に働く構成要素と、逆にそのカテゴリーに寄与しない構成要素を表にしてみました。

表 4. 「性年齢区分」の構成要素による有意性テスト

性年齢区分		男性 25 歳未満	男性 25～30 歳未満	男性 30～35 歳未満	男性 35～40 歳未満	男性 40 歳以上
カテゴリー内 サンプル数		93	161	170	118	70
カテゴリーに 寄与する 構成要素	1	ラーメン	ステーキ	トンカツ	中華	お肉
	2	丼	そば	魚	お寿司	和食
	3	カツ	唐揚げ	餃子	お肉	豆腐
	4	牛	ラーメン	カレーライス	そば	そば
	5	うどん	焼肉	フライ	めん類	餃子
	6	焼きそば	カレーライス	味噌	お好み焼き	トンカツ
	7	唐揚げ	焼き魚	焼き魚	カレーライス	納豆
	8	ファーストフード	中華	ハンバーグ	焼肉	味噌汁
	9	定食	鶏	チャーハン	豆腐	グラタン
	10	チャーハン	カツ	ハンバーガー	お刺身	海
	11	カレーライス	ハンバーガー	焼きそば	煮物	漬け物
	12	てんぷら	たまご焼き	鶏	しゃぶしゃぶ	めん類
	13	インド	味噌汁	ラーメン	インド	ご飯
	14	鶏	丼	たまご焼き	ナス	ナス
	15	焼肉	お肉	定食	魚	てんぷら
カテゴリーに 寄与しない 構成要素	15			うどん		焼肉
	14			パスタ		鶏
	13		パン	ステーキ	ハンバーグ	中華
	12	餃子	うどん	たまご	納豆	野菜
	11	お好み焼き	パスタ	野菜	チャーハン	煮物
	10	ピザ	焼きそば	和食	ピザ	唐揚げ
	9	スパゲッティ	たまご	お肉	鶏	焼き魚
	8	煮物	野菜	そば	ご飯	ハンバーグ
	7	野菜	お寿司	めん類	和食	イタリアン
	6	中華	サンドイッチ	中華	井	カレーライス
	5	お寿司	納豆	豆腐	野菜	ラーメン
	4	パスタ	めん類	井	イタリアン	スパゲッティ
	3	お刺身	豆腐	煮物	パスタ	パスタ
	2	魚	スパゲッティ	スパゲッティ	スパゲッティ	
	1	和食	餃子	イタリアン		
	1	イタリアン				

性年齢区分		女性 25 歳未満	女性 25～30 歳未満	女性 30～35 歳未満	女性 35～40 歳未満	女性 40 歳以上
カテゴリー内 サンプル数		83	188	108	33	6
カテゴリーに 寄与する 構成要素	1	スパゲッティ	パスタ	生	めん類	和食
	2	たまご	イタリアン	スパゲッティ	ボトフ	豆腐
	3	サンドイッチ	スパゲッティ	お刺身	煮込み	野菜
	4	中華	韓国	魚	スパゲッティ	お寿司
	5	納豆	パン	煮物	マーボー	スパゲッティ
	6	しゃぶしゃぶ	たこ焼き	ひじき	和食	
	7	にぎり	鳥肉	洋食	お刺身	
	8	ハンバーグステーキ	ピザ	海	サンドイッチ	
	9	煮込み	煮物	エビフライ	イタリアン	
	10	イタリアン	野菜	オムレツ	野菜	
	11	お好み焼き	たまご	トマト	鶏	
	12	ピザ	ハンバーグステーキ	ファーストフード	豆腐	
	13	野菜	グラタン	野菜	煮物	
	14	ご飯	漬け物	納豆	唐揚げ	
	15	チンジャオロース	タイ	パスタ		
カテゴリーに 寄与しない 構成要素	15	豆腐	ステーキ			
	14	煮物	ハンバーガー			
	13	唐揚げ	カツ	お寿司		
	12	お肉	めん類	チャーハン		
	11	ラーメン	丼	お好み焼き		
	10	カレーライス	納豆	餃子		
	9	そば	そば	焼肉		
	8	めん類	焼き魚	豆腐		
	7	焼き魚	鶏	カレーライス		
	6	お刺身	中華	唐揚げ		
	5	餃子	ラーメン	ハンバーグ	トンカツ	
	4	トンカツ	お肉	そば	魚	
	3	お寿司	お刺身	お肉	パスタ	
	2	魚	カレーライス	トンカツ	ラーメン	
	1	和食	トンカツ	ラーメン	中華	

7. 分析を行い、結果を観察する [構成要素×質的変数]

性年齢区分のカテゴリーに有意に働く構成要素と、逆にそのカテゴリーにあまり寄与しない構成要素をそれぞれ 15 語、要約しました（表 4）。あるカテゴリーに寄与する構成要素の程度を測る検定値、構成要素数等の数値は省略してあります。数値の詳細は[多次元データ解析パネル]－[質的変数による構成要素の有意性テスト]－[性年齢区分－質的変数の有意性テストの結果]－[頻度による有意性テスト：構成要素別一覧]を参照してください。

構成要素の編集前の結果と比較すると「すし」「鮭」等、カテゴリーに寄与する、寄与しない両方に出現していた構成要素を置換処理したことにより、統制語となった「お寿司」で統一され分析結果のゆれが少なくなったといえます。他に、「全般」「特」等、分析に必要としない構成要素が削除編集されたことによって、分析結果が読み取りやすくなりました。

編集処理をすることで変化のあった構成要素をあげると、「そば」は男性中高年層の寄与する構成要素の上位になっていることが見られ、逆に置換処理を行ったことで男性の若年層の上位であった「鶏」「鳥」は順位が下がっていることを見るができます。

「ラーメン」や「スパゲッティ」など編集処理前にも特徴のあった構成要素は大きな変化はあまりなく、「ラーメン」は男性の若年層（25 歳未満、25～30 歳未満）では上位に位置しますが、年代が上がるごとに下がっていることが見られ、男性 40 歳以上及び、女性全般では寄与しないということを見るができます。

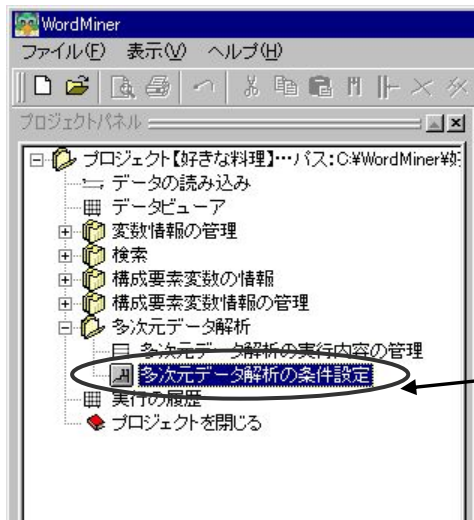
8-1 多次元データ解析を用いて

構成要素とサンプルの関連性を観察する

ある構成要素とサンプルとの関連性を観察するために、多次元データ解析の機能を用いて分析を行います。

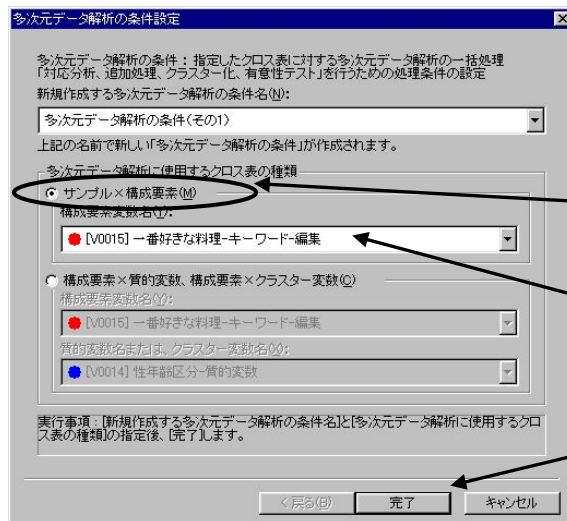
操 作 説 明

(1) 多次元データ解析を行う



① 「プロジェクトパネル」にある[多次元データ解析]－[多次元データ解析の条件設定]をダブルクリックすると、次の(2)の画面が表示されます。

(2) 多次元データ解析に用いる構成要素変数を指定する

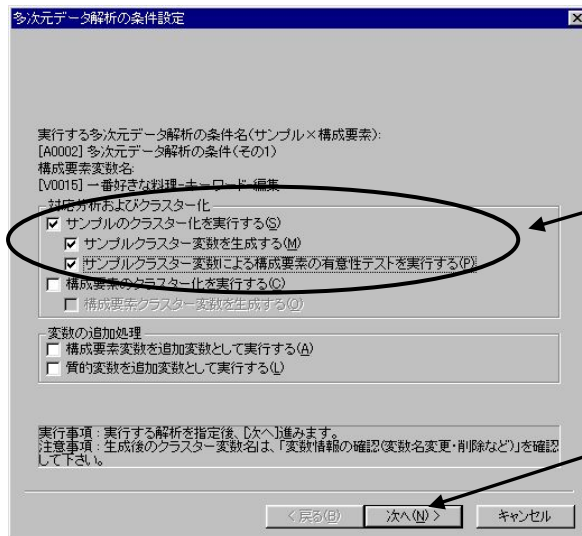


② 「サンプル×構成要素(M)」をチェックします。

③ [構成要素変数名(Y)]に[[V0015]一番好きな料理-キーワード-編集]を指定します。

④ [完了]をクリックすると、次の(3)の画面が表示されます。

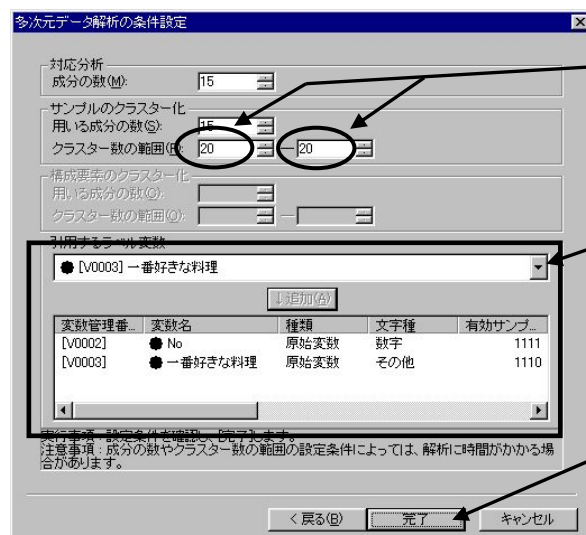
(3) 実行する多次元データ解析を指定する



⑤ [サンプルのクラスター化を実行する(S)]と[サンプルクラスター変数生成する(M)]、「サンプルクラスター変数による構成要素の有意性テストを実行する(P)」の3つにチェックを入れます。

⑥ [次へ(N)]をクリックすると、次の(4)の画面が表示されます

(4) 作成する対応分析の名前を指定する



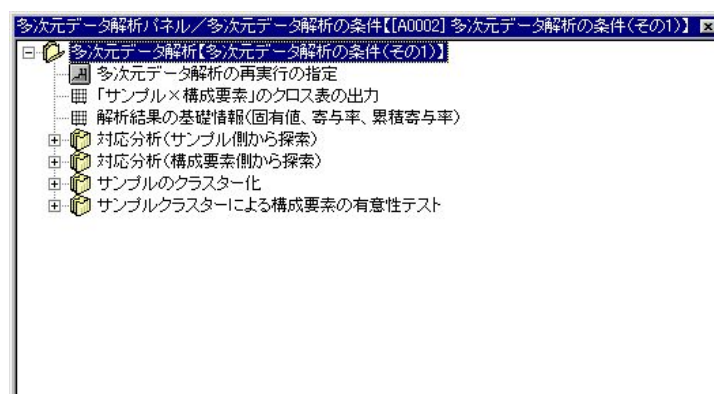
⑦ [クラスター数の範囲(P)]の両方のエディットボックスに「20」を入力します。(20群に分けます)

⑧ [引用するラベル変数]から「No」を選択し、[↓追加(A)]をクリックします。下のリストボックスに「No」が追加されたことを確認し、続けて「一番好きな料理」を追加します。

⑨ [完了]をクリックすると、解析が始まり、終了すると[多次元データ解析パネル]が表示されます。

結果の観察

以上の操作を行いますと、下の図のように[多次元データ解析パネル]が別ウィンドウで表示されます。

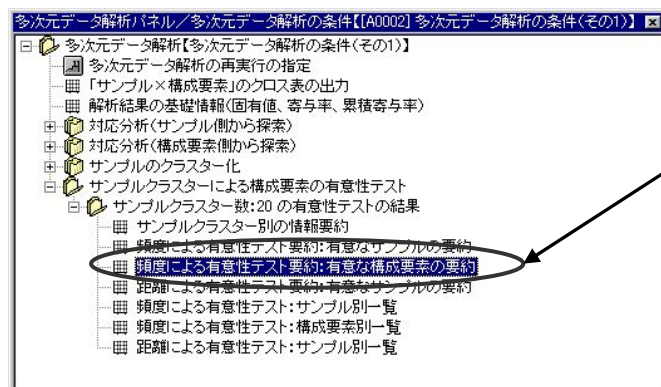


8-2 サンプルの類型化を行い、回答群の観察をする

クラスター化（自動分類）を行った結果の観察を行います。まず、サンプルの類型化（クラスター生成）を行い、どのような回答群があるのかを観察します。まず初めに各クラスターに特徴的な構成要素の観察をします。

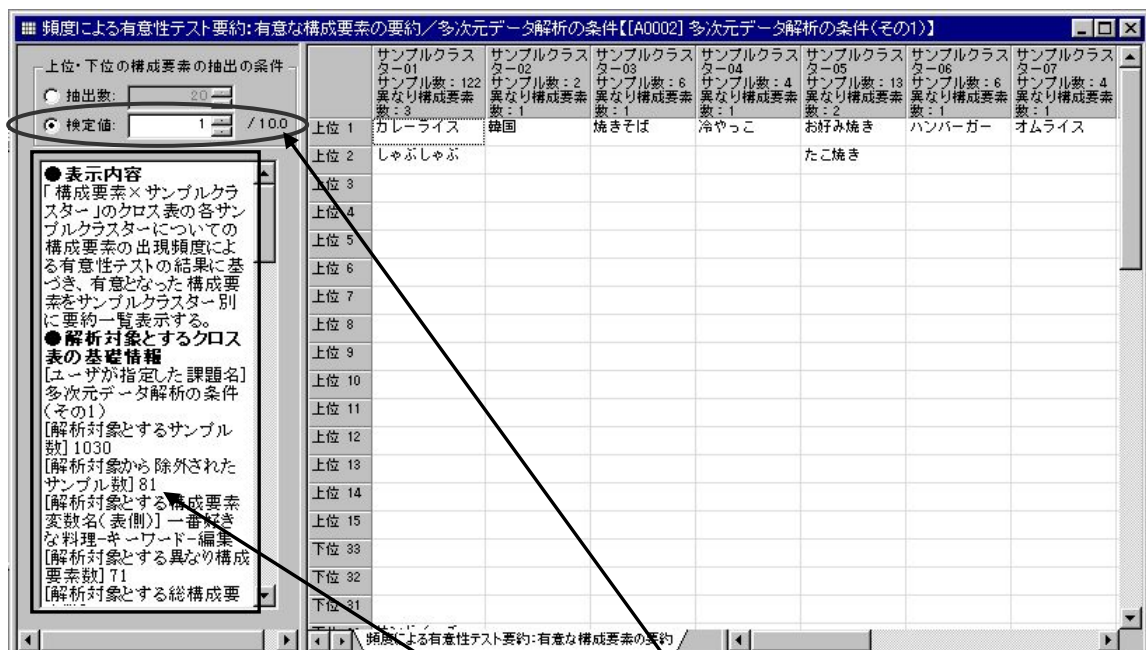
操作説明

(1) 各クラスターに有意な単語の一覧を表示する



① [サンプルクラスターによる構成要素の有意性テスト] - [サンプルクラスター数: 20 の有意性テストの結果] - [頻度による有意性テスト要約: 有意な構成要素の要約] をダブルクリックすると、(2) の画面が表示されます。

(2) 結果の一覧を観察する



② [上位・下位の構成要素の抽出の条件] - [検定値] にチェックを入れエディットボックスに「1」を入力します。

③ 解釈に必要な情報が表示されます。

結果の観察

[上位・下位の構成要素の抽出の条件]－[検定値]に「1」を入力し、クラスターごとに有意に働く構成要素と、逆にそのクラスターに寄与しない構成要素 15 語を表にしてみました。構成要素が 15 語に満たないクラスターは有意検定の基準に満たなかったということです。

表 5. クラスターごとの構成要素による有意性テスト

クラスター		クラスター 1	クラスター 2	クラスター 3	クラスター 4	クラスター 5	クラスター 6	クラスター 7	クラスター 8	クラスター 9	クラスター 10
カテゴリ内 サンプル数		122	2	6	4	13	6	4	7	8	21
カ テ ゴ リ ー に 寄 与 す る 構 成 要 素	1	カレーライス	韓国	焼きそば	冷やっこ	お好み焼き たこ焼き	ハンバーガー	オムライス	サンドイッチ	たまご タイ	唐揚げ 鶏 鳥肉 揚げ物
	2	しゃぶしゃぶ									
	3										
	4										
	5										
	6										
	7										
	8										
	9										
	10										
	11										
	12										
	13										
	14										
	15										
カ テ ゴ リ ー に 寄 与 し な い 構 成 要 素	15	焼き魚									
	14	お刺身									
	13	餃子									
	12	トンカツ									
	11	ハンバーグ									
	10	魚									
	9	野菜									
	8	中華									
	7	お寿司									
	6	イタリアン									
	5	和食									
	4	ラーメン									
	3	スパゲッティ									
	2	お肉									
	1	パスタ									
						和食 ラーメン スパゲッティ お肉 パスタ カレーライス					中華 お寿司 イタリアン 和食 ラーメン スパゲッティ お肉 パスタ カレーライス

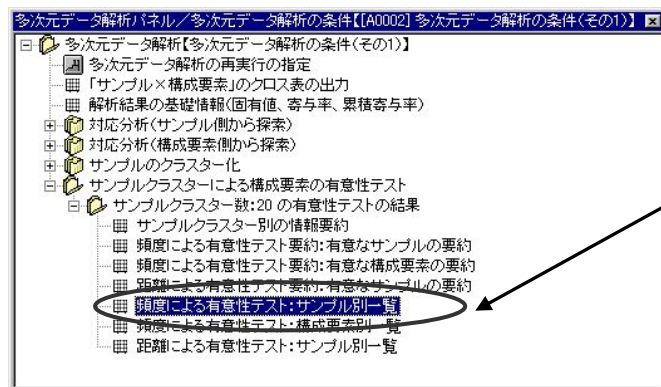
クラスター	クラスター11	クラスター12	クラスター13	クラスター14	クラスター15	クラスター16	クラスター17	クラスター18	クラスター19	クラスター20
カテゴリ内サンプル数	130	91	157	113	187	5	26	4	95	29
カテゴリに寄与する構成要素	1 和食 2 魚 3 焼き魚 4 煮物 5 鍋 6 ひじき 7 海 8 漬け物 9 定食 10 揚げ物 11 フライ 12 お刺身 13 生 14 ナス 15 味噌	お寿司 豆腐 ご飯 焼肉 お刺身 納豆 にぎり オムレツ 味噌汁 ナス マーボー 生	ラーメン 中華 餃子 丼 チャーハン カツ 牛 チンジャオロース 味噌 マーボー	スパゲッティ 野菜 てんぷら グラタン トマト	パスタ イタリアン そば うどん めん類 ピザ パン	インド たまご焼き	トンカツ ハンバーグステーキ ファーストフード フライ	エビフライ 洋食	お肉 ステーキ ポトフ	ハンバーグ 煮込み
カテゴリに寄与しない構成要素	15 唐揚げ 14 野菜 13 めん類 12 そば 11 餃子 10 トンカツ 9 ハンバーグ 8 お寿司 7 イタリアン 6 中華 5 ラーメン 4 スパゲッティ 3 お肉 2 パスタ 1 カレーライス	めん類 そば 餃子 トンカツ 魚 ハンバーグ 野菜 和食 中華 イタリアン お肉 ラーメン スパゲッティ パスタ カレーライス	唐揚げ 焼き魚 そば お刺身 餃子 トンカツ 魚 ハンバーグ 野菜 お寿司 イタリアン 和食 スパゲッティ お肉 パスタ カレーライス	焼き魚 そば お刺身 餃子 トンカツ パスタ 中華 魚 ハンバーグ お寿司 イタリアン 和食 ラーメン お肉 パスタ カレーライス	唐揚げ 焼き魚 野菜 お刺身 餃子 トンカツ お寿司 魚 ハンバーグ 中華 ラーメン スパゲッティ 和食 お肉 パスタ カレーライス		中華 お寿司 イタリアン 和食 ラーメン スパゲッティ お肉 パスタ カレーライス		唐揚げ そば めん類 焼き魚 餃子 魚 ハンバーグ 中華 お寿司 イタリアン 和食 ラーメン スパゲッティ パスタ カレーライス	中華 お寿司 イタリアン 和食 ラーメン スパゲッティ お肉 パスタ カレーライス

8. 分析を行い、結果を観察する [構成要素×サンプル]

[表5. クラスターごとの構成要素による有意性テスト]を観察しながら、各クラスター内のサンプルがどのように分類されているか観察します。一例として、クラスター5に含まれる13名、クラスター10に含まれる21名のサンプルを取り上げ表にします。

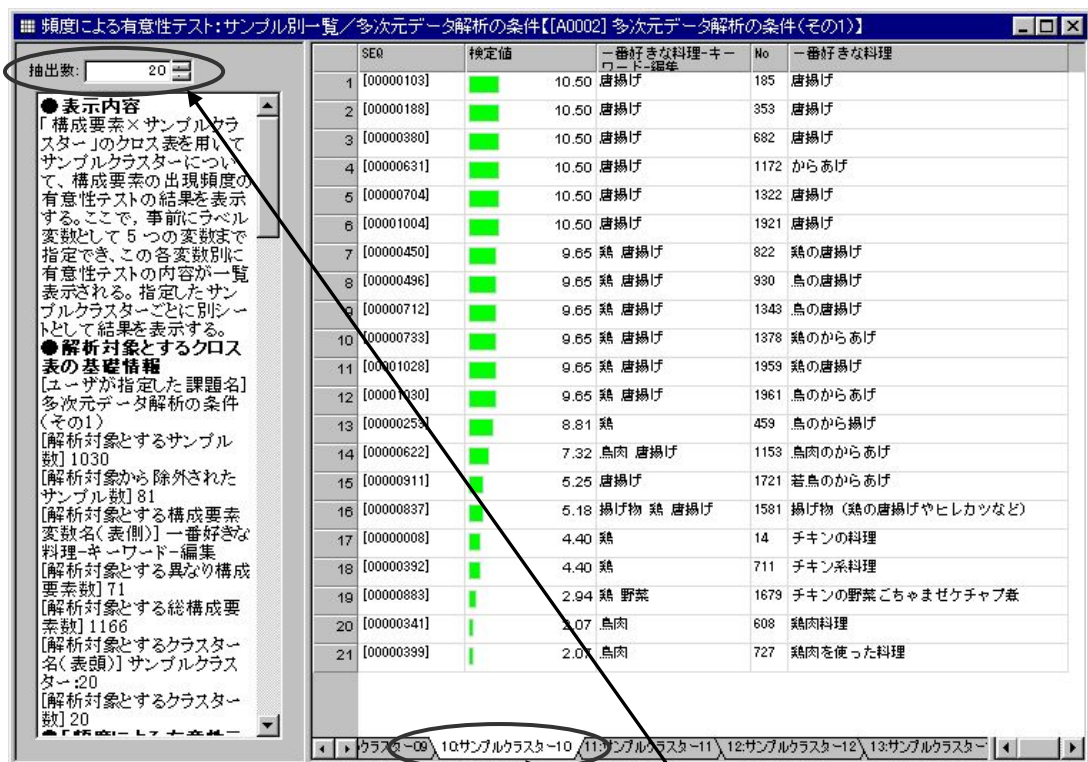
操作説明

(1) 各クラスターに有意なサンプルの一覧を表示する



① [サンプルクラスターによる構成要素の有意性テスト] - [サンプルクラスター数:20の有意性テストの結果] - [頻度による有意性テスト: サンプル別一覧]をダブルクリックすると、(2)の画面が表示されます。

(2) 結果の一覧を観察する



② [上位・下位の構成要素の抽出の条件] - [抽出数]に「20」を入力します。

③ ウィンドウ下のタブをクリックすることで各クラスターの内容を観察することができます。

結果の観察

クラスター5 とクラスター10 のサンプルを全抽出し、どのような回答があるのか観察します（表6）。

表6. クラスター5、10 のサンプル

クラスター5			クラスター10		
番号	検定値	本文	番号	検定値	本文
1	9.66	お好み焼き	1	10.50	唐揚げ
2	9.66	お好み焼き	2	10.50	唐揚げ
3	9.66	お好み焼き	3	10.50	唐揚げ
4	9.66	お好み焼き	4	10.50	からあげ
5	9.66	お好み焼	5	10.50	唐揚げ
6	9.66	お好み焼き	6	10.50	唐揚げ
7	9.66	お好み焼き	7	9.65	鶏の唐揚げ
8	9.66	お好み焼き	8	9.65	鳥の唐揚げ
9	9.66	お好み焼き	9	9.65	鳥の唐揚げ
10	9.66	お好み焼	10	9.65	鶏のからあげ
11	4.74	たこ焼き	11	9.65	鶏の唐揚げ
12	4.74	たこ焼き	12	9.65	鳥のからあげ
13	4.74	たこ焼き	13	8.81	鳥のから揚げ
			14	7.32	鳥肉のからあげ
			15	5.25	若鳥のからあげ
			16	5.18	揚げ物（鶏の唐揚げやヒレカツなど）
			17	4.40	チキンの料理
			18	4.40	チキン系料理
			19	2.94	チキンの野菜ごちやまぜけチャブ煮
			20	2.07	鶏肉料理
			21	2.07	鶏肉を使った料理

以上の操作で、クラスター5 は「お好み焼き」「たこ焼き」が一番好きであると回答したグループであり、クラスター10 は「鶏の唐揚げ」「チキン」などの「鶏肉系」が一番好きであると回答したグループであるということが分かります。

構成要素の編集前のクラスター化の結果を観察すると「唐揚げ」のクラスター10 は 16 名です。「鶏」と「鳥」、「チキン」を置換編集することによって「唐揚げ」のクラスターは置換前よりも、よりよくまとまっているということがいえます。

他に、編集前のクラスター14 は 674 名の群になっていましたが、構成要素の編集を行うことによって極端に大きなクラスターは出現しなくなったことがわかります。しかし、[表5. クラスターごとの構成要素による有意性テスト]等を観察することで細かな分析を行うことができます。

また、ここではクラスター数を 20 と指定しましたが、ユーザの指定によりクラスター数を増減することができます。

■ まとめ

この事例集で *WordMiner*®によるテキスト型データ分析の一連の流れ及び、特有の用語の意味説明をしました。しかし、これですべての機能を説明したわけではありません。ここではビジュアル面の機能を主に紹介しましたが、データを出力するにあたり用いている数値データを表示、エクスポートすることや、他にもデータの探査を行うことのできる機能を *WordMiner*®は持っています。

この事例では構成要素に「キーワード」抽出結果を使用して分析を行いましたが、「分かち書き」処理結果を構成要素として用いることで、更に細やかなテキスト型データ解析を行うことができます。

次の事例集では、用いるデータの背景から特性・傾向を調べ、分かち書きとキーワードを抽出し得られる情報の比較を行うなど詳細なテキスト型データの解析を行います。

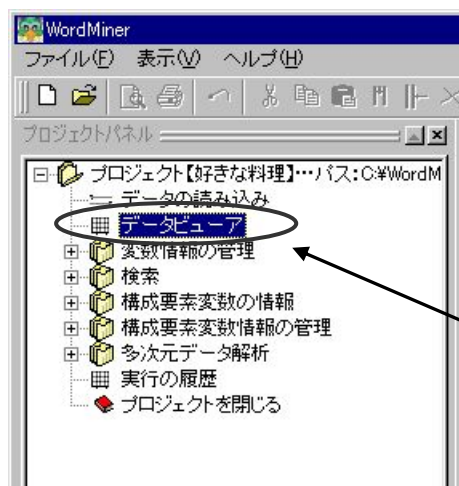
付 録

1 データビューア

データビューアを用いてファイルの読み込みが正常に行われたか、または新たに作成した変数がどのように作成されたか、データの内容を表形式で確認することができます。また、[変数情報の管理]と同様に変数一覧情報を一括して閲覧することができます。

操 作 説 明

(1) データビューアを表示する



① 「プロジェクトパネル」にある
[データビューア]をダブルクリックす
ると[データビューア]画面が表示され
ます。

結果の観察

左側のフレームの[変数名]にチェックが入っている変数が右側のフレームに表示されます。チェックをはずすとその変数は表示されません。

変数管理		データビューア							
変数番号	変数名	No	一番好きな料理	次に好きな料理	性別	年代	性別×年代	一番好きな料理-分かち書き	次に好きな料理-分かち書き
[V0001]	SEQ	1	カレー	ラーメン	1_男性	4_35歳以上40歳未満	04_男性35～40歳	カレー	ラーメン
[V0002]	No	2	ラーメン	スパゲッティ	1_男性	3_30歳以上35歳未満	03_男性30～35歳	ラーメン	スパゲッティ
[V0003]	一番好きな料理	3	カレーライス	焼肉	1_男性	4_35歳以上40歳未満	04_男性35～40歳	カレーライス	焼肉
[V0004]	次に好きな料理	4	ラーメン	炒飯	1_男性	1_25歳未満	01_男性25歳未満	ラーメン	炒飯
[V0005]	性別	5	寿司	カレー	1_男性	3_30歳以上35歳未満	03_男性30～35歳	寿司	カレー
[V0006]	年代	6	サラダ	カレーライス	2_女性	3_30歳以上35歳未満	08_女性30～35歳	サラダ	カレーライス
[V0007]	性別×年代	7	カツ	カレー	1_男性	2_25歳以上30歳未満	02_男性25～30歳	カツ	カレー
[V0008]	一番好きな料理-分かち書き	8	チキンの料理	魚料理	2_女性	4_35歳以上40歳未満	09_女性35～40歳	チキンの料理	魚料理
[V0009]	次に好きな料理-分かち書き	9	スパゲッティ	お蕎麦	2_女性	2_25歳以上30歳未満	07_女性25～30歳	スパゲッティ	お蕎麦
[V0010]	一番好きな料理-キーワード	10	カレーライス	ハンバーグ	1_男性	1_25歳未満	01_男性25歳未満	カレーライス	ハンバーグ
[V0011]	次に好きな料理-キーワード	11	お刺身	魚料理	1_男性	3_30歳以上35歳未満	03_男性30～35歳	お刺身	魚料理
[V0012]	性別-質的変数	12	納豆	焼き肉	1_男性	3_30歳以上35歳未満	03_男性30～35歳	納豆	焼き肉
[V0013]	年代-質的変数	13	カレーライス	ラーメン	1_男性	5_40歳以上	05_男性40歳以上	カレーライス	ラーメン
[V0014]	性別×年代-質的変数	14	焼き肉	納豆	1_男性	1_25歳未満	01_男性25歳未満	焼き肉	納豆
[V0015]	一番好きな料理-キーワード-編集	15	スパゲッティ	カレーライス	1_男性	3_30歳以上35歳未満	03_男性30～35歳	スパゲッティ	カレーライス
[V0016]	構成要素クラスター-20	16	ご飯と味噌汁	漬物と納豆	1_男性	2_25歳以上30歳未満	02_男性25～30歳	ご飯と味噌汁	漬物と納豆
		17	うどん	野菜炒めの定食	1_男性	1_25歳未満	01_男性25歳未満	うどん	野菜炒め
		18	パスタ	餃子	2_女性	3_30歳以上35歳未満	08_女性30～35歳	パスタ	餃子
		19	ラーメン	カレーライス	1_男性	2_25歳以上30歳未満	02_男性25～30歳	ラーメン	カレーライス
		20	ラーメン	餃子	1_男性	4_35歳以上40歳未満	04_男性35～40歳	ラーメン	餃子
		21	そば、うどん、パスタ等	魚料理(和食)	1_男性	4_35歳以上40歳未満	04_男性35～40歳	そば、うどん、パスタ等	魚料理(和食)
		22	和食	中華	1_男性	2_25歳以上30歳未満	02_男性25～30歳	和食	中華
		23	海老チャーハン	刺し身	1_男性	1_25歳未満	01_男性25歳未満	海老チャーハン	刺し身
		24	中華料理	和食	1_男性	4_35歳以上40歳未満	04_男性35～40歳	中華料理	和食
		25	ラーメン	そば	2_女性	3_30歳以上35歳未満	08_女性30～35歳	ラーメン	そば
		26	和食	エスニック	2_女性	3_30歳以上35歳未満	08_女性30～35歳	和食	エスニック

ヒント

1. データのファイル出力について

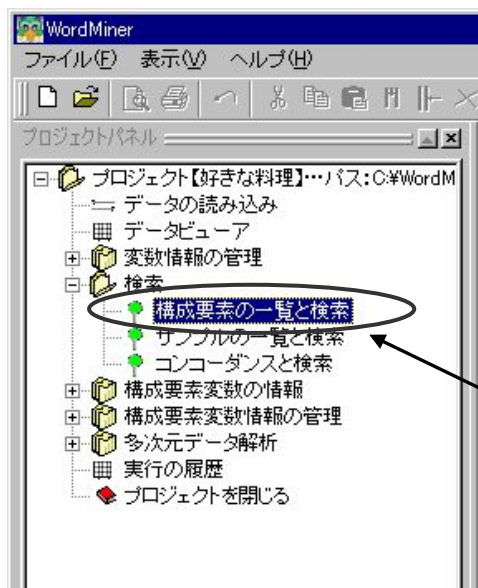
右側のフレームでコンテキストメニュー（右クリック）[すべてをエクスポート]を選択することにより、データをファイルに出力（備考4：エクスポート）することができます。

2 構成要素の検索

構成要素変数を作成した後、構成要素の一覧を表示することができます。構成要素の検索、頻度情報等を表示するために用います。

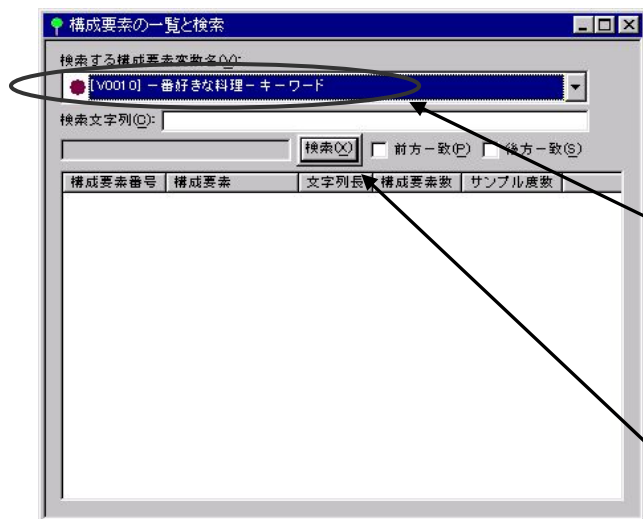
操 作 説 明

- (1) [構成要素の検索と一覧]を表示する



- ① 「プロジェクトパネル」にある
[検索]－[構成要素の一覧と検索]をダブルクリックすると、次の(2)の画面が表示されます。

- (2) 全構成要素の一覧を表示する



- ② [抽出する構成要素変数名(Y)]－
[[V0010] 一番好きな料理 - キーワード]を指定します。
- ③ [検索文字列(C)]には何も入力しません。
- ④ [検索(X)]をクリックします。

結果の観察

選択した構成要素変数の全構成要素が表示されます。

構成要素番号	構成要素	文字列長	構成要素数	サンプル度数
1	First	5	1	1
2	Mexican	7	1	1
3	pizza	5	1	1
4	あげ	2	5	5
5	あっさり	4	1	1
6	いくら	3	1	1
7	うどん	3	10	10
8	うどん等	4	2	2
9	おし	3	1	1
10	おす	2	1	1
11	お魚	2	1	1
12	お好み焼	4	2	2
13	お好み焼き	5	8	8
14	お刺し	3	3	3
15	お刺身	3	1	1
16	お寿司	3	10	10
17	お造り	3	1	1
18	お漬物	3	1	1
19	お豆腐	3	1	1

ヒント

1. 構成要素の表示

今回のように[検索文字列(C)]に何も入力せずに[検索(X)]をクリックすると選択した構成要素変数の全構成要素が表示されます。

[検索文字列(C)]に文字列を入力して[検索(X)]をクリックすることで入力した文字列に一致する構成要素が表示されます。

2. ソートについて

リスト部分のヘッダー（「構成要素番号」「文字列長」「構成要素数」等）をクリックすることで、出力されている内容をソートすることが可能です。（右図は「構成要素数」でソートした例）

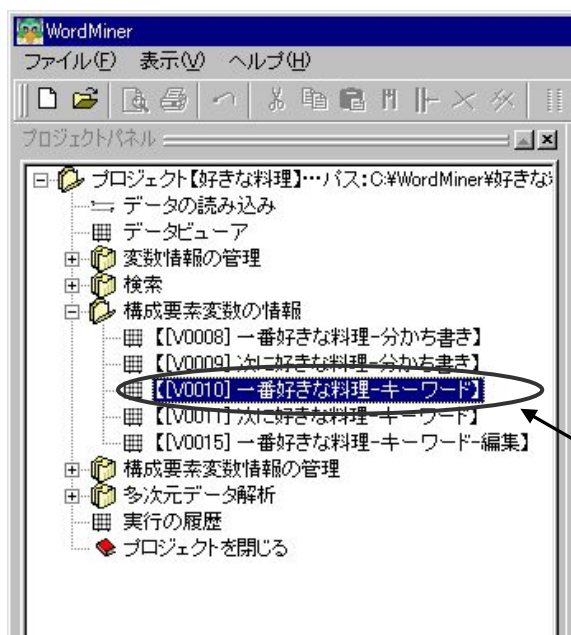
構成要素番号	構成要素	文字列長	構成要素数	サンプル度数
270	料理	2	190	190
134	パスタ	3	98	98
243	肉	1	89	89
156	ラーメン	4	78	78
81	カレーライス	6	72	72
276	和食	2	56	56
80	カレー	3	50	50
220	中華	2	42	42
98	スパゲッティ	6	38	38
62	イタリア	4	30	30
128	ハンバーグ	5	27	27
173	魚	1	27	27
100	スパゲティ	5	23	23
118	トンカツ	4	21	21
195	寿司	2	21	21
200	焼き魚	3	18	18
260	麺類	2	18	18
63	イタリアン	5	17	17
280	餃子	2	17	17

3 構成要素変数の情報

構成要素変数を作成した後、構成要素の頻度分布、異なり構成要素数、異なり構成要素率を観察するために用います。

操 作 説 明

- (1) 構成要素の検索と一覧を表示する



- ① 「プロジェクトパネル」にある
[構成要素変数の情報]－[【V0010】一番
好きな料理－キーワード]をダブルク
リックすると[構成要素変数の情報]の
画面が表示されます。

結果の観察

表示されたウインドウの下部に[構成要素変数の情報 1] [構成要素変数の情報 2]の2つのタブがありますので、これをクリックすることで表示をかえることができます。

閾値の水準	その頻度以上の総構成要素数	その頻度以上の異なり構成要素数	その頻度以上の異なり構成要素率
1 頻度 1	1541	278	18.0
2 頻度 2	1366	103	7.5
3 頻度 3	1292	66	5.1
4 頻度 4	1265	57	4.5
5 頻度 5	1245	52	4.2
6 頻度 6	1230	49	4.0
7 頻度 7	1194	43	3.6
8 頻度 8	1180	41	3.5
9 頻度 9	1148	37	3.2
10 頻度 10	1130	35	3.1
11 頻度 11	1100	32	2.9
12 頻度 12	1067	29	2.7
13 頻度 13	1031	26	2.5
14 頻度 14	1005	24	2.4
15 頻度 15	963	21	2.2
16 頻度 16	948	20	2.1
17 頻度 17	932	19	2.0
18 頻度 18	898	17	1.9
19 頻度 21	862	15	1.7
20 頻度 23	820	13	1.6
21 頻度 27	797	12	1.5
22 頻度 30	743	10	1.3

閾値の水準	各頻度区分内の構成要素数	各頻度区分内の異なり構成要素数
1 頻度 1	175	175
2 頻度 2	74	37
3 頻度 3	27	9
4 頻度 4	20	5
5 頻度 5	15	3
6 頻度 6	36	6
7 頻度 7	14	2
8 頻度 8	32	4
9 頻度 9	18	2
10 頻度 10	30	3
11 頻度 11	33	3
12 頻度 12	36	3
13 頻度 13	26	2
14 頻度 14	42	3
15 頻度 15	15	1
16 頻度 16	16	1
17 頻度 17	34	2
18 頻度 18	36	2
19 頻度 21	42	2
20 頻度 23	23	1
21 頻度 27	54	2
22 頻度 30	30	1

用語説明

1. その頻度以上の総構成要素数

設定した頻度(閾値)以上の構成要素総数のことです。例えば、頻度 5 の欄にある数値(1,245)は、閾値 5 (出現頻度 5) 以上の構成要素の総数をいいます。

2. その頻度以上の異なり構成要素数

指定した頻度(閾値)以上の構成要素の種類数のことです。

例えば「カレー」の出現回数が 3 回でも 1 つとカウントしています。

3. その頻度以上の異なり構成要素率

構成要素数に対する総異なり構成要素数の占める比率(%)のことです。

「その頻度以上の異なり構成要素率 =

$$\text{その頻度以上の異なり構成要素数} \div \text{その頻度以上の総構成要素数} \times 100$$

4. 各頻度区分内の構成要素数

例えば、頻度 5 の列では頻度が丁度 5 の構成要素が 15 語あるということです。

5. 各頻度区分内の異なり構成要素数

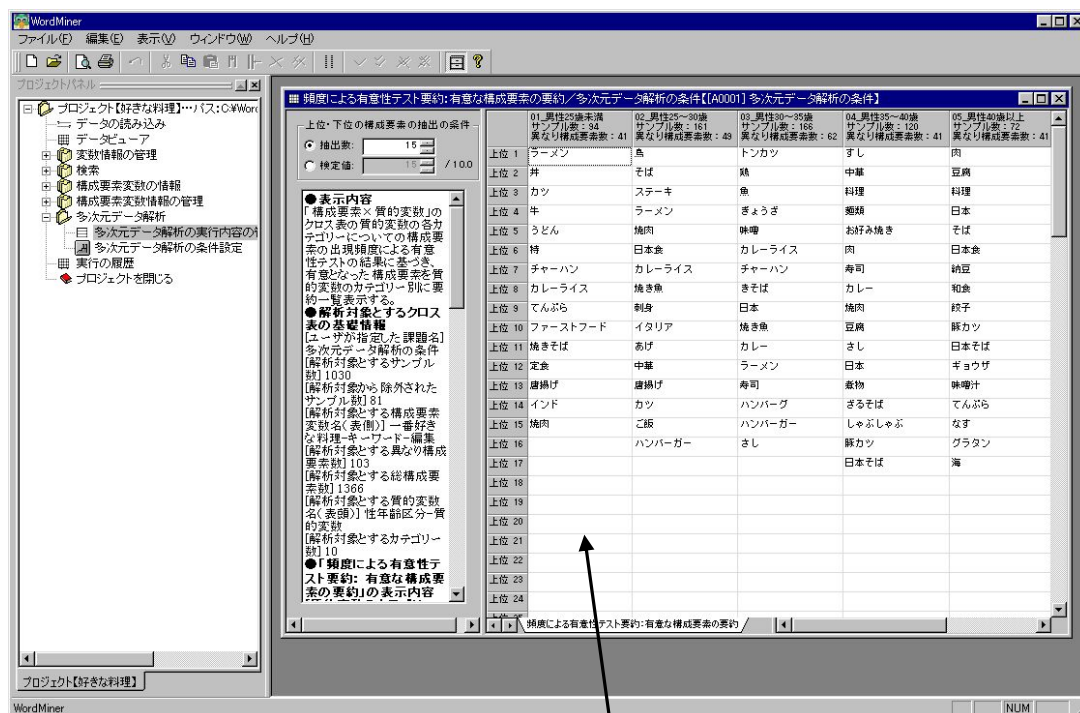
例えば、頻度 5 の列では頻度が丁度 5 の構成要素の種類が 3 語あるということです。

4 エクスポート

画面上に表示された結果をエクスポート（ファイル出力）することができます。エクスポートできる画面はリスト形式（変数情報の確認・構成要素の一覧と検索等）、あるいは表形式（データビューア・多次元データ解析の結果等）の画面です。

操作説明

（１－１）エクスポートする画面を表示する（表形式の例）



① エクスポートする画面を表示させます。（ここでは頻度による有意性テストの要約を表示しています。）

（１－２）エクスポートする



② 表部分を右クリックするとコンテキストメニューが表示されますので、[すべてエクスポート]をクリックします。

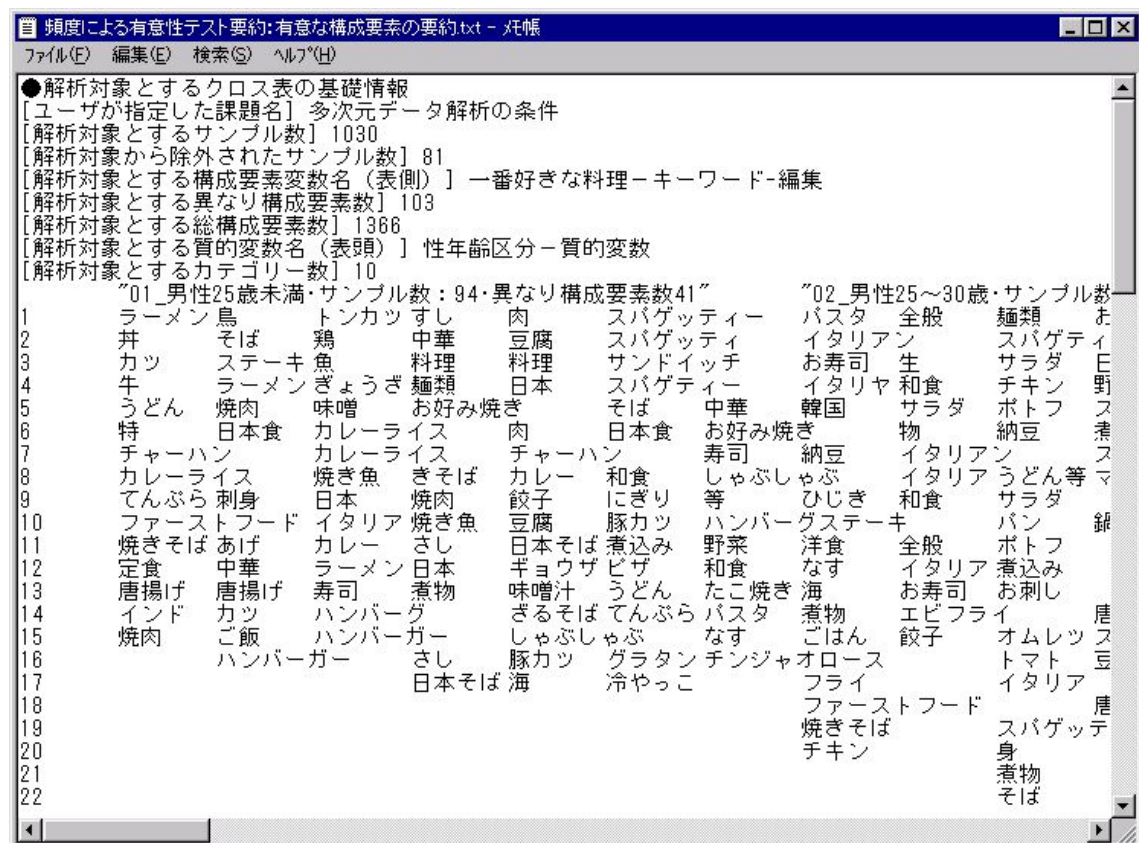
(1-3) ファイルの名前・出力場所とファイルの種類を指定する



- ③ [ファイル名(N)]と[ファイルの種類(I)]を指定して[保存(S)]をクリックすると指定した場所にファイルが作成されます。

結果の観察

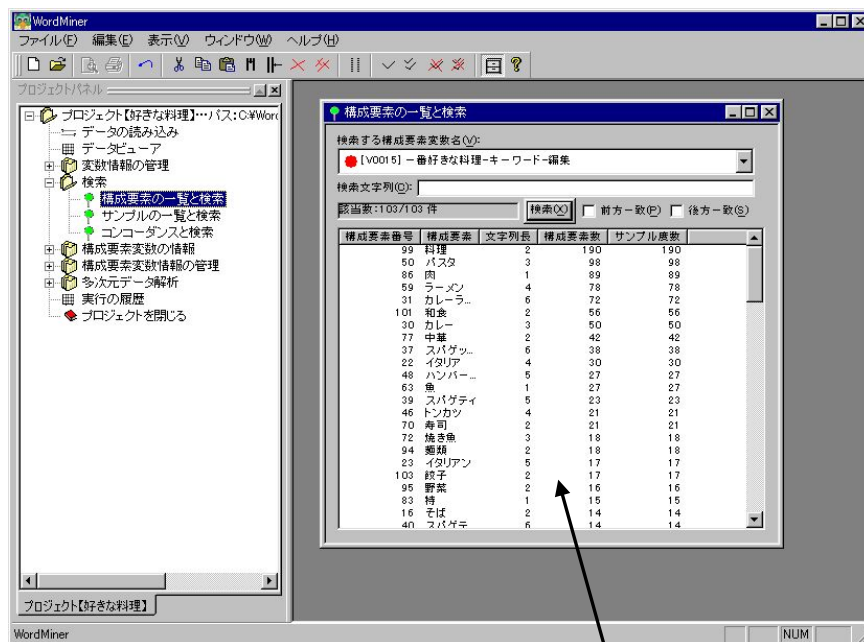
エクスポートしたファイルを開いて正しく出力されたかを確認します。



ファイルにはエクスポートした内容と、そのファイルの解析の処理条件の情報（サンプル数等）が出力されます。

操作説明

(2-1) エクスポートする画面を表示する (リスト形式の例)



① エクスポートする画面を表示させます。(ここでは構成要素の一覧を表示しています。)

(2-2) エクスポートする



② リスト部分を右クリックするとコンテキストメニューが表示されますので、[すべてエクスポート]をクリックします。

(2-3) ファイルの名前・出力場所とファイルの種類を指定する



③ [ファイル名(N)]と[ファイルの種類(I)]を指定して[保存(S)]をクリックすると指定した場所にファイルが作成されます。

結果の観察

エクスポートしたファイルを開いて正しく出力されたかを確認します。



構成要素	構成要素番号	文字列長	構成要素数	サンプル度数
269	料理	2	190	
133	パスタ	3	98	
242	肉	1	89	
155	ラーメン	4	78	
80	カレーライス	6	72	72
274	和食	2	56	
79	カレー	3	50	
219	中華	2	42	
97	スパゲッティ	6	38	38
61	イタリア	4	30	
127	ハンバーグ	5	27	27
172	魚	1	27	
99	スパゲティ	5	23	23
117	トンカツ	4	21	
194	寿司	2	21	
199	焼き魚	3	18	
259	麺類	2	18	
62	イタリアン	5	17	17
278	餃子	2	17	
261	野菜	2	16	
237	特	1	15	
39	そば	2	14	
100	スパゲティー	6	14	14

ファイルにはエクスポートした内容が出力されます。

ヒント

1. エクスポートしたファイル

上図ではメモ帳で開いていますが、テキストエディタや表計算ソフトウェアでファイルを開くとデータの編集が簡単です。

5 分析に使用するデータ

1. 分析に用いるデータの説明

この事例集で仕様するデータは 1999 年に Web 調査により得られたものです(提供: 文部科学省統計数理研究所 大隅昇教授)。分析に使用する設問の他に、多くの設問がありますが、デモンストレーション用として分かりやすいデータとして再編集したものを使用します。(有効回答数 1,111 件)

2. 分析に使用する設問

デモンストレーションで使用するデータの設問一覧です。

設 問	設問形式
あなたが一番好きな料理は何ですか？	: 自由回答
次に好きな料理は何ですか？	: 自由回答
あなたの性別は？	: 属性項目 (シングルアンサー)
あなたの年齢は？	: 属性項目 (シングルアンサー)

3. 分析結果の出力目標とするもの

データに性別と年齢がありますので性別と年齢の変数のカテゴリーと自由回答で得られた構成要素ごとの相関関係及び、カテゴリーごとに有意な構成要素を探索し、カテゴリーや構成要素ごとにどのような関係や特徴があるかを分析します。

4. データを加工する

データを加工して、新しい変数を作成します。また新しいグループを作成します。(サンプルデータは加工済み)

表 題	説 明
一番好きな料理	: 取得したデータをそのまま用いる
次に好きな料理	: 取得したデータをそのまま用いる
性 別	: 取得したデータをそのまま用いる
年 齢	: 年齢を区分 1. 25 歳未満 2. 25 歳以上 30 歳未満 3. 30 歳以上 35 歳未満 4. 35 歳以上 40 歳未満 5. 40 歳以上
性 年 齢 区 分	: 性別と年齢から性年齢区分を作成した 1. 男性 25 歳未満 2. 男性 25 歳以上 30 歳未満 3. 男性 30 歳以上 35 歳未満 4. 男性 35 歳以上 40 歳未満 5. 男性 40 歳以上 6. 女性 25 歳未満 7. 女性 25 歳以上 30 歳未満 8. 女性 30 歳以上 35 歳未満 9. 女性 35 歳以上 40 歳未満 10. 女性 40 歳以上

以上のデータを今回の導入編で使用します。ファイル名は「好きな料理.txt」です。

5. 分析に使用するデータの中身の確認と形式について

メモ帳で【好きな料理.txt】を開くと以下のようなテキストデータが表示されます。

No	一番好きな料理	次に好きな料理	性別	年代	性年齢区分
1	カレーライス	ラーメン	1_男性	4_35歳以上40歳未満	04_男性35~40歳
2	ラーメン	カレー	1_男性	3_30歳以上35歳未満	03_男性30~35歳
5	カレーライス	焼肉	1_男性	4_35歳以上40歳未満	04_男性35~40歳
7	ラーメン	炒飯	1_男性	1_25歳未満	01_男性25歳未満
8	寿司	カレー	1_男性	3_30歳以上35歳未満	03_男性30~35歳
10	サラダ	カレーライス	2_女性	3_30歳以上35歳未満	08_女性30~35歳
12	カツ	カレー	1_男性	2_25歳以上30歳未満	02_男性25~30歳
14	チキンの料理	魚料理	2_女性	4_35歳以上40歳未満	09_女性35~40歳
16	スパゲッティ	お蕎麦	2_女性	2_25歳以上30歳未満	07_女性25~30歳
17	カレーライス	ハンバーグ	1_男性	1_25歳未満	01_男性25歳未満
18	お刺身	鳥料理	1_男性	3_30歳以上35歳未満	03_男性30~35歳
20	納豆	焼肉	1_男性	3_30歳以上35歳未満	03_男性30~35歳
22	カレーライス	ラーメン	1_男性	5_40歳以上	05_男性40歳以上
23	焼き肉	納豆	1_男性	1_25歳未満	01_男性25歳未満
24	スパゲッティ	カレーライス	1_男性	3_30歳以上35歳未満	03_男性30~35歳
25	ご飯と味噌汁	漬物と納豆	1_男性	2_25歳以上30歳未満	02_男性25~30歳
26	うどん	野菜炒め	1_男性	1_25歳未満	01_男性25歳未満
28	パスタ	餃子	2_女性	3_30歳以上35歳未満	08_女性30~35歳
34	ラーメン	カレーライス	1_男性	2_25歳以上30歳未満	02_男性25~30歳
35	ラーメン	餃子	1_男性	4_35歳以上40歳未満	04_男性35~40歳
36	そば、うどん、パスタ等	魚料理(和食)	1_男性	4_35歳以上40歳未満	04_男性35~40歳
38	和食	中華	2_女性	2_25歳以上30歳未満	02_女性25~30歳
39	海老チャーハン	刺身	1_男性	1_25歳未満	01_男性25歳未満
42	中華料理	和食	1_男性	4_35歳以上40歳未満	04_男性35~40歳
44	ラーメン	そば	2_女性	3_30歳以上35歳未満	08_女性30~35歳
45	和食	エスニック	2_女性	3_30歳以上35歳未満	08_女性30~35歳
48	ラーメン	カツ丼	1_男性	3_30歳以上35歳未満	03_男性30~35歳
49	ざるそば	チャーハン	1_男性	2_25歳以上30歳未満	02_男性25~30歳
51	刺身	中華	2_女性	3_30歳以上35歳未満	08_女性30~35歳

① 1行(横)が1回答(サンプル)、
 ② 1列(縦)が1変数となります。
 ③ 回答間区切り*1は「タブ」です。

図. メモ帳で開いた「好きな料理.txt」

1	2	3	4	5	6
No	一番好きな料理	次に好きな料理	性別	年代	性年齢区分
1	カレー	ラーメン	1_男性	4_35歳以上40歳未満	04_男性35~40歳
2	ラーメン	スパゲッティ	1_男性	3_30歳以上35歳未満	03_男性30~35歳
4	5_カレーライス	焼肉	1_男性	4_35歳以上40歳未満	04_男性35~40歳
5	7_ラーメン	炒飯	1_男性	1_25歳未満	01_男性25歳未満
6	8_寿司	カレー	1_男性	3_30歳以上35歳未満	03_男性30~35歳
7	10_サラダ	カレーライス	2_女性	3_30歳以上35歳未満	08_女性30~35歳
8	12_カツ	カレー	1_男性	2_25歳以上30歳未満	02_男性25~30歳
9	14_チキンの料理	魚料理	2_女性	4_35歳以上40歳未満	09_女性35~40歳
10	16_スパゲッティ	お蕎麦	2_女性	2_25歳以上30歳未満	07_女性25~30歳
11	17_カレーライス	ハンバーグ	1_男性	1_25歳未満	01_男性25歳未満
12	18_お刺身	鳥料理	1_男性	3_30歳以上35歳未満	03_男性30~35歳
13	20_納豆	焼肉	1_男性	3_30歳以上35歳未満	03_男性30~35歳
14	22_カレーライス	ラーメン	1_男性	5_40歳以上	05_男性40歳以上
15	23_焼き肉	納豆	1_男性	1_25歳未満	01_男性25歳未満
16	24_スパゲッティ	カレーライス	1_男性	3_30歳以上35歳未満	03_男性30~35歳
17	25_ご飯と味噌汁	漬物と納豆	1_男性	2_25歳以上30歳未満	02_男性25~30歳
18	26_うどん	野菜炒め	1_男性	1_25歳未満	01_男性25歳未満
19	28_パスタ	餃子	2_女性	3_30歳以上35歳未満	08_女性30~35歳
20	34_ラーメン	カレーライス	1_男性	2_25歳以上30歳未満	02_男性25~30歳
21	35_ラーメン	餃子	1_男性	4_35歳以上40歳未満	04_男性35~40歳
22	36_そば、うどん、パスタ等	魚料理(和食)	1_男性	4_35歳以上40歳未満	04_男性35~40歳
23	38_和食	中華	2_女性	2_25歳以上30歳未満	02_女性25~30歳
24	39_海老チャーハン	刺身	1_男性	1_25歳未満	01_男性25歳未満

④ 表計算ソフトウェアでファイルを開くと、データの編集や確認が簡単です。

図. MS-Excel で開いた「好きな料理.txt」

ヒント

1. どのような形式のファイルを **WordMiner®** に読み込ませることができるか

回答間を任意の文字や記号（カンマ区切り、タブ区切り等）で区切られたテキストデータ（csv形式等）を読み込むことが可能です。

2. **WordMiner®** 内部のソートについて

WordMiner® のソートは **S-JIS** コードに従います。よって、性別の場合「男性」「女性」をソートすると「女性」「男性」の順に並びます。例として「1_男性」「2_女性」に置き換えると「1_男性」「2_女性」の順でソートすることができます。

3. 解析ファイルの表題について

ファイルの1行目に表題をつけたまま読み込ませることが可能です。表題は変数名として **WordMiner®** で使用します。表題が記述されていなくても **WordMiner®** に読み込ませることができ、分析に影響はありません。

4. データの加工について

今回の分析では左記のように年齢のカテゴリ分けを行いました。そして、「**性年齢区分**」にどのような意見が抽出されるかを探索するために、新しい変数「**性年齢区分**」を作成しています。

加工は **MS-Excel** を使用しています。

用語説明

1. 回答間区切り

性別、年齢など設問ごとに分けるために区切る文字列です。

回答間区切りの文字列を、回答の中に含めてはいけません。

「好きな料理.txt」では、タブ区切りになっています。

2. サンプルの区切り

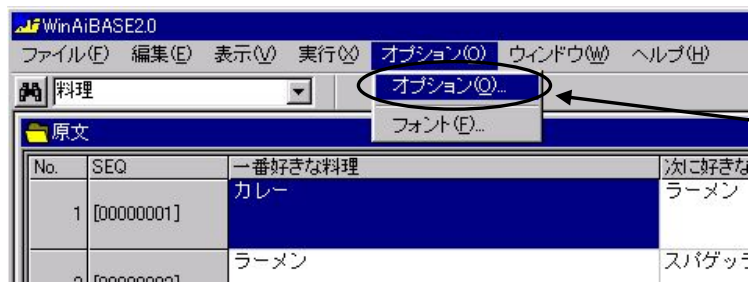
1 サンプル（1 回答）の区切りは「改行(CR+LF)」です

6 キーワード抽出

この事例集と同じ結果を出力するためには、以下の操作が必要です。(操作図、結果図は WordMiner version1.01d です)

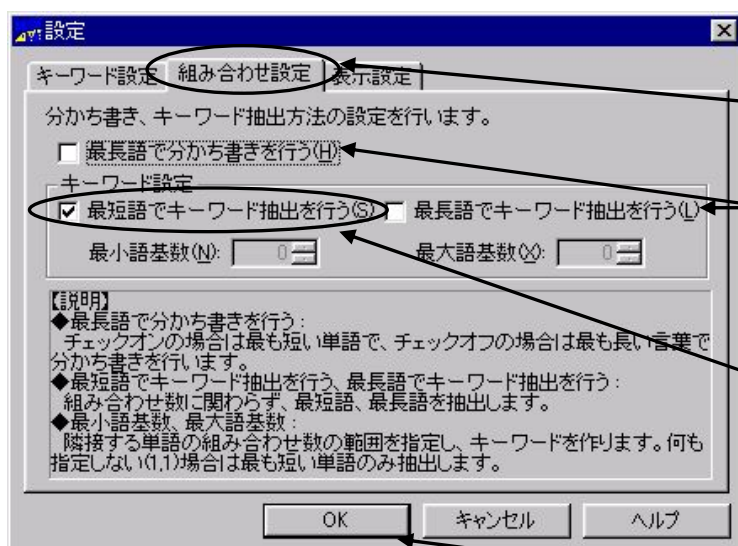
操 作 説 明

- (1) WinAiBASE のオプションを選択する



- ① WinAiBASE の[オプション(O)]-[オプション(O)]をクリックすると、次の画面が表示されます。

- (2) オプションを設定する



- ② [組み合わせ設定]タブをクリックします。
- ③ [最長語で分かち書きを行う(H)]と[最長語でキーワード抽出を行う(L)]のチェックをはずします。
- ④ [最短語でキーワード抽出を行う(S)]にチェックします。
- ⑤ [OK]をクリックすると設定が保存され、WinAiBASE の画面に戻ります。

ヒント

1. 設定を変更した後に

組み合わせ等、設定を変更した後はメニューバーの[実行]－[実行]をクリックし、分かち書き処理を行わないと実行結果に反映されません。

Word Miner® 事例集 導入編

2004年9月 発行

発行者：テキスト・マイニング研究会

監 修：大 隅 昇（統計数理研究所，名誉教授）

作 成：株式会社 平和情報センター

* 本事例集の内容のすべてについて，無断の転載，引用を禁じます。