

WordMiner Recipe 構成要素変数の生成と編集

ここでは、構成要素変数の生成や編集に関する操作のヒントをレシピ(WordMiner Recipe)としてまとめています。

詳細についてはマニュアルを参照してください。

内容は断りなく改変いたします。

新規追加・改変内容は、WordMinerのホームページに適宜アップしております。

<http://wordminer.comquest.co.jp/>

区分	Recipe ID	タイトル
構成要素 変数の生成	RD101	原始変数を分かち書きし、新たな構成要素変数を生成する
	RD102	分かち書きされた原始変数から新たな構成要素変数を生成する
	RD103	構成要素変数を併合して、あらたな構成要素変数を生成する
	RD104	原始変数を併合して、あらたな構成要素変数を生成する
	RD105	分かち書き回避記号により、分かち書きを抑制する
	RD106	構成要素変数をエクスポートし、ファイル編集後、再生成する
編集辞書 の作成	RD201	「構成要素の編集辞書の管理」から新規辞書を作成する
	RD202	「構成要素変数の編集」から新規辞書を作成する
	RD203	構成要素変数の編集に用いた辞書を修正・再適用する
	RD204	「構成要素変数の編集」で作成した辞書を別の編集で利用する
構成要素 変数の編集	RD301	「構成要素一覧と検索」から構成要素の編集辞書(削除)を作る
	RD302	「構成要素一覧と検索」から構成要素の編集辞書(置換)を作る
	RD303	構成要素の編集辞書を別のプロジェクトに複写(コピー)する
	RD304	ファイルから構成要素の編集辞書(削除)を作る
	RD305	ファイルから構成要素の編集辞書(置換)を作る
	RD306	構成要素の編集辞書(置換)をエクスポート(ファイル出力)する
構成要素 の編集	RD401	分かち書き結果を編集辞書を用いて編集(連結)する
	RD402	分かち書き結果を編集辞書を用いて編集(切断)する
構成要素 編集処理	RD501	構成要素変数の編集の適用順序
	RD502	構成要素の編集辞書の適用順序
	RD503	構成要素の編集辞書の適用結果の確認
構成要素 変数の応用	RD601	固有の構成要素からなる構成要素変数を作成する
	RD602	共通の構成要素からなる構成要素変数を作成する
	RD603	特定の構成要素からなる構成要素変数を作成する

構成要素変数の生成

原始変数を分かち書きし、新たな構成要素変数を生成する

Step1.
「変数情報の管理」>「変数の生成」から「構成要素変数を生成」の「原始変数を分かち書きし、新しい変数を生成」をチェック(選択)します。

Step2.
構成要素を生成する原始変数を選択します。

Step3.
分かち書き画面(Happiness/AiBASE)で、「実行」を選択し、分かち書き・キーワード抽出を行います。

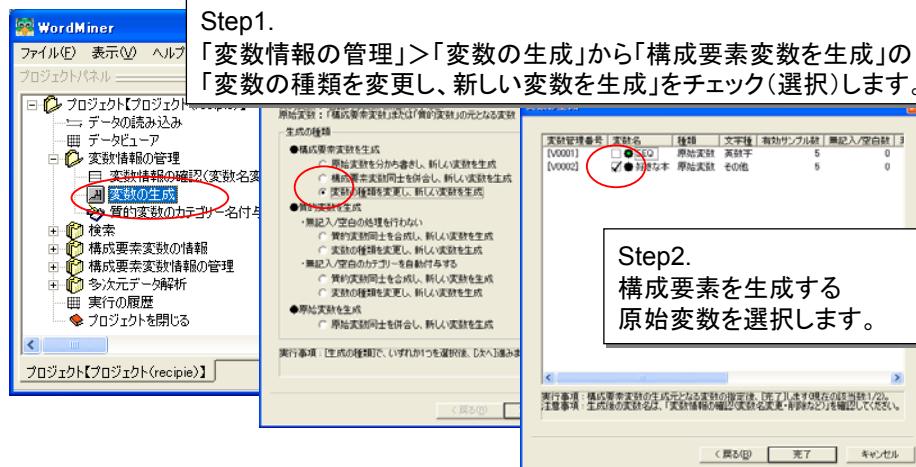
Step4.
「ファイル」から「分かち書き終了」を選択し、構成要素変数を生成します。

[狙い]
原始変数を分かち書きし、新たな構成要素変数を生成する。

[ポイント]

- (1)分かち書き処理を終了すると、分かち書き「<原始変数名>一分かち書き」とキーワード抽出「<原始変数名>-キーワード」の2つの構成要素変数が生成される。
- (2)分かち書き及びキーワード抽出方法を設定するオプションがある。切断単位やキーワードの組み合わせ、切断回避記号設定を設定することができる。
- (3)分かち書き画面の「ファイル」から「プロパティ」を選択すると、サンプル数、総処理文字数、分かち書き数、キーワード数などの処理結果情報を得ることができる。

分かち書きされた原始変数から新たな構成要素変数を生成する



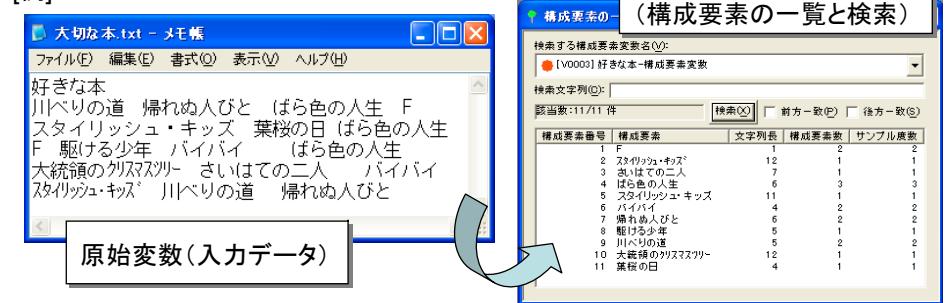
[狙い]

既に分かち書きされた原始変数(構成要素の候補が空白で区切られている)から、新たな構成要素変数を生成する。

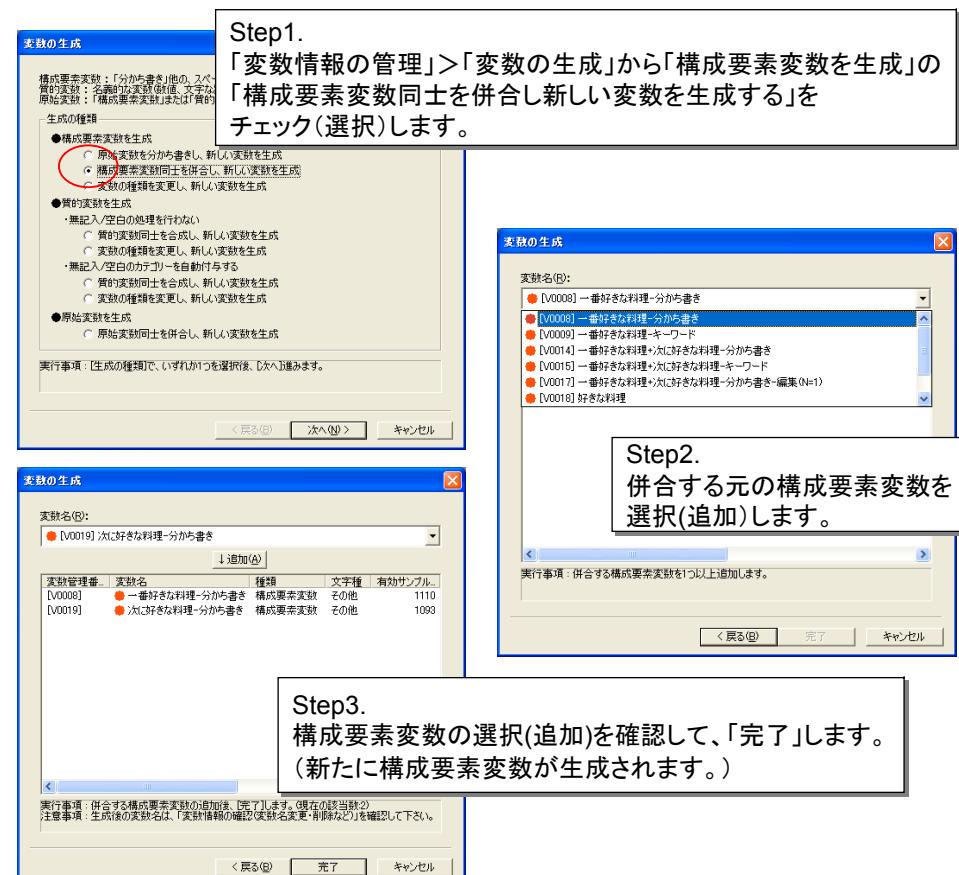
[ポイント]

- (1)区切り記号の空白は半角でも全角でもどちらでもよい。
- (2)半角と全角による表記の異なりは、異なる構成要素となる。
(分かち書き処理により構成要素を生成する場合は全角に統一される。)
- (3)分かち書きを行わずに構成要素を生成するので、WordMinerのエクスポートデータを再編集した結果、あるいは他の手段により得た分かち書き結果を読み込み、構成要素変数を生成することができる。

[例]



構成要素変数を併合して、あらたな構成要素変数を生成する



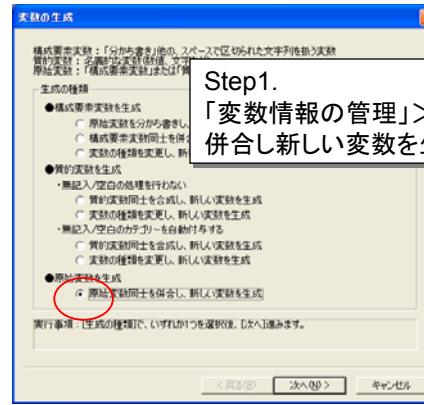
[狙い]

いくつかの構成要素変数を併合して、新たな構成要素変数を得る。

[ポイント]

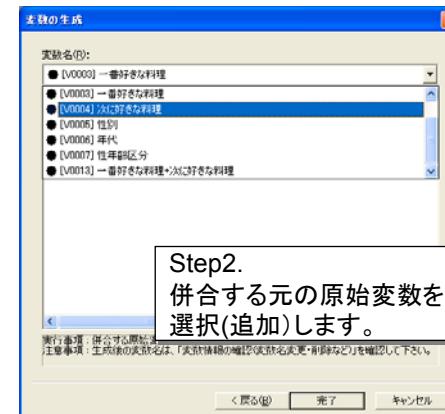
- (1)例えば、「1番大切なものの(FA)」と「その他に大切なものの(FA)」の構成要素変数から、両構成要素変数を併合した新たな構成要素変数を得る。
- (2)併合された構成要素変数のディフォルトの名称は、併合した「構成要素変数名(1)+構成要素変数名(2)+…+構成要素変数名(n)」となる。
- (3)構成要素変数の併合により得られた構成要素変数は、即座に構成要素変数として利用できる。分かち書き等、新たに構成要素変数の生成手続きを行う必要はない。

原始変数を併合して、あらたな構成要素変数を生成する



Step1.

「変数情報の管理」>「変数の生成」から「原始変数同士を併合し新しい変数を生成する」をチェック(選択)します。



Step2.

併合する元の原始変数を選択(追加)します。



Step3.

原始変数の選択(追加)を確認して、「完了」します。
(新たに原始変数が生成されます。)

Step4.

併合された(新たに生成された)原始変数を分かち書きし、構成要素変数を生成します。

[狙い]

いくつかの原始変数を併合して、新たな原始変数を得る。
この新たな原始変数から構成要素変数を生成する。

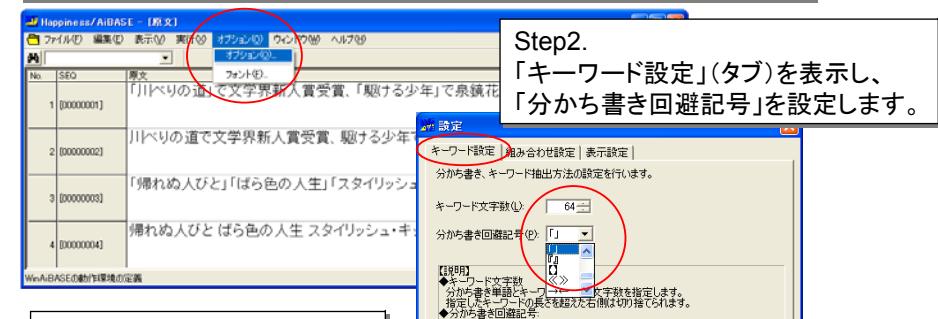
[ポイント]

- (1)例えば、「1番大切なものの(FA)」と「その他に大切なものの(FA)」の原始変数から、両原始変数を併合した新たな原始変数を得る。
- (2)併合された原始変数のデフォルトの名称は、併合した「原始変数名(1)+原始変数名(2)+…+原始変数名(n)」となる。
- (3)先に、それぞれの原始変数から構成要素変数を生成し、それらの構成要素変数を併合して新たな構成要素変数を得ることもできる。

分かち書き回避記号により、分かち書きを抑制する

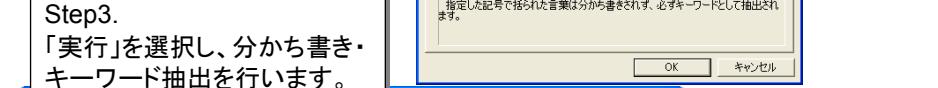
Step1.

分かち書き画面(Happiness/AiBASE)で、「オプション」を選択します。



Step2.

「キーワード設定」(タブ)を表示し、「分かち書き回避記号」を設定します。



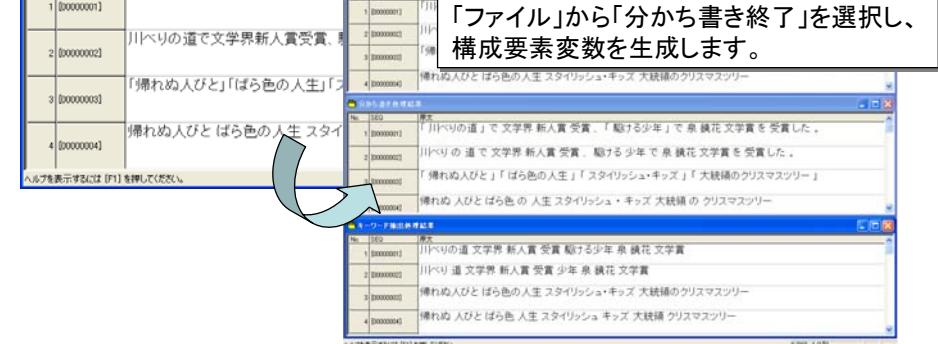
Step3.

「実行」を選択し、分かち書き・キーワード抽出を行います。



Step4.

「ファイル」から「分かち書き終了」を選択し、構成要素変数を生成します。



[狙い]

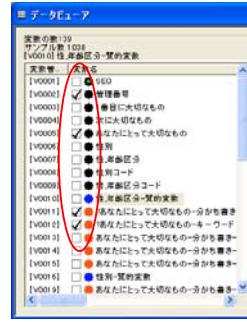
分かち書き回避記号を設定し、分かち書きを抑制し、任意の文字列を構成要素とする。

[ポイント]

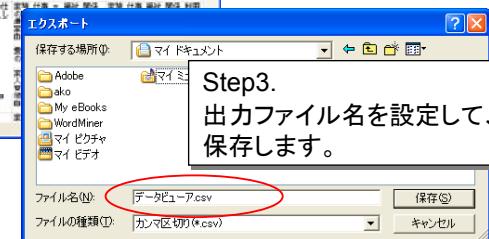
- (1)回避記号で括られた文字列中の空白(全角・半角)は削除され、空白を詰めた構成要素が生成される。例)「A B C D」→「ABCD」
- (2)キーワードによる構成要素変数生成の場合、同時に組み合わせ設定が指定されていると、組み合わせ設定の最小語基が優先される。
- (3)回避記号が対になっていない場合、エラーとなり、そのサンプル(回答)から構成要素は生成しない。以降のサンプル(回答)の処理は継続して行うことができる。

構成要素変数をエクスポートし、ファイル編集後、再生成する

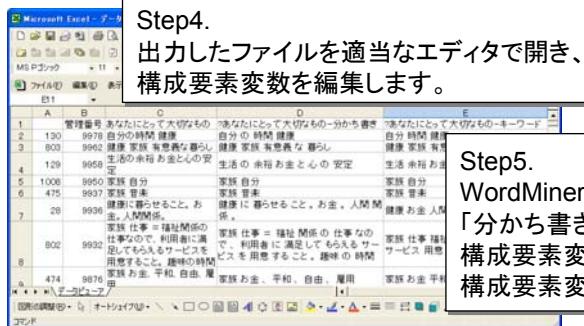
Step1.
データビューアを表示し、エクスポートする変数をチェックします。



Step2.
データビューアの表示画面で右クリックし、「すべてエクスポート」を選択します。



Step3.
出力ファイル名を設定して、保存します。



Step4.
出力したファイルを適当なエディタで開き、構成要素変数を編集します。

Step5.
WordMinerで、新たにファイルを読み込み、「分かち書きされた原始変数から新たな構成要素変数を生成する」方法により、構成要素変数を生成します。

[狙い]

WordMinerで生成した構成要素変数をエクスポートし、外部のファイルエディタ等で修正した後に、分かち書き済のデータとして読み込み、構成要素変数を生成する。

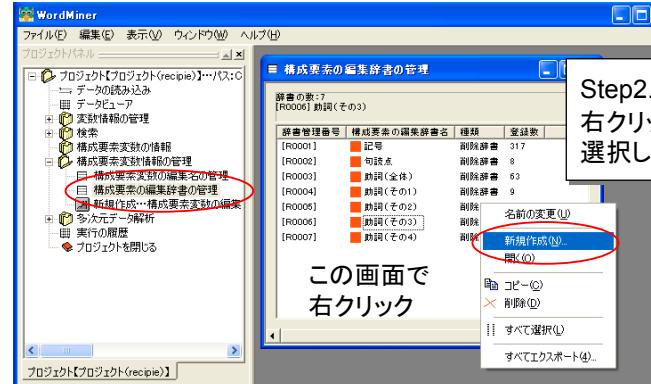
[ポイント]

- (1)データビューアから変数を選択し、エクスポートする。サンプル数かつ順序が保証される場合、変数の追加読み込みも可能だが、新たに全てのデータ(原始変数)を読み込むことを前提に必要な変数をすべて選択し、エクスポートすることが望ましい。
- (2)出力した構成要素変数は構成要素が半角空白で区切られている。
- (3)空白(半角・全角)で区切られた原始変数は、「分かち書きされた原始変数から新たな構成要素変数を生成する」方法により、構成要素変数を生成することができる。

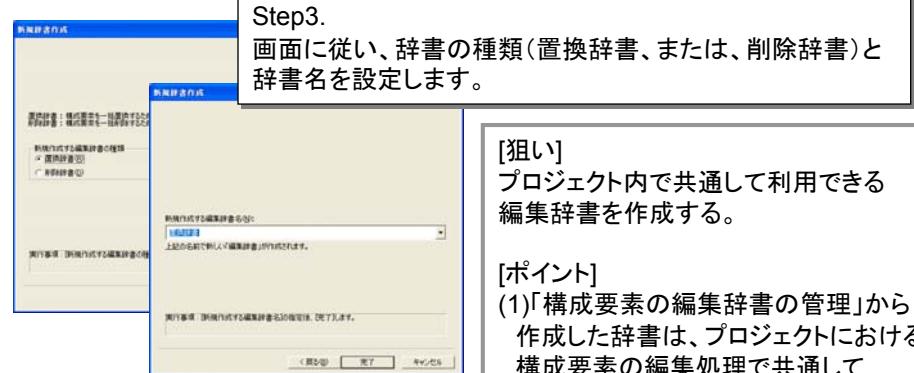
「構成要素の編集辞書の管理」から新規辞書を作成する

Step1.
プロジェクトパネルから「構成要素の編集辞書の管理」を開きます。

構成要素変数情報の管理 > 構成要素の編集辞書の管理



Step2.
右クリックして「新規作成」を選択します。



Step3.
画面に従い、辞書の種類(置換辞書、または、削除辞書)と辞書名を設定します。



Step4.
辞書の内容を登録します。

[狙い]
プロジェクト内で共通して利用できる編集辞書を作成する。

[ポイント]

- (1)「構成要素の編集辞書の管理」から作成した辞書は、プロジェクトにおける構成要素の編集処理で共通して利用することができる。
- (2)ただし、この辞書作成以前に編集した変数は、編集時の状態で専用に保存された辞書を利用するため、編集処理を再実行する際に、ここで作成した新しい辞書は選択できない。
- (3)辞書内容の登録方法については「構成要素変数の編集」についての説明を参照。

「構成要素変数の編集」から新規辞書を作成する

Step1.
プロジェクトパネルから「新規作成…構成要素変数の編集」を開きます。

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

Step2.
画面に従い、「編集する構成要素変数名」と「新規作成する構成要素変数の編集名」を設定します。

構成要素変数の編集 [R0001] [V0010] あなたにとって大切なもの

Step3.
「構成要素変数の編集」から「新規辞書作成」を開きます。

構成要素変数の編集 [R0001] [V0010] あなたにとって大切なもの

Step4.
画面に従い、辞書の種類(置換辞書、または、削除辞書)と辞書名を設定します。

置換辞書 [R0009] 置換辞書

[狙い]
特定の構成要素変数の編集処理でのみ利用できる辞書を作成する。

[ポイント]

- (1)「構成要素変数の編集」から作成した辞書は、その編集処理専用の辞書として保存される。
- (2)プロジェクトパネルや他の編集名から「構成要素の編集辞書の管理」を開いても辞書は表示されないため、特定の編集手順以外の要因で修正されることがない。
- (3)編集した変数を再編集する場合は、プロジェクトパネルから「構成要素変数の編集名の管理」を選択し、ここで(Step2.)設定した「構成要素の編集名」を選択する。
- (4)辞書内容の登録方法については別途参照。

Step5.
辞書の内容を登録します。

WordMiner recipe

構成要素変数の編集に用いた辞書を修正・再適用する

Step1.
プロジェクトパネルから「構成要素変数の編集名の管理」を開きます。

構成要素変数情報の管理 > 構成要素変数の編集名の管理

Step2.
編集する構成要素変数の編集名を選択します。

構成要素変数の編集名の管理 [R0001] [V0010] あなたにとって大切なもの

Step3.
「構成要素変数の編集」から「編集辞書の管理の表示」を押します。

構成要素変数の編集 [R0001] [V0010] あなたにとって大切なもの

Step4.
修正したい辞書を開き、内容を修正して閉じます。

置換辞書 [R0008] 置換辞書

[狙い]
辞書を編集して、構成要素変数の編集処理を再度実行する。

[ポイント]

- (1)「構成要素変数の編集」から辞書を修正し、修正した内容で編集処理を再実行できる。
- (2)構成要素変数の編集時に専用に保存された辞書を修正するため、修正内容は他の編集処理で選択する辞書には反映されない。
- (3)これらの構成要素変数の編集内容は「編集名」で管理されている。

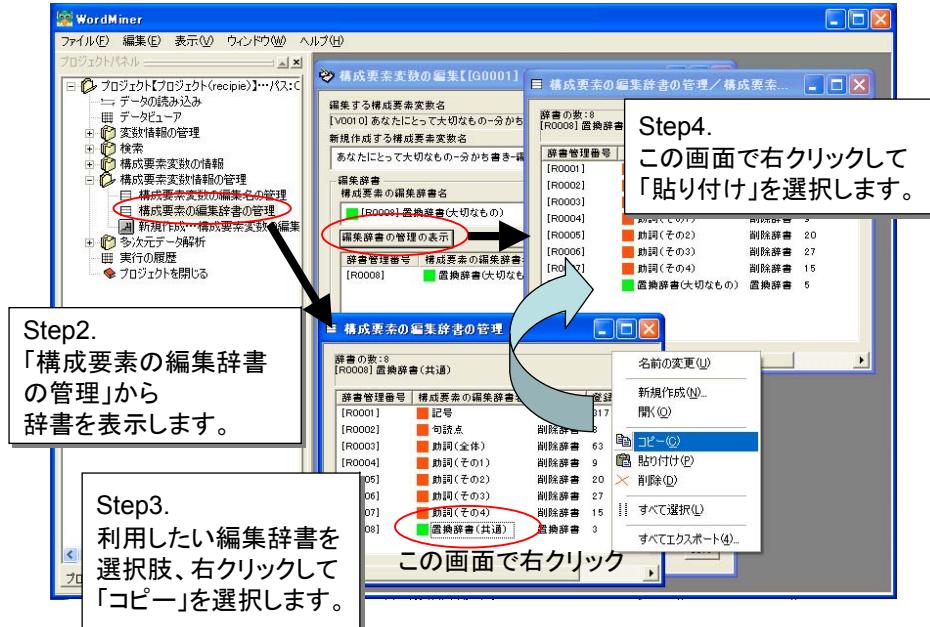
Step5.
修正した辞書の登録数が変わっていることを確認し、「実行」で編集処理を実行します。

WordMiner recipe

作成した辞書を別の編集で利用する

ここでは、「構成要素の編集辞書の管理」で作成した辞書を、既に1度は編集した構成要素変数の編集に利用する例を示します。

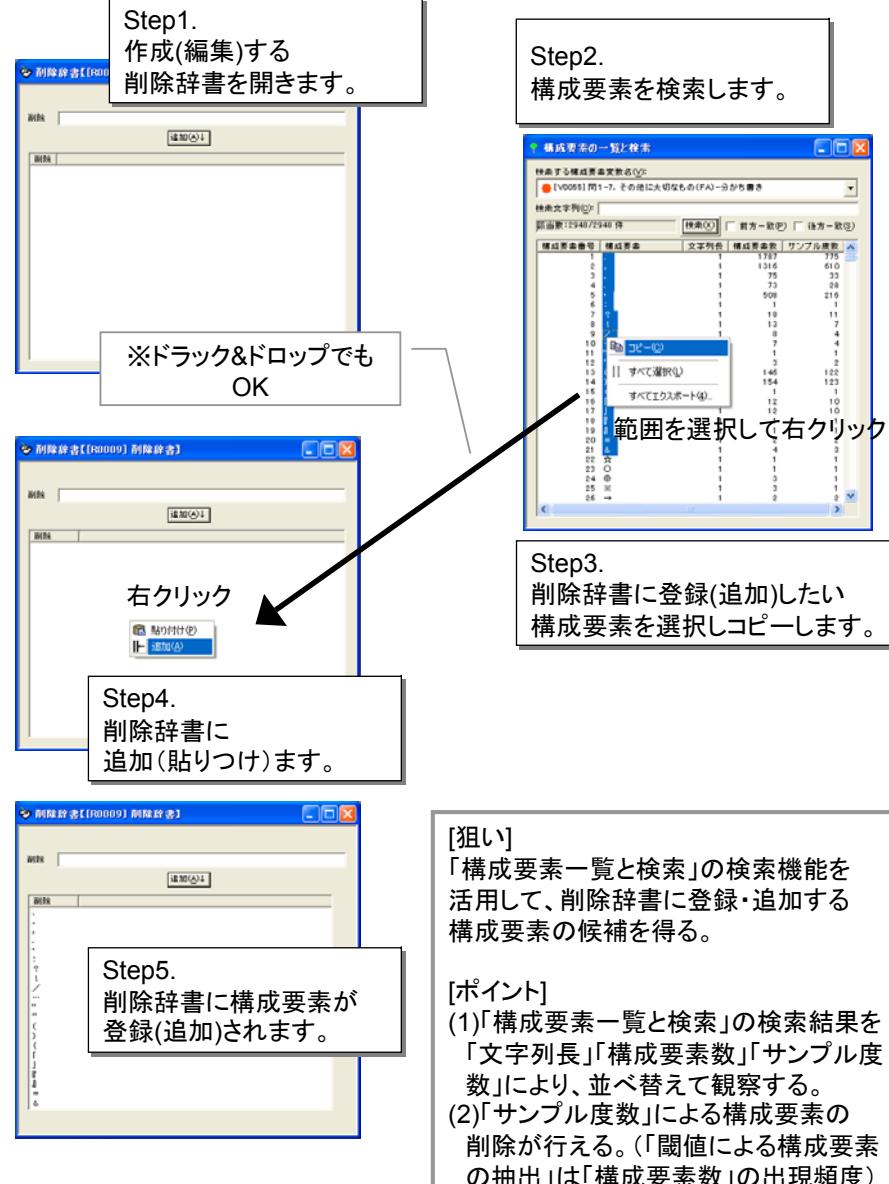
Step1.
「構成要素変数の編集」から「編集辞書の管理の表示」を押します。



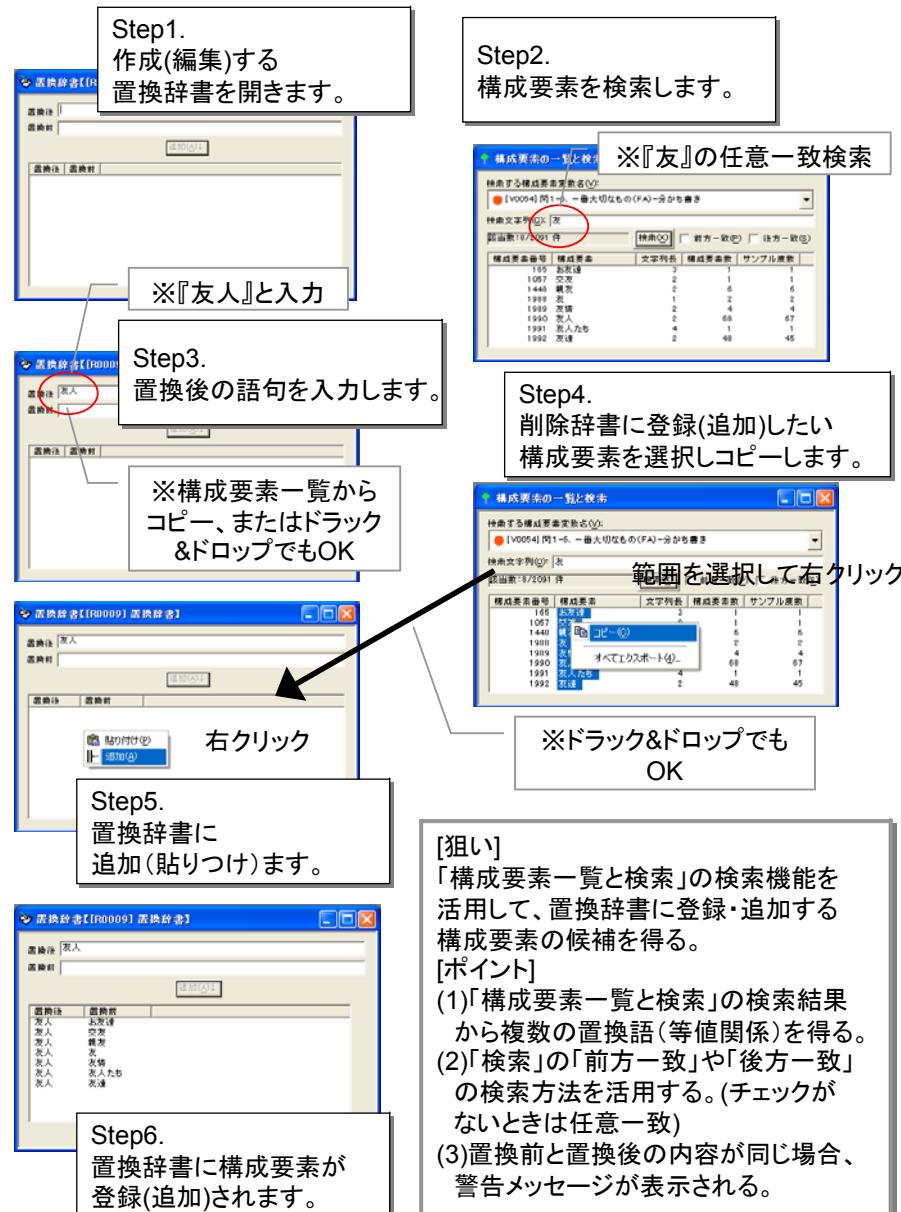
[狙い]
異なる編集処理で作成・利用した辞書を、他の編集処理で利用する。
作成した編集辞書を他の編集処理に利用するためにコピーする。

[ポイント]
(1)別の「構成要素変数の編集」に、辞書をコピーすることができる。
(2)特定の変数の編集処理において作成、修正した辞書を、
他の変数の編集処理に流用することができる。
(3)「構成要素の編集辞書の管理」画面間において、辞書のコピー・
貼り付けが行える。

「構成要素一覧と検索」から構成要素の編集辞書(削除)を作る



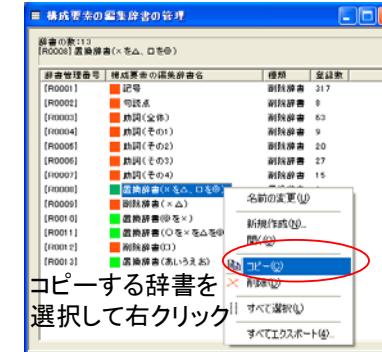
「構成要素一覧と検索」から構成要素の編集辞書(置換)を作る



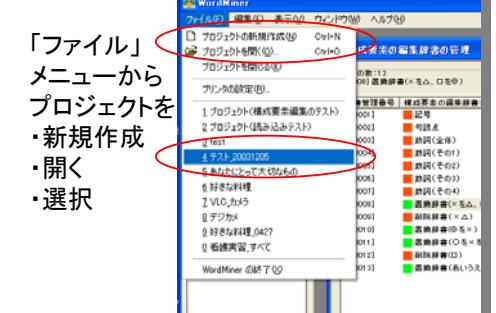
構成要素の編集辞書を別のプロジェクトに複写(コピー)する

Step1.
複写(コピー)したい編集辞書を右クリックして
「コピー」を選択します。

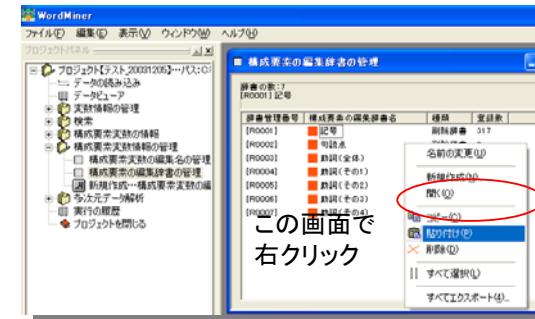
構成要素変数情報の管理 > 構成要素の編集辞書の管理



Step2.
辞書を複写するプロジェクトを開きます。

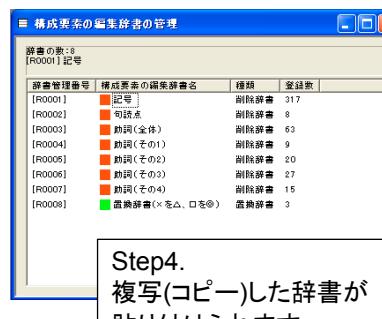


Step3.
「構成要素の編集辞書の管理」の
辞書一覧画面で右クリックして
「貼り付け」を選択します。



**この画面で
右クリック**

[狙い]
構成要素の編集辞書(削除・置換)を
別のプロジェクトで利用する。
プロジェクト間で編集辞書を複写(コピー)
する。

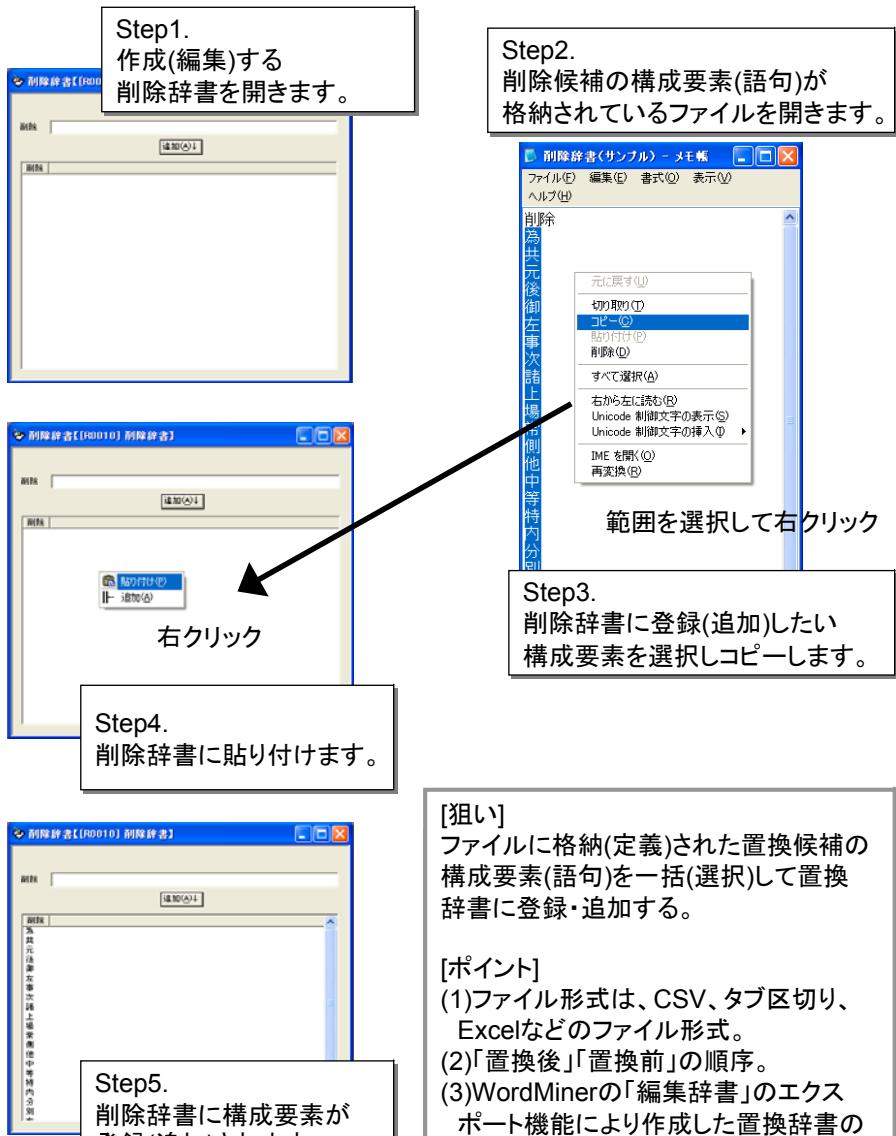


Step4.
複写(コピー)した辞書が
貼り付けられます。

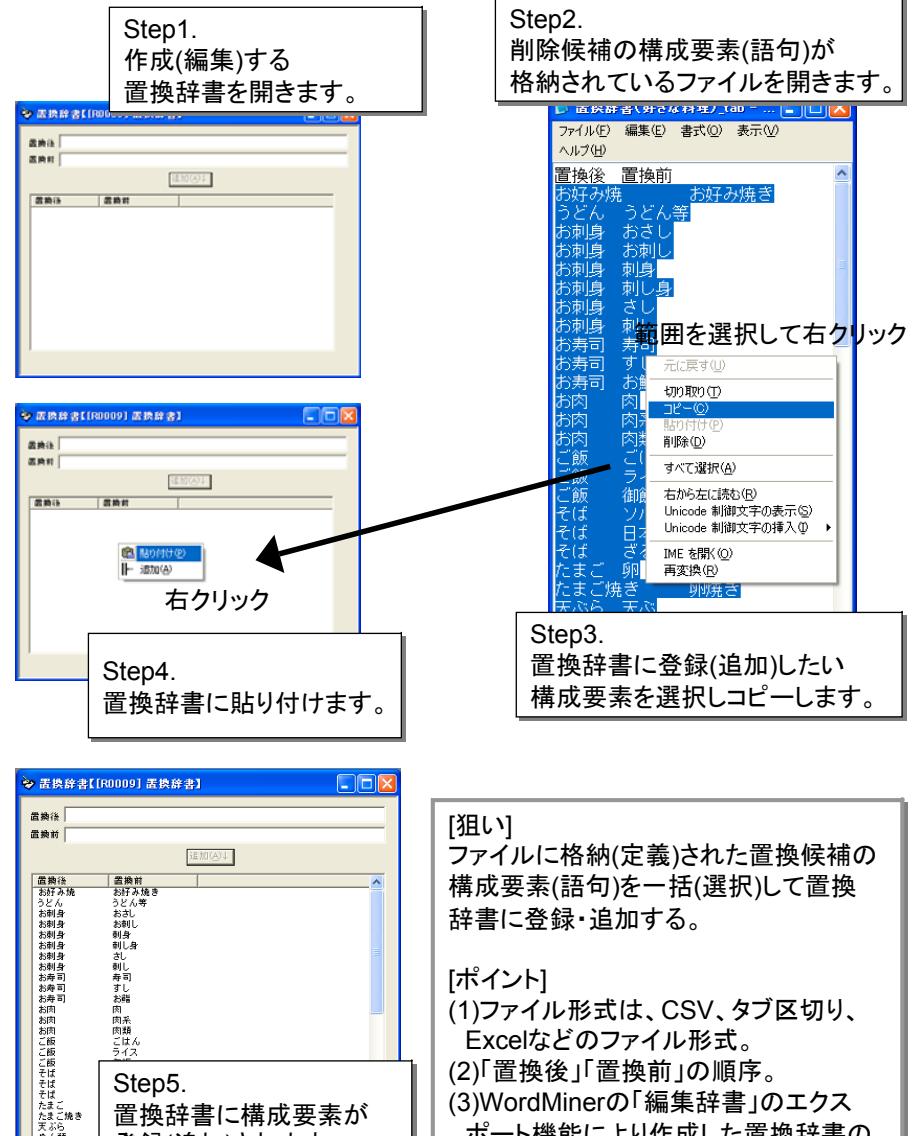
[ポイント]

- ・編集辞書を複写(コピー)した状態で
プロジェクトを切り替え(辞書を複写する
プロジェクトを開き)、辞書を貼り付ける。

ファイルから構成要素の編集辞書(削除)を作る

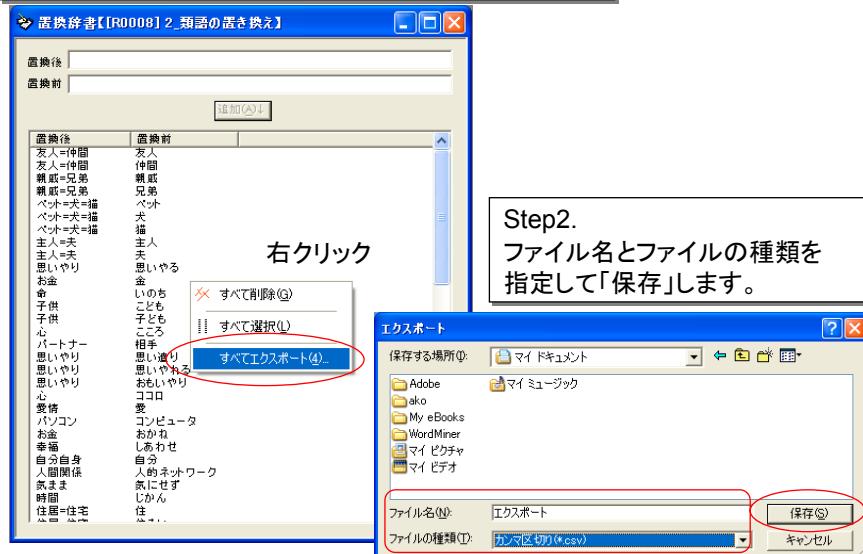


ファイルから構成要素の編集辞書(置換)を作る



構成要素の編集辞書(置換)をエクスポート(ファイル出力)する

Step1.
エクスポート(ファイル出力)したい編集辞書を開き、右クリックで「すべてエクスポート」を指定します。



	置換後	置換前
1	置換後	置換前
2	友人=仲間	友人
3	友人=仲間	仲間
4	親戚=兄弟	親戚
5	親戚=兄弟	兄弟
6	ペット=犬=猫	ペット
7	ペット=犬=猫	犬=猫
8	ペット=犬=猫	犬
9	主人=夫	主人
10	主人=夫	夫
11	思いやり	思いやる
12	お金	金
13	命	いのち
14	子供	こども
15	子供	子ども
16	心	いのち
17	バ	バ
18	思	思

Step3.
置換辞書の内容がエクスポート(ファイル出力)されます。

[狙い]
作成した構成要素の編集(置換)辞書をエクスポート(ファイル出力)する。これにより、編集辞書の流用が可能となる。

[ポイント]

- (1)ファイル形式は、CSV形式、または、タブ区切り形式。
- (2)「置換後」「置換前」の順序。
- (3)ここで作成した形式の辞書ファイルを任意のプロジェクトの編集(置換)辞書に登録(追加)することができる。これにより、作成した編集辞書を別のプロジェクトで利用することができる。

分かち書き結果を編集辞書を用いて編集(連結)する

Step1.
作成(編集)する置換辞書を開きます。

置換後: 日本庭園
置換前: 日本 庭園

Step2.
置換前に編集前の分かち書き結果を入力し、置換後に修正後分かち書き結果を入力し、置換辞書に登録します。

編集する構成要素変数名: [V0019] 発言-分かち書き(短単位)
新規作成する構成要素変数名: 短語-分かち書き(短単位)-編集
構成要素の編集辞書名: [R0013] 置換辞書(分かち書き修正)
抽出条件: 要約文によるサンプルの抽出
抽出条件となる要約文名: (指定なし)
カテゴリ: カテゴリー
関係による構成要素
構成要素表: [R0013]
Step3.
作成した編集(置換)辞書を用いて、構成要素編集を編集する。

置換後に「日本庭園」
置換前に「日本 庭園」

半角空白

「日本料理」←「日本 料理」
「中華料理」←「中華 料理」
「日本庭園」←「日本 庭園」

[狙い]

分かち書き結果を構成要素の編集辞書を用いて修正(連結)する。

[ポイント]

(1)構成要素変数の分かち書き結果は、半角空白で切断されている。

(2)構成要素の置換辞書に編集前の分かち書き結果と編集語の分かち書き結果を登録し、この辞書を用いて構成要素変数を編集することにより、分かち書き結果を修正する。

(4)2語以上の切断語列の連結も同様。

例:「日本料理教室」

←「日本 料理 教室」

(3)分かち書き結果の編集は、構成要素の生成(分かち書き処理)時に、直接入力あるいは検索・置換処理により行うことができる。

(4)分かち書き・キーワード抽出処理は、長単位(複合語連結)、短単位(複合語切断)のオプションがある。

分かち書き結果を編集辞書を用いて編集(切断)する

Step1. 作成(編集)する 置換辞書を開きます。

置換後に「日本庭園」 置換前に「日本庭園」

半角空白

「日本 料理」←「日本料理」
「中華 料理」←「中華料理」
「日本 庭園」←「日本庭園」

Step2. 置換前に編集前の分かち書き結果を入力し、置換後に修正後分かち書き結果を入力し、置換辞書に登録します。

[狙い] 分かち書き結果を構成要素の編集辞書を用いて修正(切断)する。
[ポイント]
 (1)構成要素変数の分かち書き結果は、「半角空白」で切断する。
 (2)構成要素の置換辞書に編集前の分かち書き結果と編集語の分かち書き結果を登録し、この辞書を用いて構成要素変数を編集することにより、分かち書き結果を修正する。
 (4)2語以上に語列を切断する際も同様。
 例:「日本 料理 教室」
 ←「日本料理教室」
 (3)分かち書き結果の編集は、構成要素の生成(分かち書き処理)時に、直接入力あるいは検索・置換処理により行うことができる。
 (4)分かち書き・キーワード抽出処理は、長単位(複合語連結)、短単位(複合語切断)のオプションがある。

Step3. 作成した編集(置換)辞書を用いて、構成要素編集を編集する。

構成要素変数の編集の適用順序

- 適用順序**
1. 辞書(削除・置換)による編集
 2. 質的変数によるサンプルの抽出
 3. 閾値による構成要素の抽出

1 編集に適用する編集辞書を選択(追加)します。

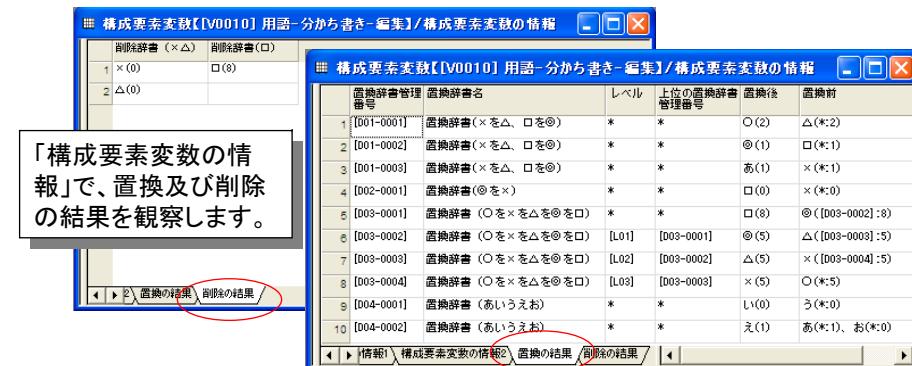
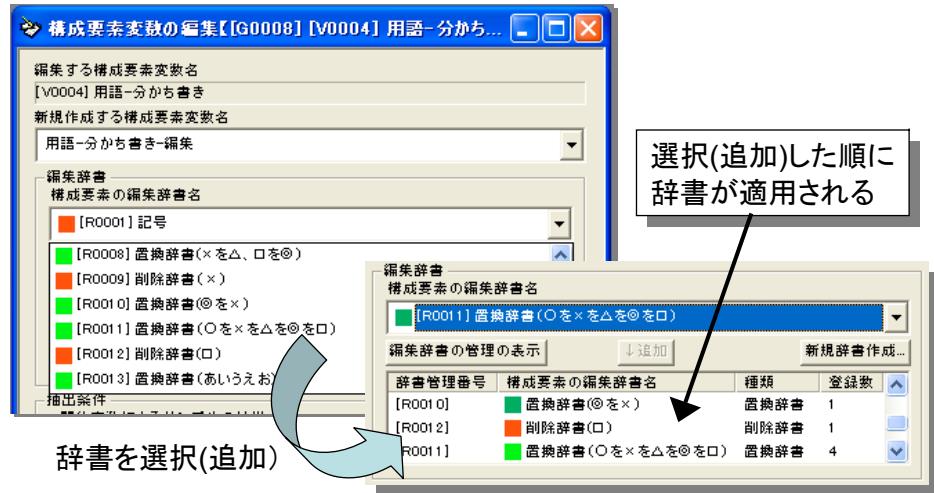
2 質的変数を選択し、サンプルから抽出する質的変数のカテゴリを選択(チェック)します。

カテ...	カテゴリー名	サンプル...
1	<input checked="" type="checkbox"/> a	3
2	<input checked="" type="checkbox"/> b	3
3	<input type="checkbox"/> c	3

3 閾値の値を指定します。(デフォルトの値は'2'

[狙い] 解析の目的や狙いに応じて、構成要素を編集する。
[ポイント]
 (1)編集の方法は、①編集(削除・置換)辞書、②質的変数によるサンプルの抽出、③閾値による抽出、の3つがある。
 (2)編集の順序は、まず、編集辞書により、構成要素の削除・置換が行われる。次に、質的変数によるサンプルの抽出が行われ、最後に閾値により構成要素が抽出される。(閾値の「値」となる構成要素数が変化する)
 -構成要素が置換辞書により置換されると構成要素数が変化(増加)する。
 -質的変数によるサンプルの抽出が行われると構成要素数が変化(減少)する。

構成要素の編集辞書の適用順序



[狙い]

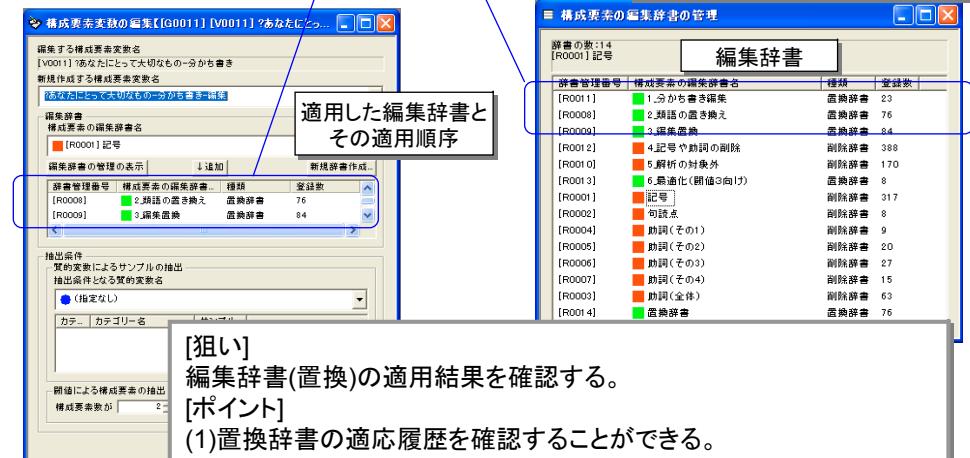
解析の目的や狙いに応じて、構成要素を編集する。

[ポイント]

- (1) 編集の方法は、①編集(削除・置換)辞書、②質的変数によるサンプルの抽出、③閾値による抽出、の3つがある。
- (2) 編集の順序は、まず、編集辞書により、構成要素の削除・置換が行わる。次に、質的変数によるサンプルの抽出が行われ、最後に閾値により構成要素が抽出される。(閾値の「値」となる構成要素数が変化する)
 - ・構成要素が置換辞書により置換されると構成要素数が変化(増加)する。
 - ・質的変数によるサンプルの抽出が行われると構成要素数が変化(減少)する。

構成要素の編集辞書の適用結果の確認

「構成要素変数の情報」で、編集後の構成要素変数を指定し、
編集辞書の適用結果を観察します。



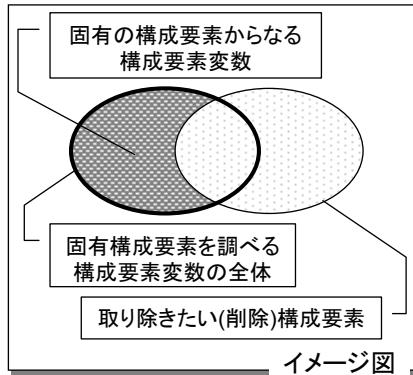
固有の構成要素からなる構成要素変数を作成する

ここでは、2つの構成要素変数から、固有の構成要素からなる構成要素変数を作成する例を示す。

Step1.
(2つの)構成要素変数を生成します。

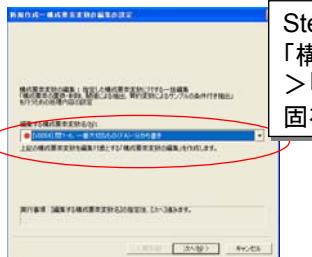
Step2.
「構成要素一覧と検索」で構成要素変数を検索して、すべての構成要素を選択・コピーする。

Step3.
(取り除きたい構成要素を)
削除辞書に貼り付け(追加)ます。

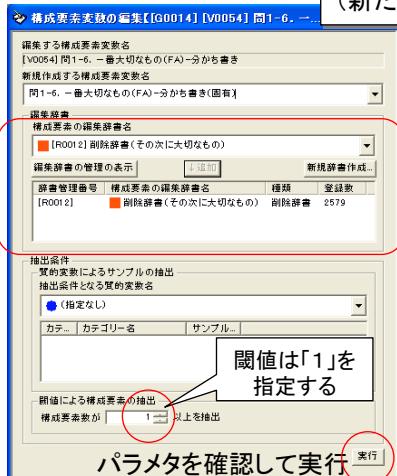


イメージ図

ここまででは、「構成要素一覧と検索」から構成要素の編集辞書(削除)を作ると同じ。



Step4.
「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」で
固有の構成要素を調べたい構成要素変数を選択します。



Step5.
編集辞書の候補から、取り除きたい構成要素の
削除辞書を選択(追加)して、実行します。
(新たに構成要素変数が生成されます。)

[狙い]

いくつかの構成要素変数において、ある構成要素変数の持つ固有の構成要素を調べる。(固有の構成要素変数を得る)

[ポイント]

- (1)例えば、「1番大切なものの(FA)」と「その他に大切なものの(FA)」の構成要素変数から、「1番大切なもの」にのみ出現した構成要素を調べたいときに操作する。
- (2)構成要素変数の編集の「閾値」の指定は「1」を設定する。(デフォルトの設定は「2」)

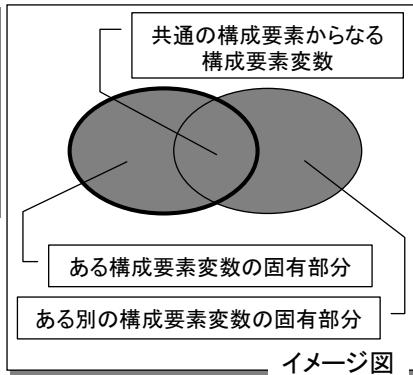
閾値は「1」を
指定する

パラメタを確認して実行

共通の構成要素からなる構成要素変数を作成する

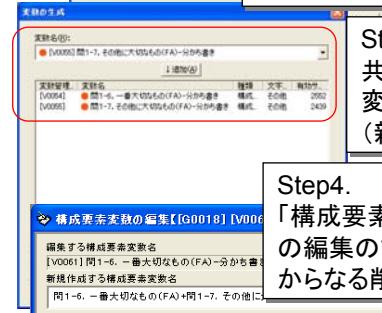
ここでは、2つの構成要素変数から、固有の構成要素からなる構成要素変数を作成する例を示す。

Step1.
共通の構成要素を調べたい(2つの)構成要素変数について、[固有の構成要素からなる構成要素変数を作成する]手順により、個々の固有の構成要素からなる構成要素変数を作成します。

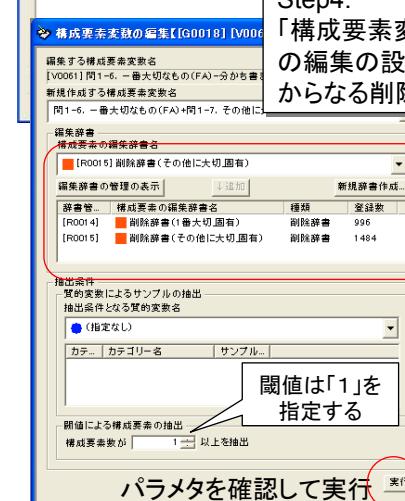


イメージ図

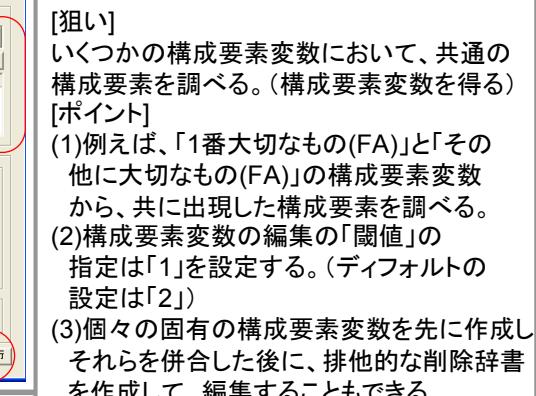
Step2.
「変数情報の管理」>「変数の生成」から「構成要素変数同士を併合し新しい変数を生成する」をチェック(選択)します。



Step3.
共通の構成要素を調べたい(2つの)元の構成要素変数を選択(追加)して、「完了」します。
(新たに構成要素変数が生成されます。)



Step4.
「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」で併合した構成要素変数を個々の固有要素からなる削除辞書を用いて編集します。



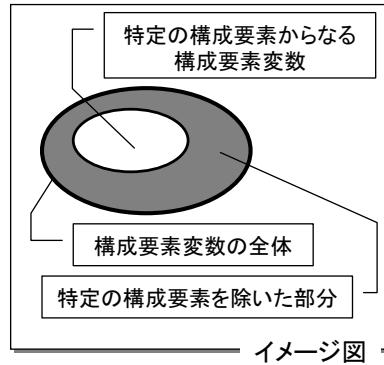
[狙い]
いくつかの構成要素変数において、共通の構成要素を調べる。(構成要素変数を得る)
[ポイント]

- (1)例えば、「1番大切なものの(FA)」と「その他に大切なものの(FA)」の構成要素変数から、共に出現した構成要素を調べる。
- (2)構成要素変数の編集の「閾値」の指定は「1」を設定する。(デフォルトの設定は「2」)
- (3)個々の固有の構成要素変数を先に作成し、それらを併合した後に、排他的な削除辞書を作成して、編集することもできる。

特定の構成要素からなる構成要素変数を作成する

ここでは、ファイルに記述(特定)された単語や語句(構成要素)のみからなる構成要素変数を作成(編集)する例を示す。

Step1.
[ファイルから構成要素の編集辞書(削除)を作る]手順により、特定の単語や語句(構成要素)からなる削除辞書を作成します。



Step2.
特定の構成要素からなる構成要素変数に編集したい元の構成要素変数を、[Step1]で作成した削除辞書を用いて編集します。
(特定の構成要素を除いた構成要素からなる構成要素変数が生成されます。)

Step3.
[Step2]で作成した「特定の構成要素を除いた構成要素変数」について、「構成要素一覧と検索」から構成要素の編集辞書(削除)を作る]手順により、削除辞書を作成します。

Step4.
特定の構成要素からなる構成要素変数に編集したい元の構成要素変数を、[Step3]で作成した特定の構成要素以外の構成要素からなる削除辞書を用いて編集します。

[狙い]
ファイルなどに定義した特定の単語や語句(構成要素)のみからなる構成要素変数を作成する。

[ポイント]
(1)例えば、別に定義されている「人間関係」や「所有物」などの用語集にもとづき、「1番大切なものの(FA)」の構成要素変数を編集する。
(2)構成要素変数の編集時の「閾値」の指定は「1」を設定する。(デフォルトの設定は「2」)