

IBM SPSS Text Analytics for  
Surveys 4.0 ユーザー ガイド



注:この情報とこの情報がサポートする製品を使用する前に、注意 p. 287 中の全般的  
情報をお読みください。

本版は、IBM® SPSS® Text Analytics for Surveys 4 およびその後のリリースと変更に、新  
しい版が別途表示されない限り、適用されます。

アドビ製品の画面コピーは、Adobe Systems Incorporated の承認を得て掲載しています。

Microsoft 製品の画面コピーは、Microsoft Corporation の承認を得て掲載しています。

Licensed Materials - Property of IBM

**© Copyright IBM Corporation 2004, 2011.**

U.S. Government Users Restricted Rights - Use, duplication or disclosure restricted  
by GSA ADP Schedule Contract with IBM Corp.

---

# はじめに

IBM® SPSS® Text Analytics for Surveys バージョン 4.0.1 の世界によるこそ。これは、自由記述式回答を有効に分析するためのアンケートテキストのコード化ツールです。この製品を使うと、アンケート調査を行う人はだれでも、構造化されていないアンケート回答を定量的なデータにすばやく変換できます。このように自由記述のテキストデータを“解放”することで、分析の質や意思決定能力が大幅に改善されます。この製品では、アンケートデータのインポート、主要キーワードの抽出、結果の調整、回答のカテゴリ化を行うことができます。使用しているデータをカテゴリ化したならば、これらのカテゴリを IBM® SPSS® Statistics などの定量的分析ツールにエクスポートし、さらに分析を行ったり、グラフを作成したりできます。

SPSS Text Analytics for Surveys は、自由記述式のアンケート回答内の主要キーワードを高い精度で抽出・分類するように、高度な言語学テクノロジーと手作業の手法を組み合わせています。しっかりした分類アルゴリズムと簡単なドラッグ&ドロップ機能を使用することで、アンケート回答をカテゴリ化するためのカテゴリ（“コード”）を作成できます。この作成されたカテゴリを再度使用することで、同一のあるいは類似した研究においても、一貫した結果が得られます。自由記述式の回答データはアンケートによって大きく異なるので、2つのプロジェクトが完全に同じになるということはありませんが、それでも分析の基本的なプロセスは同じであると想定できます。詳細は、2章 p.11 [典型的なプロセス](#) を参照してください。

## IBM Business Analytics について

IBM Business Analytics ソフトウェアは、意思決定者がビジネス パフォーマンスの改善のために使用可能な完全で整合性があり、正確な情報を提供します。[ビジネス インテリジェンス](#)、[予測解析](#)、[資金業務](#)と[戦略マネジメント](#)、そして [分析アプリケーション](#)の包括的ポートフォリオは、現行の業務に明確、迅速で、現実的な洞察をもたらし、将来的な結果を予測する能力を実現します。豊富な産業用ソリューション、証明された実践法、それに専門家によるサービスを組み合わせることにより、あらゆる規模の会社組織が、最高の生産性を推進し、信頼できる意志決定を自動化し、そして、よりよい結果を実現させることができます。

このポートフォリオの一部として、IBM SPSS Predictive Analytics ソフトウェアは、会社組織がよりよい業績結果を推進するために、未来の事象を予測し、その洞察に基づき積極的に行動できるようにお役に立ちます。世界中の民間企業、官公庁、学校関係のユーザーが、不正を削減しリスクを緩和しながら、顧客を惹き付け、維持し、そして増やすことができる競争上の優位性のために、IBM SPSS テクノロジーに信頼をおいています。IBM SPSS ソフトウェアを日常業務に取り入れることにより、会社組

織は先見性のある企業となります - ビジネスのゴールを達成し、計測可能な競争上の優位性を達成するための意志決定を直接的、自動的に行うことが可能となります。詳細な情報、または、営業担当者へのご連絡には、<http://www.ibm.com/spss>を開いてください。

## テクニカル サポート

テクニカル サポートはカスタマーのメンテナンスに使用いただけます。IBM Corp. 製品の使用方法、または対応するハードウェア環境へのインストールについてサポートが必要な場合は、テクニカル サポートにご連絡ください。テクニカル サポートのご利用には、<http://www.ibm.com/support>のIBM Corp. ウェブ サイトをご覧ください。支援を要請される場合は、事前にユーザー、会社組織、そして、サポート契約を明確にしておいていただくよう、お願いします。

---

# 内容

## パート I: まず使ってみよう

<b>1</b>	<b>テキスト分析について</b>	<b>1</b>
	新機能	1
	自由記述式のアンケートデータ	2
	テキストマイニングについて	3
	抽出の方法	4
	カテゴリ化の方法	7
	テキスト分析前の準備	8
	信頼性と微調整	9
	言語リソースの調整	10
	カテゴリの定義の修正	10
<b>2</b>	<b>まず使ってみよう</b>	<b>11</b>
	典型的なプロセス	11
	テキスト分析ウィンドウ	12
	質問ビュー	13
	プロジェクト全体ビュー	15
	リソース エディタ ウィンドウ	16
	オプションの設定	19
	オプション: システム タブ	19
	オプション: [表示] タブ	21
	オプション: [サウンド] タブ	23
	オプション: [翻訳] タブ	24
	Microsoft Internet Explorer ヘルプの設定	25

## パート II: テキスト分析

### 3 プロジェクトおよびパッケージの作成 27

プロジェクトの作成	27
使用するデータの準備	28
新しいプロジェクトの開始	30
データソースの選択	31
変数の選択	36
英語への翻訳	38
カテゴリおよびリソースの選択	40
テキスト分析パッケージの使用	44
テキスト分析パッケージの作成	44
テキスト分析パッケージの更新	46

### 4 プロジェクトの使用 50

プロジェクトを開く	50
プロジェクトのプロパティの編集	52
プロジェクト データの表示	53
変数のソート	54
変数のプロパティの編集	54
プロジェクトの保存	55
カテゴリ化の結果をエクスポート	57
IBM SPSS StatisticsまたはIBM SPSS Data Collectionにエクスポート	58
Microsoft Excelにエクスポート	60
棒グラフのエクスポート	62
データソースの変更	65
データソースの選択	66
変数の選択	71
変数の合致	74
英語への翻訳	75
データの更新	77
英語への翻訳	77
プロジェクトの共有	79
フラグ回答	79
プロジェクトのステータスバー	80

## 5 データの抽出 83

抽出結果:コンセプト、タイプ、パターン	83
データの抽出	88
抽出結果の保存	91
抽出結果の調整	91
類義語の追加	93
コンセプトのタイプへの追加	95
コンセプトの抽出からの除外	97
単語を抽出に強制投入	98

## 6 テキスト データの分類 100

[カテゴリ] パネル	102
[データ] パネル	104
カテゴリの関連性	106
カテゴリ作成の方法と戦略	108
カテゴリ作成の方法	108
カテゴリ作成の方略	109
カテゴリ作成のヒント	110
最適な記述子の選択	111
カテゴリとは	115
カテゴリのプロパティ	116
カテゴリの作成	118
詳細言語設定	122
言語学的手法について	127
頻度の詳細設定	134
カテゴリの拡張	135
手作業でのカテゴリの作成	140
カテゴリの新規作成または名前の変更	140
ドラッグ & ドロップによるカテゴリの作成	141
定義済みのインポートおよびエクスポート	142
定義済みカテゴリのインポート	143
カテゴリのエクスポート	152
カテゴリ規則の使用	156
カテゴリ規則シンタックス	157
カテゴリ規則内の TLA パターンの使用	159
カテゴリ規則におけるワイルドカードの使用	161

カテゴリ規則の例	164
カテゴリ規則の作成	166
規則の編集および削除	168
カテゴリの編集および調整	169
カテゴリのプロパティの編集	169
記述子のカテゴリへの追加	171
カテゴリ記述子の編集	172
カテゴリの移動	173
カテゴリの結合・組み合わせ	174
回答のカテゴリへの強制投入	175
カテゴリ内のテキストマッチ	176
カテゴリパネルのコピー	178
カテゴリの印刷	178
カテゴリの削除	180

## 7 グラフの視覚化

181

カテゴリ棒グラフ	182
カテゴリ Web グラフ	183
カテゴリ Web テーブル	184
グラフのツールバーおよびパレットの使用	185
視覚化の編集	186
視覚化編集の一般的な規則	187
テキストの編集および書式設定	188
色、パターン、破線化、および透明度の変更	189
点要素の形および縦横比の回転および変更	191
グラフ要素のサイズの変更	192
余白とパディングの指定	192
数値の書式設定	193
軸および尺度設定の変更	194
カテゴリの編集	196
方向パネルの変更	198
座標システムの変換	199
統計量とグラフ要素の変更	200
凡例の位置の変更	204
視覚化および視覚化データのコピー	204
キーボードのショートカット	205



## パート III: リソース エディタ

### 8 テンプレートとリソース 207

エディタのインターフェイス . . . . .	208
テンプレートの作成および更新 . . . . .	211
リソース テンプレートの切り替え . . . . .	212
テンプレートの管理 . . . . .	213
テンプレートのインポートおよびエクスポート . . . . .	215
リソースのバックアップ . . . . .	216
リソース ファイルのインポート . . . . .	218

### 9 ライブラリの使用 221

付属ライブラリ . . . . .	221
ライブラリの作成 . . . . .	223
パブリック ライブラリを追加 . . . . .	224
キーワードおよびタイプの検索 . . . . .	225
ライブラリの表示 . . . . .	225
ローカル ライブラリの管理 . . . . .	226
ローカル ライブラリの名前の変更 . . . . .	226
ローカル ライブラリを使用不可に . . . . .	227
ローカル ライブラリの削除 . . . . .	227
パブリック ライブラリの管理 . . . . .	228
ライブラリの共有 . . . . .	230
ライブラリの公開 . . . . .	232
ライブラリの更新 . . . . .	233
競合の解決 . . . . .	234

### 10 ライブラリ辞書について 236

キーワード辞書 . . . . .	236
ビルトインのタイプ . . . . .	237
タイプの作成 . . . . .	238
キーワードを追加 . . . . .	240

キーワードの強制 . . . . .	244
タイプの名前の変更 . . . . .	245
タイプの移動 . . . . .	246
タイプの無効化および削除 . . . . .	247
類義語辞書 . . . . .	247
類義語の定義 . . . . .	249
オプションの要素の定義 . . . . .	251
類義語の無効化および削除 . . . . .	252
不要語辞書 . . . . .	253

## 11 アドバンス リソースについて 256

検索 . . . . .	257
置換 . . . . .	257
Fuzzy Grouping . . . . .	258
固有表現 . . . . .	260
正規表現の定義 . . . . .	261
正規化 . . . . .	263
構成 . . . . .	264
言語処理 . . . . .	265
抽出パターン . . . . .	266
強制定義 . . . . .	266
省略形 . . . . .	267

## 付録

### A 日本語テキストの例外 268

日本語テキストの抽出およびカテゴリ化 . . . . .	268
抽出の方法 . . . . .	268
二次抽出の手順 . . . . .	271
カテゴリ化の方法 . . . . .	273
日本語テキストのリソースの編集 . . . . .	274
日本語のライブラリ ツリー、タイプ、キーワードのパネル . . . . .	275
日本語テキストで使用できるタイプ . . . . .	278
日本語のタイプのプロパティの編集 . . . . .	282

日本語テキストの類義語辞書の使用 . . . . .	283
日本語リソースの検証およびコンパイル . . . . .	285
日本語についてのその他の例外 . . . . .	285

<b>B 注意</b>	<b>287</b>
-------------	------------

<b>索引</b>	<b>290</b>
-----------	------------



パート I:  
まず使ってみよう



# テキスト分析について

IBM® SPSS® Text Analytics for Surveys バージョン 4.0.1 の世界によるこそ。これは、自由記述式回答を有効に分析するためのアンケートテキストのコード化ツールです。この製品を使うと、アンケート調査を行う人はだれでも、構造化されていないアンケート回答を定量的なデータにすばやく変換できます。このように自由記述のテキストデータを“解放”することで、分析の質や意思決定能力が大幅に改善されます。この製品では、アンケートデータのインポート、主要キーワードの抽出、結果の調整、回答のカテゴリ化を行うことができます。使用しているデータをカテゴリ化したならば、これらのカテゴリを IBM® SPSS® Statistics などの定量的分析ツールにエクスポートし、さらに分析を行ったり、グラフを作成したりできます。

SPSS Text Analytics for Surveys は、自由記述式のアンケート回答内の主要キーワードを高い精度で抽出・分類するように、高度な言語学テクノロジーと手作業の手法を組み合わせています。しっかりした分類アルゴリズムと簡単なドラッグ&ドロップ機能を使用することで、アンケート回答をカテゴリ化するためのカテゴリ（“コード”）を作成できます。この作成されたカテゴリを再度使用することで、同一のあるいは類似した研究においても、一貫した結果が得られます。自由記述式の回答データはアンケートによって大きく異なるので、2つのプロジェクトが完全に同じになるということはありませんが、それでも分析の基本的なプロセスは同じであると想定できます。詳細は、2章 p.11 [典型的なプロセス](#) を参照してください。

## 新機能

IBM® SPSS® Text Analytics for Surveys 4.0.1 には、次のような新機能があります。

**階層カテゴリ:** カテゴリが階層的な構造を持つことができ、すなわち、サブカテゴリを含むことができ、また、そのサブカテゴリにもそれ自身のサブカテゴリを更に下の階層に向かって持たせることができます。製品内にこのような階層的な分類を構築することが可能であるだけでなく、階層的な分類を持ち、以前はコード フレームと呼ばれていた、定義済みの分類構造をインポートすることも可能です。

実際に、階層カテゴリにより、1個または複数のサブカテゴリを持つツリー構造を構築して、たとえば異なるコンセプトやトピックの分野の項目をより正確にグループ化することができます。レジャー活動に関し

て簡単な例を挙げることができます。「時間があればどんな活動がしたいですか?」という質問に対する答えとして、「スポーツ」、「日曜大工」、「釣り」などをトップ カテゴリに設定し、「スポーツ」の下の階層に「球技」、「水泳」などを設定できます。

**Language Weaverアクセス** Language Weaver 翻訳インターフェースにアクセスする方法は、単一URLおよび関連セキュリティを使用することで簡略化されています。

## 自由記述式のアンケートデータ

アンケートには、通常、自由記述式の質問や選択形式の質問など、異なった種類の質問が含まれています。**選択形式の質問**では、限られた個数の選択肢を提示し、さまざまな種類の定量的分析を行うことができます。一方、**自由記述式の質問**では、さまざまな長さや細かさの構造化されていない回答となります。

質問に対して回答する際に人々が使用する言葉は、これらの人たちが何を考え、また感じているのかについて多くのことを物語ってくれます。このために、アンケートにはよく自由記述式の質問が含まれています。しかしながら、自由記述式の質問を大量に使用することは、費用の面から見て無理があるとこれまで考えられてきました。

しかしながら、自由記述式の質問を大量に使用することは、費用の面から見て無理があるとこれまで考えられてきました。これは、これらの回答を解釈するには分析の費用がかかるからです。これに加えて、これらの長い回答を細かく標準化された手法でコード化しなくてはなりません。これは困難な作業となる可能性があります。なぜなら、コード化する人によって、同じ回答でも別のカテゴリに分けられることがよくあるからです。コード化する人たちの間で一致が見られない場合、結果の信頼性は低くなります。これらのようなさまざまな理由で、自由記述式回答のコード化は、時間やお金がかかり、収集されたデータから得られるメリットにつりあわないと、しばしば考えられてきました。

IBM® SPSS® Text Analytics for Surveysは、自由記述式回答のコード化を手作業よりも圧倒的に速く行うことで、費用がかかるこれまでの方法に替わるやりかたを提供します。高度な言語学理論とテクノロジーによって、語句や文章の文法構造が回答の意味を理解するための文脈を提供しているとして、SPSS Text Analytics for Surveysは自由記述式回答テキストを分析します。テキストを分析すると、キーワードならびに語のパターンが抽出され、カテゴリに分類されます。

ビルトインのカテゴリ作成手法を使ってカテゴリを自動的に作成した後に、結果を手作業で微調整できます。抽出とカテゴリ化が一貫性のある反復可能な方法で常に行われるので、結果の信頼性は劇的に向上します。カテゴリの定義や言語学ライブラリを微調整しない限り、同じ回答は常に同じカテゴリに分類されます。



アンケート分析がうまくいくためには、1つの手法だけに頼るわけにはいきません。自由記述式の回答はその主観的特徴により、解釈のために複数の技術が必要になります。ビルトインの抽出およびカテゴリ作成技法に加え、SPSS Text Analytics for Surveysは、各アンケートでの、ユーザーによる特定のテキスト分析の理解にも依存しています。テキスト分析は反復して実行（抽出、検討、修正、再抽出）された際に最も優れたものとなり、その有用性は、抽出結果やカテゴリ定義を手作業で検討・修正した時間や労力に応じて変わってきます。詳細は、p.9 [信頼性と微調整](#) を参照してください。

繰り返されるアンケートの中で、全く同じ、あるいは類似した質問を扱う場合には、他の質問やプロジェクトで使ったカテゴリを再び使用できます。カテゴリを再使用することで、コード化の一貫性が大きく向上し、また時間と労力が大幅に節約できます。

また、他の分析を実行したい場合もあります。SPSS Text Analysis for Surveysを使用して、一連の回答に隠されているカテゴリを発見したならば、SPSS Statistics Baseのような他のプログラムにエクスポートして定量的分析を行うこともできます。SPSS Text Analytics for Surveysを使用して回答セットの根拠となるカテゴリを発見した後は、IBM® SPSS® Statistics Baseなどのプログラムによる高度な数量分析のために、カテゴリをエクスポートすることもできます。

## テキストマイニングについて

定性的分析の形式であるテキスト分析では、自由記述式の回答などテキストからの役立つ情報を抽出し、このテキスト内の主要キーワードを適切な数のカテゴリにグループ化します。テキスト分析はすべての種類および長さのテキストに実行できますが、分析へのアプローチは若干異なります。

比較的短いレコードは、それほど複雑でなく、通常不明確な単語や回答があまり含まれていないため、最も容易にカテゴリ化されます。たとえば、短い自由記述式のアンケートで好きな休日の過ごし方を3つ挙げるよう質問した場合、ビーチに行く、国立公園に行く、または何もしないなどの多くの回答が見られることが予想される場合があります。一方、比較的長い自由記述式のアンケートの回答は、特に回答者が高学歴で意欲があり、またアンケートを記入するのに十分な時間がある場合、非常に複雑で長くなる場合があります。アンケートで政治に関する考えを尋ねる、または政治に関するブログ フィールドがある場合、あらゆる種類の問題および立場について、長いコメントがいくつかあると予想されることがあります。

アンケート調査の担当者は、とても長い回答を分析することはあまりありません。多くのアンケート調査の回答は、短いあるいは中程度の長さです（1つの文から短いパラグラフまで）。IBM® SPSS® Text Analytics for Surveys は、この程度の長さのテキストを処理するよう設計されていますが、さらに長い回答も分析できます。

まず、手作業というアプローチがあります。この場合、だれかがアンケートの回答を読んでその内容を記録し、そこに含まれる主要キーワードを判断して、それにコードを割り当てます。まず、手作業というアプローチがあります。この場合、だれかがアンケートの回答を読んでその内容を記録し、そこに含まれる主要キーワードを判断して、それにコードを割り当てます。人間はテキストの理解が得意なので、このアプローチは極めて精度が高いです。しかしながらこの方法だと時間がかかり作業量も多くなります。また最近では対象となるテキストの量が膨大なので、この方法はあまり実用的ではありません。またこのアプローチはコードを設定する人の解釈に大きく影響されます。

別のアプローチとして、自動化された方法の活用があります。自動化された手法にはさまざまあり、統計に基づくものや言語学に基づくものなどがあります。SPSS Text Analytics for Surveysでは、自動化された言語学的手法と統計的手法を組み合わせることで、プロセスの各段階において最も信頼できる結果が得られるようになっていています。本製品では、回答から主要キーワードを自動抽出する際には言語学に基づいた手法を使用し、回答に割り当てるカテゴリ定義（コード）を作成する際には、言語学的手法と統計的手法の両方を使用できます。

## 抽出の方法

回答の主要キーワードの抽出時、IBM® SPSS® Text Analytics for Surveysは言語学に基づくテキスト分析に依存します。このアプローチを用いると統計に基づくシステムがもたらすようなスピードと費用対効果が得られます。また人の手を介することがほとんどないので、極めて高い精度が得られます。言語学に基づくテキスト分析は、自然言語処理、あるいは計量言語学と呼ばれる研究分野に基づいています。

抽出プロセスにおける統計的アプローチと言語学的アプローチとの違いを説明するために、reproduction of documents（ドキュメントの複製）についての質問に、それぞれがどのように回答するかを考えてみましょう。統計的ソリューションおよび言語学的ソリューションのいずれも、reproduction（複製）という単語を展開して、copy（コピー）やduplication（重複）などの類義語を含めるようにする必要があります。展開しない場合、関連情報が見落とされてしまいます。ただし、統計的ソリューションによって、こうした種類の同義語集、同じ意味を持つ他のキーワードを検索使用する場合、birth（誕生）というキーワードも加わり、関連しない多くの結果を生成する場合があります。言語について理解することによって、定義というより信頼できるアプローチによって、言語学的テキストマイニングを実行してテキストの曖昧さに切り込みます。

抽出プロセスがどうなっているのかを理解しておく、言語リソース（ライブラリ、タイプ、類義語など）を微調整する際に役に立ちます。抽出プロセスのステップには以下のものがあります。

- ソース データの標準フォーマットへの変換

- 候補のキーワードの特定
- 類義語の等価クラスおよび統合の特定
- タイプの割り当て
- インデックス化
- パターンおよびイベント抽出の合致

### 手順 1: ソース データの標準フォーマットへの変換

最初のステップでは、後続の分析に利用できるように、インポートしたデータを決まった形式に変換します。この変換は内部的に実行され、元のデータは変更されません。

### 手順 2: 候補のキーワードの特定

言語学的抽出において、候補となるキーワードを特定する際の言語リソースの役割を理解しておくのは大切なことです。言語リソースは、抽出が実行されるごとに使用されます。言語リソースは、テンプレート、ライブラリ、およびコンパイル済み辞書の形式で保存されています。ライブラリには、語のリスト、関係性、また抽出の実行や調整に使用される情報が含まれています。基幹辞書は表示・編集ができません。ただし、残りのリソース（テンプレート）をリソース エディタ で編集できます。

コンパイル済み辞書は、IBM® SPSS® Text Analytics for Surveys の抽出エンジンの主要な、内部コンポーネントです。これらのリソースには、品詞コード（名詞、動詞、形容詞、副詞、分詞、決定詞）を含む基本形のリストを収めた一般辞書が含まれています。また、リソースには、<地名>、<組織名>、または<人名>に多くの抽出されたキーワードを割り当てるために使用する、予約済みのビルトインのタイプも含まれています。詳細は、10 章 p.237 [ビルトインのタイプ](#) を参照してください。

これらコンパイル済み辞書のほか、製品にはいくつかのライブラリが付属し、それらを使用して、コンパイル済み辞書のタイプ定義およびコンセプト定義を補い、またその他のタイプや類義語を提供することができます。これらのライブラリ、および作成したユーザー指定のライブラリは、いくつかの辞書で構成されています。これらには、キーワード辞書、類義語辞書（類義語およびオプションの要素）、および不要語辞書が含まれています。詳細は、9 章 p.221 [ライブラリの使用](#) を参照してください。

データがインポートおよび変換されると、抽出エンジンは抽出の候補のキーワードの特定を開始します。候補となるキーワードとは、テキスト内の概念を特定するのに使用される語や、語の集まりのことです。テキストの処理中、コンパイル済み辞書にない単語（**ユニターム**）は、抽出の候補のキーワードとして見なされます。候補の複合語（**マルチターム**）は、品詞パターン抽出を使用して特定されます。たとえば、品詞パターンが「形容詞、名詞」のマルチターム sports car（**スポーツ カー**）は、2 つの部分に分けられます。品詞パターンが「形容詞、形容詞、名詞」のマルチターム fast sports car（**高速スポーツ カー**）は、3 つの部分に分けられます。

注: 前述のコンパイル済み一般辞書にあるキーワードは、ユニタームとして重要でないまたは言語学的にあいまいであるすべての単語を示します。これらの単語は、ユニタームを特定するときに不要語に追加されます、ただし、それらは、品詞を決定またはより長い候補の複合語（マルチターム）を参照している場合に再評価されます。

最後に、特殊なアルゴリズムを使用して、役職などの大文字の文字列を処理し、これらの特殊なパターンを抽出できるようにします。

### 手順 3: 類義語の等価クラスおよび統合の特定

候補のユニタームおよびマルチタームが特定された後、一連のアルゴリズムを使用して、ユニタームやマルチタームを比較し、等価クラスを特定します。等価クラスは、ある語句の基本形、すなわち同じ語句の2つの表現を1つの形で表わしたものです。句を等価クラスに割り当てる目的は、たとえば、`president of the company`（会社の社長）および `company president`（会社社長）を別のコンセプトとして扱わないようにすることです。等価クラスのどのコンセプトを使用するか、つまり、`president of the company`（会社の社長）または `company president`（会社社長）のどちらを主要キーワードとして使用するかを判断するために、抽出エンジンは、次の規則を順に適用します。

- ライブラリのユーザー指定の形式。
- テキスト全体で最も出現頻度の高い形式。
- テキスト全体で最も短い形式（通常、基本型に該当）。

### 手順 4: タイプの割り当て

次に、抽出されたコンセプトにタイプを割り当てます。タイプは、コンセプトの意味上のグループ化です。基幹辞書ならびにライブラリの両方がこのステップで使用されます。タイプには、上位レベルのコンセプト、肯定的な単語および否定的な単語、人名、地名、組織名などが含まれます。ユーザーがタイプを定義して追加することもできます。詳細は、10 章 p. 236 [キーワード辞書](#) を参照してください。

### 手順 5: インデックスの付与

レコードのセット全体に、テキストの位置と各等価クラスの代表キーワードの間にポインタを確定してインデックスを付けます。候補のコンセプトの活用形インスタンスはすべて、候補の基本型としてインデックスが付けられます。基本形ごとに全体の出現頻度が計算されます。

### 手順 6: パターンおよびイベント抽出の合致

IBM SPSS Text Analytics for Surveys は、タイプやコンセプトだけでなく、それらの関係性も見つけることができます。この製品ではいくつかのアルゴリズムおよびライブラリを使用でき、またタイプおよびコンセプトの間の関係性パターンを抽出する機能が用意されています。製品に対する反

応などの特定の意見、または政治的グループやゲノムのリンクなど、人々またはオブジェクトの間の関係性リンクを探す場合に特に役立ちます。

## カテゴリ化の方法

いくつかの手法から選択して、カテゴリを作成できます。すべてのデータセットが一意であるため、手法の数やそれらを適用する順序は、変わる場合があります。結果の解釈が、他の人とは異なる場合があるため、テキスト データにとってどの手法が最良の結果を生み出すか、それぞれの手法を検証する必要があります。

このガイドの場合、**カテゴリの作成**は、カテゴリ定義の生成および、1つまたは複数のビルトインの手法を使用した分類を指し、また**カテゴリ化**は、スコアリング、またはラベル付け、一意の識別子（名前/ID/値）を各レコードのカテゴリ定義に割り当てるプロセスのことを指します。

カテゴリ作成時、抽出されたコンセプトおよびタイプはカテゴリの構築ブロックとして使用されます。カテゴリを作成すると、カテゴリ定義の要素に一致するテキストが含まれる場合、レコードが自動的にカテゴリに割り当てられます。

IBM® SPSS® Text Analytics for Surveys には、自動カテゴリ作成手法がいくつか用意されており、レコードを迅速にカテゴリ化することができます。

### グループ化手法

使用できるそれぞれの手法は、特定の種類のデータおよび状況に適していますが、同じ分析で手法を組み合わせ、全範囲のレコードをキャプチャすると役に立つ場合があります。複数のカテゴリのコンセプトを表示したり、重複するカテゴリを見つけることができます。

**派生関係のコンセプトの語幹:** コンセプト コンポーネントが形態的に関連するか、または語幹を共有するかどうかを分析するとき、コンセプトを取得し、そのコンセプトに関連するその他のコンセプトを検索することによって、カテゴリを作成します。この手法は、生成された各カテゴリのコンセプトが類義語または意味の上で密接に関連しているため、類義語の複合語コンセプトを特定するのに非常に役立ちます。長さの異なるデータを処理し、コンパクトなカテゴリをより少なく生成します。たとえば、コンセプト `opportunities to advance` は、コンセプト `opportunity for advancement` および `advancement opportunity` とグループ化されます。詳細は、6 章 p.128 [派生関係のコンセプトの語幹](#) を参照してください。

**セマンティック ネットワーク:** 各コンセプトの考えられうる意味を、単語の関係の拡張インデックスから特定することによって開始し、関連するコンセプトをグループ化することによってカテゴリを作成します。この手法は、コンセプトがセマンティック ネットワークに認識され、あまり曖昧でない場

合に最も適しています。テキストに、ネットワークが認識していない特殊な用語または専門用語が含まれている場合はあまり役に立ちません。たとえば、コンセプト granny smith apple は、granny smith と横の関係があるため、gala apple および winesap apple とグループ化されます。またあるいは、コンセプト animal (動物) は、その下位語である cat (ネコ) および kangaroo (カンガルー) とグループ化されます。このリリースでは、英語テキストにのみ使用できます。詳細は、6 章 p.131 [セマンティック ネットワーク](#) を参照してください。

**内包関係のコンセプト:** この手法では、一方の共通の文字列である単語を含むかどうかに基づき、マルチタームのコンセプト (複合語) をグループ化することによってカテゴリを作成します。たとえば、コンセプト seat (シート) は、コンセプト safety seat (セーフティ シート)、seat belt (シート ベルト)、および seat belt buckle (シート ベルトのバックル) とグループ化されます。詳細は、6 章 p.130 [内包関係のコンセプト](#) を参照してください。

**共起:** この手法では、テキスト内の共起関係のコンセプトからカテゴリを作成します。ドキュメントおよびレコードでコンセプトまたはコンセプト パターンがいっしょに出現することが多いとき、共起関係のコンセプトはおそらくカテゴリ定義の値のものである基底となる関連を反映します。単語が頻繁に共起する場合、共起規則が作成され、新しいサブカテゴリのカテゴリの記述子として使用できます。たとえば、多くのレコードに単語 price (価格) および availability (有効性) が含まれている場合 (ただし、一方を含み、もう一方を含まないレコードはほとんどない)、これらのコンセプトを共起規則 (price & available) にグループ化し、たとえばカテゴリ price のサブカテゴリに割り当てることができます。詳細は、6 章 p.132 [共起規則](#) を参照してください。

- **最小レコード数:** 共起関係のコンセプトの重要性を判断できるようにするため、カテゴリの記述子として使用されるよう、指定の共起関係のコンセプトを含む必要のあるレコードの最小数を定義します。

## テキスト分析前の準備

テキスト分析は、抽出とカテゴリ化だけではありません。テキスト分析を成功させるためには、以下の点に注意してください。

- アンケートの設計時、IBM® SPSS® Text Analytics for Surveys にインポートする回答の質はカテゴリ化の結果に直接影響します。一般的に言って、曖昧ではっきりしない質問の回答はまとまりのないものになり、分析がとてつもない困難です。
- 統計分析の場合と同様に、テキスト分析ははっきりとした目標に基づいて実行されるべきです。分析をする前にはいつも、行おうとしている研究について熟考し、何を知らうとしているのかを決める必要があります。

たとえばある学区内で、子どもが受けている教育の質に対する親の態度を調べるアンケートが実施されたとします。分析をする際には、教師の名前や、学校のカリキュラムなどに焦点を絞ることができます。あるいは肯定的な意見や否定的な意見の特定とグループ化に焦点を絞ることもできます。同様に、この分析をどのくらい細かく行うのかについても定義しておく必要があります。たとえば助成金に対する意見をすべてひとまとめにするのか、あるいはこれを各助成金ごとに分けるのかということです。作成されるコードやカテゴリは、分析の焦点や目的を反映したものでなくてはなりません。

- 統計分析の場合以上に、テキスト分析は厳密には科学ではありません。なぜなら1つの「正しい」結果があるわけではないからです。テキスト分析はある目的を想定して実行されるものですが、回答者が伝えようとしたメッセージは分析者の解釈によって影響を受けてしまう主観的なものです。たとえば、皮肉な態度を特定し分類する場合には主観的に判断せざるを得ません。その目的や焦点によっては、同じデータを分析したとしても、人によって別の結論が導き出されることもあります。
- テキスト分析は極めて「反復的なプロセス」です。アンケート回答を扱う際には、別のカテゴリ定義（つまりコード化スキーム）や、別のカテゴリ定義ならびに類義語定義、別のグループ化手法を使って再度抽出し、再度カテゴリ化することが必要になるでしょう。

テキストからコンセプトを抽出し、カテゴリを作成したならば、結果を注意深く検討してください。調整したい要素がある場合は、カテゴリ定義と言語ライブラリ定義を微調整して分析を調整してください。これによって、再抽出時に回答が自動的に再度カテゴリ化されます。分析結果に満足できるようになるまで、このプロセスを何度でも繰り返すことになります。

注: データのインポート前の注意に関する詳細は以下を参照してください。 [使用するデータの準備 p. 28](#)

## 信頼性と微調整

データをコード化する際には、結果として得られるカテゴリが信頼できるものでなくてはなりません。これは自由記述式の回答のコード化においては、コード化を別々の人が行っても、ある回答に対しては同じ規則（コード化フレーム）を使用して、全く同じようにコード化がされなくてはならないということです。テキスト分析を手作業で行う場合には、これは極めて重要な問題です。有効な複数のカテゴリを作成できたとしても、回答への適用が信頼できる方法で行われなかったら、その価値は著しく損なわれることになります。IBM® SPSS® Text Analytics for Surveysが、同じ言語リソースと共に、同じデータに適用されると、以前の分析を常に完璧に再現します。この信頼性は100%です。

しかしながら、これは分析には間違いが全くないということではなく、コード化で注目している対象が「微調整」という部分に移ったということです。手動でコード化を行う場合、操作者は回答を読んですべての記述のニュアンスを理解できます（カテゴリのコード化が難しい場合でも）。SPSS Text Analytics for Surveysはカテゴリにコード化を適用できますが、カテゴリはニュアンスや特徴が理解できるように定義される必要があります。微調整を行うには2つの方法があります。

- 言語リソースの調整
- カテゴリ定義の調整

## 言語リソースの調整:

IBM® SPSS® Text Analytics for Surveysでは、われわれが何も手を下さなくても簡単にカテゴリを作成しますが、回答に含まれるすべての情報を把握できるというわけではありません。このソフトウェアが使用する言語学的基盤を改善することで、個々のテキストの特徴により特化したカテゴリを作成できるようになります。この基盤を改善するために、テキストの抽出に使用される言語リソースをカスタム化し、微調整することができます。

ここで言う微調整には、語や語句のさまざまな言語学ライブラリや辞書への追加や、分析に不要な語の指定、類義語の定義、特定の目標を想定したカスタムライブラリの作成などの作業が含まれます。ここでの目標は、テキストにおける回答者の考えを正確に把握し、結果から曖昧さを取り除くことにあります。

## カテゴリの定義の修正

言語リソースを修正するだけでなく、カテゴリを検討してこれらを結合したり、定義を整理する必要があります。またカテゴリ化された回答を個別にチェックする必要があります。自動カテゴリ作成の手法を使ってカテゴリを作成することもできますが、いくつかの点でこれらの定義を調整することが必ず必要になるでしょう。何らかの手法が適用されると、ウィンドウ内にたくさんの新規カテゴリが表示されます。カテゴリを展開すると各カテゴリに含まれるコンセプトが表示されます。カテゴリ内の回答を検討して、満足が行くまでカテゴリの定義を調整してください。

自動的手法では、データを完全にカテゴリ化できません。そのため、データに合った1つまたは複数の自動的手法を見つけ、適用することをお勧めします。手作業で若干の調整を行い、間違った分類を除去したり、把握されなかったレコードや語を追加することもできます。



# まず使ってみよう

ここでは、IBM® SPSS® Text Analytics for Surveys で実行できる作業や、回答をカテゴリ化する手法について紹介します。ここに示されている情報は、一番最初に行う分析の手順を示すものです。データの分析に必要なプロセスはすべて説明していますが、1つ1つのデータセットはそれぞれ異なっているので、自分の分析をどの時点で完了とするのかは、最終的には自分で判断しなくてはなりません。

この章では、テキスト分析を行う際に、ユーザーが体験する典型的なプロセスについて説明します。インターフェイスの概要や、実際に行う主な作業や要素についても説明しています。

## 典型的なプロセス

IBM® SPSS® Text Analytics for Surveysの使用にあたっては、次の典型的なワークフロー プロセスにしたがってください。

- アンケートデータをインポートして**プロジェクトの作成**を行います。データには自由記述形式の回答、SPSS Text Analytics for SurveysのID変数や参照変数が含まれます。IBM® SPSS® Statisticsデータファイル、Microsoft Excel、ODBC準拠のデータベースプログラム、IBM® SPSS® Data Collectionデータソースからデータを読み込むことができます。定義済みのカテゴリや特化リソースから得られるテキスト分析パッケージを選択して、起動と実行を素早く行うこともできます。
- **抽出**：インポートした自由記述式の質問からコンセプトやパターンを抽出します。内部抽出エンジンは、テキスト内のキーワードを自動的に識別して収集します。これらのキーワードはメイン コンセプトの下にグループ分けされています。これらのキーワードは、タイプにもまとめられます。タイプは、組織、製品、肯定的な意見などの類似したキーワードの集合体です。またパターンも抽出されます。パターンはキーワードやタイプが組み合わせられて、意見や関係性（ある組織に関する肯定的なコメントなど）を表現しているものです。
- **修正**：語のタイプ、キーワード、類義語、不要語辞書を含む、1つあるいは複数のライブラリを直接扱うことで、抽出されたキーワードを修正したり、抽出を微調整したりできます。前述のように、テキスト分析は反復作業であり、ライブラリや辞書を修正することで、各データに合わせて調整された結果を得ることができます。

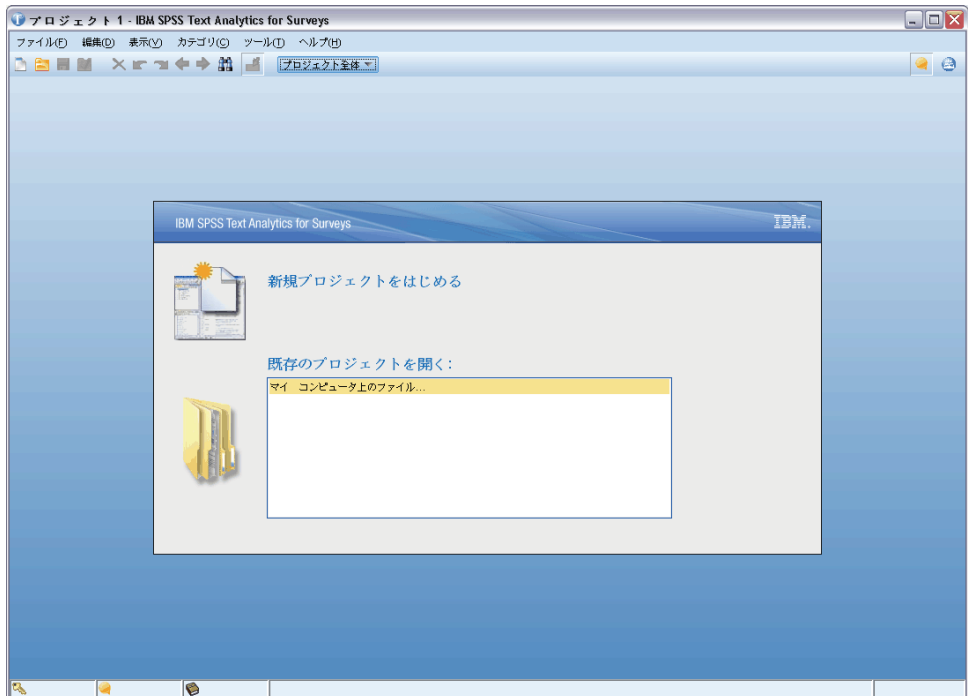
- **カテゴリ化** カテゴリ ルール、コード フレーム、カテゴリ自動作成技法を利用した、カテゴリの手動作成および編集による回答カテゴリとは、回答者が示した主な考え方や態度を捉えた、大まかな概念のことであり、
- **エクスポート** : カテゴリとID変数を一般的なファイル形式にエクスポートして、他のアプリケーションでさらに分析したりグラフを作成したりできます。出力形式としては、SPSS StatisticsファイルやMicrosoft Excelファイル形式の、一連のマルチアンサー（多重回答）変数などがあります。

## テキスト分析ウィンドウ

本製品のインターフェイスは2つのウィンドウから構成されています。このウィンドウでデータ内の各質問を分析します。それぞれの質問に関してキーワード、タイプ、パターンを抽出し、それから回答をカテゴリ化します。このウィンドウでは、データ内の各質問を分析できます。各質問では、各コンセプト、タイプ、パターンを抽出し、回答をカテゴリ化できます。

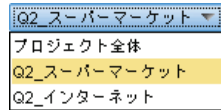
アプリケーションをスタートすると、既存のプロジェクトと新規作成されたプロジェクトを開くことができる画面が表示されます。新しいプロジェクトを作成すると、ウィザードが開き、プロジェクト作成プロセスに進むことができます。詳細は、3章 p.27 [プロジェクトの作成](#) を参照してください。

図 2-1  
製品起動時のテキスト分析ウィンドウ



データをインポートすると、以下を確認できます。2つのビューとは、[質問] ビューと[プロジェクト全体] ビュー。ビューを変更するには、テキスト分析ウィンドウ内のツールバーのドロップダウンリストから選択するか、[表示] メニューからいずれかのビューを選択します。リストボックスに表示されるテキストは、各質問の変数ラベルから取られたものです。

図 2-2  
ビューセクター ドロップダウンリスト付のアプリケーション ツールバー



## 質問ビュー

[質問] ビューでは、特定の質問に対する回答を分析し、カテゴリ化します。新しいプロジェクトの作成後、インポートした最初の自由記述変数に対する回答と質問ビューが表示されます。ドロップダウンリストからこのビューを選択するか、ビュー>質問>メニューから質問名を選択できます。

図 2-3  
質問ビュー

カテゴリ	種グラフ	選択 %	回答者	合計 %
size		9.1	2	9.1
photos		4.5	1	0.7
music		18.2	4	18.5
computer network		4.5	1	1.2
listening		9.1	2	6.7
design		4.5	1	4.2
consumer electronics		4.5	1	0.0
place of business		4.5	1	3.5
memory device		36.5	21	7.2
function		4.5	1	0.7
playlists		4.5	1	1.7
color		4.5	1	2.5
screen		4.5	1	0.5
radio		4.5	1	2.2
songs		13.6	3	7.4

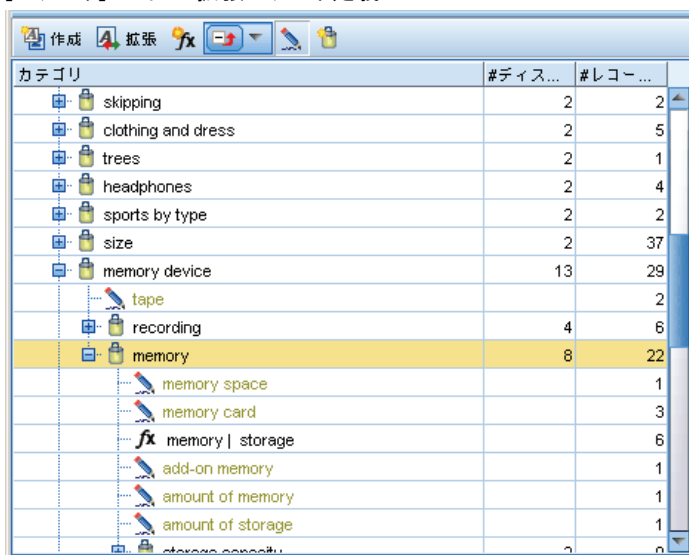
id	回答	カテゴリ
1	I like that Product A has a lot of storage. Also, the interface is very easy to use.	memory device/memory
2	Large storage capacity	...vice/memory/storage capacity
3	Small size. It has 512mb of <b>RAM</b> , so it is quick to load and play music. It can also encode directly from external devices from the radio or a CD player.	...sumer electronics/home audio memory device/memory music radio size
4	storage capacity	...vice/memory/storage capacity
5	Big storage capacity, also does video	...vice/memory/storage capacity memory device/recording/video
6	Large storage capacity and a good LCD screen for viewing digital photos	...vice/memory/storage capacity photos screen
7	Storage capacity (40GB)	...vice/memory/storage capacity

質問ビューで実行する操作は3つの要素を中心に展開しています。抽出結果、カテゴリ、回答データ。これらの要素それぞれに対する作業や分析を行いやすいように、このウィンドウは4つの部分に分かれています。

## [カテゴリ] パネル

左上にあるこの領域は、構築したカテゴリを管理できる表が表示されます。テキスト データからコンセプトとタイプを抽出した後、セマンティック ネットワークや内包関係のコンセプトなど、自動的な方法を使用して、または手動で作成してカテゴリを構築できます。カテゴリをクリックして拡張子、コンセプト、タイプ、ルールなどを定義するすべての記述子を確認できます。カテゴリまたは記述子を選択すると、[データ] パネルと [視覚化] パネルで対応するレコードの情報を確認することができます。詳細は、6 章 p.102 [カテゴリ] パネル を参照してください。

図 2-4  
[カテゴリ] パネル:拡張カテゴリ定義



カテゴリ	#ディス...	#レコー...
skipping	2	2
clothing and dress	2	5
trees	2	1
headphones	2	4
sports by type	2	2
size	2	37
memory device	13	29
tape		2
recording	4	6
memory	8	22
memory space		1
memory card		3
memory   storage		6
add-on memory		1
amount of memory		1
amount of storage		1
storage capacity		0

## [抽出結果] パネル

左下のこの領域には、抽出結果が表示されます。抽出を実行すると、抽出エンジンがテキスト データを読み込み、関連コンセプトを特定し、それぞれにタイプを割り当てます。コンセプトは、テキスト データから抽出した単語や句です。タイプは、キーワード辞書の形式で保存されたコンセプトのセマンティック グループです。抽出が完了すると、コンセプトとタイプが [抽出結果] パネルに表示されます。コンセプトおよびタイプはカラーコード化され、タイプを特定することができます。ここで収集したコンセプト、タイプ、パターンを使用して、カテゴリを作成できます。詳細は、5 章 p.83 抽出結果:コンセプト、タイプ、パターン を参照してください。

テキスト マイニングは、抽出結果をテキスト データのコンテキストに従ってレビューし、新しい結果を作成するよう調整、そして再評価するインタラクティブ プロセスです。言語リソースを修正することによって、抽出結果を修正できます。この調整は、[抽出結果] パネルまたは [データ] パ

ネルから部分的に直接、またリソース エディタ ビューから直接実行できます。詳細は、p. 16 [リソース エディタ ウィンドウ](#) を参照してください。

### [データ] パネル

このビューの右側にあります。ここには選択された自由記述式の質問に対する回答データが表形式で表示されます。デフォルトでは、データパネルには列が3つあります（レコードID、テキスト形式の回答、割り当てられたカテゴリ）。このパネルに表示される回答の個数は、他のパネルでの選択によって制限されています。インポートされたデータはこのパネルで見えることはできますが、編集、削除、レコードに印を付ける（フラグを立てる）ことはできません。詳細は、6章 p.104 [\[データ\] パネル](#) を参照してください。

### [視覚化] パネル

このビューの右上にあります。デフォルトでは、これは表示されません。このパネルはメニューから表示できます（[\[表示\]>\[視覚化\]](#)）。このパネルには、回答がどのようにカテゴリに適合しているのか、あるいは複数のカテゴリがどのように回答を共有しているのかに関するビューが3つあります（webグラフ、棒グラフ、テーブル）。これは他のパネルでの選択に基づいています。詳細は、7章 p.181 [グラフの視覚化](#) を参照してください。

新しいプロジェクト ウィザードで抽出オプションを選択するかどうかに応じて、左下のパネルに抽出結果が表示されるかが決まります。抽出結果パネルの[\[抽出\]](#) ボタンをクリックすると、抽出が始まります。抽出が終わったら、類義語を1つのキーワードにまとめたり、一般的すぎるキーワードをリストから除外するなどの微調整が必要かどうかを検討します。

抽出結果が満足できるものになったならば、手作業でキーワードをカテゴリにドラッグ&ドロップしたり、あるいはキーワードの内包などの分類手法を使って、回答をカテゴリ化できます。

## プロジェクト全体ビュー

プロジェクト全体ビューには、このプロジェクトにインポートされたすべての変数の概要が表示されます。このビューはドロップダウンリストあるいは[\[表示\]>\[プロジェクト全体\]](#) から選択できます。このビューでは、インポートされたデータの検討、変数の役割の変更（たとえば、質問から参照変数に）、値やラベルの変数への割り当てなどができます。詳細は、4章 p.53 [プロジェクト データの表示](#) を参照してください。インポートされたデータはこのビューで見えることはできますが、編集、修正、削除、レコードに印を付ける（フラグを立てる）ことはできません。

注: このビューでセルの内容を見るには、セル上にカーソルを移動させます。ツールチップにセルの内容が表示されます。

図 2-5  
プロジェクト全体ビュー

項目ID	質問	回答	性別	年齢
1	1	平日に時間があまりないときや休日のんびりしたい時	男	20代
2	2	休日の朝、昼	男	40代
3	3	休日、母が毎に食事をこ馳走したくなったとき	女	30代
4	4	コミュニケーションのため	男	40代
5	5	実度するが面倒くさい程度する時間がない自分では作...	女	40代
6	6	行き先をお店を見つけたときや料理を採りたい時など	女	30代
7	7	友人と久しぶりに会う時や、気の合う仲間と気軽に情報...	女	30代
8	8	家で作るようなものが食べなくなった時	女	30代
9	9	旧知の贈り物、送別会、O.B会、ビジネス関係者	男	50代
10	10	お昼	男	30代
11	11	外食がメインの用事ではない外出時	男	30代
12	12	気が向いたとき	男	30代
13	13	家で作るのが面倒なとき、食事を友人ととるとき	男	30代
14	14	気分転換、時間に余裕があるとき。久しぶりに友人と再...	女	30代
15	15	お金があるとき	男	30代
16	16	友人と出かけたとき	女	20代
17	17	いつもの家庭料理の味に飽きた時や、気分を変えたいと...	女	40代
18	18	友人と会う時やデートの時	女	20代
19	19	友人に誘われたとき	男	20代
20	20	焼かない時は殆ど外食です	男	20代
21	21	休日に外出し、帰りが遅くなった場合に家族と利用する...	男	30代
22	22	TVでお勧めの店の特徴と見た時や料理をするのが面倒な時	男	30代
23	23	仕事帰りの夕食。日常の食事の代わり	男	20代
24	24	友人と会うときに。家族と食事をとるときに。	女	30代
25	25	夜飲みに行くとき	女	20代
26	26	家でご飯を作るのが面倒な時。土日に外出した時に、昼...	男	40代
27	27	外で会う約束をした時	男	30代
28	28	家にも食べれるものがないとき。夕食を作るのが面倒く...	男	40代
29	29	職場の同僚とコミュニケーションを図りたいとき。友人...	女	30代
30	30	週末に友人と遊ぶ際に、外食します	男	20代
31	31	友人と会うとき。金曜日、土日。	女	20代

## リソース エディタ ウィンドウ

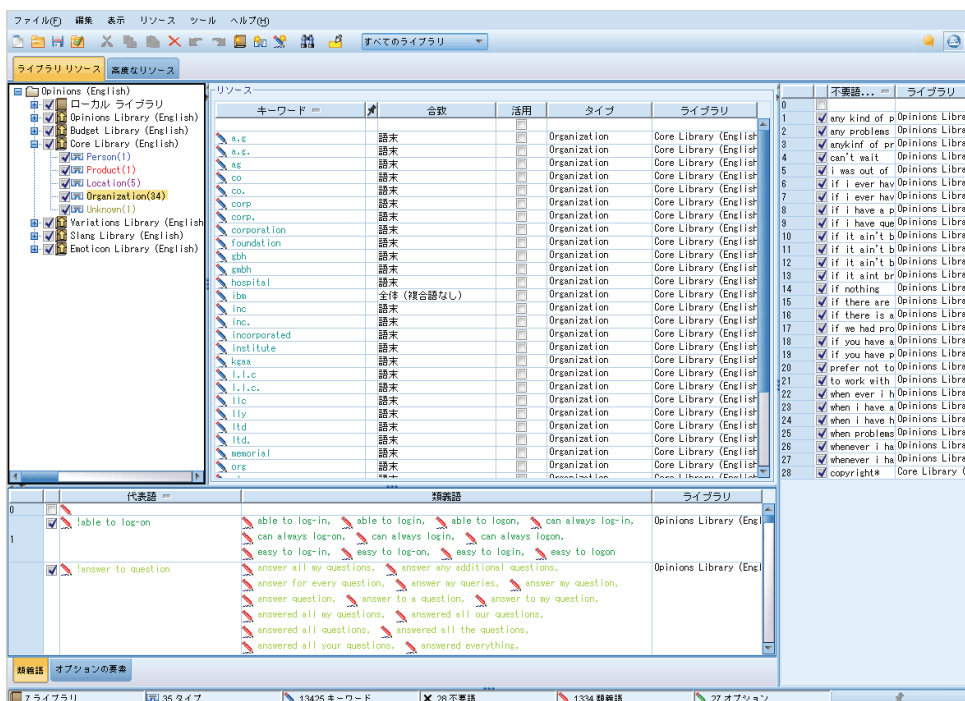
IBM® SPSS® Text Analytics for Surveys は、頑健な抽出エンジンを使用して、主要なコンセプトをテキストデータから迅速かつ正確にキャプチャします。このエンジンは、大容量の非構造化テキスト データをどのように分析および解釈するかを示す言語リソースによって大きく異なります。

リソース エディタ ビューでは、コンセプトの抽出、タイプに基づいたグループ化、テキスト データでのパターンの検出などに使用する言語リソースを表示および調整できます。IBM® SPSS® Text Analytics for Surveys では、事前設定されたリソース テンプレートをいくつか用意しています。また、一部の言語では、テキスト分析パッケージのリソースを使用することもできます。詳細は、3 章 p.44 [テキスト分析パッケージの使用](#) を参照してください。

これらのリソースは、常にデータのコンテキストに完全に対応しているとは限らないため、リソース エディタ で特定のコンテキストまたはドメインの独自のリソースを作成、編集および管理できます。詳細は、9 章 p.221 [ライブラリの使用](#) を参照してください。製品に付属するリソースで十分であるため、一部のユーザーはこのウィンドウをあまり使用しません。また、実行する場合がある単純なライブラリの作業の多くは、テキスト分析ウィンドウの [抽出結果] パネルから直接実行できます。

言語リソースの調整プロセスを簡略化するために、一般的な辞書タスクを、[抽出結果] パネルおよび [データ] パネルのコンテキストメニューを使用して、テキスト分析ビューから直接実行できます。詳細は、5 章 p. 91 [抽出結果の調整](#) を参照してください。

図 2-6  
リソース エディタ ビュー



リソース エディタ で実行する操作は、言語リソースの管理および調整を中心に展開しています。これらのリソースは、テンプレートおよびライブラリの形で保存されています。リソース エディタビューは次のような4つの部分で構成されています：ライブラリ ツリー パネル、キーワード辞書 パネル、不要語辞書パネル。

インターフェイスは、次のような 4 つの部分で構成されています。

**1. [ライブラリ ツリー] パネル:** 左上部分にあり、ここにはツリーが表示されます。このツリー内のライブラリを使用可あるいは使用不可にしたり、ツリー内でライブラリを選択することで、表示内容を制約することもできます。このツリーでは、右クリックをしてポップアップメニューを使って、さまざまな操作を実行できます。ツリー内のライブラリを展開すると、そこに含まれるタイプが表示されます。特定のライブラリのみにも焦点を当てたい場合、[表示] メニューからこのリストをフィルタリングすることもできます。

**2. [キーワード辞書のキーワード リスト] パネル:** ライブラリ ツリーの右側にあるこのパネルには、ツリーで選択されたライブラリのキーワード辞書のキーワード リストが表示されます。**キーワード辞書**は、ラベル (タイプ) ごとにまとめられた語を集めたものです。抽出エンジンがテキストデータを読み込むとき、テキスト内の単語をキーワード辞書内で定義されている語と比較します。抽出されたキーワードがキーワード辞書内にある場合、このタイプ名が割り当てられます。キーワード辞書は、(タイプという) 共通点でキーワードをまとめた個別の辞書と考えていいでしょう。たとえば、コア ライブラリの <Location> タイプには、new orleans、great britain、および new york などのコンセプトが含まれます。これらのキーワードはすべて、地名を示します。ライブラリには、1 つまたは複数のキーワード辞書が含まれます。詳細は、10 章 p.236 **キーワード辞書** を参照してください。

**3. [不要語辞書] パネル:** 右側にあるこのパネルには、最終的な抽出結果から除外されるキーワードの集合が表示されます。ここに登録されたキーワードは [抽出結果] パネルには表示されません。不要語キーワードは選択するライブラリに保存できます。ただし、[不要語辞書] パネルには、ライブラリ ツリーに表示されるすべてのライブラリに不要語登録されたすべてのキーワードが表示されます。詳細は、10 章 p.253 **不要語辞書** を参照してください。

**4. [類義語辞書] パネル:** 左下にあるこのパネルには、類義語およびオプションの要素がそれぞれのタブに表示されます。類義語およびオプションの要素を使用すると、最終的な抽出結果の代表語に基づいて類似したキーワードをグループ化できます。この辞書には既知の類義語やユーザー定義の類義語および要素、そして一般的なスペルミスと正しいスペルのペアが含まれています。類義語の定義およびオプションの要素は、選択するライブラリに保存できます。ただし、[類義語辞書] パネルには、ライブラリ ツリーに表示されるすべてのライブラリのすべての内容が表示されます。このパネルにはすべてのライブラリのすべての類義語またはオプションの要素が表示されますが、ツリーのすべてのライブラリの類義語は、このパネルでいっしょに表示されます。1 つのライブラリに含まれる類義語辞書は 1 つだけです。詳細は、10 章 p.247 **類義語辞書** を参照してください。

注:

- 1 つのライブラリに関する情報のみ表示されるようフィルタリングしたい場合、ツールバーのドロップダウン リストを使用して、ライブラリビューを変更できます。[すべてのライブラリ] という上位レベルのエントリおよび各ライブラリの追加エントリが含まれます。詳細は、9 章 p.225 **ライブラリの表示** を参照してください。



## オプションの設定

[オプション] ダイアログ ボックスで IBM® SPSS® Text Analytics for Surveys の一般的なオプションを設定できます。このダイアログボックスには、以下のようなタブがあります。

**システム**タブ：ライブラリ リスト、自動保存、抽出結果の保存、区切り文字、インターフェイス言語のデフォルトについてのオプションがあります。

**表示**タブ：インターフェイスで使用される色についてのオプションがあります。

**サウンド**タブ：サウンド キューについてのオプションがあります。

**翻訳**タブ：翻訳接続に関連するオプションがあります。

**オプションを編集するには：**

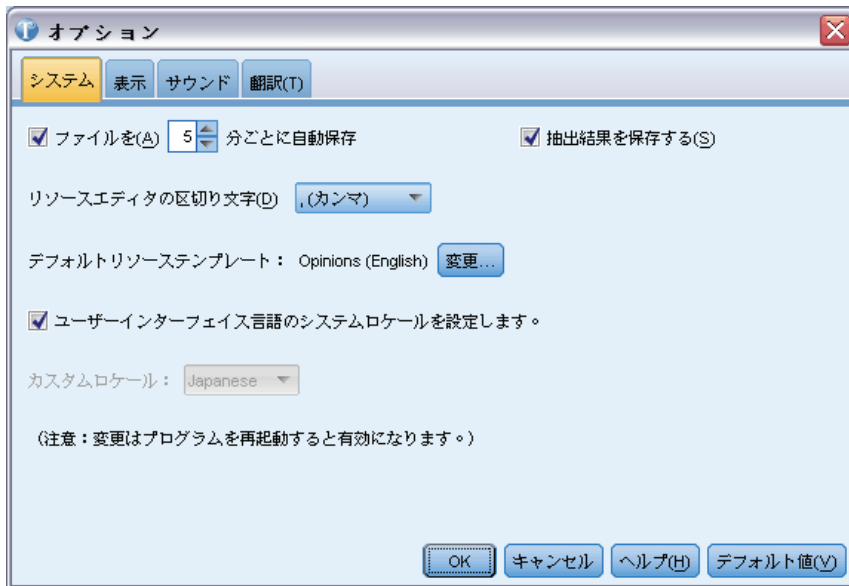
- ▶ メニューの [ツール] > [オプション] を選択します。[オプション] ダイアログボックスが開きます。
- ▶ 変更する情報を含むタブを選択します。
- ▶ オプションのいずれかを変更します。
- ▶ [OK] をクリックして、変更を保存します。

### オプション:システム タブ

このタブでは、プロジェクトのさまざまな設定を定義できます。以下のようなものがあります。

- ライブラリの追加または削除は、デフォルトではすべての新しいプロジェクトで表示されます。
- 自動保存・復元機能の有効化または無効化
- 抽出結果の自動保存の有効化または無効化
- リソース エディタで要素の分割に使用するグローバルな区切りの定義

図 2-7  
[オプション] ダイアログ ボックス:システム タブ



**ファイルをn分ごとに自動保存:** このオプションを選択すると、マシンに問題が発生したときのために、現在のプロジェクトのファイルの一時保存バージョンが自動的に作成されます。また保存する間隔を分で設定します。この機能を使うと、本プログラムが予期しないときに終了したり、マシンに問題があったりしたときに、次に起動した際に、ファイルの復元バージョンを開いて作業することが可能になります。

**抽出結果を保存:** このオプションを選択すると、プロジェクトの抽出結果を保存できます。まだカテゴリを作成している途中である場合には、これで時間を節約できます。しかしながら、これによって起動時間が長くなり、プロジェクトのサイズが大きくなります。セキュリティのために、これらの抽出結果は保存の過程で暗号化されてからデータベースに置かれます。この手続きによって、熟練したユーザーにとってもデータベース内でデータを見つけることは困難になっています。それだけでなく、抽出結果は、データソースが特定されない限り、IBM® SPSS® Text Analytics for Surveysには絶対に表示されません。このため、データがパスワードで保護されている場合、抽出結果を画面に表示するには、このデータソースに関してユーザーがユーザー名とパスワードを入力する必要があります。

- 保存は時間効率上もっとも有効です。データセットが大きいと抽出プロセスに時間がかかります。保存することによって、いつでもプロジェクトを開いて、結果にすぐにアクセスできるようになります。しかし、プロジェクトを開く時に少し時間がかかることに注意してください。
- 保存しないは、セキュリティ手段が整っていても、回答テキストをオリジナル データ ファイル以外の場所におきたくない場合に使います。

**リソース エディタ 区切り** コンセプト、類義語、オプション要素などの要素をリソース エディタ ビューで入力する場合に区切り文字として使用する文字を選択しますリソース エディタ。

**リソース テンプレート:**テキスト分析パッケージを選択しなかった場合は、デフォルト リソースのセットが使用されます。これらのリソースはテンプレートに保存されます。変更をクリックして、別のデフォルトリソーステンプレートを選択します。その後、テンプレート変更ダイアログで、使用したいテンプレートのテーブルの線を選択してOKをクリックします。

**ユーザーインターフェース言語のシステムの設定言語の使用** このオプションSPSS Text Analytics for Surveysを選択するとシステムの設定言語をユーザーインターフェースで使用できます。または、このオプションの選択を解除して別のインターフェース言語を選択できます。たとえば、これによって別の言語で情報分析を記録した場合に、システムの設定言語の同言語でSPSS Text Analytics for Surveysを実行できます。

SPSS Text Analytics for Surveysを再起動するまで変更が有効にならないことに注意してください。

## オプション:[表示] タブ

このタブでは、全体的な外観に関するオプションや、要素を区別するための色を編集できます。

図 2-8  
[オプション] ダイアログ ボックス:[表示] タブ



**標準的なフォントと色(再起動後に反映)**このコントロールボックス内のオプションは、カラー スキームおよび表示確認のために使用します。ここで選ばれたオプションは、アプリケーションを閉じて再起動するまで有効になりません。以下から選択してください。

- **SPSS Standard (デフォルト):**SPSS ブランド (IBM Corp. の一部) 製品共通のデザインです。
- **SPSS Classic:** このアプリケーションの以前のバージョンのユーザーが使いやすいデザインです。
- **Microsoft Windows:** 豊富になったストリーム キャンバスおよびパレットで使いやすい Microsoft Windows デザインです。

**ユーザー定義の色:** 画面上に表示される要素の色を編集します。表内のそれぞれの要素に関して、色を変更できます。ユーザー定義の色を指定するには、変更したい要素の右側にある色をクリックし、色のドロップダウンリストから色を変更します。

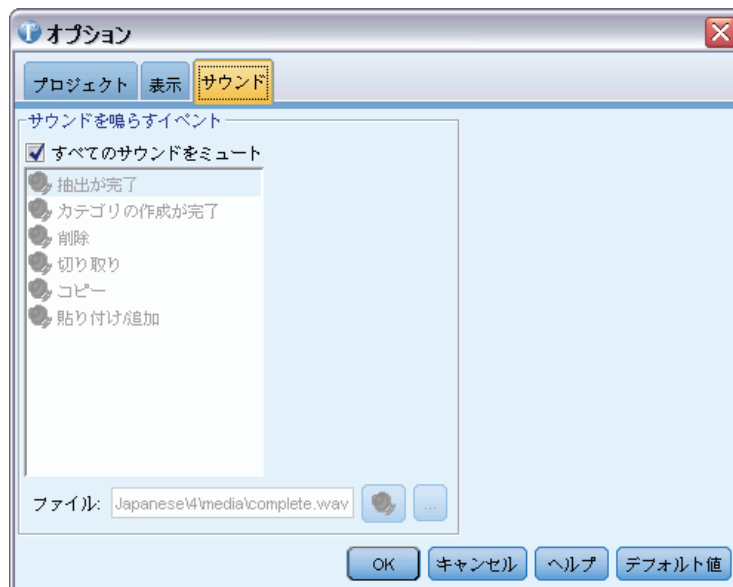
- **未抽出のテキスト:**抽出されていないけれどもデータパネル内に表示されている回答テキスト。
- **強調背景:** パネルの要素または [データ] パネルのテキストを選択する場合のテキスト選択の背景色。
- **抽出が必要な背景:**抽出結果パネルの背景色で、ライブラリに変更が加えられたので、抽出を行う必要があることを示しているもの。
- **カテゴリ フィールド バック背景:**回答のドラッグ&ドロップや、データパネルの回答をカテゴリパネルに強制的に移動させるなどの操作を行った後のカテゴリの背景色。
- **デフォルト タイプ:** [データ] パネルおよび [抽出結果] パネルに出現するタイプおよびキーワードのデフォルト色。たとえば、**名詞タイプ** やそれに関連したキーワードが出現すると、インターフェイス内にこの色が表示されます。この色は、リソース エディタ 内で作成したカスタムのタイプについても適用されます。リソース エディタ でこれらのキーワード辞書のプロパティを編集し、カスタム キーワード辞書のこのデフォルト色を上書きします。詳細は、10 章 p.238 **タイプの作成** を参照してください。
- **テーブルの縞 1:**[強制キーワードを編集] ダイアログボックス内の表で、各行を見やすくするために交互に色が変わっているものの 1 つめの色。
- **テーブルの縞 2:**[強制キーワードを編集] ダイアログボックス内の表で、各行を見やすくするために交互に色が変わっているものの 2 つめの色。
- **エラー用前景**[コードフレームマネージャ] 内で重複している項目のテキスト色。エラーを表わしています。
- **エラー用背景**[コードフレームマネージャ] 内で重複している項目の背景色。エラーを表わしています。

**視覚化用の色の順番:** 視覚化パネル内でカテゴリ棒グラフを使用し、参照変数を選択した場合、パネルの下の部分に表示される凡例に、この参照変数の各値が表示されます。これらの値は、棒グラフ内で視覚的に区別できるように色分けされています。これらのデフォルト色はここで変更できます。詳細は、7章 p.181 [グラフの視覚化](#) を参照してください。

## オプション:[サウンド] タブ

このタブでは、製品で仕様するサウンドを編集できます。サウンドを鳴らすイベントの部分で、あるイベントが起こった際にこれを知らせるサウンドを指定できます。サウンドのオン / オフまたは特定のサウンドの割り当てによって、ソフトウェアで特定の操作をした時の警告方法を調整することができます。たとえば、抽出プロセスや自動カテゴリ化、またはカット、ペースト、コピー、削除などの一般的な操作の終了時にサウンドを鳴らすことができます。

図 2-9  
[オプション] ダイアログ ボックス:[サウンド] タブ



使用できるサウンドはたくさんあります。サウンドを参照して選択する場合は、[...] ボタンを使用します。IBM® SPSS® Text Analytics for Surveys のサウンドを作成するために使われる .wav ファイルは、インストール ディレクトリ中の /media サブディレクトリにあります。サウンドを鳴らしたくない場合、[すべてのサウンドをミュート] を選択します。デフォルトでは、音が鳴らないようになっています。

## オプション:[翻訳] タブ

**重要!**英語への翻訳のみが可能です。

このダイアログで、翻訳するときいつでも再利用できる Language Weaver 翻訳サーバーを定義および管理できます。接続をここで定義すると、すべての接続設定を再度入力することなく、翻訳時に言語ペア接続をすばやく選択できます。

言語ペア接続により、サーバーへの接続の詳細 (LOC、WAN、または HTTP) のほか、ソース言語や翻訳言語を特定します。たとえば、[中国語 - 英語] は、ソース テキストが中国語で、翻訳テキストが英語であることを示します。Language Weaver オンライン サービスで、アクセスする Language Weaverサーバーへの接続を手動で定義できます。

翻訳結果はこのダイアログで定義されたディレクトリに保存されます。このディレクトリ内、またはここで別に指定したディレクトリ内で翻訳ファイルを管理できます。

図 2-10  
[オプション] ダイアログ ボックス:[翻訳] タブ

The screenshot shows a dialog box titled 'オプション' (Options) with a '翻訳(T)' (Translation) tab selected. The dialog contains the following fields and buttons:

- 接続URL:** A text field containing 'https://mwaccess.languageweaver.com' and a 'テスト' (Test) button.
- ユーザー ID:** A text field containing '94476'.
- APIキー:** A text field containing 'abc121efg345678nopqrs789012dcba224466880C'.
- 翻訳ディレクトリ:** A text field containing 'C:\Program Files\IBM\SPSS\Text Analytics for Surveys\4\Translation' and a 'ファイルの場所...' (File location...) button.
- At the bottom, there are buttons for 'OK', 'キャンセル' (Cancel), 'ヘルプ(H)' (Help), and 'デフォルト値(Y)' (Default value).

**接続URL :** Language Weaver オンライン サーバーへのサーバー URL または Web アドレスを入力します。

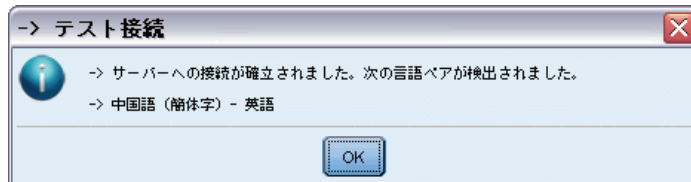
**ユーザー ID. :** Language Weaverから入手した一意のIDを入力します。

**API キー :** Language Weaverから入手した一意のIDを入力します。

**テスト**接続を選択して [テスト] をクリックすると、接続が正しく設定されていることを検証し、その接続の言語ペアを表示します。

**翻訳精度:** ブラウズをクリックして別のディレクトリに変更するか、フィールドにフォルダパスディレクトリを入力します。

図 2-11  
正常な接続のメッセージ



## Microsoft Internet Explorer ヘルプの設定

### Microsoft Internet Explorer の設定

このアプリケーションのほとんどのヘルプ機能では、Microsoft Internet Explorer に基づいたテクノロジーが使用されています。Internet Explorer のバージョンによっては (Microsoft Windows XP、Service Pack 2 と共に提供されるバージョンも含む)、ローカル コンピュータ上の [Internet Explorer] ウィンドウ内で「アクティブなコンテンツ」と見なされる対象が、デフォルトにより封鎖されます。このデフォルトの設定により、ヘルプ機能内である種のコンテンツが表示されなくなります。すべてのヘルプ コンテンツを表示するために、Internet Explorer のデフォルトの動作を変更できます。

- ▶ Internet Explorer のメニューから次の項目を選択します。  
ツール > インターネット オプション
- ▶ [詳細] タブをクリックします。
- ▶ [セキュリティ] セクションまで下方へスクロールします。
- ▶ [マイコンピュータのファイルでのアクティブ コンテンツの実行を許可する] を選択 (チェック) します。

# パート II: テキスト分析



# プロジェクトおよびパッケージの作成

## プロジェクトの作成

IBM® SPSS® Text Analytics for Surveysでは、アンケートデータを使用し、カテゴリ化できます。そのためには、データソースからデータをインポートするプロジェクトを作成し、変数を選択し、カテゴリとリソースを選択する必要があります。プロジェクトを作成後は、カテゴリのセットが納得のいくものになるまで、リソースやカテゴリーを微調整することができます。プロジェクトには以下のような要素を含めることができます。アンケート データ、言語リソース、抽出結果、カテゴリ

### アンケートデータ

プロジェクト内では、インポートしたアンケートデータファイルを参照します。しかしアンケートデータそのものはプロジェクト内に保存されません。コンテキスト内の最も重要な変数が自由記述式のテキスト変数である一方で、（人口統計量変数などの）参照変数はデータのインポート時に指定されるため、一意のID変数は保持されます。これらすべての変数の値は、テキスト分析ウィンドウ内の[質問] ビューにある[データ] パネルか、テキスト分析ウィンドウ内の[プロジェクト全体] ビューで確認できます。

### 言語リソース

製品に同梱されている汎用ライブラリ、サンプルライブラリ、およびユーザーがカスタム化したライブラリでは、キーワード・類義語・不要語のリストの作成やタイプの定義ができます。これはプロジェクト内含まれており、編集できます。これに加えて、基幹辞書が内蔵されています。これはバイナリ化されていてユーザーが参照することはできません。ライブラリを公開してデータベース内でパブリックなものにすれば、他のプロジェクトでも使用できます。公開されたライブラリは他のプロジェクトに追加できます。

### 抽出結果

抽出結果とは、テキスト形式の回答内で抽出された、キーワードと語句（コンセプト）、そのグループ（「タイプ」）、その関係性（「パターン」）です。これらの抽出結果はプロジェクトの一部であり、カテゴリ作成の基

礎です。デフォルトでは、抽出結果はプロジェクト内に保存されます。プロジェクトのファイルサイズがあまりに大きくなる場合には、保存機能をオフにして、次回このプロジェクトを開いた際に、再度抽出することもできます。詳細は、5 章 p.91 [抽出結果の保存](#) を参照してください。

## カテゴリ

テキスト形式の回答はカテゴリに分けられます。カテゴリは、ビルトインのカテゴリ作成手法による自動作成、ドラッグ&ドロップによる手作業の方法、カテゴリ定義ファイルのインポート、コードフレームマネージャの使用などの方法によって作成されます。抽出結果を保存しないように選択した場合、プロジェクトを再度開くと、カテゴリの定義は保持されています。この定義の一部である、回答の抽出結果からの度数は、再抽出を行うまで疑問符（「?」）として表示されます。

## 使用するデータの準備

IBM® SPSS® Text Analytics for Surveysにデータをインポートする前に、以下に注意してください。

- **データのインプット** SPSS Text Analytics for Surveysにインポートするためには、データには基本的な要素、すなわちID変数ならびに少なくとも1つの自由記述式の質問が必要です。ID変数は一意的な値だけを取ることができます。重複があった場合には、インポートが失敗します。複数の自由記述式の質問や参照変数をインポートできます。詳細は、p. 36 [変数の選択](#) を参照してください。
- **スペルミス** 本製品では若干のつづり間違いは許容されますが、インポートする前にそれらの間違いを直しておくことをお勧めします。テキスト分析においては、つづり間違いは人間のみならずソフトウェアプログラムにとっても問題となります。事前につづり間違いを直しておけばおおよそ、結果として得られるカテゴリの信頼性がより高いものになります。また、よく間違われる語とその正しい表記をプログラム内で類義語として作成しておくこともできます。よくあるつづり間違いのいくつかの例が、コアライブラリ内ですでに定義されています。つづり間違いを校正するのにどのくらいの労力がかかるのかよくわからない場合には、小さなサンプルの回答を使用して、つづり間違いによってこの分析がどのくらい影響を受けるのかを試してください。
- **空白の回答** 自由記述式アンケートデータにおいては、回答が空白になっていることは少なくありません。たとえば、何人の人がある質問に答えなかったのか、あるいは、どのような人が答えなかったのかについて調べたいこともあります。たとえば、何人の人がある質問に答えなかったのか、あるいは、どのような人が答えなかったのかについて調べたいこともあります。しかしながら、SPSS Text Analytics for Surveysでは

テキストを使ってキーワードを抽出し、回答をカテゴリ化するので、これらの空白の回答をカテゴリ化するのは容易ではありません。

1つの方法としては、インポートする前に、データ内ですべての空白の回答を「空白」あるいは他の適切なキーワードで置き換えるという方法が考えられます。それからデータをインポートして抽出した後に、「空白」という語（あるいは指定した他の語）をあるカテゴリを表わすキーワードすることによって、空白の回答を表わす新しいカテゴリを作成できます。

別の方法としては、空白の回答を1つのカテゴリに強制的にまとめるというものがあります。回答をカテゴリ化した場合、空白の回答は、最初はカテゴリ化されません。カテゴリパネル内で右クリックすることで、空白のカテゴリを手作業で新しく作成することができます。それからすべての空白の回答をすべて選択して、これらを新規の空白カテゴリに強制投入することができます。

- **マルチアンサーの質問:**自由記述式の質問は通常はそれだけで独立していますが、いつもそうだとは限りません。場合によっては、自由記述式の質問がマルチアンサー（複数選択可の選択肢）として使用されることもあります。たとえば、「当ホテルにおいて改善すべき点を3つお教えてください」という質問を尋ねて、回答欄（変数）を3つ提示している場合、これはマルチアンサーの質問となります。

SPSS Text Analytics for Surveysは、質問変数ごとに別々に分析しますが、1つめの回答を分析するために作成したカテゴリや言語リソースを2つめ、3つめの回答に使用することもできます。しかしながら、これは効率的な方法ではありません。データをプログラムにインポートする前に、これら3つの回答変数を1つの変数にまとめることを検討してみてください。これらをまとめる場合には、1つめの回答の最後の語と次の回答の最初の語との間にスペース（あるいはできればピリオド）があるようにしてください。大きなデータセットの場合、作業には時間がかかるので、データファイルを作成する時点で（後からではなく）回答をまとめておくことをお勧めします。

- **回答サンプル**回答の数と平均の長さが増えると、D抽出またはカテゴリ化にかかる時間は長くなります。効率よく作業するためには、ファイルサイズが大きくなった時に（約15000ケース以上）、最初にランダム サンプルを採取して回答の小さなサブセットとして使用し、分析の第一段階としてください。

通常小さなサンプルは、言語リソースの編集方法の決定に適していません。小さなデータファイルに対してカテゴリ化を一度行った後に、ファイル全体を読み込んで再度抽出を行うと、回答の多くが自動的にカテゴリ化されます。それから先に作成したカテゴリに適合していない回答を探して、必要であれば修正を加えます。ランダムサンプルの大きさはいろいろ考えられますが、300ケースぐらいで十分でしょう。

**重要!**テキスト分析のプロセス全体に関しては、他にも考慮すべき点があります。詳細は、1 章 p.8 [テキスト分析前の準備](#) を参照してください。

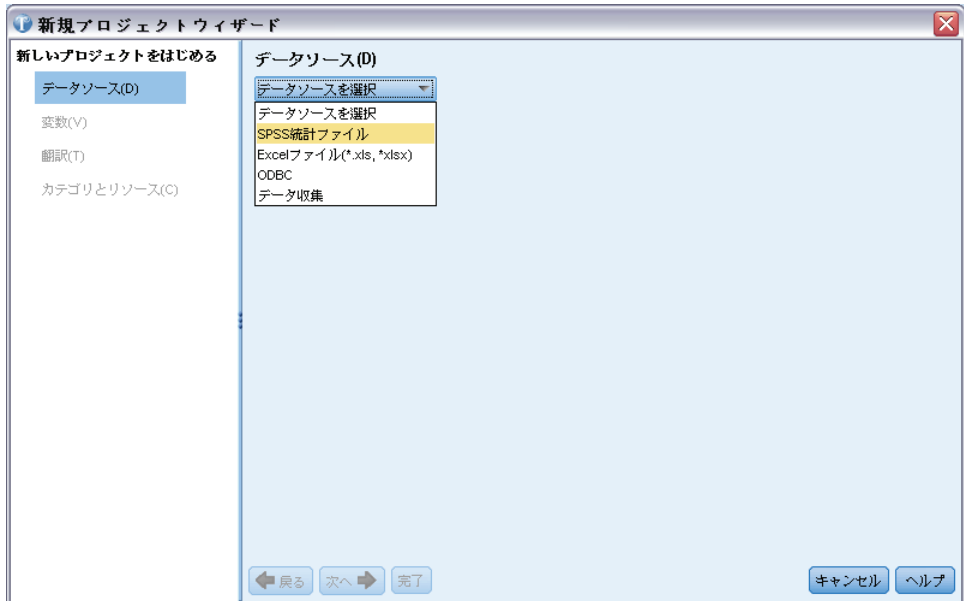
## 新しいプロジェクトの開始

回答データのカテゴリ化を開始するために、最初にプロジェクトを作成する必要があります。ウィザードによって、データソースから変数選択、さらにカテゴリおよびリソースの指定などに進みます。プロジェクトを作成する前には、データを準備する必要があります。

**新しいプロジェクトを開始するために:**

- ▶ [ファイル] メニューから、[ファイル > 新しいプロジェクト] を選択します。または、プロジェクトが開いていない場合は、スタートアップ画面から **新しいプロジェクトを開始** をクリックします。新しいプロジェクト ウィザードが表示されます。

図 3-1  
新しいプロジェクト ウィザード



- ▶ データソースを選択ドロップダウン リストからデータソースのタイプを選択して開始します。詳細は、p.31 [データソースの選択](#) を参照してください。

## データソースの選択

ウィザードが開いたら、最初にデータソースを選択します。IBM® SPSS® Text Analytics for Surveysは1レコードあたり500文字（全角）、1000件のレコードを持つデータセットを処理することを標準的な使用方法と想定し設計されています。しかし、レコードに含まれるテキストの量に応じてパフォーマンスは変わってきます。パフォーマンスや推奨環境に関しては、インストールガイドに記述されているので参考にしてください。

**重要!**このデータをインポートするには、各レコードごとに一意の値を持つID変数が必要です。

データソースは以下の中から1つを選択します。

- **SPSS Statistics ファイル** (\*.sav)
- **Microsoft Excel ファイル** (\*.xls / \*.xlsx)
- **ODBC** (Microsoft Open Database Connectivity プロトコル) データベース
- **Data Collection** データモデルこのオプションはデータモデルをインストールしている場合に限り利用できます。

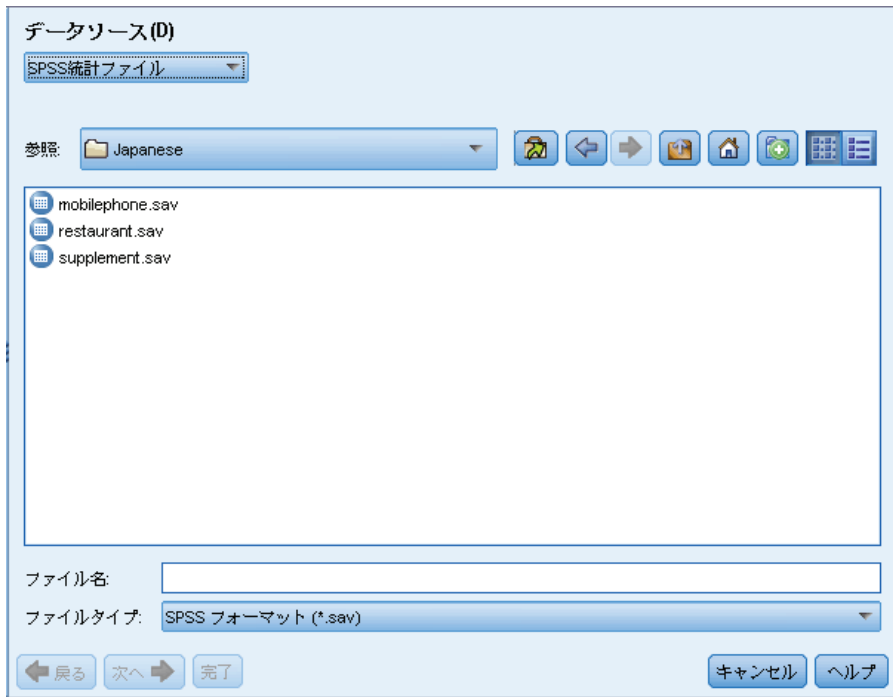
## IBM SPSS Statistics ファイルの使用

IBM® SPSS® Statistics (.sav) ファイルをIBM® SPSS® Text Analytics for Surveysにインポートできます。このデータをインポートするには、各レコードごとに一意の値を持つID変数が必要です。

**重要!**SPSS Statistics (.sav) ファイルは 4000 文字を超えるレコードとインポートすることはできません。

注:SPSS Text Analytics for Surveysは1レコードあたり500文字（全角）、1000件のレコードを持つデータセットを処理することを標準的な使用方法と想定し設計されています。しかし、レコードに含まれるテキストの量に応じてパフォーマンスは変わってきます。パフォーマンスや推奨環境に関しては、インストールガイドに記述されているので参考にしてください。

図 3-2  
IBM SPSS Statistics ファイルのデータソース オプション



### IBM SPSS Statistics からデータを取得するために

- ▶ ウィザードの最初の画面で、ドロップダウン リストから SPSS Statistics ファイルを選択します。ウィザードが SPSS Statistics ファイルのオプションを表示します。
- ▶ [参照] ドロップダウン リストから、ファイルを投入するドライブとフォルダを選択します。
- ▶ リストからファイルを選択します。[ファイル名] テキストボックス内にその名前が表示されます。
- ▶ [次へ]をクリックして変数を選択します。 詳細は、 p. 36 [変数の選択](#) を参照してください。

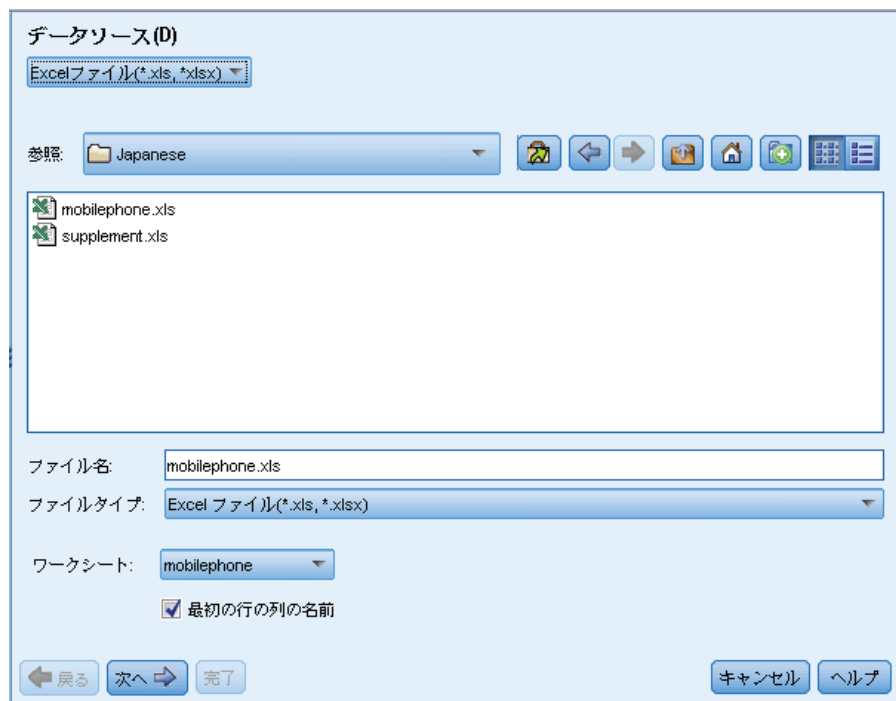
### Microsoft Excel ファイルの使用

Microsoft Excel (.xls / .xlsx) ファイルを IBM® SPSS® Text Analytics for Surveys にインポートできます。このデータをインポートするには、各レコードごとに一意の値を持つID変数が必要です。

**重要!**Microsoft Excelファイルをインポートする際に、[1行めに列名] オプションを選択できます。このオプションを使用するには、データが開始される上の行ではなく、ワークシートの最初の行に列名を入力する必要があります。たとえば、データおよび列名が行 7 で始まる場合、ファイルを正しくインポートするためには、インポート前に 1 ~ 6 行めを削除する必要があります。

注:SPSS Text Analytics for Surveysは1レコードあたり500文字（全角）、1000件のレコードを持つデータセットを処理することを標準的な使用方法と想定し設計されています。しかし、レコードに含まれるテキストの量に応じてパフォーマンスは変わってきます。パフォーマンスや推奨環境に関しては、インストールガイドに記述されているので参考にしてください。

図 3-3  
Microsoft Excel ファイルのデータソース オプション



### Microsoft Excel からデータを取得するために

- ▶ ウィザードの最初の画面で、ドロップダウン リストからExcelを選択します。ウィザードが Microsoft Excel ファイルのオプションを表示します。
- ▶ [参照] ドロップダウン リストから、ファイルを投入するドライブとフォルダを選択します。
- ▶ リストからファイルを選択します。[ファイル名] テキストボックス内にその名前が表示されます。

- ▶ ドロップダウン リストからワークシートを選択します。1 つのワークシートからのみカテゴリをインポートできます。複数のワークシートに含まれているデータを扱う際には、複数のプロジェクトを作成する必要があります。
- ▶ このワークシートの最初の行に列の見出しを指定する場合、[最初の行に列名]を選択します。このオプションを使用するには、データが開始される上の行ではなく、ワークシートの最初の行に列名を入力する必要があります。たとえば、データおよび列名が行 7 で始まる場合、ファイルを正しくインポートするためには、インポート前に 1 ～ 6 行目を削除する必要があります。アプリケーションはこれら（または、列の見出しが IBM® SPSS® Statistics の変数名規則に従っていない場合は変換されたバージョン）を変数名として使用できます。そうでない場合には、ワークシートの列名（アルファベット）が識別用に使用されます。
- ▶ [次へ]をクリックして変数を選択します。詳細は、p. 36 [変数の選択](#) を参照してください。

## ODBC からのデータの使用

データベース ソース（データベース）からのデータは簡単に IBM® SPSS® Text Analytics for Surveys にインポートできます。ODBC ドライバを使用するデータベースならば、SPSS Text Analytics for Surveys がインストールされたコンピュータに適切なドライバをインストールした後に直接読み込むことができます。このデータをインポートするには、各レコードごとに一意の値を持つ ID 変数が必要です。

注: SPSS Text Analytics for Surveys は 1レコードあたり 500 文字（全角）、1000 件のレコードを持つデータセットを処理することを標準的な使用方法と想定し設計されています。しかし、レコードに含まれるテキストの量に応じてパフォーマンスは変わってきます。パフォーマンスや推奨環境に関しては、インストールガイドに記述されているので参考にしてください。



図 3-4  
ODBC のデータソース オプション

データソース(D)

ODBC

ソース(DSN)

名前	説明
MS Access Database	Microsoft Access Driver (*.mdb, *.accdb)
Excel Files	Microsoft Excel Driver (*.xls, *.xlsx, *.xlsm, *.xlsb)
dBASE Files	Microsoft Access dBASE Driver (*.dbf, *.ndx, *.m...

ODBC

ユーザー:

パスワード:

テーブル

SQL

戻る 次へ 完了 キャンセル ヘルプ

### ODBC から使用するために

- ▶ ウィザードの最初の画面で、ドロップダウン リストから ODBC を選択します。ウィザードに ODBC オプションが表示されます。
- ▶ 登録された ODBC ソースのリスト内でデータソースを特定するか、あるいは [ソース (DSN)] テキストボックスにデータソースの名前を入力します。このリストに表示されていないデータソースを新しく登録するには、[ODBC] をクリックします。これで、ほとんどの Microsoft Windows コンピュータに付属の [ODBCデータソースアドミニストレータ] が開きます。これがない場合には、ODBCインポートを使用できません。詳細については、Microsoft Windows のヘルプを参照してください。
- ▶ 使用するデータソースがパスワードで保護されている場合には、ユーザー名とパスワードを入力する必要があります。これはプロジェクトを開くたびに要求されます。セキュリティ上の理由で、ユーザー名ならびにパスワードはプロジェクト内には保存されません。
- ▶ データの選択方法には2種類あり、Table または SQL. テーブルを直接選択するか、あるいはSQLコマンドを使用してデータを選択することができます。

- ▶ [次へ]をクリックして変数を選択します。詳細は、p. 36 [変数の選択](#) を参照してください。

## IBM SPSS Data Collection データの使用

### IBM SPSS Data Collectionからインポートするために

- ▶ ウィザードの最初の画面で、ドロップダウン リストからData Collectionを選択します。IBM® SPSS® Data Collectionデータモデルのオプションは、他の製品でこのデータモデルがインストールされている場合にのみ使用可能です。

## 変数の選択

データソースの選択後のステップは、インポートする変数の指定です。プロジェクトにインポート可能な変数は3種類あります。

### 一意のID変数(必須)

ID変数は数値や文字列のキーで、各回答者を特定するものです。データファイルを読み込むためには、一意のID変数に基づいて並んでいる必要はありません。プログラムに読み込んだ後に、レコードをさまざまな基準でソートできます。詳細は、4 章 p. 54 [変数のソート](#) を参照してください。データをインポートするには、この変数が必要です。インポートされた各レコード（ケース）は一意のID値を持つ必要があります。

インポートが失敗する原因は2つ考えられます。

- 重複するID値が検知された場合
- ID値が空白のレコード

注: 重複するIDが検知された場合、使用中のコンピュータにIBM® SPSS® Statisticsがインストールされているならば、[重複ケースの特定] 手続きを使用できます。[データ] メニューから [重複ケースの特定] を選択し、[オプション] を使ってどのレコードを保持するのか（プライマリケースとするのか）を指定します。

### 自由記述変数(必須)

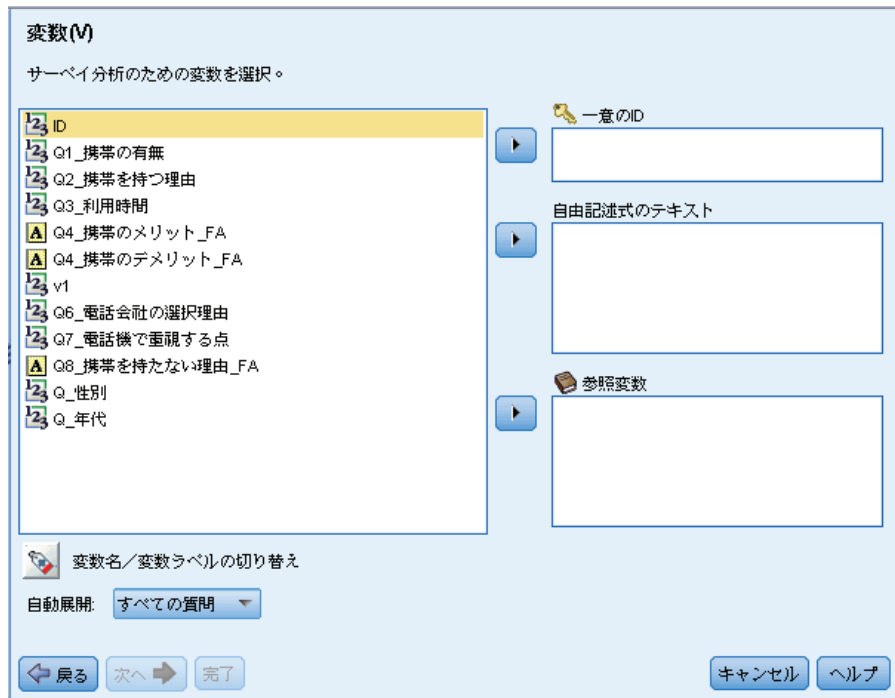
自由記述変数は、アンケート内の質問に対するテキスト形式の回答を表わしています。データをインポートする際には、この変数が少なくとも1つは必要です。これらの変数としてはSPSS Statistics内の文字列あるいは長い文字列型の変数、Microsoft Excel内の一般のセルやテキストセルを含む列、データベースのテキストあるいはノートフィールドがあります。自由記述変数はそれぞれ独立したものとして分析されます。SAVファイルからインポートする変数のサイズ（幅）の上限は4,000文字です。

## 参照変数(オプション)

参照変数はオプションです。これは通常カテゴリ型で、参照目的でインポートされます。参照変数はテキスト分析には使用されませんが、理解や解釈を助けるような、回答者に関する補足的な情報を提供します。参照変数としては、人口統計量変数がよく使用されます。これはどのキーワードやカテゴリがどの集団や個人によって使用されたのかを理解する助けになるからです。たとえば学生やトレーニングの評価においては、性別、部署、職業、学部などが考えられます。インポートされたすべての参照変数は、プロジェクト全体ビューに表示されます。参照変数を質問ビューのデータパネルに表示させることもできます。また視覚化パネルの棒グラフ内で参照変数を選択し、回答者のサブセットを詳細に検討することもできます。

注:SPSS Statisticsデータファイルから読み込んだ参照変数では、データパネルの列見出しに変数ラベル(あらかじめ設定されている場合)が表示され、その値ラベル(あらかじめ設定されている場合)がセルに表示されます。

図 3-5  
変数の選択



### 変数および抽出オプションを選択するために

- ▶ 使用可能な変数のリストから、データセット内のID変数に該当する変数を選択します。次に矢印ボタンをクリックし、[一意のID] ボックスに移動し

ます。IDは一意の数字あるいは英数文字列で、あるレコードを他のレコードと区別するためのものです。使用するデータに重複するIDが含まれている場合、エラーメッセージが表示されます。このような場合、データを整理してから、再度インポートを行ってください。

- ▶ 使用可能な変数のリストから、自由記述式の回答変数に対応する変数を、1つあるいは複数選択します。次に矢印ボタンをクリックし、この変数（複数可）を[自由記述式のテキスト] リストに移動します。インポートされたそれぞれの変数は独立した質問で、その回答レコードが分析やカテゴリ化の対象となります。
- ▶ 使用可能な変数のリストから、参照変数あるいはカテゴリ変数に該当する変数を、1つあるいは複数選択します。次に矢印ボタンをクリックし、この変数を[参照変数] リストに移動します。参照変数は自動カテゴリ作成手法では使用されません。しかしながら、回答をカテゴリ化する際にこの内容を見ることで、より詳しい情報に即した決断を行うことができます。
- ▶ 変数名の変わりに変数ラベルを表示したい場合は、左の変数リストの下にあるボタンをクリックします。
- ▶ 抽出設定を変更するには、ドロップダウン リストから選択します。デフォルトでは、**最初の質問のみ**が選択されています。これは自由記述変数を複数選択すると、ウィザード終了後に、最初の質問で抽出プロセスが自動的に開始するという意味です。抽出は大きなデータセットからも行えます。そのため、時間に応じてなした**すべての質問**の抽出設定を選択できます。
- ▶ 変数を全て選択してから、[次へ]をクリックします。

## 英語への翻訳

英語以外のテキストを使用している場合は、Language Weaverに接続して英語に翻訳することができます。英語への翻訳のみが可能です。Language Weaverを適切に設定し、接続を翻訳に定義する必要があります。詳細は、2 章 p.24 [オプション：\[翻訳\] タブ](#) を参照してください。

図 3-6  
翻訳オプション



### 英語に翻訳するには

- ▶ テキスト データをライセンスされた言語から英語に翻訳するためには、英語に翻訳チェックボックスを選択します。
- ▶ 言語ペア接続リストから、使用する Language Weaver 言語ペアの接続を選択します。ローカル コンピュータで Language Weaver を設定した場合、これらの言語ペアが自動的にこのリストに表示されます。[オプション] ダイアログ ボックスの [翻訳] タブ内で、オンライン サービス接続を追加、変更、テストできます。詳細は、2 章 p.24 オプション: [翻訳] タブ を参照してください。
- ▶ 必要に応じて、**翻訳精度**を指定してください。精度と競合する速度を指定する 1 ~ 3 の数値を選択してください。値が小さいほど、翻訳結果が速く作成されますが、精度は劣ります。値が大きいと、精度の高い結果が作成されますが、時間がかかります。時間を最適化するには、低いレベルから始め、結果を見た後でより高い精度が必要だと感じた場合にのみ、レベルを高くすることをお勧めします。
- ▶ 以前に Language Weaver のカスタム辞書を作成している場合は、翻訳時に使用することができます。カスタム辞書を使用するには、**カスタム辞書**を使用

チェックボックスを洗濯し、辞書名を入力します。複数の辞書を使用する場合は、辞書名をコンマで区切ってください。

- ▶ 新しいプロジェクトウィザードで、次へ>をクリックしてカテゴリーとリソースの選択を開始します。詳細は、p. 40 [カテゴリおよびリソースの選択](#) を参照してください。
- ▶ データセットの変更ウィザードで、終了をクリックしてデータセットを終了し、翻訳プロセスを開始します。

#### 翻訳をスキップするには

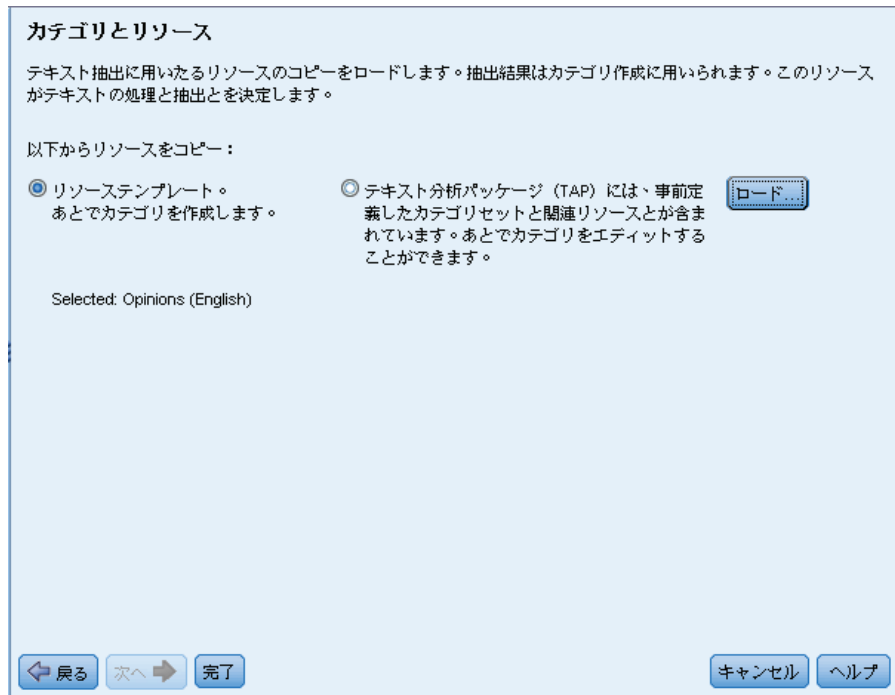
- ▶ 英語に翻訳オプションの選択を解除します。
- ▶ 新しいプロジェクトウィザードで、次へ>をクリックしてカテゴリーとリソースの選択を開始します。詳細は、p. 40 [カテゴリおよびリソースの選択](#) を参照してください。
- ▶ データセットの変更ウィザードで、終了をクリックしてデータセットを終了します。

## カテゴリおよびリソースの選択

この最後のステップでは、テキストから目立ったコンセプトやパターンを抽出する際に使用する言語リソースを選択します。または、テキスト分析パッケージ (TAP) を読み込むことができます。TAPには、言語リソースだけでなく、強化コード フレームを表す1つまたは複数の事前定義されたカテゴリーセットが含まれます。詳細は、p. 44 [テキスト分析パッケージの使用](#) を参照してください。IBM® SPSS® Text Analytics for Surveysには事前作成された複数の英語用TAPファイルが必要です。製品に付属する各TAPファイルは、特定のアンケートタイプ（従業員、製品、顧客満足度など）のために微調整されています。製品のサポートするテキスト言語に合わせて独自のTAPを作成することもできます。

デフォルトでは、リソース テンプレートが事前に読み込まれています。オプション ダイアログの最初のタブに示されるデフォルトのリソース テンプレートは変更できます。詳細は、2 章 p. 19 [オプションの設定](#) を参照してください。別のリソース テンプレートを読み込むことも、TAPを選択することもできます。

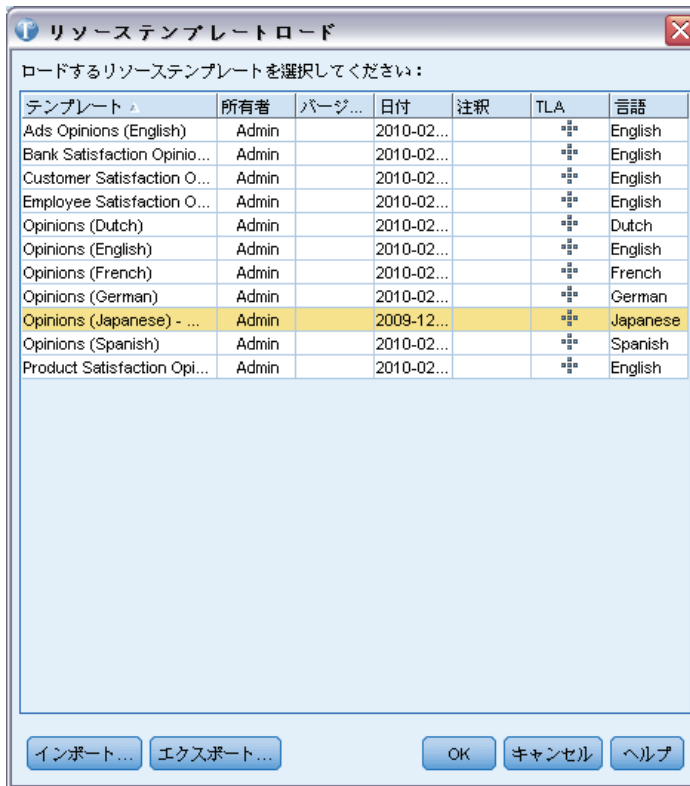
図 3-7  
リソースの選択



**別のリソース テンプレートを選択するために：**

- ▶ 別のリソース テンプレートを読み込むために、リソース テンプレートオプションを選択し、読み込みをクリックしてください。[リソース テンプレートを開く] ダイアログ ボックスが開きます。

図 3-8  
リソース テンプレートの読み込み



- ▶ リソース テンプレートの読み込みダイアログで、使用したテンプレートを  
選択してOKをクリックします。ダイアログが閉じ、ウィザードが選択した  
新しいテンプレートを表示します。

注意：ライセンスのない言語のテンプレートがある場合、チェックボック  
スはダイアログの下部に表示され、ライセンスのない言語のテンプレート  
を非表示にすることができます。

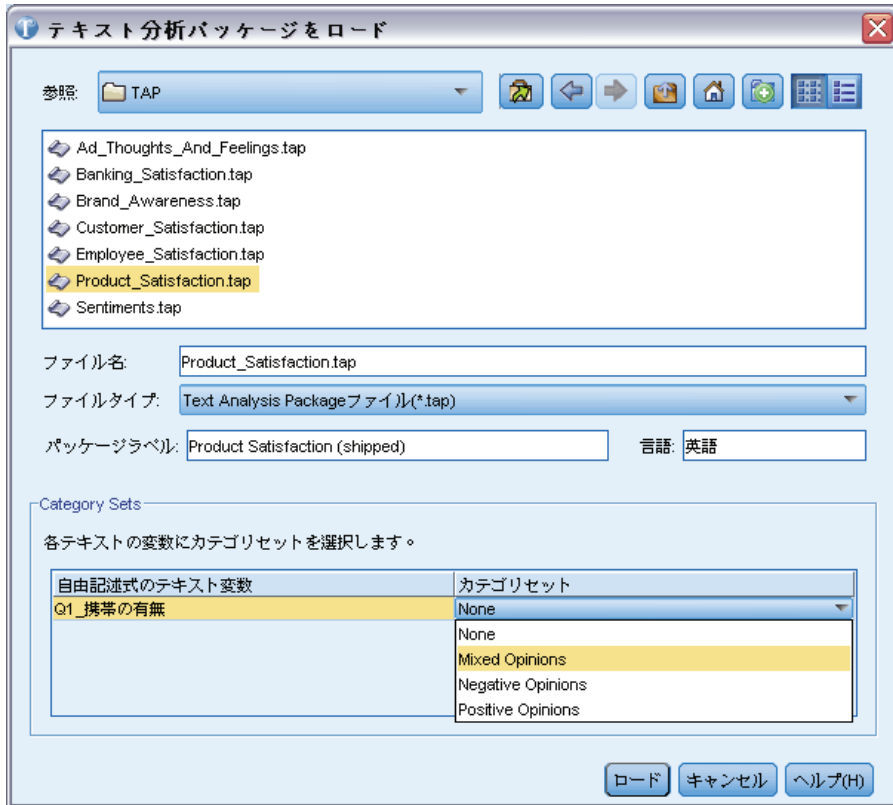
- ▶ [完了] をクリックすると、データがインポートされます。このオプション  
を選択すると、プロジェクトの結果はデフォルトのリソースを使用し、抽  
出の実行後には、カテゴリを手動で、または自動作成手法で作成できま  
す。詳細は、6 章 p.100 [テキスト データの分類](#) を参照してください。

#### テキスト分析パッケージを選択するために：

- ▶ テキスト分析パッケージを読み込むために、テキスト分析パッケージオプション  
を選択し、読み込みをクリックしてください。テキスト分析パッケージの読  
み込みダイアログが表示されます。



図 3-9  
テキスト分析パッケージの読み込み



- ▶ ダイアログで、使用したいTAPを選択します。デフォルトで保存されているパッケージでのみ、<installation\_directory>\TAPがリストに直接表示されず、以下のフィールドの詳細が選択したTAPに応じて更新されます。
- ▶ カテゴリセットテーブルでは、カテゴリセットを各テキスト変数に割り当てることができます。カテゴリセット列で、各セルのドロップダウンリストをクリックし、利用可能なカテゴリセットを選択します。なしを選択した場合、後で作成するまで、テキスト変数にはカテゴリは割り当てられません。
- ▶ [OK] をクリックします。ダイアログが閉じ、ウィザードが選択した新しいTAPを表示します。TAPとカテゴリセットを選択後、ウィザードは終了し、事前作成されたカテゴリにコード化されたレコードがすぐに表示されます。ここから、結果をエクスポートする、またはカテゴリを分析の開始点として使用することができます。
- ▶ 終了をクリックしてダイアログボックスを閉じ、プロジェクトを作成します。終了すると、プロジェクト内の1つめの自由記述式の質問に関する質問ビューが自動的に開きます。抽出を選択した場合は、抽出の進行ダイアログが表示されます。抽出が完了するまでは、しばらく時間がかかります。

す。これで質問の分析を始めることができます。他の質問に切り替えるには、メニューから [表示] > [質問] を選択します。

## テキスト分析パッケージの使用

TAP とも呼ばれるテキスト分析パッケージは、テキスト回答のカテゴリ化を行うためのテンプレートとして機能します。TAP には多くのレコードを迅速かつ自動的にコード化するために必要なコード フレームおよび言語リソースが含まれているため、TAP を使用すると、最小限の介入でテキスト データをカテゴリ化できます。言語リソースを使用して、テキスト データを分析およびマイニングし、主要キーワードを抽出します。テキストの主要キーワードおよびパターンに基づき、レコードを TAP で選択したカテゴリ セットにカテゴリ化できます。独自の TAP を作成または TAP を更新できます。

TAP は、次の要素で構成されています。

- **カテゴリ セット**: カテゴリ セットは、定義済みカテゴリ、カテゴリコード、各カテゴリの記述子、カテゴリ セット全体の名前で構成されています。記述子とは、キーワード安いやパターン高価などの言語学的要素です。記述子を使用してカテゴリを定義し、テキストがカテゴリ記述子に一致すると、レコードがカテゴリに投入されます。
- **言語リソース**: 言語リソースは、一連のライブラリ、および主要キーワードおよびパターンを抽出するために調整された高度なリソースです。これらの抽出コンセプトおよびパターンは、レコードをカテゴリセットのカテゴリに投入できる記述子として使用されます。

テキスト分析パッケージを作成および更新できます。

TAP を選択してカテゴリ セットを新しいプロジェクト ウィザードの各テキスト変数に選択した後、IBM® SPSS® Text Analytics for Surveys でレコードを抽出およびカテゴリ化できます。そこから、結果をエクスポート、または希望する結果を取得するまでカテゴリ化の調整を続行できます。

注: TAP を作成し、SPSS Text Analytics for Surveys および IBM® SPSS® Text Analytics の製品間で交互に使用できます。

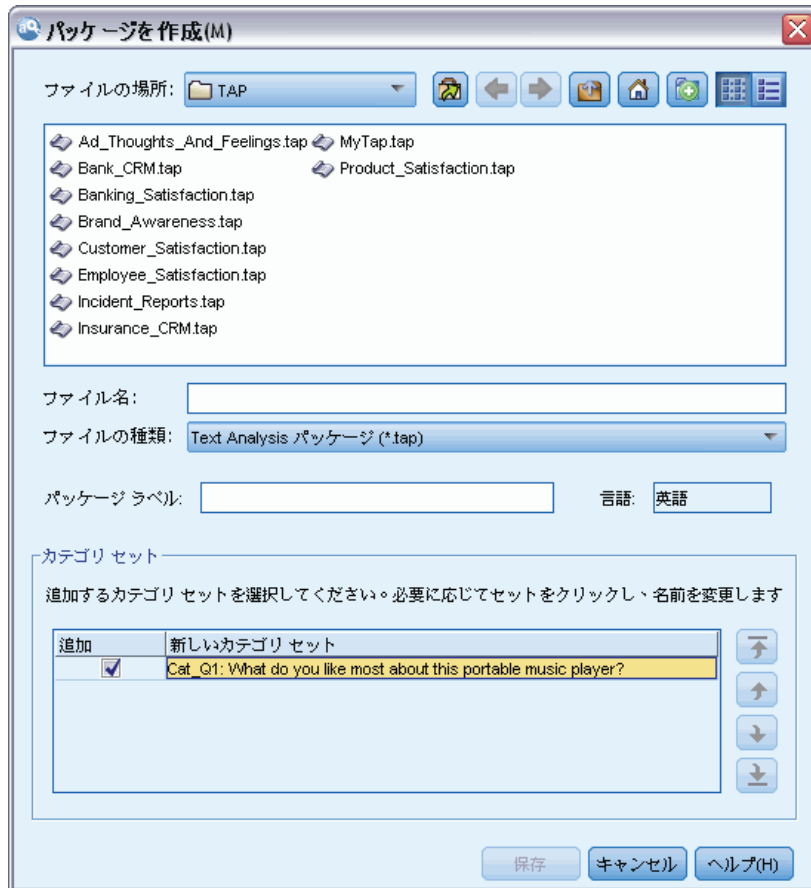
## テキスト分析パッケージの作成

少なくとも 1 つのカテゴリといくつかのリソースを含むプロジェクトがある場合、オープン プロジェクトのコンテンツからテキスト分析パッケージ (TAP) を作成できます。各質問のカテゴリおよび記述子 (コンセプト、タイプ、条件規則または TLA パターン出力) のセットをリソース エディタで開かれたすべての言語リソースと共に使用して TAP を作成できます。

リソースが作成された言語を表示できます。言語は、リソース エディタビューの [高度なリソース] タブで設定します。

**重要!** カテゴリにテキスト マッチ、強制されたレコード、またはフラグが含まれる場合、これらはデータ固有のものであり、多くが他のデータに使用できないものであるため、カテゴリ セットには保存されません◆◆ただし、ラベルおよびカテゴリ コードは保存されます。

図 3-10  
[テキスト分析パッケージの作成] ダイアログ



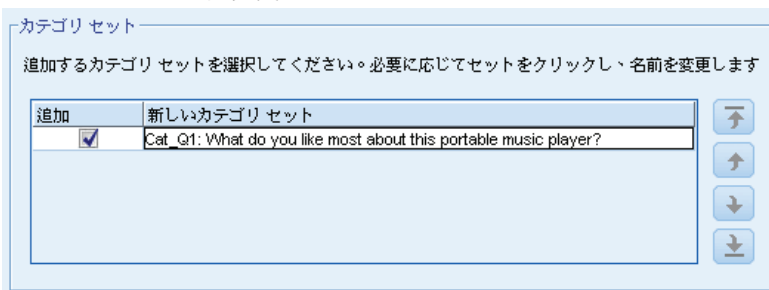
### テキスト分析パッケージを作成するには

- ▶ メニューで、[ファイル] → [テキスト分析パッケージ] → [パッケージを作成] を選択します。[パッケージを作成] ダイアログが表示されます。
- ▶ TAP を保存するディレクトリを参照します。デフォルトでは、TAP は製品のインストール ディレクトリの \TAP サブディレクトリに保存されます。
- ▶ [ファイル名] フィールドに TAP の名前を入力します。
- ▶ [パッケージラベル] フィールドにラベルを入力します。ファイル名を入力すると、この名前がラベルとして自動的に表示されますが、このラベルは変更

できます。このデフォルトのディレクトリに TAP を保存すると、パッケージ ラベルが新しいプロジェクト ウィザードのドロップダウン リストにオプションとして表示されます。

- ▶ TAP からカテゴリ セットを除外するには、[追加] チェックボックスをオフにします。カテゴリ セットを除外すると、カテゴリセットがパッケージに追加されなくなります。デフォルトでは、質問ごとに 1 つのカテゴリ セットが TAP に追加されます。TAP には、1 つ以上のカテゴリ セットが必要です。
- ▶ カテゴリ セットの名前を変更します。[新しいカテゴリ セット] 列にはデフォルトで一般名が入力されています。一般名はテキスト変数名に **Cat\_** 接頭辞を追加して生成されます。セルを 1 回クリックすると、名前を編集できます。入力して他の場所をクリックすると、名前の変更が適用されます。カテゴリ セットの名前を変更すると、TAP のみの名前が変更され、オープン プロジェクトの編集名は変わりません。

図 3-11  
カテゴリ セットの名前変更

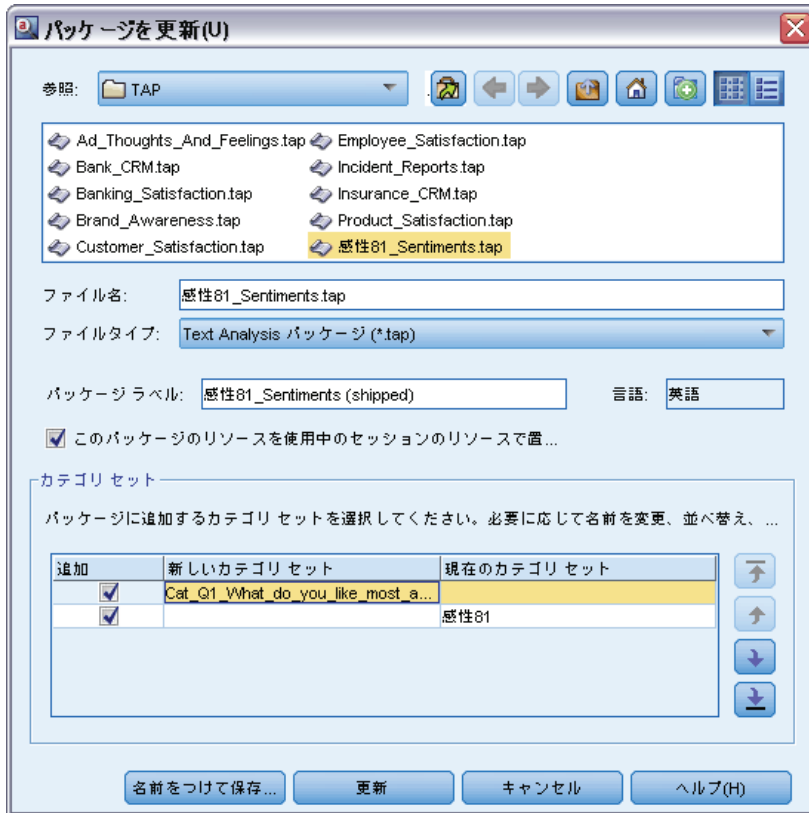


- ▶ 必要に応じて、カテゴリ セット テーブルの右側にある矢印キーを使用して、カテゴリ セットを並べ替えます。
- ▶ [保存] をクリックして、テキスト分析パッケージを作成します。ダイアログ ボックスが閉じます。

## テキスト分析パッケージの更新

カテゴリセット、言語リソースを改善するか、新しいカテゴリ セットを作成する場合、テキスト分析パッケージ (TAP) を更新して、これらの改善点を後で再利用しやすくすることができます。TAP を更新するには、TAP に追加したい情報を含むオープン プロジェクトで作業する必要があります。更新する場合、カテゴリ セットの追加、リソースの置き換え、パッケージ ラベルの変更、またはカテゴリ セットの名前変更/並べ替えを選択できます。

図 3-12  
[テキスト分析パッケージの更新] ダイアログ



### テキスト分析パッケージを更新するには

- ▶ メニューで、[ファイル] → [テキスト分析パッケージ] → [パッケージを更新] を選択します。[テキスト分析パッケージの更新] ダイアログが表示されます。
- ▶ 更新したいテキスト分析パッケージを含むディレクトリを参照します。
- ▶ [ファイル名] フィールドに TAP の名前を入力します。
- ▶ TAP 内の言語リソースと現在のプロジェクトの言語リソースと置き換えるには、[このパッケージのリソースを使用中のセッションのリソースで置き換える] オプションを選択します。通常、言語リソースはカテゴリ定義の作成に使用される主要キーワードおよびパターンを抽出するために使用されるため、言語リソースの更新は重要です。最新の言語リソースがあれば、レコードのカテゴリ化において最善の結果を得ることができます。このオプションを選択しない場合、パッケージ内にすでにある言語リソースは変更されません。

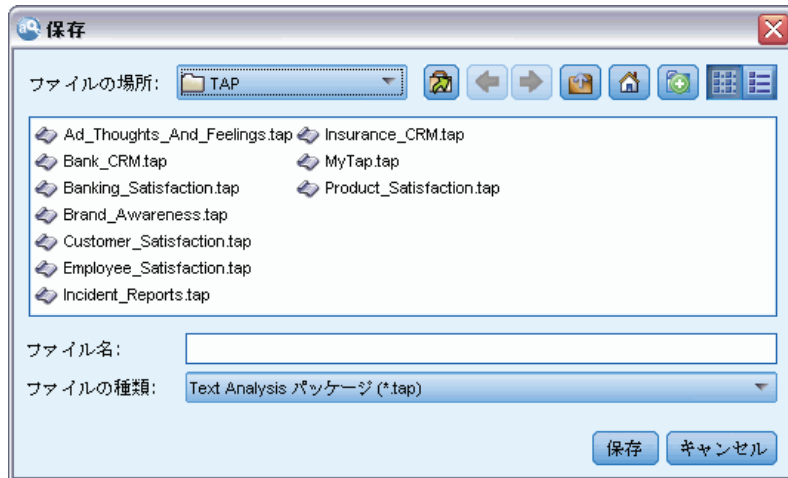
- ▶ 言語リソースのみを更新するには、かならず [このパッケージのリソースを使用中のセッションのリソースで置き換える] オプションを選択し、TAP 内にすでにある現在のカテゴリ セットののみを選択してください。
- ▶ 使用中のプロジェクトの新しいカテゴリ セットを TAP に追加するには、追加する各カテゴリのチェックボックスをオンにしてください。1 つ、複数のカテゴリ セットを追加できますし、追加しなくてもかまいません。
- ▶ TAP からカテゴリ セットを削除するには、対応する [追加] チェックボックスをオンにします。改善されたカテゴリ セットを追加するため、すでに TAP にあるカテゴリ セットを削除する場合があります。カテゴリ セットを削除するには、[現在のカテゴリ セット] 列の該当するカテゴリ セットの [追加] チェックボックスをオフにします。TAP には、1 つ以上のカテゴリ セットが必要です。
- ▶ 必要に応じて、カテゴリ セットの名前を変更します。セルを 1 回クリックすると、名前を編集できます。入力して他の場所をクリックすると、名前の変更が適用されます。カテゴリ セットの名前を変更すると、TAP のみの名前が変更され、オープン プロジェクトの編集名は変わりません。2 つのカテゴリ セットに同じ名前が付いている場合、重複を修正するまで名前が赤で表示されます。

図 3-13  
重複する名前

追加	新しいカテゴリ セット	現在のカテゴリ セット
<input checked="" type="checkbox"/>	Cat_Q1_1_外食頻度	
<input checked="" type="checkbox"/>		Cat_Q1_1_外食頻度

- ▶ 選択した TAP の内容と結合されたセッションの内容で新しいパッケージを作成するには、[名前をつけて保存] をクリックします。[テキスト分析パッケージの名前を付けて保存] ダイアログが表示されます。次の説明を参照してください。
- ▶ [更新] をクリックすると、選択した TAP に行われた変更が保存されます。

図 3-14  
[テキスト分析パッケージの名前を付けて保存] ダイアログ



### テキスト分析パッケージを保存するには

- ▶ TAP ファイルを保存するディレクトリを参照します。デフォルトでは、TAP ファイルはインストール ディレクトリの TAP サブディレクトリに保存されます。
- ▶ [ファイル名] フィールドに TAP ファイルの名前を入力します。
- ▶ [パッケージ ラベル] フィールドにラベルを入力します。ファイル名を入力すると、この名前がラベルとして自動的に表示されます。ただし、このラベルの名前は変更できます。ラベルはかならず指定する必要があります。このデフォルトのディレクトリに TAP を保存すると、パッケージ ラベルが新しいプロジェクト ウィザードのドロップダウン リストにオプションとして表示されます。
- ▶ [保存] をクリックして新しいパッケージを作成します。

# プロジェクトの使用

IBM® SPSS® Text Analytics for Surveysでは、アンケートデータを使用し、カテゴリ化できます。こその一連の作業を行うために、プロジェクトを作成してカテゴリを保存し、カテゴリの定義ならびにそれに対応する回答を作成・保存します。プロジェクトには以下のような要素を含めることができます。

- **データの調査**自由記述式の質問に対するテキスト形式の回答を含んだ変数、一意のID変数、参照変数のことです。アンケートデータはプロジェクト内に保存されるのではなく、プロジェクトを開いた際に元のデータソースから読み込まれます。
- **言語リソース**:テキストからコンセプトやパターンを抽出する際に使用される、製品に同梱されているライブラリならびにユーザーがカスタム化したライブラリ（類義語、不要語、キーワード辞書）です。
- **抽出結果**抽出の実行後に提示されるものです。回答データから抽出されたキーワードや語句です。これらのキーワードを使ってカテゴリを作成します。
- **カテゴリ型**:TAPカテゴリセットに由来する、手動作成および／または自動カテゴリ作成手法アンケート回答はこれらのカテゴリに割り当てられます。

## プロジェクトを開く

既存のプロジェクトに戻りたい場合には、そのプロジェクトを開きます。一度に開くことができるプロジェクトの数は1つだけです。すでにプロジェクトが開いている状態で他のプロジェクトを開こうとすると、必要であれば前のプロジェクトを保存するというメッセージが表示されます。

プロジェクトを開くと、IBM® SPSS® Text Analytics for Surveysは、言語リソースをチェックして、プロジェクト内のローカルライブラリと、パブリックライブラリのバージョンの新旧を比較します。パブリックライブラリにより新しいバージョンがある場合は、このライブラリを更新するというメッセージが表示されます。その際、現プロジェクトでは更新せずに以前のバージョンのままにしておくのか、あるいは更新されたものに置き換えるのかを選択できます。詳細は、9章 p.233 [ライブラリの更新](#) を参照してください。

**重要!** SPSS Text Analytics for Surveysがソースデータをプロジェクト内に物理的に保存することはありません。その代わりに、マシン上に保存されているデータへの参照情報が、プロジェクト内に保存されます。デー



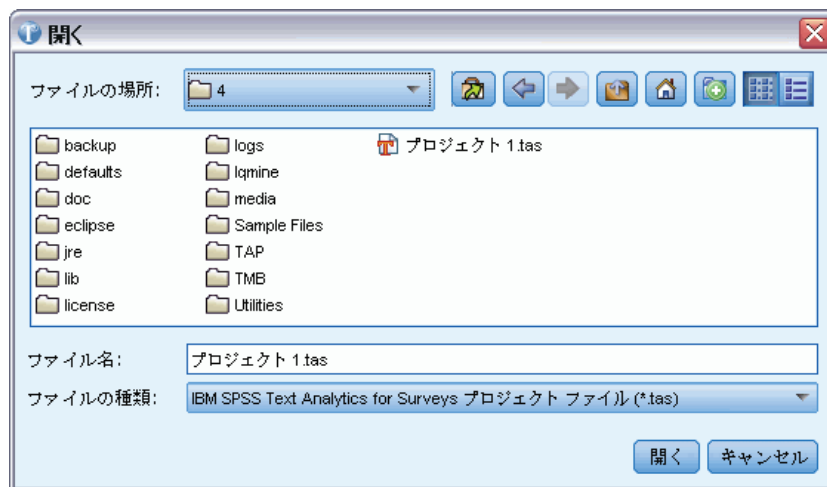
タソースに含まれるインポートされた元の変数が、その後変更した場合、「データが見つからない」という警告が表示されます。この場合は、データを再インポートして「新しい」質問を元の変数名に合致させ、プロジェクトを続行します。一般的に言って、ソースデータ内の列見出しの名前を変更することはお勧めしません。

注: 抽出結果は、デフォルト設定では、プロジェクトといっしょに保存されます（[ツール] > [オプション] で変更可能です）。プロジェクトを閉じるとカテゴリの定義は保存されますが、抽出結果パネルの内容は消去されます。このプロジェクトを開いて、回答のカテゴリ化を継続して行いたい場合には、抽出を実行する必要があります。既存のカテゴリの定義では回答の度数の代わりに疑問符 (?) が表示されます。再抽出を行うと、回答の度数が再度表示されます。

### プロジェクトを開くには:

- ▶ [ファイル] メニューから、[ファイル > 新しいプロジェクト] を選択します。[開く] ダイアログボックスが開きます。

図 4-1  
ダイアログボックスを開く



- ▶ リストから、開きたいプロジェクトのディレクトリと名前を選択します。このダイアログボックス内では、ファイルの大きさや日付などでソートすることはできません。
- ▶ [OK] をクリックします。メインウィンドウ内でこのプロジェクトが開きます。
- ▶ プロジェクト内にパスワードで保護されたデータベースからのデータが含まれている場合、このプロジェクトを開くたびにパスワードを入力するように指示されます。

- ▶ プロジェクトに製品の以前のバージョンが含まれている場合、リソースを新しいフォーマットに転換を進めるメッセージが表示されます。これはプロジェクトの保存後、製品の以前のバージョンではプロジェクトを開くことができないことを意味します。
- ▶ プロジェクトを最後に開いてからプロジェクト内のパブリック ライブラリが変更されている場合は、警告によってこの変更が通知されます。

**重要!**何らかのプロジェクトを開くたびに、それに対応するデータセットが開きます。データが見つからない場合は、エラーメッセージが表示されます。引き続きデータを使用する場合は、データを再インポートする必要があります。詳細は、 p. 65 [データ ソースの変更](#) を参照してください。

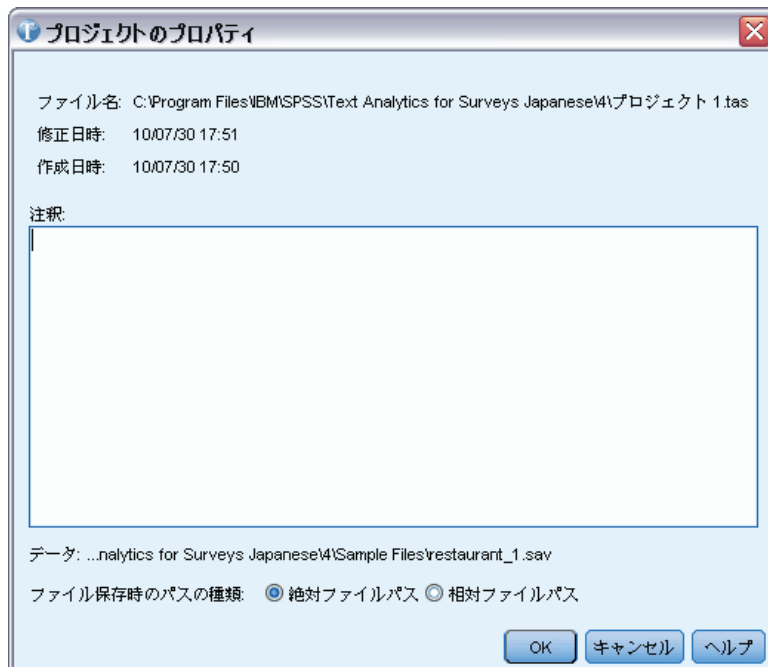
## プロジェクトのプロパティの編集

デフォルトでは、新規プロジェクトは常に「プロジェクト1」という名前になります。このプロジェクトの基本的なプロパティを検討したり、またこのプロジェクトの注釈を追加・修正できます。

**プロジェクトのプロパティを編集するには：**

- ▶ [ファイル] メニューから、[ファイル > 新しいプロジェクト] を選択します。プロジェクトのプロパティダイアログボックスが開きます。

図 4-2  
[プロジェクトのプロパティ] ダイアログ ボックス



- ▶ 必要であれば、テキストボックスにこのプロジェクトに関するコメントや説明を入力します。

注:データファイルの名前がこのダイアログボックス内に表示されます。ここでは今プロジェクトを作成しようとしており、データもインポートされていないので、データファイルの名前はわかりません。インポートすると、データファイル名の末尾60文字のみがここに表示されます。名前がこれ以上長い場合には、カーソルを名前の上に移動してください。名前がツールチップ内に全部表示されます。

- ▶ [OK] をクリックして、新しい名前を承認します。ダイアログボックスが閉じ、プロジェクトのプロパティが適用されます。

## プロジェクト データの表示

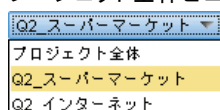
データファイルをインポートすると、このプロジェクト内にある最初の自由記述式の質問に関する[質問] ビューが表示されます。しかしながら、インポートしたデータをすべて見る必要がある場合もあります。その場合には、[プロジェクト全体] ビューを使用します。これはデータの包括的なビューです。このビューにアクセスするには、メニューから[表示]>[プロジェクト全体] を選択します。このビューでは、以下のことができます。

- インポートしたすべての変数の内容を確認できます。
- 変数に値やラベルを割り当てます。
- 変数のタイプを変更します。
- 変数をソートします。
- 隣接するセルのデータをコピーし、他のアプリケーションに貼り付けます。
- 変数列の幅を変えます。

**重要!**プロジェクトにインポートしたデータは読み取り専用です、IBM® SPSS® Text Analytics for Surveysで編集することはできません。

これでカテゴリを作成するこれらの回答からコンセプトを抽出することができます。詳細は、6章 p.100 [テキスト データの分類](#) を参照してください。

図 4-3  
プロジェクト全体ビューへのアクセス



## 変数のソート

[プロジェクト全体] ビューのデータは、Unicodeコード順あるいはデータの長さでソートできます。

**[プロジェクト全体] ビュー内でデータをソートするには:**

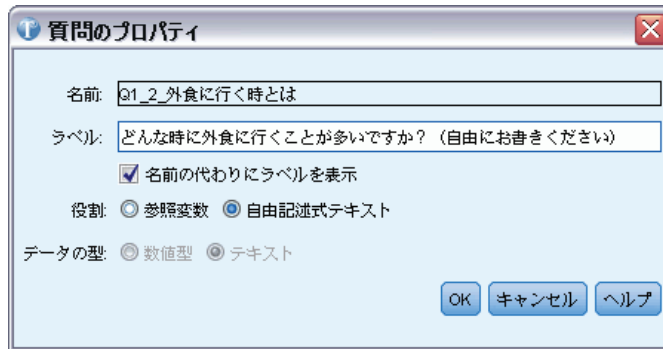
- ▶ ソートしたい列の列見出しを右クリックします。ポップアップメニューが開きます。
- ▶ 以下のソートオプションから好きなものを選択します。
  - **作成順でソート**:インポート中に読み込んだ順番で結果をソートします。
  - **昇順でソート:A-Z**:Unicodeコード順でソートします。
  - **昇順でソート:A-Z**:Unicodeコード順でソートします。
  - **昇順でソート:長さ**:長さでソートします。一番短い回答が一番上に表示されます。
  - **昇順でソート:長さ**:長さでソートします。一番長い回答が一番上に表示されます。

## 変数のプロパティの編集

インポートの際にデータを定義しようとする時、一意のIDを表わす変数、分析の対象となる質問を表わす変数、また場合によっては参照変数を指定するように指示されます。データをインポートした後に、これらの変数のプロパティに情報を追加したり、プロジェクトにおけるその役割を変更したいことがあります。たとえば、参照変数としてインポートした変数を分析の対象にしたい場合が考えられます。変数の以下のプロパティを変更できます。

- 変数名や変数ラベルを追加・変更します。
- 参照変数を自由記述変数に変更します。
- 自由記述変数を参照変数に変更します。
- IDや参照変数のデータの型を変更します。

図 4-4  
[タイプのプロパティ] ダイアログ ボックス



### カテゴリのプロパティを編集するには

- ▶ [プロジェクト全体] ビュー内で、プロパティを修正したい変数の列を選択します。この列見出しを右クリックし、ポップアップメニューを開きます。
- ▶ メニューからプロパティを選択します。[プロパティ] ダイアログ ボックスが開きます。
- ▶ 必要であれば、変数名や変数ラベルを追加・修正します。
- ▶ 変数名の代わりに変数ラベルを使用したい場合には、[名前代わりにラベルを表示] オプションを選択します。
- ▶ E 必要であれば、分析における変数の役割を [参照変数] あるいは [自由記述式テキスト] のいずれかに変更します。ID変数の役割は変更できません。自由記述式テキスト変数（質問）として作業した後にこの変数の役割を参照変数に変更すると、それまでのカテゴリ化の作業は失われます。
- ▶ 変数の [データの型を] [テキスト] あるいは [数値型] に変更します。

## プロジェクトの保存

プロジェクトを閉じたり、セッションを終了する際には、必要であれば変更を保存するというメッセージが表示されます。プロジェクトは、\*.tas という拡張子が付いたファイルに保存されます。

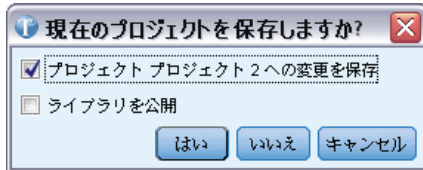
### プロジェクトを保存するために:

- ▶ [ファイル] メニューから、[ファイル > プロジェクトを保存] を選択します。このプロジェクトが保存されます。

### 終了時にプロジェクトを保存するために:

- ▶ プロジェクトを閉じようとする時、ダイアログボックスが開きます。このプロジェクトへの新しい変更を保存するかどうか、またこのライブラリを公開するかどうかを聞いてきます。

図 4-5  
現在のプロジェクト ダイアログボックスを保存

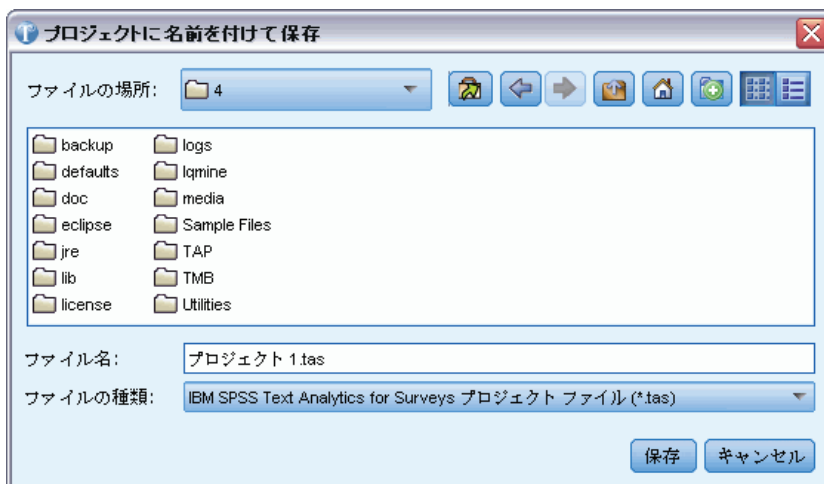


- ▶ [プロジェクト(プロジェクト名)への変更を保存] を選択します。
- ▶ ライブラリを公開して後で使いたい場合や、パブリックライブラリのバージョンを更新したい場合には、[ライブラリを公開] も選択します。公開する必要があるライブラリがない場合には、このオプションは使用できません。詳細は、9 章 p.232 [ライブラリの公開](#) を参照してください。
- ▶ [はい] をクリックして、保存します。ライブラリを公開を選択すると、別のダイアログボックスが開きます。詳細は、9 章 p.230 [ライブラリの共有](#) を参照してください。

### プロジェクトを別名保存するために:

同じ名前のプロジェクトがすでに存在しているという警告を受けた場合、またプロジェクトを別の名前で保存したい場合には[プロジェクトに名前を付けて保存] ダイアログボックスを開きます。

図 4-6  
ダイアログボックスとしてプロジェクトを保存



- ▶ [ファイル名] テキストボックスに、このプロジェクトのファイル名を新規に入力します。
- ▶ [保存] をクリックし、保存します。

## カテゴリ化の結果をエクスポート

場合によっては、カテゴリが作成された時点で分析が終わり、ということもあります。回答者が表わしている主要なテーマは何かということを知り、また回答者の何人が各テーマについて言及しているかどうかを知るだけで、テキスト分析の目的を達成したということもあるでしょう。一方、結果を表示するために表やグラフを作成して、レポートを実行したり、あるカテゴリに対してさらに分析を行っていくこともよくあります。アンケート内の他の変数を使用して、各カテゴリに含まれる回答者の特徴を調べたり、あるいは他の回答を検討するためにこれらのカテゴリを使用することもあります。

新しくカテゴリ化された結果を使って作業を続けたい場合には、定量的分析用アプリケーション（IBM® SPSS® Statistics Baseなど）へ、カテゴリをエクスポートできます。結果として得られたファイルには、回答IDならびに、カテゴリの名前とラベルが含まれています。しかし、参照変数の値や、自由記述式の回答は含まれていません。

注: 上位 5 つのカテゴリの棒グラフなど、集計グラフを生成することもできます。これらのグラフを HTML にエクスポートして、プレゼンテーションに使用できます。詳細は、p. 62 [棒グラフのエクスポート](#) を参照してください。

### エクスポートされたファイルの種類

エクスポートする際には、いくつかの種類 of ファイルを作成できます。

- SPSS Statistics ファイル (\*.sav) 詳細は、p. 58 [IBM SPSS Statistics](#) または [IBM SPSS Data Collectionにエクスポート](#) を参照してください。
- Microsoft Excel ファイル (\*.xls / \*.xlsx) 詳細は、p. 60 [Microsoft Excelにエクスポート](#) を参照してください。
- IBM® SPSS® Data Collection : 詳細は、p. 58 [IBM SPSS Statistics](#) または [IBM SPSS Data Collectionにエクスポート](#) を参照してください。Data Collection データモデルのData Collection 開発ライブラリも参照してください。

### 2値型出力vsカテゴリ型出力

IBM® SPSS® Text Analytics for Surveysでコード化されたテキストデータは、マルチアンサーのセット（多重回答形式）と等しいものになります。これは 1 人の回答者が複数の回答をするので、1 つの質問において複数の

カテゴリに割り当てられる可能性があるからです。このため、エクスポートするためには、このデータを特別な形式でコード化する必要があります。エクスポートでは2種類のアウトプットフォーマットが利用できます。

**2値型** および **カテゴリ型**です。

**2値型:**各回答IDのカテゴリ メンバーシップ フラッグを扱ったものです。データ内の各カテゴリ、各回答者（IDごと）バイナリ フラグを利用したカテゴリに属しているか、属していないかに関わりなく、真または偽でコード化されます。左の列がIDで、それから各カテゴリごとに右に1列ずつとなります。このデータタイプでは、回答ごとに無制限の個数のカテゴリを設定できます。10のカテゴリがある場合、10の新しい変数があります。

**カテゴリ型:**回答が所属する一連のカテゴリを扱ったものです。個々の回答に割り当てられる各カテゴリは、値（SPSS Statisticsの場合）あるいはカテゴリ自体（Microsoft Excelの場合）として表示されます。エクスポートされたカテゴリは表形式で構造化されています。左の列がIDで、それから各カテゴリ（少なくとも回答が1つ含まれるもの）ごとに1列ずつとなります。これらの列は特定のカテゴリではなく、割り当てられたカテゴリコードを記録するための入れ物のようなものです。回答ごとに、この回答が属するカテゴリコードがこの個々の入れ物に入れられます。回答に割り当てられるカテゴリの最大個数によって、作成されるべき変数の個数が決まってきます。カテゴリは10個あるけれども、そのうち回答者が含まれているのは4つである場合、これを表現するのに必要な変数の個数は4つになります。

- **SPSS Statistics/Data Collectionのために**データ内の各回答IDが割り当てられる各カテゴリには1からN（カテゴリコードの値としてはNが最大として）までの離散的な値が割り当てられています。コードフレームマネージャでコードを割り当てていない場合には、カテゴリが作成される際に、自動的にコードが割り当てられます。全カテゴリ数に満たないカテゴリ数を持つレコード（回答ID）の未使用部分のセルは、SPSS Statisticsのシステム欠損値（ピリオド）としてコード化されます。
- **Microsoft Excelのために**データ内の各回答のために、割り当てられた各カテゴリは、プロダクト インターフェイスの使用方法に応じて、カテゴリ名またはカテゴリ ラベルとして表示されます。全カテゴリ数に満たないカテゴリ数を持つレコード（回答ID）の未使用部分のセルは、Microsoft Excel内では空白としてコード化されます。

## IBM SPSS StatisticsまたはIBM SPSS Data Collectionにエクスポート

回答をカテゴリ化した後には、統計処理を用いて結果を分析する必要があります。IBM® SPSS® Text Analytics for Surveysでは、別の製品で使用できるようにフォーマットされたデータファイルを作成できます。以下はIBM® SPSS® Statistics（統計分析プログラム）や様々なIBM® SPSS® Data Collection製品でエクスポートする際の説明です。SPSS Text



Analytics for Surveysは、エクスポートファイルにマルチアンサーの変数を自動的に作成します。ファイルの形式は選択されたデータタイプ（2値型かカテゴリ型）によって変わってきます。

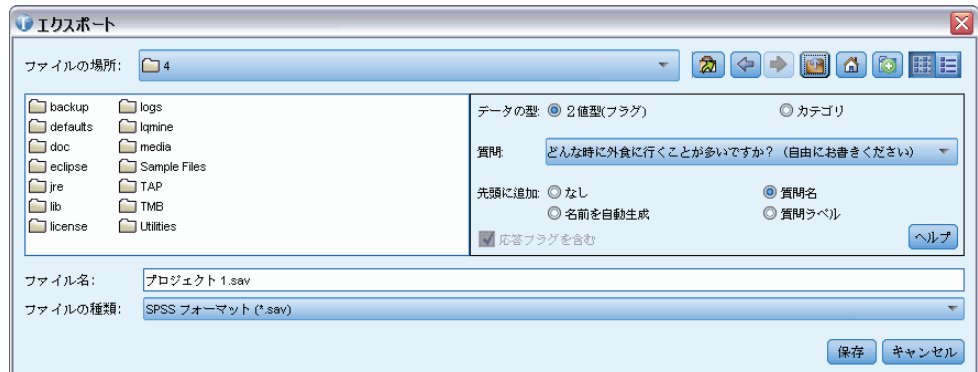
注:結果として得られたファイルには、回答IDならびに、カテゴリの名前とラベルが含まれています。しかし、参照変数の値や、自由記述式の回答は含まれていません。

**SPSS Statistics のみ**アウトプットでは、特定の質問に回答者が回答しなかったことによるデータやケースの欠如がデータセットにあると、アプリケーションはシステム欠損値をこれらのケースに割り当てます。SPSS Text Analytics for SurveysでエクスポートされるSPSS StatisticsファイルはSPSS Statistics の 7.5以前のバージョンではサポートされていません。

### データをエクスポートするには:

- ▶ ファイル > 結果の抽出メニューから、以下のオプションの1つを選択して抽出ダイアログボックスを開きます。
  - SPSS Statistics ファイル
  - Data Collection ファイル

図 4-7  
ダイアログボックスをIBM SPSS Statisticsファイル フォーマットに抽出



- ▶ [保存] ドロップダウンリストから、このファイルを保存するドライブならびにフォルダを選択します。
- ▶ [データの型] オプションを選択します。詳細は、 p. 57 [カテゴリ化の結果をエクスポート](#) を参照してください。
  - **2値型:**
  - **カテゴリ型:**Data Collectionデータファイルでは、このオプションは使用できません。デフォルトで2値型が選択されます。
- ▶ [質問] のドロップダウンリストから、エクスポートしたい質問を選択します。カテゴリ化の結果を1つの質問についてだけ、あるいはプロジェクト

全体でエクスポートするのかが選択できます。各質問を個別にエクスポートしたい場合には、各質問を1つずつ選択してこれをエクスポートする必要があります。あるいは、[プロジェクト全体]を選択し、すべての自由記述式の質問の結果をすべてエクスポートできます。

- ▶ [先頭に追加] オプションを選択すると、プロジェクト全体のカテゴリをエクスポートする際に、名前に接頭辞を加えることができます。このオプションは、複数の質問のデータをエクスポートする際に特に便利です。このオプションによって、元の変数ラベルあるいはカテゴリ名に接頭辞を加えることで、プロジェクト全体をエクスポートして複数の質問の結果を組み合わせる場合に、重複がないことを確認できます。以下のようなオプションがあります。
  - なし名前のおり、接頭辞は追加されません。
  - 質問名自由記述変数（質問）の名前を、出力されるカテゴリ変数名（プロジェクトでカテゴリ名かカテゴリラベルのどちらを使用していたのかによって変わってきます）の接頭辞として使用します。質問変数の名前は、元のデータソースにあるものです。出力されるカテゴリ変数の名前が、変数の命名規則に合致しない場合、また40文字を超える場合には、デフォルトの名前が作成されます（[自動生成された名前] オプションによる）。
  - 自動生成名カテゴリ名に自動的にQ1、Q2、Q3などと接頭辞を付けていきます。エクスポートされる一番最初の質問がQ1となります。
  - 質問ラベル自由記述変数（質問）ラベルを、出力されるカテゴリ変数名（プロジェクトでカテゴリ名かカテゴリラベルのどちらを使用していたのかによって変わってきます）の接頭辞として使用します。質問変数の名前は、元のデータソースにあるものです。出力されるカテゴリ変数の名前が、変数の命名規則に合致しない場合、また40文字を超える場合には、デフォルトの名前が作成されます（[自動生成された名前] オプションによる）。
- ▶ データ内に回答フラグがある場合には、これもエクスポートするかどうかを選択できます。回答フラグをエクスポートするには、このオプションを選択します。詳細は、p. 79 [フラグ回答](#) を参照してください。
- ▶ [ファイル名] テキストボックスで、表示されているデフォルトのプロジェクト名を選択するか、あるいはこのファイルに別の名前を入力します。
- ▶ [保存] をクリックし、結果をエクスポートします。

## Microsoft Excelにエクスポート

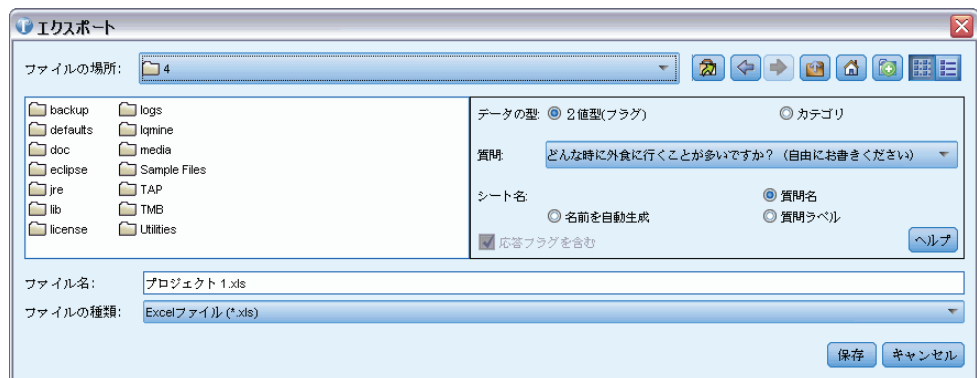
回答をカテゴリ化した後は、統計処理を用いて結果を分析する必要があります。IBM® SPSS® Text Analytics for Surveysでは、別の製品で使えるようにフォーマットされたデータファイルを作成できます。以下

は、Microsoft Excel形式でエクスポートする際の説明です。SPSS Text Analytics for Surveys は、エクスポートするファイル内に、自動的にマルチアンサーの変数を作成します。ファイルの形式は選択されたデータタイプ（2 値型かカテゴリ型）によって変わってきます。結果として得られたファイルには、回答IDならびに、カテゴリの名前とラベルが含まれています。しかし、参照変数の値や、自由記述式の回答は含まれていません。

### データをエクスポートするには：

- ▶ メニューの **ファイル** > **結果のエクスポート** > **Microsoft Excel ファイル** を選択します。  
[エクスポート] ダイアログボックスが開きます。

図 4-8  
ダイアログボックスをMicrosoft Excelファイルに抽出



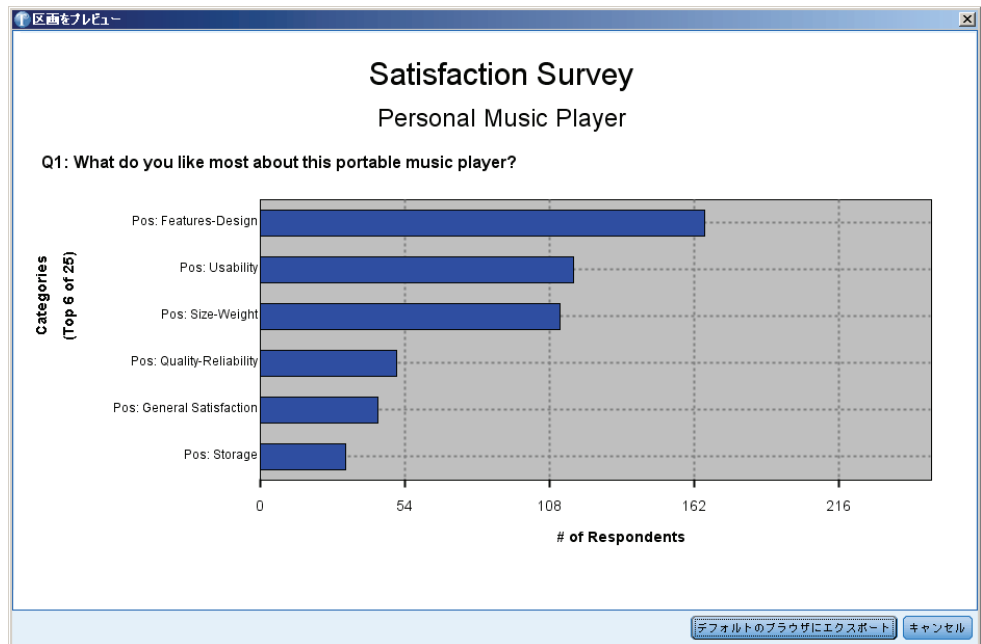
- ▶ [保存] ドロップダウンリストから、このファイルを保存するドライブならびにフォルダを選択します。
- ▶ [データの型] オプションを選択します。詳細は、 p. 57 [カテゴリ化の結果をエクスポート](#) を参照してください。
  - **2 値型**：
  - **カテゴリ型**：IBM® SPSS® Data Collectionデータファイルでは、このオプションは使用できません。デフォルトで2 値型が選択されます。
- ▶ [質問] のドロップダウンリストから、エクスポートしたい質問を選択します。カテゴリ化の結果を1つの質問についてだけ、あるいはプロジェクト全体でエクスポートするのかが選択できます。各質問を個別にエクスポートしたい場合には、各質問を1つずつ選択してこれをエクスポートする必要があります。あるいは、[プロジェクト全体]を選択し、すべての自由記述式の質問の結果をすべてエクスポートできます。
- ▶ [シート名] のオプションを選択し、エクスポートされたMicrosoft Excelファイル内の各ワークシートにどのように名前が付けられるのを設定できます。以下のようなオプションがあります。

- **質問名自由記述変数（質問）**の名前をワークシートの名前として使用します。質問の名前は元のデータソースにあるものです。出力されるカテゴリ変数の名前が、変数の命名規則に合致しない場合、また40文字を超える場合には、デフォルトの名前が作成されます（[自動生成された名前] オプションによる）。
  - **自動生成名カテゴリ名**に自動的にQ1、Q2、Q3などと接頭辞を付けていきます。エクスポートされる一番最初の質問がQ1となります。
  - **質問ラベル自由記述変数（質問）**のラベルがある場合、これをワークシートの名前として使用します。出力されるカテゴリ変数の名前が、変数の命名規則に合致しない場合、また40文字を超える場合には、デフォルトの名前が作成されます（[自動生成された名前] オプションによる）。
- ▶ データ内に回答フラグがある場合には、これもエクスポートするかどうかを選択できます。回答フラグをエクスポートするには、このオプションを選択します。詳細は、 p. 79 [フラグ回答](#) を参照してください。
- ▶ [ファイル名] テキストボックスで、表示されているデフォルトのプロジェクト名を選択するか、あるいはこのファイルに別の名前を入力します。

## 棒グラフのエクスポート

カテゴリとデータの処理を終了した後、棒グラフ レポートをエクスポートして、分析結果を共有することができます。アウトプットは質問1つごとに1本の棒グラフを作成します。各グラフに使用するトップ カテゴリの数を選択し、任意の質問に対するトップ5またはトップ10カテゴリを表示することができます。グラフはデフォルトのブラウザにエクスポートでき、他の製品やプレゼンテーションで使用できるように画像を保存することができます。

図 4-9  
ブラウザ ウィンドウ内でのサンプル棒グラフ



### 棒グラフのエクスポート

- ▶ メニューの **カテゴリ** > **棒グラフのエクスポート** を選択します。[棒グラフのエクスポート] ダイアログボックスが開きます。
- ▶ このトピック内に説明されたオプションを利用して、棒グラフを設定します。
- ▶ **作成** をクリックしてグラフを作成し、プレビューパネルで表示します。
- ▶ **デフォルト ブラウザにエクスポート** をクリックして、ブラウザ ウィンドウでグラフを確認します。

図 4-10  
[棒グラフのエクスポート]ダイアログボックス

## レポート オプション

**グラフのエクスポート:** 棒グラフのエクスポートを、プロジェクトの全質問で行うか、現在選択中の質問だけで行うかを選択します。

**グラフに使用するトップ カテゴリの数** グラフで表示するカテゴリの最大数を選択します。これらのカテゴリのうちレコード数が最大のものが最初に使用されます。

**日付の古い抽出結果がある場合は、再抽出** このオプションを選択すると、日付の古い抽出結果がある場合に、グラフの作成前に再抽出を自動的に行います。

**メインタイトル** グラフのメインタイトルを入力たとえば、アンケートのタイトルを使用します。

**サブタイトル** グラフのサブタイトルを入力たとえば、会社名やアンケートの実施年を入力します。

**質問タイトル** 各チャートの確認を容易にするために、タイトルは質問に由来します。質問の変数名、ラベルを使うか、あるいは名前を使用しないかを選択します。

**X軸ラベル** グラフのX軸のラベルを定義します。ラベルはデフォルトで提示されます。

**Y軸ラベル** グラフのY軸のラベルを定義します。ラベルはデフォルトで提示されます。

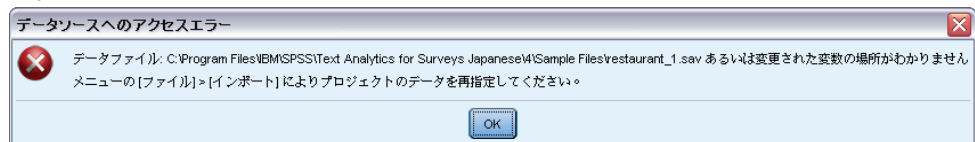
**バーの色** 棒グラフのバーの色を選択します。この色はすべての質問に適用されます。

**フォント**。グラフのタイトルとラベルのフォントを選択します。

## データソースの変更

何らかのプロジェクトを開くたびに、それに対応するデータセットが開きます。データが見つからない場合は、エラーメッセージが表示されます。データが見つからない理由としては、別の場所に移動されている、偶然に削除されている、名前が変更されているなどが考えられます。または、データソースを切り替える必要姓があります。

図 4-11  
欠損データのエラーメッセージ

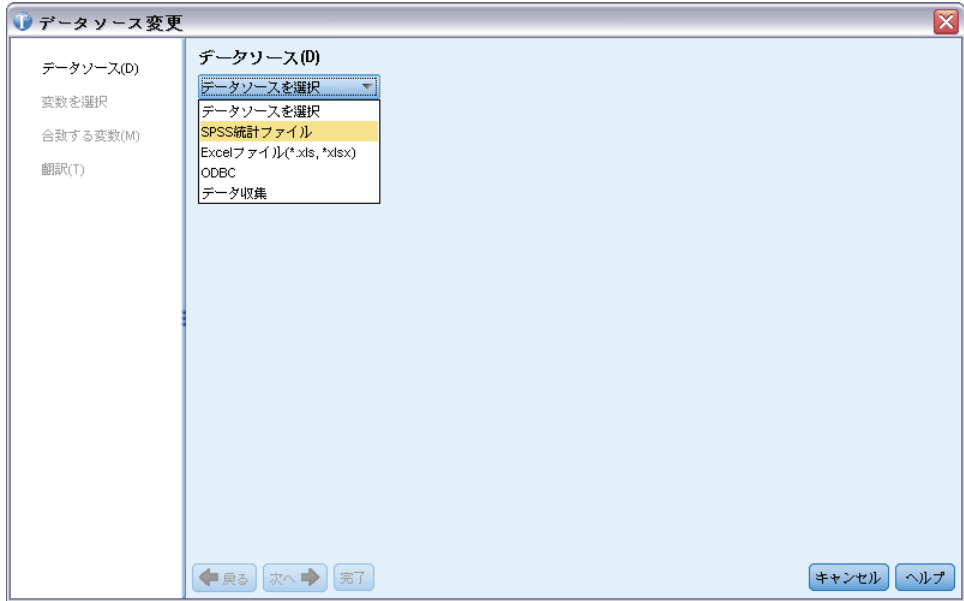


データを引き続き使用するためには、データ位置を変更して適切なデータソースを選択する必要があります。新しい変数、変数の名前変更、変数の欠損など、データ内の変数に変更されている場合は、以前にインポートした変数を新しい変数に合致させるように求められます。

### データソースを変更するために:

- ▶ このエラーメッセージが表示された場合はOKをクリックします。
- ▶ [ファイル] メニューから、[ファイル > データソースの変更] を選択します。データソースの変更ウィザード ダイアログが表示されます。

図 4-12  
データソースの変更ウィザード



## データソースの選択

ウィザードが開いたら、最初にデータソースを選択します。IBM® SPSS® Text Analytics for Surveysは1レコードあたり500文字（全角）、1000件のレコードを持つデータセットを処理することを標準的な使用方法と想定し設計されています。しかし、レコードに含まれるテキストの量に応じてパフォーマンスは変わってきます。パフォーマンスや推奨環境に関しては、インストールガイドに記述されているので参考にしてください。

**重要!** このデータをインポートするには、各レコードごとに一意の値を持つID変数が必要です。

データソースは以下の中から1つを選択します。

- **SPSS Statistics ファイル** (\*.sav)
- **Microsoft Excel ファイル** (\*.xls / \*.xlsx)
- **ODBC** (Microsoft Open Database Connectivity プロトコル) データベース
- **Data Collection** データモデルこのオプションはデータモデルをインストールしている場合に限り利用できます。



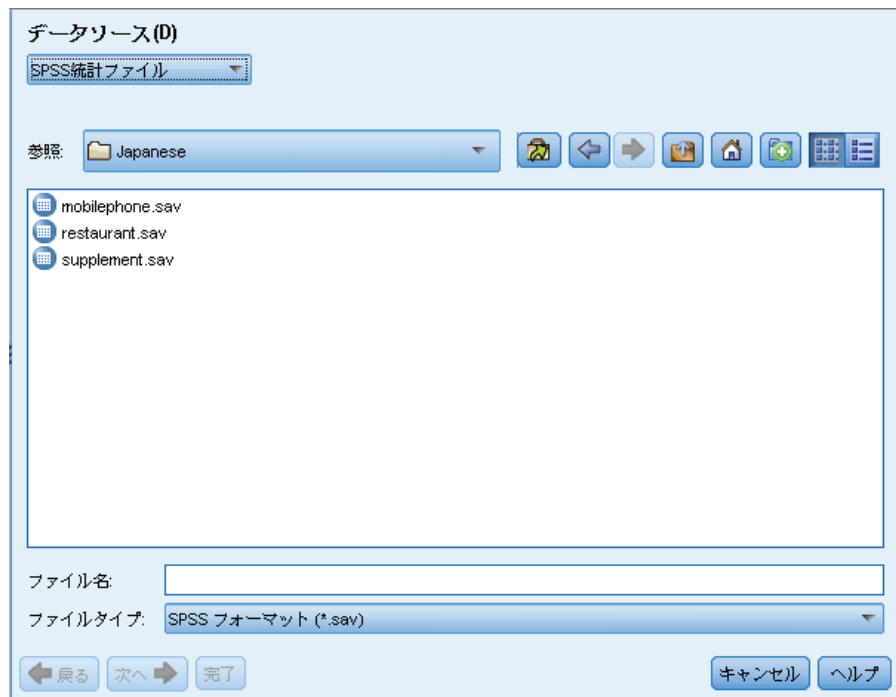
## IBM SPSS Statistics ファイルの使用

IBM® SPSS® Statistics (.sav) ファイルをIBM® SPSS® Text Analytics for Surveysにインポートできます。このデータをインポートするには、各レコードごとに一意の値を持つID変数が必要です。

**重要!**SPSS Statistics (.sav) ファイルは 4000 文字を超えるレコードとインポートすることはできません。

注:SPSS Text Analytics for Surveysは1レコードあたり500文字（全角）、1000件のレコードを持つデータセットを処理することを標準的な使用方法と想定し設計されています。しかし、レコードに含まれるテキストの量に応じてパフォーマンスは変わってきます。パフォーマンスや推奨環境に関しては、インストールガイドに記述されているので参考にしてください。

図 4-13  
IBM SPSS Statistics ファイルのデータソース オプション



### IBM SPSS Statistics からデータを取得するために

- ▶ ウィザードの最初の画面で、ドロップダウン リストから SPSS Statistics ファイルを選択します。ウィザードが SPSS Statistics ファイルのオプションを表示します。
- ▶ [参照] ドロップダウン リストから、ファイルを投入するドライブとフォルダを選択します。

- ▶ リストからファイルを選択します。[ファイル名] テキストボックス内にその名前が表示されます。
- ▶ [次へ]をクリックして変数を選択します。 詳細は、 p. 36 [変数の選択](#) を参照してください。

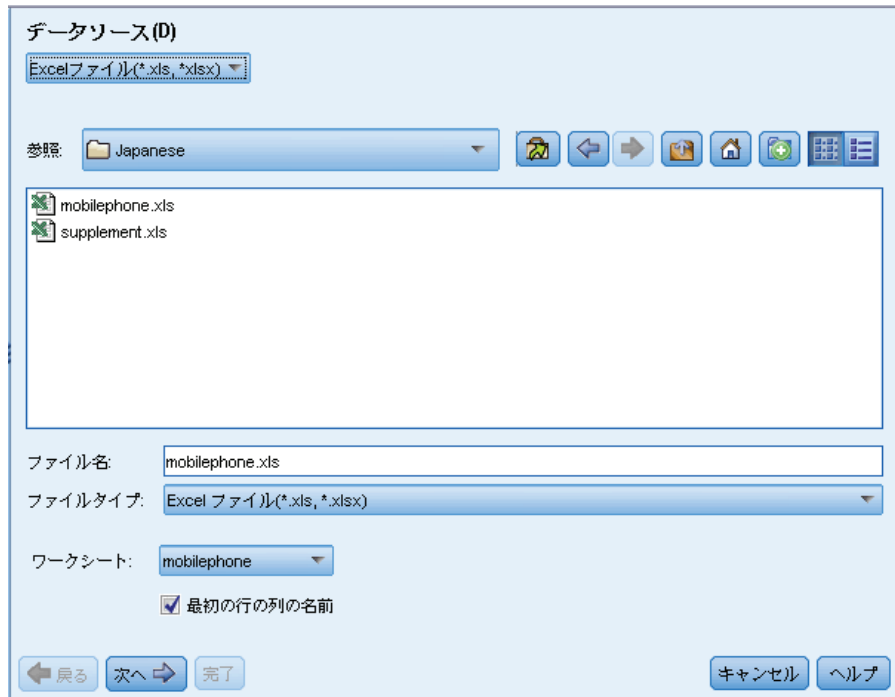
## Microsoft Excel ファイルの使用

Microsoft Excel (.xls / .xlsx) ファイルを IBM® SPSS® Text Analytics for Surveys にインポートできます。このデータをインポートするには、各レコードごとに一意の値を持つID変数が必要です。

**重要!**Microsoft Excelファイルをインポートする際に、[1行めに列名] オプションを選択できます。このオプションを使用するには、データが開始される上の行ではなく、ワークシートの最初の行に列名を入力する必要があります。たとえば、データおよび列名が行 7 で始まる場合、ファイルを正しくインポートするためには、インポート前に 1 ~ 6 行目を削除する必要があります。

注:SPSS Text Analytics for Surveysは1レコードあたり500文字（全角）、1000件のレコードを持つデータセットを処理することを標準的な使用方法と想定し設計されています。しかし、レコードに含まれるテキストの量に応じてパフォーマンスは変わってきます。パフォーマンスや推奨環境に関しては、インストールガイドに記述されているので参考にしてください。

図 4-14  
Microsoft Excel ファイルのデータソース オプション



#### Microsoft Excel からデータを取得するために

- ▶ ウィザードの最初の画面で、ドロップダウン リストからExcelを選択します。ウィザードが Microsoft Excel ファイルのオプションを表示します。
- ▶ [参照] ドロップダウン リストから、ファイルを投入するドライブとフォルダを選択します。
- ▶ リストからファイルを選択します。[ファイル名] テキストボックス内にその名前が表示されます。
- ▶ ドロップダウン リストからワークシートを選択します。1 つのワークシートからのみカテゴリをインポートできます。複数のワークシートに含まれているデータを扱う際には、複数のプロジェクトを作成する必要があります。
- ▶ このワークシートの最初の行に列の見出しを指定する場合、[最初の行に列名]を選択します。このオプションを使用するには、データが開始される上の行ではなく、ワークシートの最初の行に列名を入力する必要があります。たとえば、データおよび列名が行 7 で始まる場合、ファイルを正しくインポートするためには、インポート前に 1 ~ 6 行目を削除する必要があります。アプリケーションはこれら（または、列の見出しが IBM® SPSS® Statistics の変数名規則に従っていない場合は変換されたバージョン）

を変数名として使用できます。そうでない場合には、ワークシートの列名（アルファベット）が識別用に使用されます。

- ▶ [次へ]をクリックして変数を選択します。 詳細は、 p. 36 **変数の選択** を参照してください。

## ODBC からのデータの使用

データベース ソース（データベース）からのデータは簡単にIBM® SPSS® Text Analytics for Surveysにインポートできます。ODBCドライバを使用するデータベースならば、SPSS Text Analytics for Surveys がインストールされたコンピュータに適切なドライバをインストールした後に直接読み込むことができます。このデータをインポートするには、各レコードごとに一意の値を持つID変数が必要です。

注:SPSS Text Analytics for Surveys は1レコードあたり500文字（全角）、1000件のレコードを持つデータセットを処理することを標準的な使用方法と想定し設計されています。しかし、レコードに含まれるテキストの量に応じてパフォーマンスは変わってきます。パフォーマンスや推奨環境に関しては、インストールガイドに記述されているので参考にしてください。

図 4-15  
ODBC のデータソース オプション

データソース(D)

ODBC

ソース(DSN)

名前	説明
MS Access Database	Microsoft Access Driver (*.mdb, *.accdb)
Excel Files	Microsoft Excel Driver (*.xls, *.xlsx, *.xlsm, *.xlsb)
dBASE Files	Microsoft Access dBASE Driver (*.dbf, *.ndx, *.m...

ODBC

ユーザー:

パスワード:

テーブル

SQL

戻る 次へ 完了 キャンセル ヘルプ

### ODBC から使用するために

- ▶ ウィザードの最初の画面で、ドロップダウン リストから ODBC を選択します。ウィザードに ODBC オプションが表示されます。
- ▶ 登録された ODBC ソースのリスト内でデータソースを特定するか、あるいは [ソース (DSN) ] テキストボックスにデータソースの名前を入力します。このリストに表示されていないデータソースを新しく登録するには、[ODBC] をクリックします。これで、ほとんどの Microsoft Windows コンピュータに付属の [ODBCデータソースアドミニストレータ] が開きます。これがない場合には、ODBCインポートを使用できません。詳細については、Microsoft Windows のヘルプを参照してください。
- ▶ 使用するデータソースがパスワードで保護されている場合には、ユーザー名とパスワードを入力する必要があります。これはプロジェクトを開くたびに要求されます。セキュリティ上の理由で、ユーザー名ならびにパスワードはプロジェクト内には保存されません。
- ▶ データの選択方法には2種類あり、Table または SQL. テーブルを直接選択するか、あるいはSQLコマンドを使用してデータを選択することができます。
- ▶ [次へ]をクリックして変数を選択します。 詳細は、 p. 36 [変数の選択](#) を参照してください。

## IBM SPSS Data Collection データの使用

### IBM SPSS Data Collectionからインポートするために

- ▶ ウィザードの最初の画面で、ドロップダウン リストからData Collectionを選択します。IBM® SPSS® Data Collectionデータモデルのオプションは、他の製品でこのデータモデルがインストールされている場合にのみ使用可能です。

## 変数の選択

データソースの選択後のステップは、インポートする変数の指定です。プロジェクトにインポート可能な変数は3種類あります。

### 一意のID変数(必須)

ID変数は数値や文字列のキーで、各回答者を特定するものです。データファイルを読み込むためには、一意のID変数に基づいて並んでいる必要はありません。プログラムに読み込んだ後に、レコードをさまざまな基準でソートできます。詳細は、 p. 54 [変数のソート](#) を参照してください。データをインポートするには、この変数が必要です。インポートされた各レコード (ケース) は一意のID値を持つ必要があります。

インポートが失敗する原因は2つ考えられます。

- 重複するID値が検知された場合
- ID値が空白のレコード

注: 重複するIDが検知された場合、使用中のコンピュータにIBM® SPSS® Statisticsがインストールされているならば、[重複ケースの特定] 手続きを使用できます。[データ] メニューから [重複ケースの特定] を選択し、[オプション] を使ってどのレコードを保持するのか (プライマリケースとするのか) を指定します。

### 自由記述変数(必須)

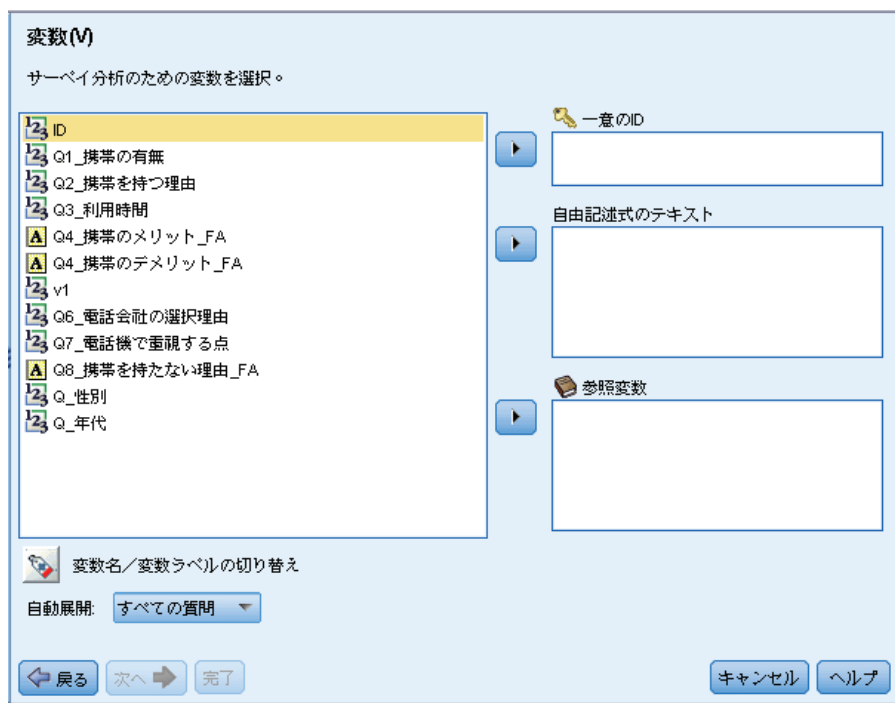
自由記述変数は、アンケート内の質問に対するテキスト形式の回答を表わしています。データをインポートする際には、この変数が少なくとも1つは必要です。これらの変数としてはSPSS Statistics内の文字列あるいは長い文字列型の変数、Microsoft Excel内の一般のセルやテキストセルを含む列、データベースのテキストあるいはノートフィールドがあります。自由記述変数はそれぞれ独立したものとして分析されます。SAVファイルからインポートする変数のサイズ (幅) の上限は4,000文字です。

### 参照変数(オプション)

参照変数はオプションです。これは通常カテゴリ型で、参照目的でインポートされます。参照変数はテキスト分析には使用されませんが、理解や解釈を助けるような、回答者に関する補足的な情報を提供します。参照変数としては、人口統計量変数がよく使用されます。これはどのキーワードやカテゴリがどの集団や個人によって使用されたのかを理解する助けになるからです。たとえば学生やトレーニングの評価においては、性別、部署、職業、学部などが考えられます。インポートされたすべての参照変数は、プロジェクト全体ビューに表示されます。参照変数を質問ビューのデータパネルに表示させることもできます。また視覚化パネルの棒グラフ内で参照変数を選択し、回答者のサブセットを詳細に検討することもできます。

注: SPSS Statisticsデータファイルから読み込んだ参照変数では、データパネルの列見出しに変数ラベル (あらかじめ設定されている場合) が表示され、その値ラベル (あらかじめ設定されている場合) がセルに表示されます。

図 4-16  
変数の選択



### 変数および抽出オプションを選択するために

- ▶ 使用可能な変数のリストから、データセット内のID変数に該当する変数を選択します。次に矢印ボタンをクリックし、[一意のID] ボックスに移動します。IDは一意の数字あるいは英数文字列で、あるレコードを他のレコードと区別するためのものです。使用するデータに重複するIDが含まれている場合、エラーメッセージが表示されます。このような場合、データを整理してから、再度インポートを行ってください。
- ▶ 使用可能な変数のリストから、自由記述式の回答変数に対応する変数を、1つあるいは複数選択します。次に矢印ボタンをクリックし、この変数（複数可）を[自由記述式のテキスト] リストに移動します。インポートされたそれぞれの変数は独立した質問で、その回答レコードが分析やカテゴリ化の対象となります。
- ▶ 使用可能な変数のリストから、参照変数あるいはカテゴリ変数に該当する変数を、1つあるいは複数選択します。次に矢印ボタンをクリックし、この変数を[参照変数] リストに移動します。参照変数は自動カテゴリ作成手法では使用されません。しかしながら、回答をカテゴリ化する際にこの内容を見ることで、より詳しい情報に即した決断を行うことができます。

- ▶ 変数名の変わりに変数ラベルを表示したい場合は、左の変数リストの下にあるボタンをクリックします。
- ▶ 抽出設定を変更するには、ドロップダウン リストから選択します。デフォルトでは、**最初の質問のみ**が選択されています。これは自由記述変数を複数選択すると、ウィザード終了後に、最初の質問で抽出プロセスが自動的に開始するという意味です。抽出は大きなデータセットからも行えます。そのため、時間に応じて**なしたはずすべての質問**の抽出設定を選択できます。
- ▶ 変数を全て選択してから、[次へ]をクリックします。

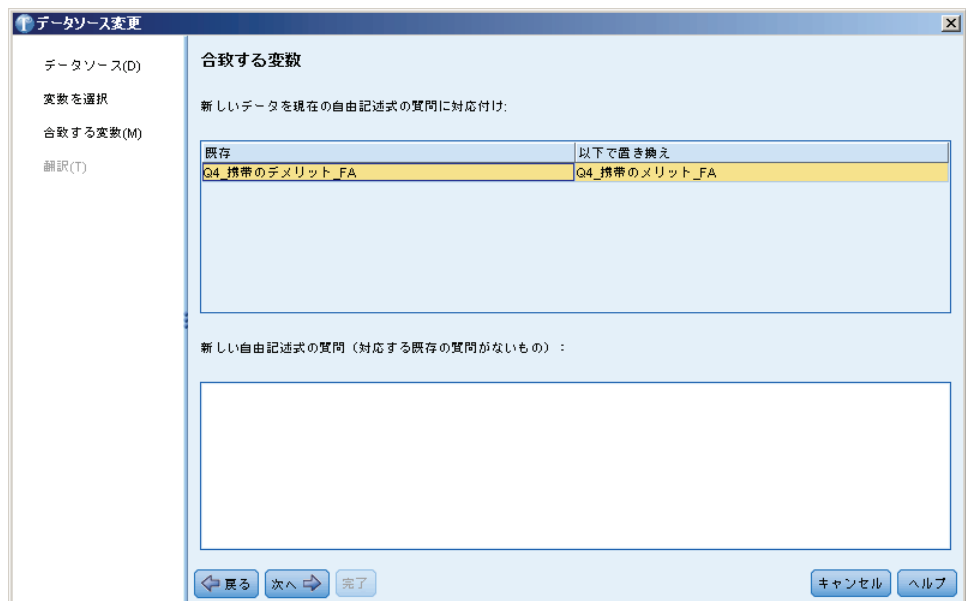
## 変数の合致

以前のステップで変数を選択した後、IBM® SPSS® Text Analytics for Surveysは以前にインポートした変数と選択した変数の合致を試みます。合致は自動的に行われますが、**以下で置換列内**をクリックして別の変数を選択しても変数の合致を行うことができます。

以前のプロジェクトに含まれていた自由記述変数が新規データファイルに含まれていない場合は、リストから「なし」を選択します。これで、昔の質問に関連するデータはこのプロジェクトから廃棄されます。

既存のプロジェクトには対応するものがない変数が新規データファイル内にあった場合には、これらは下の[新しい自由記述式の質問]に表示されます。データセットの変更後は、これらの残りの変数は、プロジェクトの新しい質問に表示されます。

図 4-17  
既存のプロジェクトダイアログボックスへの内容の合致





**既存の変数に合致させるために：**

- ▶ 新しい変数をデータファイル内で以前に見つかった変数へ合致させる処理は自動的に試みられますが、**以下で置換列内をクリックして別の変数を選択すれば**、変数を合致させる方法を変更することができます。以前のステップで、**以下で置換列内で指定されるよりも多くの変数を選択している場合は**、**[新しい自由記述式の質問]**（既存の質問に対応するものがない）に表示されます。データセットの変更後は、これらの残りの変数は、プロジェクトの新しい質問に表示されます。
- ▶ 既存の変数が新しいデータセットに合致しない場合は、**なし**を選択し、既存の質も文を破棄します。
- ▶ プロジェクト内に、カテゴリに強制投入・排除された回答やフラグが含まれている場合、これら回答IDに固有の結果を保持するのがあるいは破棄するのを聞かれます。通常は、別のデータ（新しい質問や他の回答者等）をインポートする場合はこういった情報を破棄し、間違った結果にならないようにします。同じデータファイルをインポートする場合には回答IDが前のデータと対応しているので、この情報を保持するのが普通です。

**英語への翻訳**

英語以外のテキストを使用している場合は、Language Weaverに接続して英語に翻訳することができます。英語への翻訳のみが可能です。Language Weaverを適切に設定し、接続を翻訳に定義する必要があります。詳細は、2章 p.24 [オプション：\[翻訳\] タブ](#) を参照してください。

図 4-18  
翻訳オプション



### 英語に翻訳するには

- ▶ テキスト データをライセンスされた言語から英語に翻訳するためには、英語に翻訳チェックボックスを選択します。
- ▶ 言語ペア接続リストから、使用する Language Weaver 言語ペアの接続を選択します。ローカル コンピュータで Language Weaver を設定した場合、これらの言語ペアが自動的にこのリストに表示されます。[オプション] ダイアログ ボックスの [翻訳] タブ内で、オンライン サービス接続を追加、変更、テストできます。詳細は、2 章 p.24 オプション: [翻訳] タブ を参照してください。
- ▶ 必要に応じて、**翻訳精度**を指定してください。精度と競合する速度を指定する 1 ~ 3 の数値を選択してください。値が小さいほど、翻訳結果が速く作成されますが、精度は劣ります。値が大きいと、精度の高い結果が作成されますが、時間がかかります。時間を最適化するには、低いレベルから始め、結果を見た後でより高い精度が必要だと感じた場合にのみ、レベルを高くすることをお勧めします。
- ▶ 以前に Language Weaver のカスタム辞書を作成している場合は、翻訳時に使用することができます。カスタム辞書を使用するには、**カスタム辞書**を使用

チェックボックスを洗濯し、辞書名を入力します。複数の辞書を使用する場合は、辞書名をコンマで区切ってください。

- ▶ 新しいプロジェクトウィザードで、次へをクリックしてカテゴリーとリソースの選択を開始します。詳細は、3章 p.40 [カテゴリおよびリソースの選択](#) を参照してください。
- ▶ データセットの変更ウィザードで、終了をクリックしてデータセットを終了し、翻訳プロセスを開始します。

#### 翻訳をスキップするには

- ▶ 英語に翻訳オプションの選択を解除します。
- ▶ 新しいプロジェクトウィザードで、次へをクリックしてカテゴリーとリソースの選択を開始します。詳細は、3章 p.40 [カテゴリおよびリソースの選択](#) を参照してください。
- ▶ データセットの変更ウィザードで、終了をクリックしてデータセットを終了します。

## データの更新

プロジェクト データを使用する場合、元のデータ ソースを変更する必要があります。たとえば、レコードの追加や削除が必要になります。データの更新機能を使ってデータを更新することができます。しかしながら、変数の定義やファイル名を変更した場合などには、データをすべて再インポートする必要があります。詳細は、p.65 [データ ソースの変更](#) を参照してください。

#### データを更新するために:

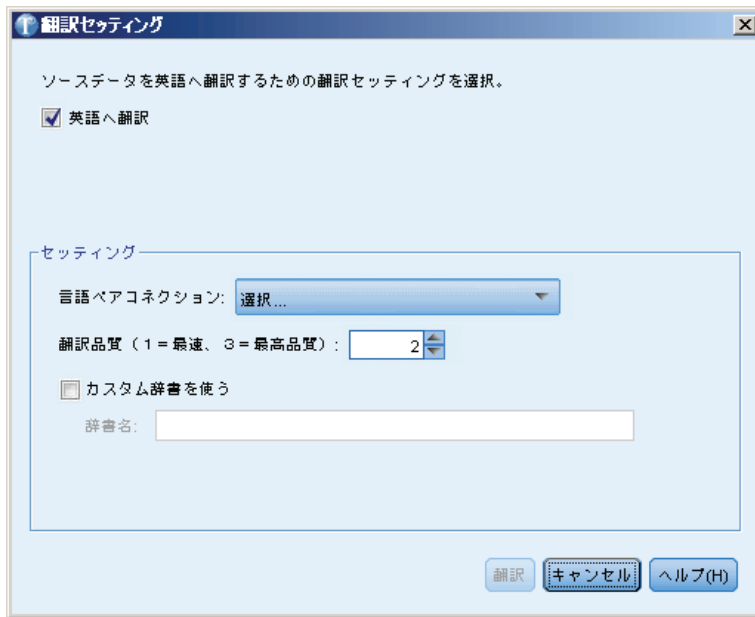
- ▶ メニューから、[ファイル]>[データの更新]を選択します。データが再度読み込まれ、変更が反映されます。
- ▶ 以前に英語に翻訳されている場合は、翻訳設定ダイアログが表示され、言語ペアを選択して再翻訳ができます。詳細は、p.77 [英語への翻訳](#) を参照してください。

## 英語への翻訳

翻訳はいつでも更新することができます。翻訳後は、翻訳結果が新しい翻訳と同期していないため、再抽出が必要になります。

注: 新しいデータを翻訳する場合は、新しいプロジェクトを作成する時に、新しいプロジェクトウィザードから直接翻訳できます。詳細は、p.38 [英語への翻訳](#) を参照してください。

図 4-19  
翻訳設定



### 英語に翻訳するには

- ▶ メニューの **ツール > 翻訳設定** を選択します。翻訳設定ダイアログが表示されます。
- ▶ テキスト データをライセンスされた言語から英語に翻訳するためには、英語に翻訳チェックボックスを選択します。
- ▶ 言語ペア接続リストから、使用する Language Weaver 言語ペアの接続を選択します。ローカル コンピュータで Language Weaver を設定した場合、これらの言語ペアが自動的にこのリストに表示されます。[翻訳設定] ダイアログでネットワーク (WAN) またはオンライン サービス (HTTP) 接続を追加またはテストできます。詳細は、2 章 p.24 **オプション: [翻訳] タブ** を参照してください。
- ▶ 必要に応じて、**翻訳精度**を指定してください。精度と競合する速度を指定する 1 ~ 3 の数値を選択してください。値が小さいほど、翻訳結果が速く作成されますが、精度は劣ります。値が大きいと、精度の高い結果が作成されますが、時間がかかります。時間を最適化するには、低いレベルから始め、結果を見た後でより高い精度が必要だと感じた場合にのみ、レベルを高くすることをお勧めします。
- ▶ 以前に Language Weaver のカスタム辞書を作成している場合は、翻訳時に使用することができます。カスタム辞書を使用するには、**カスタム辞書を使用**

チェックボックスを洗濯し、辞書名を入力します。複数の辞書を使用する場合は、辞書名をコンマで区切ってください。

- ▶ [翻訳] をクリックして、翻訳プロセスを開始します。翻訳進行ダイアログが表示されます。

## プロジェクトの共有

自分のプロジェクトを他のユーザーと共有できます。また共有することで、別のコンピュータ上でこのプロジェクトを開いて作業することもできます。

### プロジェクトを共有するには:

- ▶ [ファイル] メニューから、[ファイル > プロジェクトを保存] を選択します。このプロジェクトが保存されます。
- ▶ プロジェクトファイルを他のコンピュータあるいは人に送ります。このプロジェクトファイルには、すでにインポートされているデータファイルへの参照情報が含まれています。ユーザーがプロジェクトファイルを開く際、IBM® SPSS® Text Analytics for Surveysでは、プロジェクトファイル内のローカルライブラリと、パブリックライブラリのどちらを使用するのかが選択できます。
- ▶ 他のユーザーがSPSS Text Analytics for Surveysでファイルを開いている場合、同ユーザーはプロジェクト内に含まれるローカル ライブラリを使用するか、既に持っているパブリック バージョンのライブラリを使用するかを選択できます。通常は、同じ結果が得られるようにするためには、ローカルバージョンを使用してください。
- ▶ SPSS Text Analytics for Surveys がデータを見つけられない場合、メッセージが表示され、ユーザーにデータの再インポートが必要であることが警告されます。詳細は、 p. 65 [データ ソースの変更](#) を参照してください。

## フラグ回答

アンケートを分析する際の進行状況をモニターするために、データパネル内で回答（印）にフラグを付けることができます。回答にフラグを付ける必要性が出る場合としては、以下のようなものがあります。

- 自分で確認した回答にフラグを付けることで、どこまで確認したのかわかるようにしたい。
- どう扱ったらいいかわからない回答にフラグを付けておきたい。
- フラグを付けて、他のプログラムにエクスポートしたい。

フラグを付けた回答を使って作業を続けることができます。フラグは純粋に記録のためのものです。以下のようなフラグを選択できます

テーブル 4-1  
フラグの説明




フラグ	説明
	完了フラグ。終了したと思う回答を表わします。
	重要フラグ。重要だと思う回答を表わします。

図 4-20  
[データ] パネル内の回答フラグ

	Id	回答	回答	カテゴリ
1	36		美味しい物が食べたい時や、友人等と会う時。	なる時 会う友人
2	53		1. 美味しいお店の評判を聞いて 2. 友人と話すため	話する友人
3	81		美味しいものが食べたいとき、休日、何かの記念日、気分を変えたいとき、久しぶりの友人と会うとき。	休日 会うとき 日 友人
4	84		友人と語りたい時。ストレス発散したい時。美味しいお店を見つけたい時。	ストレス 時 友人
5	89		美味しいものが食べたい時	時

#### 回答にフラグを付けるには：

- ▶ [データ] パネル内で、フラグを付けたい回答を選択します。
- ▶ E メニューから、編集 > 回答にフラグを付ける を選択し、使用したいフラグのタイプ（[重要フラグ] か [完了フラグ]）を選択します。選択したフラグが割り当てられます。データパネル内でフラグ列が表示されていない場合でも、これで[フラグ] 列が表示されるようになります。ステータスバーが更新され、フラグの付いた回答の個数が表示されます。

#### フラグを消去するには：

- ▶ データパネル内で、フラグを除去したい回答を右クリックします。
- ▶ 右クリックしてポップアップメニューから、[回答にフラグを付ける] > [フラグを消去] を選択します。選択されたフラグが除去されます。

## プロジェクトのステータスバー

作業中のウィンドウやビューごとに、異なるステータスバーがあります。デフォルトでは、プロジェクトを開くたびにステータスバーが表示されます。このステータスバーにはプロジェクトならびにそこに含まれ

る要素の要約情報が表示されます。ステータスバーは希望に応じて表示したり非表示にしたりできます。

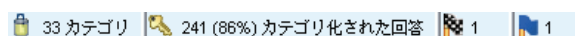
### ウィンドウ内でステータスバーを表示/非表示にするには:

- ▶ メニューから、[表示]>[ステータスバー]を選択します。

### テキスト分析ウィンドウ

ステータスバーには、プロジェクト内の質問と回答に関する要約情報が表示されます。テキスト分析ウィンドウ内のどこを使用しているのかによって、ステータスバー内の情報が変わってきます。重要フラグあるいは完了フラグの付いた回答の個数も表示されます。

図 4-21  
テキスト分析ウィンドウ内のステータスバー質問ビュー



質問ビューを開いているときは、この質問のカテゴリの個数と回答に対するカテゴリ化のパーセントが表示されます。プロジェクト全体ビューにいるときは、プロジェクト全体に関する情報が表示されます。

図 4-22  
テキスト分析ウィンドウ内のステータスバープロジェクト全体ビュー



ステータスバーの各要素は以下の表のようになります。

テーブル 4-2  
テキスト分析ウィンドウ:ステータスバーの説明

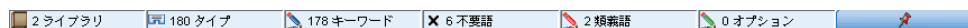
要素	説明
レコード	データ内のレコードの個数。
質問	プロジェクト内の質問の個数。
参照変数	インポートされた参照変数の個数。参照変数は、参照目的でインポートされた変数で、分析の対象ではありません。
カテゴリ	各質問ごとのカテゴリの個数。空白のカテゴリが存在する場合、空白のカテゴリの数がカッコ内に表示されます。
カテゴリ化	辞書エディタウィンドウ

### リソース エディタ ウィンドウ

ステータスバーは、プロジェクト内の言語リソースに関する情報を表示します。このバー上の[強制キーワード]の部分は操作可能で、ここをクリックして作業を行うことができます。ライブラリを使用している際には、その中である要素を使用不可にすることで、処理から除外できます。詳細は、9章 p.227 ローカル ライブラリを使用不可に を参照してください。プロジェクト内に使用不可の要素がある場合、ステータスバーに数

が2つ表示されます。1つめは存在する要素の個数で、2つめは使用可能な要素の個数です。たとえば、ステータスバーに5(2) ライブラリとある場合、これは、このプロジェクトにはライブラリが5つあるけれども、そのうちの2つだけが使用可能だということです。

図 4-23  
辞書エディタ ウィンドウ内のステータスバー



ステータスバーの各要素は以下の表のようになります。

テーブル 4-3  
リソース エディタ ウィンドウ:ステータスバーの説明

要素	説明
ライブラリ	プロジェクト内のライブラリの個数。
タイプ	プロジェクト全体に含まれているタイプの個数。
キーワード	すべてのライブラリに含まれているキーワードの個数。不要語辞書に含まれている場合でも、キーワードはカウントされます。あるタイプが使用不可の場合、そのタイプに含まれるキーワードはすべて使用不可となるので注意してください。
不要語	プロジェクト内のすべてのライブラリにおいて除外されたキーワードの個数。
類義語	プロジェクト内のすべてのライブラリに含まれている類義語の代表語の個数。
オプション	プロジェクト内のすべてのライブラリにおいて除外されたキーワードの個数。カウント数にはセルごとに区切られた各ワードが含まれていることに注意してください。
強制キーワード	プロジェクトのライブラリ内に強制キーワードがある場合に使用可能になるボタン。このボタンをクリックすると、[強制キーワードの編集]ダイアログボックスが表示されます。詳細は、10 章 p.244 キーワードの強制 を参照してください。



# データの抽出

新しいプロジェクト ウィザードでプロジェクトを作成した場合、デフォルトの選択では最初の質問で抽出を行います。抽出の更新や新しい質問での抽出を行いたい場合は、手動で（ツール>抽出）またはカテゴリ作成時に抽出を選択して行うことができます。この抽出の最終的な結果として、一連のキーワード、タイプ、パターンが得られます。[抽出結果] パネル内で、これらの結果を表示して作業することができます。

プロジェクトの作成時に抽出を行わなかった、または抽出結果を保存しない選択をした場合は、使用したい質問に進み（ビュー>質問>“質問”）抽出を行うことができます。

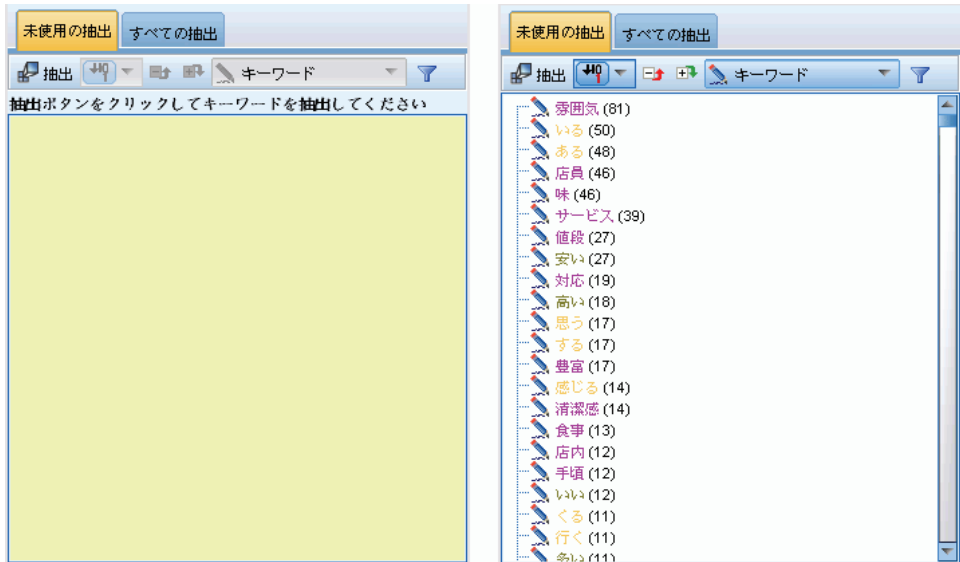
抽出が終わったら、結果を検討し、必要な変更を加えます。詳細は、p. 91 [抽出結果の調整](#) を参照してください。それから再度抽出を行い、新しい結果を確認します。回答を手作業でコード化する場合、人によって回答のまとめかたが若干違ってきます。しかしながら、アンケートをカテゴリ化するには、精度と一貫性が極めて重要です。

IBM® SPSS® Text Analytics for Surveysの強みはカテゴリ定義を矛盾なく再適用できる機能にあります。最初から抽出結果を微調整することで、再抽出をした場合にも、データの文脈に対応して、カテゴリの定義に関していつでも全く同じ結果が得られるようになります。その結果、より正確で反復可能な方法で、回答をカテゴリの定義に割り当てることができます。

## 抽出結果:コンセプト、タイプ、パターン

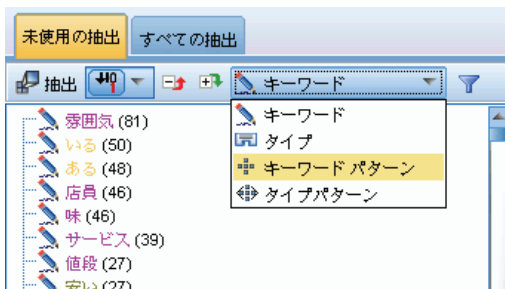
プロジェクトを作成した後、インポートした最初の自由記述式の質問が自動的にウィンドウに表示されます。[抽出結果] パネルは、質問ビューの左下隅にあります。このビューは、[表示] メニュー（[表示] → [質問] → [質問]）から開きます。抽出結果がない場合は、抽出して作業を開始する必要があります。詳細は、p. 88 [データの抽出](#) を参照してください。

図 5-1  
抽出前後の [抽出結果] パネル



[抽出結果] パネルが空白または古い場合、黄色で表示されます。[抽出] をクリックして、抽出プロセスを開始します。抽出後、ドロップダウン リストから表示したい内容を選択して、結果を表示できます。

図 5-2  
[抽出結果] パネルのドロップダウン リスト



抽出されたコンセプト、タイプ、および TLA パターンは、まとめて**抽出結果**と呼ばれ、カテゴリの記述子、または構築ブロックとして機能します。また、カテゴリ規則でコンセプト、タイプ、およびパターンを使用することもできます。さらに、自動的手法では、コンセプトおよびタイプを使用してカテゴリを作成します。

テキスト分析は、抽出結果をテキスト データのコンテキストに従ってレビューし、新しい結果を作成するよう調整、そして再評価するインタラクティブ プロセスです。抽出後、結果を表示し、必要に応じて言語リソースを修正することによって、結果を変更する必要があります。[抽出結果] パネルまたはコンテキスト メニューの [データ] パネルから直接、リソース

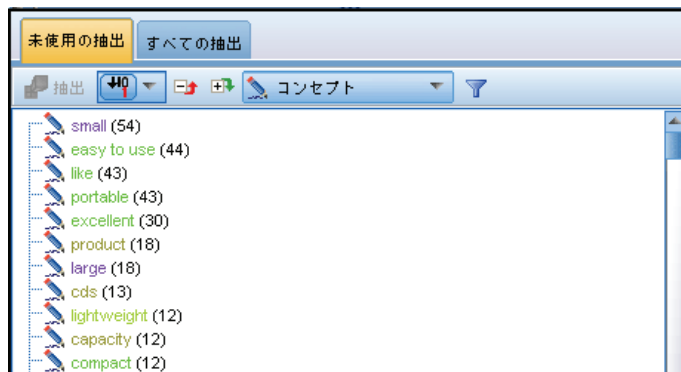
をある程度調整できます。詳細は、 p.91 [抽出結果の調整](#) を参照してください。リソース エディタ ビューで直接調整することもできます。詳細は、 2 章 p.16 [リソース エディタ ウィンドウ](#) を参照してください。

調整した後、再抽出して新しい結果を表示できます。最初から抽出結果を微調整することで、再抽出をした場合にも、データの文脈に完全に対応して、カテゴリの定義に関していつでも全く同じ結果が得られるようになります。このようにして、レコードをより正確で、繰り返し可能な方法で、カテゴリ定義に割り当てます。

## コンセプト

抽出プロセスで、テキスト データをスキャンして分析し、テキスト内の関心のあるまたは関連する 1 つの単語 (election または peace など) や句 (presidential election, election of the president、または peace treaties など) を特定します。これらの単語や句を、まとめて「キーワード」と呼びます。言語リソースを使用して、関連キーワードを抽出し、類似したキーワードを**コンセプト**と呼ばれる代表語でグループ化します。

図 5-3  
[抽出結果] パネル:コンセプト ビュー



デフォルトでは、コンセプトが見つかるレコード数に従って、降順で並べられます。コンセプトが抽出されると、それらをタイプに割り当てて、同様のコンセプトをグループ化します。それらはこのタイプに従ってカラーコード化されます。F, I、リソース エディタ のタイプのプロパティで定義されます。詳細は、 10 章 p.236 [キーワード辞書](#) を参照してください。

カテゴリ定義でコンセプト、タイプ、またはパターンが使用されている場合、テーブル内でイタリックで表示されます。

## タイプ

**タイプ**は、コンセプトの意味上のグループ化です。コンセプトが抽出されると、それらをタイプに割り当てて、同様のコンセプトをグループ化します。<Location>、<Organization>、<Person>、<Positive>、

<Negative>など、いくつかのビルトインのタイプが IBM® SPSS® Text Analytics for Surveys に付属しています。たとえば、<Location> のタイプは、地理的なキーワードや地名をグループ化します。このタイプは、chicago、paris、および tokyo などのコンセプトに割り当てられます。キーワード辞書がなく、テキストから抽出されたコンセプトは、自動的に <Unknown> のタイプとなります。詳細は、10 章 p.237 **ビルトインのタイプ** を参照してください。

タイプ ビューを選択すると、デフォルトでは抽出したタイプが頻度の高い順に表示されます。ツリーが展開されると、そのタイプに抽出されたコンセプトが表示されます。タイプは区別できるよう、カラー コード化されます。色は、タイプのプロパティの一部です。詳細は、10 章 p.238 **タイプの作成** を参照してください。独自のタイプを作成することもできます。

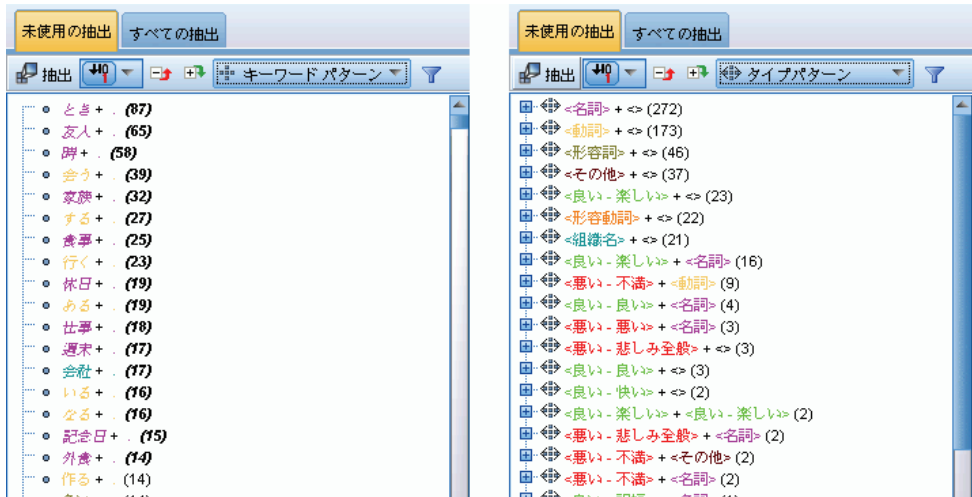
図 5-4  
[抽出結果] パネル:タイプ ビュー



## パターン

パターンは2つの部分からできています。コンセプトと、識別子と形容詞を表すビルトインのタイプが組み合わさっています。パターンは、特定のサブジェクトに関する意見を探索する場合に最も役立ちます。競合他社の製品名を抽出しても、重要でない場合があります。この場合、抽出したパターンを参照し、回答者が製品を良い、悪い、または高いと感じた例があるかどうかを確認することができます。表示には2種類のパターン、コンセプト パターンとタイプ パターンがあります。

図 5-5  
[抽出結果] パネル:コンセプト パターン ビューおよびタイプ パターン ビュー



**コンセプト パターン:** このビューでは、[抽出結果] パネル内のツリーのトップ レベルには、次の構造を持つパターンが表示されます。コンセプト 1 + <タイプ1>、たとえばtext analysis + <Positive> または cost + <Negative> ツリーを展開すると、テキスト分析 + 強力またはコスト + 高いなどの正確なパターンが表示されます。

パターンは2つめの部分がなくても重要な意味があります。たとえば、回答者が特に主題を限定せずに、否定的あるいは肯定的な意見を言った場合が考えられます。この場合、コンセプト 1 + . と表します。. は空識別子を指定します。たとえば、質問「レンタカーを選ぶ際の重要な要素は何ですか」への回答時に回答者が「コスト、場所」と回答した場合、抽出では、空パターンとしてコスト + . および場所 + . が生成されます。パターンが表示される際には、そのタイプに応じてパターンの各要素が色分けされます。

**タイプパターン:** このビューでは、[抽出結果] パネル内のツリーのトップレベルには、次の構造を持つパターンが表示されます。<タイプ> + <タイプ>、たとえば <Budget> + <Positive>。ツリーをさらに展開すると、コンセプト パターン ビューで説明および表示されているように関連性が表示されます。パターンが表示される際には、そのタイプに応じてパターンの各要素が色分けされます。

### 未使用の抽出タブとすべての抽出タブ

抽出結果パネルには、抽出の結果が出力されます。カテゴリを作ると、抽出結果のいくつか（コンセプト、タイプ、パターン）が、カテゴリの定義の一部となります。このため、SPSS Text Analytics for Surveysはタブを利用した2種類の方法でこの情報を表示します。既にカテゴリ定義または抽出コ

ンセプト全体のセットで使用された要素の表示を切り替えることができます。切り替えは [未使用の抽出] タブおよび [すべての抽出] タブをクリックして行います。[未使用の抽出] タブには現在カテゴリの記述子の一部になっていないすべての要素を表示します。[すべての抽出] タブは抽出されたすべての項目を表示し、使用された項目をイタリック体で表示します。

## データの抽出

抽出が必要な場合、[抽出結果] パネルが黄色で表示され、メッセージ[抽出ボタンをクリックしてキーワードを抽出してください] というメッセージが、このウィドゥ枠のツールバーの下に表示されます。

抽出結果がない場合、言語リソースに変更を行い抽出結果を更新する必要がない場合、または抽出結果を保存していないプロジェクトを開く場合は、抽出が必要な場合があります ([ツール] > [オプション])。

注: 抽出結果を [セッション作業を使用] オプションを使用してキャッシュした後、ストリームの入力ノードを変更する際、抽出結果を更新する場合にインタラクティブ ワークベンチ セッションが起動したら新しい抽出を実行する必要があります。

抽出実行中には進行状況が表示されます。抽出している間、抽出エンジンはテキスト データをすべて読み込み、関連キーワードおよびパターンを特定し、それらを抽出して、タイプに割り当てます。そして、エンジンは、1 つの主要なキーワード、コンセプトに類義語のキーワードをグループ化します。プロセスが完了すると、生成されたコンセプト、タイプ、パターンが [抽出結果] パネルに表示されます。結果を処理および確認できるようになります。

注: データセットのサイズと、抽出プロセスを完了するためにかかる時間の間には、関連性があります。パフォーマンスの統計および推奨事項については、インストールの説明書を参照してください。

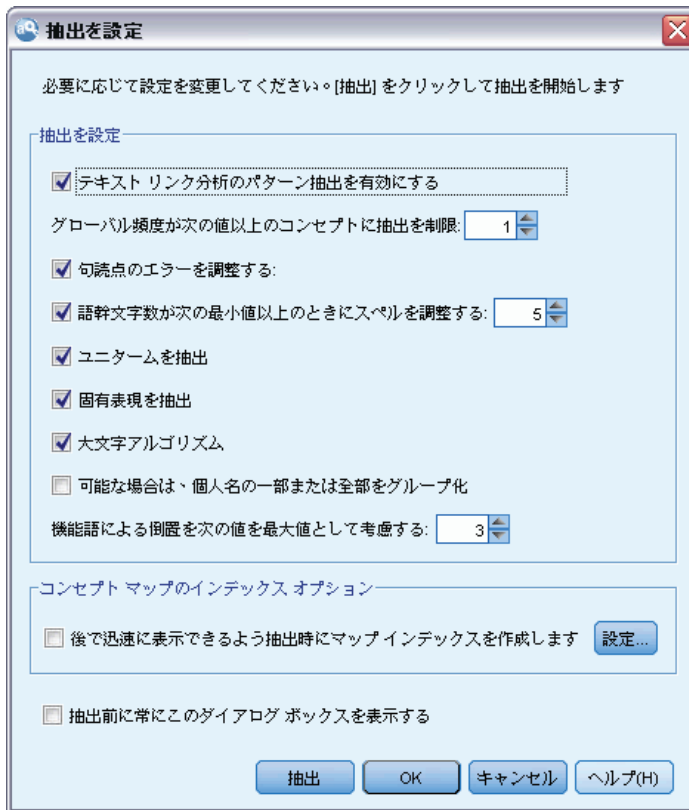
### データを抽出するには

- ▶ メニューの [ツール] > [抽出] を選択します。または、[抽出] ツールバーボタンをクリックします。
- ▶ [抽出設定] ダイアログの表示を選択すると必ず、ダイアログが表示され、変更を行うことができます。各設定の記述子については、このトピックの後半を参照してください。
- ▶ [抽出] をクリックして、抽出プロセスを開始します。抽出が始まると、進捗状況のダイアログ ボックスが表示されます。抽出後、結果が [抽出結果] ウィンドウに表示されます。デフォルトでは、コンセプトが見つかるレコード数に従って、降順で並べられます。

ツールバー オプションを使用して結果を確認し、結果を並べ替える、結果を絞り込む、または異なるビュー（コンセプト、パターン、またはタイプ）に切り替えることができます。言語リソースを処理して、抽出結果を調整することもできます。詳細は、p. 91 [抽出結果の調整](#) を参照してください。

[抽出設定] ダイアログ ボックスには、基本的な抽出オプションがいくつか表示されます。

図 5-6  
[抽出設定] ダイアログ ボックス



**句読点のエラーを調整する:** 抽出時に句読点エラー（不適切な使用方法など）を含むテキストを一時的に正規化し、コンセプトの抽出可能性を向上させます。自由記述式アンケートの回答、電子メール、CRM データなど、テキストが短く品質が悪い場合、またはテキストに略語が多く含まれている場合に特に役立ちます。

**語幹文字数が次の最小値以上のときにスペルを調整する:** Fuzzy Grouping の手法を適用し、共通してスペルミスのある単語またはスペルの近い単語を 1 つのコンセプトにグループ化できるようにします。Fuzzy Grouping アルゴリズムでは、最初の母音を除くすべての母音を一時的に抜き取った後抽出した単語から 2 つ/3 つの子音を抜き取り、それらを比較して、それらが

同じで modeling と modelling が同じグループに分けられるかどうかを確認します。ただし、各キーワードが <Unknown> タイプを除いて、別のタイプに割り当てられた場合、Fuzzy Grouping 手法は適用されません。

Fuzzy Grouping を使用する前に必要な、語幹文字数の制限を定義することもできます。キーワード内の語幹文字数は、すべての文字を合計し、活用語尾、複合語キーワードの場合は区切り文字および前置詞を形成する文字を差し引いて計算します。たとえば、キーワード exercises の語幹文字数は「exercise」という形式で 8 文字と数えられます。語末の s は活用語尾（複数形）であるためです。同様に、apple sauce の語幹文字は 10 文字（「apple sauce」）、そして manufacturing of cars の語幹文字は 16 文字（「manufacturing car」）となります。この算出方法は、Fuzzy Grouping を適用するべきかどうかを確認するためにのみ使用されますが、単語がどのように一致するかについては影響を与えません。

注：特定の単語が不正に後ろのグループに振り分けられている場合は、[アドバンス リソース] タブの Fuzzy Grouping: 例外セクションで明示的に宣言し、単語のペアを除外できます。詳細は、11 章 p.258 [Fuzzy Grouping](#) を参照してください。

**ユニタームを抽出：** 単語が複合語の一部でない限り、または名詞、またはスピーチ内の認識できない部分である場合、このオプションは単一の単語（ユニターム）を抽出します。

**固有表現を抽出：** 電話番号、セキュリティ番号、時間、日付、通貨、数字、パーセント、電子メールアドレス、HTTP アドレスなどの固有表現を抽出します。[アドバンス リソース] タブの **固有表現：設定** で表現を含めたり除外したりできます。不要なエンティティを無効にすることにより、抽出エンジンは処理時間を節約できます。詳細は、11 章 p.264 [構成](#) を参照してください。

**大文字アルゴリズム：** キーワードの最初の文字が大文字である場合、組み込み辞書にない単純キーワードおよび複合キーワードを抽出します。このオプションには、最も適切な名詞を抽出するのに優れた方法があります。

**可能な場合は、個人名の一部または全部をグループ化：** テキスト内で別々の形式で同時に出現する名前をグループ化します。名前はテキストの始めでは完全な形式で、後は短い形式でのみ参照されるため、この機能が役立ちます。このオプションでは、タイプが <Unknown> のユニタームが、タイプ <Person> の複合キーワードの最後の単語に一致するようにします。たとえば、doe があり、最初タイプが <Unknown> である場合、抽出エンジンは、<Person> タイプの複合キーワードに最後の単語として doe が含まれているかどうか（例：john doe）を確認します。ほとんどがユニタームとして抽出されることがないため、人の名前に適用されることはありません。

**機能語による倒置を次の値を最大値として考慮する：** 倒置手法を適用する場合に指定されている場合がある非機能的単語の最大数を指定します。この倒置手法では、活用語尾に関係なく、含まれる非機能的単語（of や the など）



によってお互いに異なる類似した句をグループ化します。たとえば、この値を最大 2 単語に設定し、company officials および officials of the company が抽出されたとします。この場合、両方の抽出キーワードは、of the が無視されると同じであるとみなされるため、最終コンセプトリストに共にグループ化されます。

**抽出前に常にこのダイアログ ボックスを表示する:** [ツール] メニューを選択しない限り表示したくない場合、抽出ごとに [抽出設定] ダイアログを表示するかどうか、または抽出設定を編集する場合、抽出ごとに表示するかどうかを尋ねるかどうかを指定します。

## 抽出結果の保存

抽出を行うと、[抽出結果] パネルに結果が表示されます。これを使って、回答をカテゴリ化できます。IBM® SPSS® Text Analytics for Surveys のセッションの間は、これらの抽出結果はメモリ内に保持されており、これを使った作業ができます。デフォルトでは、抽出結果はプロジェクト内に保存されます。プロジェクトを閉じる際に抽出結果を保存するかどうかは、グローバルに設定されており、オプションダイアログボックスでいつでも変更できます(ツール>オプション)。詳細は、2 章 p.19 [オプション: システム タブ](#) を参照してください。

セキュリティのために、これらの抽出結果は保存の過程で暗号化されたからデータベースに置かれます。この手続きによって、熟練したユーザーにとってもデータベース内でデータを見つけることは困難になっています。それだけでなく、抽出結果は、データソースが特定されない限り、SPSS Text Analytics for Surveysには絶対に表示されません。このため、データがパスワードで保護されている場合、抽出結果を画面に表示するには、このデータソースに関してユーザーがユーザー名とパスワードを入力する必要があります。

## 抽出結果の調整

抽出とは反復可能なプロセスで、結果を抽出、確認、結果を変更、および再抽出して結果を更新できます。テキスト マイニングおよびカテゴリ化には精度および継続性が不可欠であるため、最初から抽出結果を調整することによって、再抽出ごとに、カテゴリ定義でまったく同じ結果を取得します。このようにして、レコードをより正確で、繰り返し可能な方法で、カテゴリに割り当てます。

抽出結果は、カテゴリを作成するための要素となります。これらの抽出結果を使用してカテゴリを作成すると、1 つまたは複数のカテゴリ記述子に一致するテキストが含まれる場合、レコードが自動的にカテゴリに割り当てられます。言語リソースを調整する前にカテゴリ化を開始できますが、開始前に少なくとも 1 回、抽出結果を確認しておく役立ちます。

結果を確認すると、抽出エンジンが異なる方法で処理する必要のある要素が見つかる場合があります。以下のような例があります。

- **認識されない類義語:** 賢い、知的、頭脳明晰、博識など、類義語と考えられるいくつかのコンセプトが見つかり、抽出結果に個別のコンセプトとしてすべて表示されたとします。知的、頭脳明晰、博識がすべて代表コンセプト 賢い の名ですべてグループ化されるよう、類義語定義を作成できます。こうすることにより、これらのコンセプトをすべて賢いとグループ化し、グローバル出現頻度も高くなります。詳細は、 p.93 [類義語の追加](#) を参照してください。
- **ミスタイプ コンセプト:** 抽出結果のコンセプトがあるタイプに出現し、別のタイプに割り当てたい場合があります。また、抽出結果に 15 種類の野菜のコンセプトがあり、それらすべてを <Vegetable> という新しいタイプに追加したい場合もあります。キーワード辞書になく、テキストから抽出されたコンセプトは、自動的に <Unknown> のタイプとなります。コンセプトをタイプに追加できます。詳細は、 p.95 [コンセプトのタイプへの追加](#) を参照してください。
- **重要でないコンセプト:** 抽出されたコンセプトで非常に頻度の高い、つまり多くのレコードで見つかる場合があります。ただし、このコンセプトは分析には重要でないと見なされます。このコンセプトを抽出から除外できます。詳細は、 p.97 [コンセプトの抽出からの除外](#) を参照してください。
- **不正な合致:** 特定のコンセプトを含むレコードを確認する場合、faculty (能力) と facility (施設) のように 2 つの単語が誤ってグループ化されているのを発見する場合があります。この合致は Fuzzy Grouping という、内部アルゴリズムによるものであり、2 つまたは 3 つの子音および母音を一時的に無視して、一般的なスペルミスをグループ化します。これらの単語を無視する必要のある単語のペアのリストに追加できます。詳細は、11 章 p.258 [Fuzzy Grouping](#) を参照してください。
- **未抽出コンセプト:** 特定のコンセプトが抽出されるのを期待しているにもかかわらず、レコード テキストを確認しているときに一部の単語または句が抽出されていないことに気づく場合があります。これらの単語は重要でない動詞または形容詞でない場合が多くあります。ただし、抽出されなかった単語または句をカテゴリ定義として使用したい場合があります。コンセプトを抽出するために、キーワードをキーワード辞書に強制投入できます。詳細は、 p.98 [単語を抽出に強制投入](#) を参照してください。

こうした変更の多くは、1 つまたは複数の要素を選択して右クリックし、コンテキスト メニューを使用することによって、[抽出結果] パネルまたは [データ] パネルから直接実行できます。

変更を行った後、パネルの背景色が変わり、変更を表示するには再抽出が必要であることを示します。詳細は、 p.88 [データの抽出](#) を参照してください。大きいデータセットを扱う場合、変更ごとではなく、複数の変更を行った後に再抽出を行うとより効率的です。

注:リソース エディタ ビューで抽出結果を作成するために使用する、編集可能な言語リソースのセット全体を表示できます ([表示]→[リソース エディタ])。これらのリソースは、このビューにライブラリおよび辞書の形式で表示されます。ライブラリおよび辞書内で直接コンセプトおよびタイプをカスタマイズできます。詳細は、9 章 p.221 [ライブラリの使用](#) を参照してください。

## 類義語の追加

**類義語**は、同じ意味を持つ 2 つ以上の単語と関連があります。類義語はキーワードとその短縮形をまとめるのにもよく使用されます。またよくつづり間違いが起こる語を正しい書き方のもので置き換えるのにも使用されます。類義語を使用すると、代表コンセプトの頻度が高くなり、テキスト データ内のさまざまな方法で表示される類似した情報を見つけやすくなります。

製品に付属する言語リソース テンプレートおよびライブラリには、事前定義された多くの類義語が含まれています。ただし、認識されていない類義語が見つかった場合、次回抽出するときに認識されるよう、類義語を定義することができます。

まず、代表コンセプトまたは主要キーワードを決定します。**代表コンセプト**は、最終的な結果ですべての類義語のキーワードをグループ化したい単語または句です。抽出時、類義語は、この代表コンセプトの下でグループ化されます。次に、このコンセプトのすべての類義語を特定します。代表コンセプトは、最終的な抽出で、すべての類義語と置き換えられます。類義語となるためにはキーワードは抽出されなければなりません。ただし、代表コンセプトを、置換えを行うために抽出する必要はありません。たとえば、**知的**という単語を**賢い**という単語に置き換えたい場合、**知的**が類義語となり、**賢い**が代表コンセプトとなります。

新しい類義語定義を作成する場合、新しい代表コンセプトが辞書に追加されます。その後、類義語をその代表コンセプトに追加する必要があります。類義語を作成または編集する場合、これらの変更が **リソース エディタ** の類義語辞書に記録されます。これらの類義語辞書の内容全体を表示したい場合、またはかなりの数の変更を行いたい場合、**リソース エディタ** で直接作業することが必要な場合があります。詳細は、10 章 p.247 [類義語辞書](#) を参照してください。

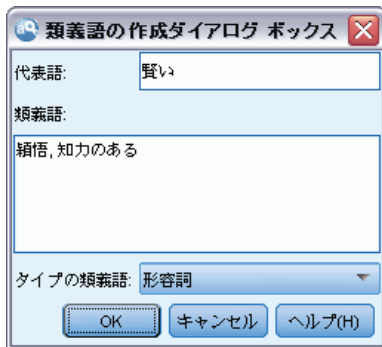
新しい類義語は、**リソース エディタ** ビューのライブラリ ツリーの表示された最初のライブラリに自動的に保存されます。デフォルトでは、これがローカル ライブラリとなります。

注:類義語定義を探していても、コンテキスト メニューまたは **リソース エディタ** で直接見つからない場合、内部の Fuzzy Grouping 手法によって、合致が発生している場合があります。詳細は、11 章 p.258 [Fuzzy Grouping](#) を参照してください。

### 新しい類義語を作成するには

- ▶ [抽出結果] パネルまたは [データ] パネルで、新しい類義語を作成したいコンセプトを選択します。
- ▶ メニューから [編集] → [類義語への追加] → [新規] を選択します。[類義語の作成] ダイアログ ボックスが開きます。

図 5-7  
[類義語の作成] ダイアログボックス

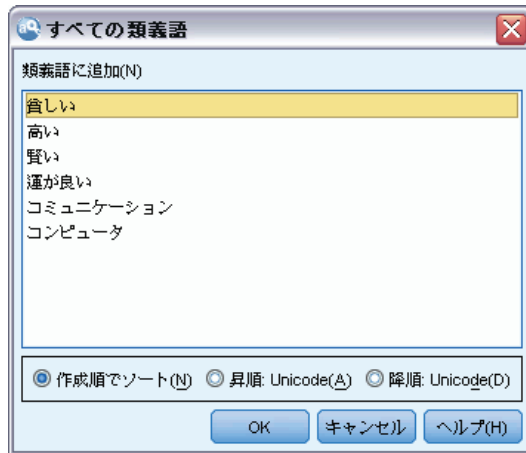


- ▶ [代表語] テキスト ボックスに、代表語を入力します。これが、すべての類義語がグループ化されるコンセプトとなります。
- ▶ さらに類義語を追加したい場合、[類義語] リスト ボックスにそれらを入力します。辞書エディタの区切り文字（デフォルトでは「, 」(カンマ)）を使って、各類義語を分けます。詳細は、2 章 p.19 [オプション：システム タブ](#) を参照してください。
- ▶ [OK] をクリックし、変更を適用します。ダイアログ ボックスが閉じて、[抽出結果] パネルの背景色が変わり、変更を表示するには再抽出が必要であることを示します。変更が複数ある場合には、再抽出する前に変更を終わらせてください。

### 類義語に追加するには

- ▶ [抽出結果] パネルまたは [データ] パネルで、既存の類義語定義に追加したいコンセプトを選択します。
- ▶ メニューから [編集] → [類義語への追加] を選択します。メニューには、一番最後に作成された代表語がリストの一番上に表示されます。選択したコンセプトを追加したい類義語の名前を選択します。探している類義語が表示されたら、その名前を選択します。選択したコンセプトがその類義語の定義に追加されます。メニューに類義語が表示されない場合、[もっと表示] を選択すると [すべての類義語] ダイアログ ボックスが表示されます。

図 5-8  
[すべての類義語] ダイアログ ボックス



- ▶ [すべての類義語] ダイアログ ボックスで、リストを自然な並び順（作成順）の昇順または降順で並べ替えることができます。選択したコンセプトを追加したい類義語の名前を選択し、[OK] をクリックします。ダイアログ ボックスが閉じ、コンセプトが類義語の定義に追加されます。

## コンセプトのタイプへの追加

抽出を実行している場合、共通点を持つキーワードをグループ化するために、抽出されたコンセプトがタイプに割り当てられます。IBM® SPSS® Text Analytics for Surveys は、多くのビルトインのタイプに付属しています。詳細は、10 章 p.237 [ビルトインのタイプ](#) を参照してください。キーワード辞書がなく、テキストから抽出されたコンセプトは、自動的に <Unknown> のタイプとなります。

結果を確認しているとき、あるタイプに割り当てたいコンセプトがいくつか異なるタイプに割り当てられていたり、単語のグループが実際は新しいタイプに割り当てられていたりすることが分かる場合があります。こうした場合、コンセプトを別のタイプに割り当てなおすか、まとめて新しいタイプを作成する場合があります。

たとえば、自動車に関連する調査データを扱っており、車のさまざまな領域に焦点を当ててのカテゴリ化に関心があるとします。<Dashboard> というタイプを作成し、車のダッシュボードにある計測器およびノブに関連するすべてのコンセプトをグループ化することができます。そして、gas gauge、gas gauge、radio、および odometer などのコンセプトを、この新しいタイプに割り当てることができます。

またたとえば、大学に関連する調査データ、そして <Organization> ではなく <Person> というタイプとしての抽出 Johns Hopkins (大学) を扱っているとします。この場合、このコンセプトを <Organization> タイプに追加することができます。

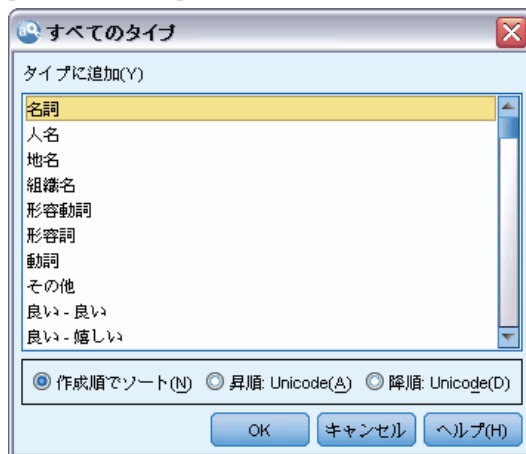
タイプを作成、またはコンセプトをタイプのキーワード リストに追加する場合、これらの変更は リソース エディタ の言語リソース ライブラリのキーワード辞書に記録されます。これらのライブラリの内容を表示したい場合、またはかなりの数の変更を行いたい場合、リソース エディタ で直接作業することが必要な場合があります。詳細は、10 章 p. 240 [キーワードを追加](#) を参照してください。

### コンセプトをタイプに追加するには

- ▶ [抽出結果] パネルまたは [データ] パネルで、既存のタイプに追加したいコンセプトを選択します。
- ▶ 右クリックしてコンテキスト メニューを開きます。
- ▶ メニューから [編集] → [タイプへの追加] を選択します。タイプ名が見つかったなら、これを選択します。選択したコンセプトを追加したいタイプの名前を選択します。探しているタイプの名前が表示されたら、その名前を選択します。選択したコンセプトがそのタイプに追加されます。メニューに類義語が表示されない場合、[もっと表示] を選択すると [すべてのタイプ] ダイアログ ボックスが表示されます。

図 5-9

#### [すべてのタイプ] ダイアログ ボックス

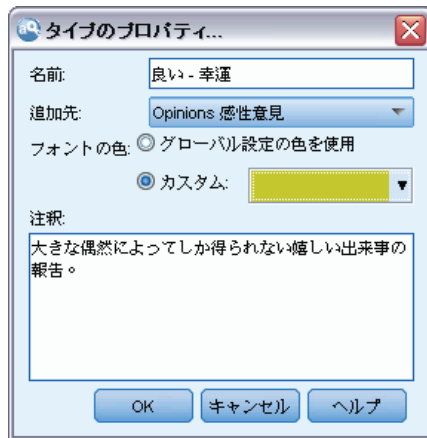


- ▶ [すべてのタイプ] ダイアログ ボックスで、リストを自然な並び順（作成順）の昇順または降順で並べ替えることができます。選択したコンセプトを追加したいタイプの名前を選択し、[OK] をクリックします。ダイアログ ボックスが閉じ、コンセプトがタイプに追加されます。

### 新しいタイプを作成するには

- ▶ [抽出結果] パネルまたは [データ] パネルで、新しいタイプを作成したいコンセプトを選択します。
- ▶ メニューから [編集] → [タイプへの追加] → [新規] を選択します。[タイプのプロパティ] ダイアログ ボックスが開きます。

図 5-10  
[タイプのプロパティ] ダイアログ ボックス



- ▶ [名前] テキスト ボックスにこのタイプの新しい名前を入力し、他のフィールドを変更します。詳細は、10 章 p.238 [タイプの作成](#) を参照してください。
- ▶ [OK] をクリックし、変更を適用します。ダイアログ ボックスが閉じて、[抽出結果] パネルの背景色が変わり、変更を表示するには再抽出が必要であることを示します。変更が複数ある場合には、再抽出する前に変更を終わらせてください。

## コンセプトの抽出からの除外

結果を確認しているときに、自動カテゴリ作成手法によって必要のないコンセプトが抽出または使用されていることが分かることがあります。これらのコンセプトの頻度が非常に高く、分析には全く重要でない場合もあります。この場合、コンセプトを不要なものとしてマークすることができます。通常、このリストに追加するコンセプトは、穴埋めのための単語または句で、一貫性を維持するためにテキストで使用され重要でなく、抽出結果を混乱させる場合があります、コンセプトを不要語辞書に追加することによって、それらが確実に抽出されないようにします。

不要語に追加することにより、次回抽出するときには除外コンセプトのすべての変化が抽出結果から除外されています。このコンセプトがカテゴリの記述子として出現している場合、再抽出後は 0 度数でカテゴリ内に残ります。

不要語に追加すると、これらの変更が リソース エディタ の不要語辞書に記録されます。不要語辞書のすべてを表示し、それらを編集したい場合、リソース エディタ で直接作業することが必要な場合があります。詳細は、10 章 p.253 [不要語辞書](#) を参照してください。

### コンセプトを不要語に追加するには

- ▶ [抽出結果] パネルまたは [データ] パネルで、不要語に追加したいコンセプトを選択します。
- ▶ 右クリックしてコンテキスト メニューを開きます。
- ▶ [不要語に追加] を選択します。コンセプトが リソース エディタ の不要語辞書に追加され、[抽出結果] パネルの背景色が変わり、変更を表示するには再抽出が必要であることを示します。変更が複数ある場合には、再抽出する前に変更を終わらせてください。

注：不要語に追加した単語は、リソース エディタ のライブラリ ツリーに表示された最初のライブラリに自動的に保存されます。デフォルトでは、これがローカル ライブラリとなります。

## 単語を抽出に強制投入

抽出後、[データ] パネルでテキストデータを確認するとき、一部の単語または句が抽出されていないことが分かる場合があります。これらの単語は重要でない動詞または形容詞でない場合が多くあります。ただし、抽出されなかった単語または句をカテゴリ定義として使用したい場合があります。

これらの語や語句を抽出するには2つのオプションがあります。

- キーワードをタイプ ライブラリに強制投入する：詳細は、10 章 p.244 [キーワードの強制](#) を参照してください。
- 単語を既存のカテゴリ定義に直接追加します。この方法は通常、上記の方法で、望ましい結果が得られなかった場合に使用されます。詳細は、6 章 p.176 [カテゴリ内のテキストマッチ](#) を参照してください。

**重要!** 辞書のキーワードを強制することが、絶対に確実というわけではありません。強制することによって、キーワードを辞書に明示的に追加している場合でも、再抽出後に [抽出結果] パネルに出現しない場合、あるいは出現しても宣言したとおりのものではない場合があります。抽出のプロセスにおいてある語や語句がより長い語句の一部として既に抽出されてしまっている場合や、語が品詞（名詞、動詞、形容詞、前置詞など）に分割されている場合が考えられます。これを回避するために、**全体**（複合語なし）



---

マッチ オプションをキーワード辞書のこのキーワードに適用します。詳細は、10 章 p.240 [キーワードを追加](#) を参照してください。

# テキスト データの分類

IBM® SPSS® Text Analytics for Surveys で、テキスト内の主要なキーワード、情報、属性をキャプチャする高いレベルのコンセプトまたはトピックを示す**カテゴリ**を作成できます。

カテゴリが階層的な構造を持つことができ、すなわち、サブカテゴリを含むことができ、また、そのサブカテゴリにもそれ自身のサブカテゴリを更に下の階層に向かって持たせることができます。製品内にこのような階層的な分類を構築することが可能であるだけでなく、階層的な分類を持ち、以前はコード フレームと呼ばれていた、定義済みの分類構造をインポートすることも可能です。

実際に、階層的カテゴリにより、1個または複数のサブカテゴリを持つツリー構造を構築して、たとえば異なるコンセプトやトピックの分野の項目をより正確にグループ化することができます。レジャー活動に関して簡単な例を挙げることができます。「時間があればどんな活動がしたいですか？」という質問に対する答えとして、「スポーツ」、「日曜大工」、「釣り」などをトップ カテゴリに設定し、「スポーツ」の下の階層に「球技」、「水泳」などを設定できます。

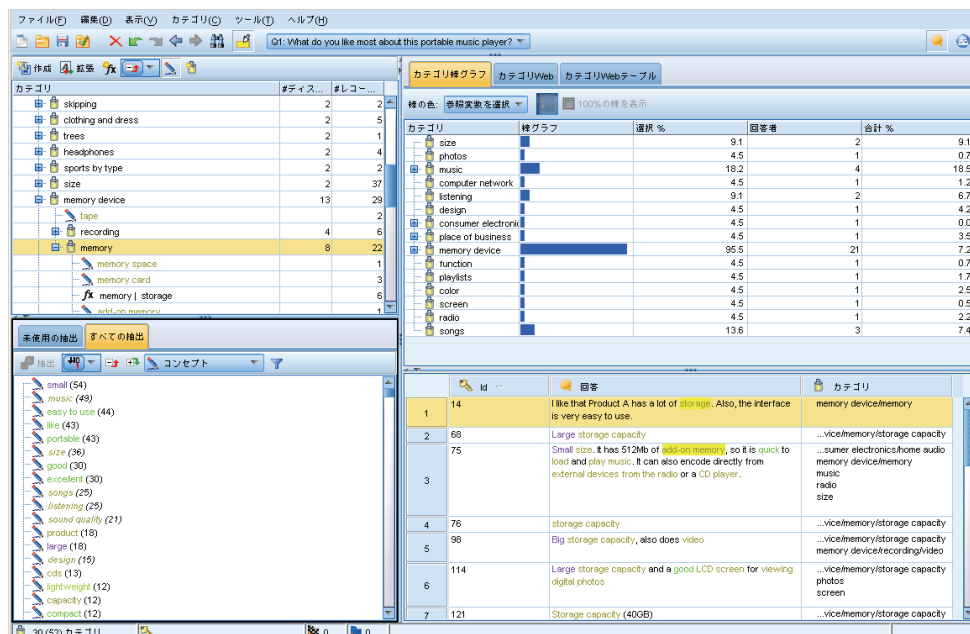
カテゴリは、コンセプト、タイプ、パターンおよびカテゴリ規則などの一連の記述子で構成されています。また、これらの記述子を共に使用して、レコードが指定されたカテゴリに属するかどうかを特定します。レコード内のテキストをスキャンして、テキストが記述子に一致するかどうかを確認することができます。一致が見つかった場合は、レコードはそのカテゴリに割り当てられます。このプロセスを、**カテゴリ化**といいます。

4 つのパネルに表示されたデータを使用して、カテゴリを処理、作成および視覚的に検証することができます。また、[表示] メニューからその名前を選択して隠したり表示したりできます。

- **[カテゴリ] パネル:** このパネルでカテゴリを作成し、管理します。詳細は、p. 102 [\[カテゴリ\] パネル](#) を参照してください。
- **[抽出結果] パネル:** このパネルで抽出したコンセプトおよびタイプを検証および処理します。詳細は、5 章 p. 83 [抽出結果: コンセプト、タイプ、パターン](#) を参照してください。

- **[視覚化] パネル:** このパネルでカテゴリについて、またカテゴリがどのように相互作用するかを視覚的に検証します。詳細は、7章 p. 181 [グラフの視覚化](#) を参照してください。
- **[データ] パネル:** このパネルでの選択に対応するレコード内に含まれるテキストを検証および確認できます。詳細は、p. 104 [\[データ\] パネル](#) を参照してください。

図 6-1  
質問ビュー



テキスト分析パッケージ (TAP) の一連のカテゴリからはじめることができますが、独自のカテゴリを作成する、または定義済みカテゴリ ファイルからインポートする必要もあります。コンセプト、タイプ、パターンの抽出結果を使用する製品の自動化手法の頑健なセットを使用して、カテゴリおよびそれらの記述子を自動的に作成することができます。データに関する追加の洞察を使用して、カテゴリを手動で作成することもできます。抽出結果をカテゴリに手動でドラッグ アンド ドロップして、カテゴリ定義を作成できます。カテゴリ規則をカテゴリに追加し、独自の事前定義済みカテゴリを使用、抽出されなかった単語または句 (テキスト マッチ) を追加、回答をカテゴリ、または組み合わせに直接強制投入して、これらのカテゴリまたは空のカテゴリの品質を向上させることができます。

それぞれの手法は、特定の種類のデータおよび状況に適していますが、同じ分析で手法を組み合わせると、全範囲のレコードをキャプチャすると役に立つ場合があります。またカテゴリ化を行う際には、言語リソースに対して変更を加えたほうがよい場合もあります。

## [カテゴリ] パネル

[カテゴリ] パネルでは、カテゴリを作成および管理できます。このパネルは、[表示] メニューから使用できる質問ビュー（[表示]>[質問]>「Your\_Question」）の左上に表示されます。テキスト データからコンセプトとタイプを抽出した後、内包関係のコンセプト、共起など、自動的な方法を使用して、または手動で作成してカテゴリの作成を開始できます。詳細は、p. 118 [カテゴリの作成](#) を参照してください。

図 6-2

カテゴリがない場合とある場合の [カテゴリ] パネル

カテゴリ	#ディス...	#レコー...
すべてのレコード	-	259
未カテゴリ化	-	259
コンセプトは展開されていません	-	259

カテゴリ	#ディス...	#レコー...
すべてのレコード	-	405
未カテゴリ化	-	155
コンセプトは展開されていません	-	4
music	25	75
child's music		1
music catalogue		1
play music		2
music choice		1
amounts of music		1
library of music		1
collection of music		1
bank of music		1
digital music hands		1
share music		1
music quality		1
music		49
collection of music		4

カテゴリを作成または更新するごとに、レコードが自動的にスキャンされ、fJfefSfŠ?è< ‘, Ì<LqŽq, ÉfefLfXfg, a?î?ž, ·, é, ©, Ç, □, ○, ðŠm” F, μ, Û, ·。一致が見つかった場合は、レコードはそのカテゴリに割り当てられます。この最終結果は、レコードのすべてではなくともその多くが、カテゴリの記述子に基づいて、カテゴリに割り当てられます、

### カテゴリ ツリー一覧

このパネルのツリー一覧には、一連のカテゴリ、サブカテゴリ、記述子が表示されます。ツリーには、各ツリー項目の情報を示す列があります。表示できるのは次の列です。

- **コード:** 各カテゴリのコード値を表示します。この列は、デフォルトでは非表示になっています。この列はツリーテーブル内をクリックして表示>カテゴリコードを選択すると表示されます。
- **カテゴリ:** カテゴリ ツリーには、カテゴリおよびサブカテゴリの名前が表示されます。また、記述子のツールバー アイコンをクリックすると、一連の記述子も表示されます。

- **記述子:** その定義を構成する記述子の数が表示されます。この度数には、サブカテゴリの記述子数は含まれません。[カテゴリ] 列に記述子名が表示されている場合、度数は表示されません。この列はツリーテーブル内をクリックして**表示 > すべての記述子**を選択すると表示されます。
- **ドキュメント:** スコアリング後、この列には、該当するカテゴリとすべてのサブカテゴリにカテゴリ化されているレコードの数が表示されます。つまり、5 つのレコードが記述子に基づいて上位カテゴリに合致し、7 つの異なるレコードがその記述子に基づいてサブカテゴリに合致する場合、上位カテゴリのドキュメント数の合計は、この 2 つの数値の合計、この場合は 12 となります。ただし、同じレコードが上位カテゴリとそのサブカテゴリに合致する場合、度数は 11 となります。

カテゴリがない場合でも、テーブルには 2 つの行が表示されます。上の行の**すべてのレコード**は、レコード数の合計です。2 番目の行**[未カテゴリ化]**には、カテゴリがされていないドキュメント/レコード数が表示されます。

パネルの各カテゴリについて、小さい黄色のバケツのアイコンがカテゴリ名の前に表示されます。カテゴリをクリック、ツリーテーブル内をクリックして**カテゴリ定義**、**カテゴリ定義**ダイアログボックスが開き、**記述子**と呼ばれるすべての要素が表示されます。記述子はコンセプト、タイプ、パターン、**カテゴリルール**などの定義を決定します。詳細は、p. 115 **カテゴリとは**を参照してください。デフォルトでは、**カテゴリ ツリー**一覧には、カテゴリの記述子は表示されません。[**カテゴリ定義**] ダイアログボックスではなくツリーで直接記述子を表示する場合、ツールバーの鉛筆のアイコンが表示された切り替えボタンをクリックします。この切り替えボタンを選択すると、ツリーが展開され、記述子が表示されます。

### カテゴリのスコアリング

カテゴリ ツリー一覧の**[ドキュメント]**列には、特定のカテゴリにカテゴリ化されているレコード数が表示されます。数値が過去のものまたは計算されていない場合、アイコンがその列に表示されます。大きいデータセットを使用する場合、スコアリング プロセスには時間がかかる場合があります。

### ツリー内のカテゴリの選択

ツリー内で選択すると、横グループのカテゴリのみ選択できます。つまり、上位レベルのカテゴリを選択すると、サブカテゴリは選択できません。または指定されたカテゴリの 2 つのサブカテゴリを選択すると、同時に別のカテゴリのサブカテゴリを選択できません。不連続なカテゴリを選択すると、以前の選択内容が失われます。

### [データ] パネルおよび [視覚化] パネルの表示

テーブル内の行を選択すると、[視覚化] および [データ] パネルが、選択に応じた情報に自動的に更新されます。

## カテゴリの調整

カテゴリ化を行っても、最初から完全な結果が得られるとは限りません。削除したいカテゴリや、他のカテゴリとまとめたいカテゴリもあるでしょう。また、抽出結果を確認して、役立つと思われるいくつかのカテゴリが作成されていないことが分かる場合があります。その場合、結果を手動で変更し、特定の状況に対して結果を調整することができます。詳細は、p. 169 [カテゴリの編集および調整](#) を参照してください。

- カテゴリを移動、結合、または削除するほか、カテゴリ定義を編集またはカテゴリ定義に追加します。詳細は、下記の「[カテゴリの編集および調整](#)」を参照してください。
- 特定の回答 ID をカテゴリに追加またはカテゴリから削除します。詳細は、p. 175 [回答のカテゴリへの強制投入](#) を参照してください。
- テキスト マッチをカテゴリに追加して、同じテキストを含む回答をカテゴリにキャプチャします。詳細は、p. 176 [カテゴリ内のテキストマッチ](#) を参照してください。
- カテゴリ規則をカテゴリに追加して、論理式に基づき、回答をカテゴリに自動的に分類します。詳細は、p. 156 [カテゴリ規則の使用](#) を参照してください。
- カテゴリ間の関係を視覚化する。詳細は、7 章 p. 181 [グラフの視覚化](#) を参照してください。
- カテゴリ化の結果をエクスポートします。詳細は、4 章 p. 57 [カテゴリ化の結果をエクスポート](#) を参照してください。

## [データ] パネル

カテゴリを作成した場合、作業しているテキスト データを確認したい場合があります。たとえば、640 件のレコードがカテゴリ化されているカテゴリを作成する場合、実際にどのようなテキストが書かれているのかを確認するため、これらのレコードの一部またはすべてに目を通したい場合があります。右下の [データ] パネルでレコードを確認することができます。デフォルトで表示されない場合は、メニューから [表示] → [パネル] → [データ] を選択してください。

このパネルは、自由記述データのレスポンス レコードをテーブルで表示します。ビュー内の他のパネルで選択されているものに応じて、一致するレコードのみがパネルに表示されます。たとえば、[抽出結果] パネルのコンセプトを選択した場合、コンセプト（および関連キーワード）を含むレコードのみが、[データ] パネルに表示されます。

もう一方のパネルでコンセプトまたはカテゴリを選択してデータを表示すると、レコード内に含まれるコンセプト（および関連キーワード）は色つきで強調表示され、テキスト内で簡単に識別できます。カラー コードは、コンセプトが属するタイプに対応します。カラーコード化された項目

上でマウス ポインタを停止させて、項目が抽出されたコンセプトと、項目が割り当てられたタイプを表示することもできます。抽出されていないテキストは、黒で表示されます。通常、こうした抽出されていない単語は接続詞（「および」または「と」）、代名詞（「私」または「彼ら」）および動詞（「いる」、「持つ」、または「取る」）のケースが多くあります。

図 6-3  
[データ] パネル

	Id	回答	カテゴリ
1	14	I like that Product A has a lot of storage. Also, the interface is very easy to use.	memory device/memory
2	45	Everything! Product A rules! I can't wait to get a video one!	memory device/recording/video
3	68	Large storage capacity	...vice/memory/storage capacity
4	75	Small size. It has 512Mb of add-on memory, so it is quick to load and play music. It can also encode directly from external devices from the radio or a CD player.	...sumer electronics/home audio memory device/memory music radio size
5	76	storage capacity	...vice/memory/storage capacity
6	89	Small but lots of space (60 GB). Video is a bit of a toy but cool.	memory device/recording/video
7	98	Big storage capacity, also does video	...vice/memory/storage capacity memory device/recording/video
8	114	Large storage capacity and a good LCD screen for viewing digital photos	...vice/memory/storage capacity photos screen
9	120	The sound quality and the ability to record and mix my own play lists.	...nics/audio/sound/sound quality memory device/recording

注:[データ] パネル内で選択された質問に対応するすべてのレコードを表示するためには、[カテゴリ] パネルの上にある [すべてのレコード] ノードをクリックします。

デフォルトでは、データパネルには列が3つあります（ID、レスポンス、カテゴリ）。このパネル内に列を追加することもできます。追加できる列には次のものがあります。

- **ID:** レコードまたはドキュメントの識別子（ID）がインポートされている場合に、それをリスト表示します。
- **レスポンス:** コンセプトおよびタイプを抽出したテキストデータをリスト表示します。
- **カテゴリ:** レコードが属するカテゴリがそれぞれ表示されます。この列を表示する場合、最新の情報を示すため、[データ] パネルの更新に少し時間がかかる場合があります。この列ではレコードとの関連性に基づいてカテゴリがリストされます。詳細は、 p. 106 [カテゴリの関連性](#) を参照してください。
- **強制投入:** 当該レコードの回答を強制的に投入したカテゴリをリスト表示します。回答を強制的にカテゴリに含めるには、[カテゴリ]>[回答を強制投入]を選択します。詳細は、 p. 175 [回答のカテゴリへの強制投入](#) を参照してください。

- **強制排除:**当該レコードの回答を強制的に排除したカテゴリをリスト表示します。回答を強制的にカテゴリから排除するには、[カテゴリ]>[回答を強制排除]を選択します。たとえば皮肉の意味で使われた回答が誤ってカテゴリ化される場合にこれを使います。詳細は、 p.175 [回答のカテゴリへの強制投入](#) を参照してください。
- **合致するテキスト:**各回答で見つかったテキスト マッチをリスト表示します。ストリングが抽出されたかどうかに関わらず、ストリングを定義して特定のテキストをカテゴリ定義の一部にすることができます。詳細は、 p.176 [カテゴリ内のテキストマッチ](#) を参照してください。
- **カテゴリの個数:**質問において、1回答（レコード）内に含まれるカテゴリの総数が表示されます。
- **適合度順位:**1つのカテゴリの各レコードの順位が表示されます。この適合度順位は、カテゴリ内の他のレコードと比較して、レコードがカテゴリにどれだけ適合しているかを示します。[カテゴリ] パネル（左上のパネル）でカテゴリを選択すると、順位が表示されます。詳細は、 p.106 [カテゴリの関連性](#) を参照してください。
- **回答フラグ:**使用している回答フラグ用の列を追加します。この列の中をクリックして、各回答に割り当てるフラグの種類を変更できます。
- **<参照変数名>:**選択した参照変数の列を [データ] パネルに追加します。まだ参照変数がインポートされていない場合、ここには何も表示されません。参照変数1つにつき、1列となります。詳細は、 3 章 p.36 [変数の選択](#) を参照してください。

#### データパネルのその他の列を表示するために

- ▶ [データ] パネル内で、列見出しを右クリックするとポップアップメニューが開きます。
- ▶ メニューから [列を表示] をクリックし、データパネルに表示したい列を選択します。新しい列がパネル内に表示されます。

注: 回答をカテゴリに投入またはカテゴリから除外すると、実際のカテゴリ定義を変更することなく、自動カテゴリ作成手法で作成されたカテゴリに定義を上書きできます。詳細は、 p.175 [回答のカテゴリへの強制投入](#) を参照してください。

## カテゴリの関連性

より良いカテゴリを作成するため、各カテゴリのレコードの関連性のほか、レコードが属するすべてのカテゴリの関連性を確認できます。



## カテゴリのレコードに対する関連性

[データ] パネルにレコードが表示されると、それらのすべてのカテゴリが [カテゴリ] 列に表示されます。レコードが複数のカテゴリに含まれる場合、この列のカテゴリは、関連性が最も大きなものから小さなものの順に表示されます。最初に表示されたカテゴリは、このレコードに最も対応すると考えられます。詳細は、 p. 104 [データ] パネル を参照してください。

## レコードのカテゴリに対する関連性

カテゴリを選択すると、[データ] パネルの [適合度順位] に各レコードの関連性が表示されます。この適合度順位は、レコードが選択したカテゴリにどれだけ適合しているかを、そのカテゴリのほかのレコードと比較して示します。単一のカテゴリのレコードの順位を確認するには、左上の [カテゴリ] パネルでそのカテゴリを選択します。レコードの順位が列に表示されます。デフォルトでは、この列は表示されませんが、列が表示されるよう選択することができます。詳細は、 p. 104 [データ] パネル を参照してください。

レコードの順位が低いほど、このレコードの、選択したカテゴリに対する適合度または関連性が大きくなり、1 が最も適合度が高くなります。複数のレコードの関連性が同じ場合、それぞれが同じ順位で表示され、その後に関連性が同じであることを示す等号 (=) が追加されます。たとえば、1=、1=、3、4 などのようになります。このカテゴリに最もマッチするレコードが 2 つあることを意味します。

**ヒント:**最も関連性の高いレコードのテキストをカテゴリの注釈に追加して、カテゴリの説明をより分かりやすくすることができます。テキストを選択し、メニューから [カテゴリ] > [注釈に追加] を選択して、[データ] パネルからテキストを直接追加します。

図 6-4  
カテゴリおよび適合度順位を示す [データ] パネル

	Q1_6_もう一度来たい店とは (18)	カテゴリ	適合度順位
1	旨いのは当たり前ですが、廉な思いをさせない対応を常にしてくれるお店	残りのコンセプト する 対応 店 思う 旨い	18
2	おいしくて、雰囲気が良いお店。雰囲気・値段に合わせて離れないサービスを提供 のお店。値段を気にし過ぎなくてもいいお店。	残りのコンセプト する 店 サービス 値段 雰囲気	15
3	まず、料理がうまいということが第一条件だが、料理を楽しむのは、そのテーブルという空間を一定の時間買っているのでお店の雰囲気や従業員の教育もしっかりとされていて料理を楽しむことができるかによって、次も行きたい店になると思う	残りのコンセプト する 店 うまい 思う 雰囲気 いる 料理	14
4	食べ物美味しい事も大事ですが、やはりその店のサービスや店員の対応です。あと、店の設備がちゃんとされているかどうか大事なポイントです。例えば、トイレはちゃんと掃除されているのかだとか、床にゴミなどは落ちていないかなど。	残りのコンセプト する 店 いる サービス 対応 おいしい 店員	13
5	テーブルの間隔がゆったりとしていて寛げる店。	する 店	12

## カテゴリ作成の方法と戦略

また抽出していない、または抽出結果が古い場合、カテゴリ作成方法または拡張方法のいずれかを使用すると、抽出についてのプロンプトが自動的に表示されます。手法を適用した後、カテゴリにグループ化したコンセプトおよびタイプはその他の手法で構築したカテゴリに使用できます。つまり、再利用しないことを選択しないかぎり、複数のカテゴリのコンセプトを表示することができます。

最適なカテゴリを作成するために、次のことを確認してください。

- カテゴリ作成の方法
- カテゴリ作成の戦略
- カテゴリ作成のヒント

## カテゴリ作成の方法

すべてのデータセットが一意であるため、カテゴリ作成方法の数やそれらを適用する順序は、時間によって変わる場合があります。また、テキストマイニングの目標が、データセットによって異なる場合があるため、それぞれの手法を検証して、指定したテキスト データにとってどの手法が最良

の結果を生み出すかを確認する必要があります。 自動的手法では、データを完全にカテゴリ化できません。そのため、データに合った 1 つまたは複数の自動的手法を見つけ、適用することをお勧めします。

事前作成されたカテゴリ セットを持つテキスト分析パッケージ (TAP、\*.tap) を使用するほか、次の方法を組み合わせて回答をカテゴリ化することもできます。

- **自動作成手法:** いくつかの言語ベースおよび頻度ベースのカテゴリオプションを使用して、カテゴリを自動的に作成できます。 詳細は、p. 118 [カテゴリの作成](#) を参照してください。
- **自動拡張手法:** いくつかの言語的手法を使用し、記述子を追加および拡張することによって既存のカテゴリを展開し、より多くのレコードをキャプチャすることができます。 詳細は、 p. 135 [カテゴリの拡張](#) を参照してください。
- **手動による手法:** ドラッグアンドドロップなど、手動による手法がいくつかあります。 詳細は、 p. 140 [手作業でのカテゴリの作成](#) を参照してください。
- **コードフレーム:** 独自のコード フレームをインポートするか、コードをコード フレーム マネージャにコピー/ペーストします。 詳細は、p. 143 [定義済みカテゴリのインポート](#) を参照してください。

## カテゴリ作成の方略

次のリストの方略は包括的ではありませんが、カテゴリの作成方法について、いくつかのキーワードが用意されています。

- プロジェクトを開始する場合、テキスト分析パッケージからカテゴリを選択し、いくつかの作成済みカテゴリを使用して分析を開始します。これらのカテゴリは、テキストを最初から十分にカテゴリ化することができます。ただし、カテゴリを追加する場合、カテゴリ作成設定 ([カテゴリ] → [設定を行う]) を編集することができます。 [詳細設定: 言語ダイアログを開き、カテゴリ入力オプションの未使用の抽出結果を選択し、追加カテゴリを作成します。](#)
- プロジェクトを開始するときに、TAP からカテゴリを選択します。次に、未使用のコンセプトまたはパターンを適切なカテゴリにドラッグアンドドロップします。そして、編集した既存のカテゴリを展開 ([カテゴリ] → [カテゴリを展開]) し、既存のカテゴリ記述子に関連するより多くの記述子を取得します。
- 詳細言語設定を使用して、自動的にカテゴリを作成します ([カテゴリ] → [カテゴリを作成])。生成されるカテゴリが適切なものとなるまで、記述子を削除、カテゴリを削除、または同様のカテゴリを結合してカテゴリを手動で調整します。また、元来 [\[可能な場合ワイルドカードを使用して一般化\]](#) オプションを使用しないでカテゴリを作成する場合、[\[一般化\]](#) オプ

ションをオンにして [カテゴリを展開] を使用し、自動的にカテゴリを簡略化することもできます。

- 非常に説明的なカテゴリ名および注釈を持つ事前定義済みカテゴリ ファイルをインポートします。また、元来そのオプションを選択しないでインポートし、カテゴリ名から記述子をインポートまたは生成する場合、後で [カテゴリを拡張] ダイアログを使用して、[カテゴリ名から生成された記述子を使用して空白のカテゴリを拡張する] オプションを選択することができます。そして、これらのカテゴリの 2 回目の展開を行います。今回はグループ化手法を使用します。
- コンセプトまたはコンセプト パターンを頻度によって並べ替え、最も関心のあるコンセプトまたはコンセプト パターンを [カテゴリ] パネルにドラッグ アンド ドロップすることによって、カテゴリの最初のセットを手動で作成します。カテゴリの最初のセットを作成したら、展開機能 ([カテゴリ] → [カテゴリを展開]) を使用して、すべてのカテゴリを展開し、その他の関連記述子を含めてより多くのレコードに一致するように選択したカテゴリを調整します。

これらの手法を適用した後、作成されたカテゴリを確認、手動による手法を使用して小さな調整を行い、誤分類を削除、または欠損したと考えられるレコードまたは単語を追加することをお勧めします。また、さまざまな手法を使用すると、重複したカテゴリを作成する場合もあるため、必要に応じてカテゴリを結合または削除することもできます。詳細は、p. 169 [カテゴリの編集および調整](#) を参照してください。

## カテゴリ作成のヒント

より良いカテゴリを作成できるよう、方法を決定できるヒントをいくつか確認できます。

### カテゴリ-to-回答比率のヒント

[最後にアウトレット ストアに行ったのはいつですか?] という、選択形式の質問のコードを作成する場合、回答が振り分けられるカテゴリは相互に排他的で、包括的である必要があります。しかしこの原則は定性的なテキスト分析においては必ずしも要求されません。これには少なくとも 2 つの理由があります。

- 1 つめの理由は、一般的に、テキスト レコードが長いほど、表されるキーワードおよび意見がより明確なものとなることです。そのため、レコードに複数のカテゴリが割り当てられるという機会が大幅に増えます。
- 2 つめの理由は、論理的に分けられていないテキスト レコードをグループ化および解釈するさまざまな方法があるということです。回答者の政治的な信念に関する、自由記述式の質問を含んだ調査の場合、「リベ

ラル」および「保守的」、または「共和党」および「民主党」のようなカテゴリのほか、「社会的にリベラル」、「財政的に保守的」など、より特定のなカテゴリも作成できます。これらのカテゴリは相互に重複部分がなかったり、すべてをカバーしたりしている必要はありません。

### 作成するカテゴリ数のヒント

極端に単純な自由記述式の質問の場合を除いて、カテゴリをいくつ作成したらよいか直感的にわかるということはまずありえません。実際のところカテゴリの個数はさほど問題にはなりません。ただ、カテゴリ作成は、データから直接生じることはありません。この調査の目的について何か興味深いものがあった場合、カテゴリを作成してその属性およびキーワードを示すことができます。

- **カテゴリ度数:** カテゴリが役立つものになるには、最低限のレコードが必要です。1 つまたは 2 つのレコードには非常に興味深いものが含まれている場合がありますが、それが 1,000 件のレコードのうちの 1 つまたは 2 つである場合、含まれる情報は、実際に役立つほど母集団の中では頻繁ではありません。
- **複雑さ:** 作成したカテゴリが多いほど、分析が完了した後確認および要約する必要がある情報が多くなります。ただし、カテゴリ数が多すぎる場合、複雑さが増しても、役立つ詳細は増えません。

カテゴリ数が多すぎることを判断する規則、またはカテゴリあたりの最小レコード数を決定する規則はありません。個々の状況の必要性に応じて分析者が判断する必要があります。

しかしながら、基本的なアドバイスはあります。まずカテゴリの個数は多すぎてもいいませんが、分析の早い段階においては、カテゴリが少なすぎるよりは多すぎるほうがいいでしょう。比較的類似したカテゴリをまとめるほうが、ケースを分けて新しいカテゴリに細分化するよりも簡単なので、多くのカテゴリからより少ない個数のカテゴリになるように作業していくほうが、一般に容易だといえます。テキスト マイニングの反復性およびこのソフトウェア プログラムを使用して達成できる容易さにより、より多くのカテゴリを作成することが、開始時点では推奨されます。

## 最適な記述子の選択

次に、カテゴリに最適な記述子（コンセプト、タイプ、TLA パターンおよびカテゴリ規則）の選択または作成におけるガイドラインをいくつか示します。記述子とは、カテゴリの構築ブロックです。レコードの一部またはすべてのテキストが記述子に合致する場合、レコードはカテゴリに合致します。

記述子が抽出されたコンセプトまたはパターンを含まないまたは対応しない場合、レコードに合致しません。そのため、次で説明しているように、コンセプト、タイプ、パターン、およびカテゴリ規則を使用します。

コンセプトが、それ自体だけでなく複数/単数形、類義語、およびスペルの変異形にいたる一連の基本キーワードも示すため、コンセプト自体は、記述子、または記述子の一部として使用する必要があります。指定されたコンセプトの基本キーワードについての詳細を知るには、[抽出結果] パネルのコンセプト名をクリックします。コンセプト名でマウスポインタを停止すると、ツールヒントが表示され、そこに最後の抽出時にテキストで検出された基本キーワードが表示されます。すべてのコンセプトに基本キーワードがあるわけではありません。たとえば、car と vehicle は類義語ですが car がコンセプトとして、vehicle が基本キーワードとして抽出された場合、vehicle を含むレコードに合致するため、記述子には car だけを使用します。

### 記述子としてのコンセプトとタイプ

コンセプト（または基本キーワードのいずれか）を含むすべてのレコードを検出する場合、そのコンセプトを記述子として使用します。この場合、正確なコンセプト名で十分であるため、より複雑なカテゴリ規則を使用する必要はありません。意見を抽出するリソースを使用する場合、文の真の意味を抽出する TLA パターン抽出時にコンセプトが変更される場合があるので注意してください（TLA に関する次の項の例を参照してください）。

たとえば、「Apple and pineapple are the best」のような、各個人の好きな果物を示す調査の回答によって、apple と pineapple が抽出されます。コンセプト apple をカテゴリに記述子として追加すると、コンセプト apple（または基本キーワードのいずれか）を含むすべての回答がそのカテゴリに合致します。

ただし、とにかく apple について言及する回答を知りたい場合、カテゴリ規則を \* apple \* のように作成すると、apple、apple sauce、または french apple tart のようなコンセプトを含む回答もキャプチャできます。

また、<Fruit> のようにタイプを記述子として直接指定することによって、同じようにタイプ指定されたコンセプトを含むすべてのレコードをキャプチャすることもできます。タイプには \* は使用できませんので注意してください。

詳細は、5 章 p.83 [抽出結果:コンセプト、タイプ、パターン](#) を参照してください。

### 記述子としてのテキストリンク分析 (TLA) パターン

より詳細で微妙なアイデアをキャプチャする場合、TLA パターン結果を記述子として使用します。テキストが TLA 抽出中に分析されると、テキスト全体（レコード）を処理するのではなく、一度に 1 文、または 1 句ずつテキストが処理されます。1 つの文の全部分を考慮することによって、TLA は意見、2 つの要素間の関係性、または否定的な表現を特定し

て、真の意味を理解できます。コンセプト パターンまたはタイプ パターンを記述子として使用できます。

たとえば、「the room was not that clean」というテキストがある場合、room および clean というコンセプトが抽出されます。ただし、抽出設定で TLA 抽出が有効になっている場合、TLA で clean が否定文で使用されており、実際は not clean に対応し、コンセプト dirty の類義語であることを検出できます。ここで、記述子であるコンセプト clean がこのテキストに合致しますが、清潔さについて示すその他のレコードのキャプチャできることが確認できます。そのため、このテキストに合致し、より適切な記述子となるため、dirty を出力コンセプトに指定した TLA コンセプト パターンを使用することをお勧めします。

### 記述子としてのカテゴリ ビジネス規則

カテゴリ規則とは、抽出したコンセプト、タイプ、パターン、およびブール型演算子を使用した論理式に基づいて、レコードを自動的に分類するステートメントです。たとえば、「このカテゴリに、アルゼンチンではなく、大使館という抽出したコンセプトを含む」という内容を意味する式を作成することができます。

カテゴリ規則をカテゴリの記述子として記述および使用し、&、|、および !() によってさまざまなアイデアを表現することができます。ブール値。これらの規則のシンタックスおよびそれらの作成、編集方法の詳細は、「カテゴリ規則の使用」( p.156 ) を参照してください。

- & (AND) ブール演算子を含むカテゴリ規則を使用して、2 つ以上のコンセプトが出現するレコードを検出します。 & 演算子でつながった 2 つ以上のコンセプトは、同じ文またはフレーズで出現する必要はありませんが、同じレコードのどこかに出現するとカテゴリに合致すると見なされます。たとえば、記述子にカテゴリ規則 food & cheap を作成すると、テキストに food と cheap が含まれているため、food が cheap という名詞でないにもかかわらず、「the food was pretty expensive, but the rooms were cheap」というテキストを含むレコードに合致します。
- (NOT) ブール演算子を含むカテゴリ規則を使用して、いくつかのコンセプトまたはタイプのいずれかを含むレコードを検出します。コンテキストではなく、単語に基づいて関連すると思われる情報がグループ化されないようにします。たとえば、カテゴリ規則 <Organization> & !(ibm) を記述子として作成すると、「SPSS Inc. was a company founded in 1967」というテキストには合致しますが、「the software company was acquired by IBM.」には合致しません。
- | (OR) ブール演算子を含むカテゴリ規則を使用して、いくつかのコンセプトまたはタイプのいずれかを含むレコードを検出します。たとえば、カテゴリ規則 (personnel|staff|team|coworkers) & bad を記

述子として作成すると、これらの名詞がコンセプト bad と共に検出されるレコードと合致します。

- カテゴリ規則にタイプを使用すると、その規則はより一般的になり、より展開しやすくなる場合があります。たとえば、ホテルのデータを扱っている場合、ホテルのスタッフに対して顧客がどう思っているかについて、非常に興味があります。関連するキーワードには、receptionist、waiter、waitress、reception desk、front desk（受付、ウェイター、ウェイトレス、受付デスク、フロント デスク）などがあります。この場合、<HotelStaff> という新しいタイプを作成し、上記のキーワードをすべてこのタイプに追加します。[\* waitress \* & nice]、[\* desk \* & friendly]、[\* receptionist \* & accommodating] のようなすべての種類のスタッフに 1 つのカテゴリ規則を作成することができますが、タイプ <HotelStaff> を使用して 1 つの、より一般的なカテゴリ規則を作成して[<HotelStaff> & <Positive>] の形式でホテル スタッフに対して好意的な意見の回答をすべてキャプチャできます。

注： カテゴリ規則に TLA パターンを含む場合、これらの規則に + と & を両方使用できます。詳細は、p.159 [カテゴリ規則内の TLA パターンの使用](#) を参照してください。

### 記述子であるコンセプト、TLA、またはカテゴリ規則がどのように合致するかについての例

次の例では、記述子としてコンセプト、カテゴリ規則、TLA パターンを使用するとレコードがどのようにカテゴリ化されるのかについて説明します。次のような 5 つのレコードがあるとします。

- A: “awesome restaurant staff, excellent food and rooms comfortable and clean.”（素晴らしいレストラン スタッフ、食べ物もおいしく、部屋は快適で清潔）
- B: “restaurant personnel was awful, but rooms were clean.”（レストラン スタッフはひどいが、部屋はきれいだった）
- C: “Comfortable, clean rooms.”（快適で清潔な部屋）
- D: “My room was not that clean. “（私の部屋はきれいではなかった）
- E: “Clean.”（きれいだった）

レコードに「clean」という単語を含み、この情報をキャプチャしたいため、次の表に示す記述子のいずれかを作成します。キャプチャしようとしている核心に基づき、ある種類の記述子を別の記述子に対して使用すると、どのように異なる結果を生成するかを確認できます。



テーブル 6-1  
レコードが記述子に合致する例

記述子	A	B	C	D	E	説明
clean	合致	合致	合致	合致	合致	記述子は抽出されたコンセプトです。TLA のないレコード D も含めてすべてのレコードがコンセプト clean を含み、TLA 規則によって「not clean」が「dirty」を意味することは自動的に認識されません。
clean + .	-	-	-	-	合致	記述子は clean を示す TLA パターンです。clean が TLA 抽出時に関連するコンセプトなしで抽出されたレコードとのみ合致します。
[clean]	合致	合致	合致	-	合致	記述子は、それ自体またはその他の規則に clean を含む TLA 規則を探すカテゴリ規則です。clean を含む TLA 出力が、clean が room のような別のコンセプトにリンクするかどうか、別のスロット位置にあるかどうかに関係なく、検出されたすべてのレコードに合致します。

## カテゴリとは

**カテゴリ**は、密接に関連したコンセプト、オピニオン、または属性のグループのことを言います。短いことば（ラベル）を付けて、カテゴリの内容が簡単にわかるようにしておくのが便利です。

たとえば、新しい洗濯用洗剤について消費者からのアンケートの回答を分析する場合、製品の香りを示すすべての回答を含む「香り」というラベルの付いたカテゴリを作成できます。ただし、そのようなカテゴリは良い香りと感じた消費者と、香りが強いと感じた消費者とを差別化するものではありません。IBM® SPSS® Text Analytics for Surveys は適切なリソースを使用する場合意見の抽出ができるため、2 つのカテゴリを他に作成して、「香りが良い」と答えた回答者と「香りは好みではない」と答えた回答者を特定することができます。






テキスト分析ウィンドウの左上のパネルの [カテゴリ] パネルで、カテゴリを作成したり作業することができます。各カテゴリは、1 つまたは複数の記述子で定義されます。**記述子**は、カテゴリを定義するために使用されているコンセプト、タイプ、パターンおよびカテゴリ規則です。

指定のカテゴリを構成する記述子を表示する場合、ツリー内のカテゴリを展開します。アイコンがツリー内に表示されるので、各要素を容易に特定できます。最初のレベルのみがカテゴリを定義します。この定義をさらに展開すると、データ内で発見されるものが表示されます。

内包関係のコンセプトなどのカテゴリ作成手法を使用してカテゴリを自動的に作成する場合、その手法ではコンセプトおよびタイプを記述子として使用し、カテゴリを作成します。これらのパターンまたはパターンの一部をカテゴリ記述子として追加することもできます。最後に、カテゴリ規則を手動で作成して、カテゴリの記述子として使用することができます。詳細は、 p. 156 [カテゴリ規則の使用](#) を参照してください。

たとえば、タイプをカテゴリ定義に追加すると、この時点でデータに少数しかない場合であっても、そのタイプに割り当てられたコンセプトが自動的に追加されます。これは、新しいデータに対して、カテゴリの定義を再利用する際に役に立ちます。詳細は、 p. 178 [カテゴリパネルのコピー](#) を参照してください。手動でカテゴリ規則を作成し、カテゴリの記述子として使用することもできます。詳細は、 p. 156 [カテゴリ規則の使用](#) を参照してください。

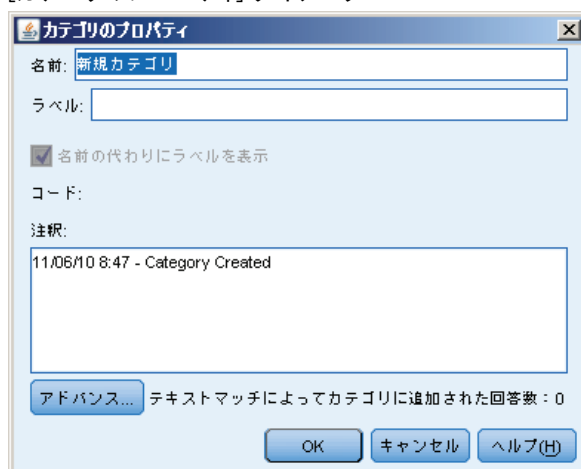
テーブル 6-2  
定義の要素を特定するアイコン

アイコン	説明
	コンセプト:
	タイプ (拡張して、含まれるコンセプトを確認できる)。
	コンセプト パターン (拡張して、パターン内の特定コンセプトを確認するできる)。
	タイプ パターン (拡張して、コンセプト パターン レベルを確認できる)。
	カテゴリのカテゴリ規則。規則名を右クリックして規則を編集します。

## カテゴリのプロパティ

記述子のほかカテゴリには、カテゴリの名前を変更、ラベルまたは注釈を追加、またはテキスト マッチ ダイアログを使用するために編集できるプロパティもあります。

図 6-5  
[カテゴリのプロパティ] ダイアログ



以下のようなプロパティがあります。

- **名前:** この名前は、デフォルトでツリーに表示されます。自動化の手法でカテゴリを作成した場合、名前は自動的に付けられます。
- **ラベル:** ラベルの使用は、他の製品または他のテーブルまたはグラフで使用する、より重要なカテゴリ記述子を作成する場合に役立ちます。[名前の代わりにラベルを表示]を選択すると、インターフェイス内でカテゴリを特定する際にラベルが使用されます。IBM® SPSS® Statistics では、変数ラベルとしてエクスポートされます。Microsoft Excel では、個別の行としてエクスポートされます。
- **コード:** コード番号は、このカテゴリのコード値に対応します。コードフレーム マネージャを使用してこのコードを編集します。詳細は、p. 106 “コードフレームによるカテゴリの作成” を参照してください。
- **注釈:** このフィールドで各カテゴリの短い説明を追加できます。[カテゴリを作成] ダイアログでカテゴリを生成した場合、この注釈にメモが自動的に追加されます。テキストを選択し、メニューから [カテゴリ] → [注釈に追加] を選択して、[データ] パネルからサンプル テキストを注釈に直接追加することもできます。
- **詳細テキスト マッチ:** [詳細] ボタンを使用して、カテゴリ定義に単語または句を追加できます。しばしば、これはうまくいかなかった抽出を無効にするのに使用されます。詳細は、p. 176 [カテゴリ内のテキスト マッチ](#) を参照してください。

## カテゴリの作成

テキスト分析パッケージのカテゴリがある場合でも、さまざまな言語学的手法および出現頻度に基づく手法を使用して、カテゴリを自動的に作成することもできます。[カテゴリ作成設定] ダイアログを使用して、自動的な言語学的手法および出現頻度に基づく手法を適用し、コンセプトまたはコンセプトパターンのいずれかよりカテゴリを作成することができます。

図 6-6  
[カテゴリを作成] ダイアログ ボックス



一般的に、カテゴリはさまざまな記述子（タイプ、コンセプト、TLA パターン、カテゴリ規則）で構成されます。自動化されたカテゴリ作成方法を使用してカテゴリを作成する場合、作成されたカテゴリは、選択した入力に応じてコンセプトまたはコンセプトのパターンから命名され、それぞれ一連の記述子を使用します。これらの記述子は、カテゴリ規則またはコンセプトの形式で、手法によって発見されたすべての関連コンセプトを含む場合があります。

カテゴリを作成した後、[カテゴリ] パネルで確認またはグラフや図表を使用して検討することにより、カテゴリについて多くのことを学ぶことができます。手作業で若干の調整を行い、間違った分類を除去したり、把握

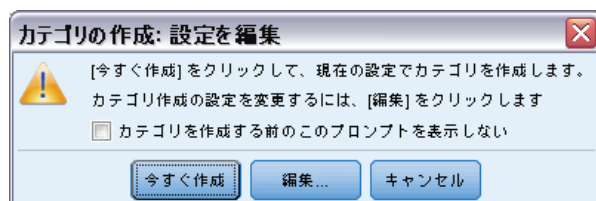
されなかったレコードや語を追加することもできます。手法を適用した後、カテゴリにグループ化したコンセプト、種類、パターンは他の手法にも使用できます。また、さまざまな手法を使用すると、重複したカテゴリまたは不適切なカテゴリを作成する場合もあるため、カテゴリを結合または削除することもできます。詳細は、p. 169 [カテゴリの編集および調整](#) を参照してください。

**重要!** 以前のリリースでは、共起規則および類義語規則は、大カッコで囲まれていました。このリリースの場合、大カッコはパターン結果を示します。その代わりに、共起規則や類義語規則は、(スピーカー システム | スピーカー) のように、カッコで囲まれます。

### カテゴリを作成するには

- ▶ メニューから、[カテゴリ]>[カテゴリを作成] を選択します。プロンプトが表示されないよう設定している場合、メッセージ ボックスが表示されます。

図 6-7  
作成する前のプロンプト



- ▶ 今すぐ作成するか、左記に設定を編集するかを選択します。
  - [今すぐ作成] をクリックすると、現在の設定でカテゴリの作成が開始されます。デフォルトで選択されている設定で、十分カテゴリ化プロセスを開始できます。カテゴリ作成のプロセスが開始し、進捗状況のダイアログが表示されます。
  - [編集] をクリックして、ビルド設定を確認し、変更します。

### 入力

カテゴリは、タイプ パターンまたはタイプのいずれかより派生した記述子から作成されます。デフォルトでは、タイプ パターンはドロップダウン リストより選択します。表内で、カテゴリ作成プロセスに使用する各タイプまたはパターンを選択できます。

**タイプ パターン** :タイプ パターンを選択すると、タイプおよびコンセプトではなくパターンからカテゴリが独自に作成されます。このように、選択したタイプ パターンに属するコンセプト パターンを含むレコードがカテゴリ化されます。そのため、表で <Budget タイプ パターンおよび <Positive タイプ パターンを選択した場合、 cost & <Positive > または rates & excellent などのカテゴリを作成することができます。

表に <Location + <Positive や <Positive> + <Location など、タイプを組み合わせた行を 1 つだけ表示する場合、それらの順序は、カテゴリの生成方法に対して重要ではありません。

図 6-8  
使用できるタイプ パターンを示す [カテゴリを作成] ダイアログ



自動カテゴリ作成でタイプ パターンを入力として使用すると、その手法によってカテゴリ構造を形成する複数の方法が特定される場合があります。技術的には、カテゴリを作成する適切な方法はありませんが、分析により適した構造がある場合があります。この場合の出力をカスタマイズするために、タイプを優先的に指定できます。パターンタイプの構造カテゴリフィールドでこのタイプを選択すると、テーブルが更新され、選択したタイプを含む利用可能なパターンのみが表示されます。多くの場合、<Unknown>が事前に選択されています。この結果、タイプ <Unknown> が選択されます。表には、タイプ、レコード数 の最も多いものから降順で表示されます。

**タイプ** :タイプを選択すると、カテゴリは選択したタイプに属するコンセプトから作成されます。そのため、表で <Budget タイプを選択した場合、cost および price は <Budget に割り当てられたコンセプトであるため、cost または price のようなカテゴリを作成できます。

デフォルトでは、最も多いレコードをキャプチャするタイプのみが選択されます。このように事前選択すると、最も関心の高いタイプにすばやく焦点をあて、関心の低いカテゴリが作成されないようにすることができます。表には、タイプ、レコード数 の最も多いものから降順で表示されます。意見ライブラリのタイプは、デフォルトではタイプ テーブルで選択されていません。

入力の選択は、取得するカテゴリに影響します。入力としてタイプを選んだ場合は、明確に関連付けられたコンセプトをより簡単に見ることができます。たとえば、タイプを入力として使用してカテゴリを作成する場合、**果物** というカテゴリを、リンゴ、梨、柑橘類、オレンジなどのコンセプトとともに取得できます。他方、タイプ パターンを入力として選び、パターンとしてたとえば <不明 + <肯定的 を選択した場合には、**果物 + <肯定的** というカテゴリに、**果物 + おいしい** や **リンゴ + 良い**などの 1、2 種類の

果物を伴ったものを取得することになるでしょう。この第 2 の結果は、2 つのコンセプト パターンを示すのみです。それは、他の果物の出現が必ずしも肯定的に評価されるものとは限らないからです。また、現在手元にあるテキスト データについてはこれで十分であったとしても、異なるドキュメント セットを使用する経時的な研究においては、**柑橘類 + 肯定的**のような別の記述子を手動で追加したり、タイプを使いたいと考えることもあるでしょう。タイプだけを入力として使用することは、すべての可能な果物を見つけ出すのに役に立ちます。

図 6-9  
使用できるタイプを示す [カテゴリを作成] ダイアログ



## 手法

すべてのデータセットが一意であるため、手法の数やそれらを適用する順序は、時間によって変わる場合があります。テキスト マイニングの目標が、データセットによって異なる場合があるため、それぞれの手法を検証して、指定したテキスト データにとってどの手法が最良の結果を生み出すかを確認する必要があります。

これらの設定について詳しく知らなくても、使用することができます。デフォルトでは、最も一般的で平均的な設定がすでに選択されています。そのため、高度な設定のダイアログを省略して、カテゴリをすぐに作成することができます。同様に、ここで変更を行うと、最新の設定が常に保持されるため、設定ごとに設定ダイアログに戻る必要はありません。

言語的手法または出現頻度に基づく手法のいずれかを選択し、[詳細設定] ボタンをクリックして、選択した手法の設定を表示します。自動的手法では、データを完全にカテゴリ化できません。そのため、データに合った 1 つまたは複数の自動的手法を見つけ、適用することをお勧めします。言語的手法および出現頻度に基づく手法を同時に使用して作成することはできません。

- **詳細言語法**: 詳細は、 p. 122 を参照してください。
- **高度な出現頻度に基づく手法**: 詳細は、 p. 134 を参照してください。

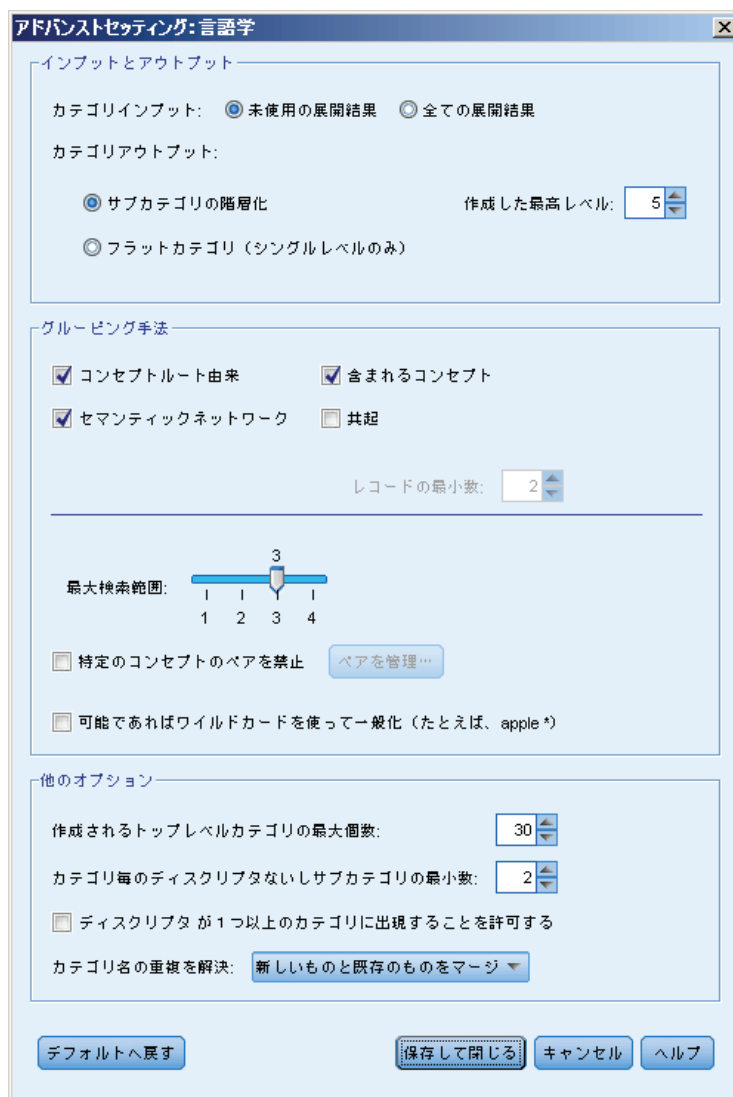
## 詳細言語設定

カテゴリを作成する場合、「派生関係のコンセプトの語幹」、「内包関係のコンセプト」、「セマンティック ネットワーク」（英語テキストのみ）、および「共起規則」など、さまざまな詳細言語カテゴリ作成手法から選択することができます。これらの手法は個別に、またはそれぞれを組み合わせることでカテゴリを作成することができます。

すべてのデータセットが一意であるため、手法の数やそれらを適用する順序は、時間によって変わる場合がありますので、注意してください。テキストマイニングの目標が、データセットによって異なる場合があるため、それぞれの手法を検証して、指定したテキストデータにとってどの手法が最良の結果を生み出すかを確認する必要があります。自動的手法では、データを完全にカテゴリ化できません。そのため、データに合った1つまたは複数の自動的手法を見つけ、適用することをお勧めします。



図 6-10  
詳細設定:カテゴリ作成のための言語ダイアログボックス



## 入力および出力

**カテゴリ入力:** カテゴリが作成される内容を選択します。

- **未使用の抽出結果:** 既存のカテゴリで使用されていない抽出結果からカテゴリを作成できます。レコードが、複数のカテゴリと合致する傾向が最も小さくなり、作成されるカテゴリの数が制限されます。
- **すべての抽出結果:** 抽出結果のいずれを使用してもカテゴリを作成できます。カテゴリがないまたは少ない場合に最も役立ちます。

**カテゴリ出力:** 作成されるカテゴリの一般的な構造を選択します。

- **サブカテゴリによる階層:** サブカテゴリおよびサブサブカテゴリの作成を有効にします。作成できる最大数のレベル（[作成するレベルの最大個数] フィールド）を選択して、カテゴリの深度を設定できます。3 を選択すると、カテゴリ内にサブカテゴリを作成でき、またこれらのサブカテゴリ内にもサブカテゴリを作成できます。
- **フラットなカテゴリ(単一レベルのみ):** 1 レベルのみのカテゴリを作成できます。つまり、サブカテゴリは生成できません。

### グループ化手法

使用できるそれぞれの手法は、特定の種類のデータおよび状況に適していますが、同じ分析で手法を組み合わせると、全範囲のレコードをキャプチャすると役に立つ場合があります。複数のカテゴリのコンセプトを表示したり、重複するカテゴリを見つけることができます。

**派生関係のコンセプトの語幹:** コンセプト コンポーネントが形態的に関連するか、または語幹を共有するかどうかを分析するとき、コンセプトを取得し、そのコンセプトに関連するその他のコンセプトを検索することによって、カテゴリを作成します。この手法は、生成された各カテゴリのコンセプトが類義語または意味の上で密接に関連しているため、類義語の複合語コンセプトを特定するのに非常に役立ちます。長さの異なるデータを処理し、コンパクトなカテゴリをより少なく生成します。たとえば、コンセプト *opportunities to advance* は、コンセプト *opportunity for advancement* および *advancement opportunity* とグループ化されます。詳細は、p. 128 [派生関係のコンセプトの語幹](#) を参照してください。

**セマンティック ネットワーク:** 各コンセプトの考えられうる意味を、単語の関係の拡張インデックスから特定することによって開始し、関連するコンセプトをグループ化することによってカテゴリを作成します。この手法は、コンセプトがセマンティック ネットワークに認識され、あまり曖昧でない場合に最も適しています。テキストに、ネットワークが認識していない特殊な用語または専門用語が含まれている場合はあまり役に立ちません。たとえば、コンセプト *granny smith apple* は、*granny smith* と横の関係があるため、*gala apple* および *winesap apple* とグループ化されます。またあるいは、コンセプト *animal* (動物) は、その下位語である *cat* (ネコ) および *kangaroo* (カンガルー) とグループ化されます。このリリースでは、英語テキストにのみ使用できます。詳細は、p. 131 [セマンティック ネットワーク](#) を参照してください。

**内包関係のコンセプト:** この手法では、一方の共通の文字列である単語を含むかどうかに基づき、マルチタームのコンセプト（複合語）をグループ化することによってカテゴリを作成します。たとえば、コンセプト *seat* (シート) は、コンセプト *safety seat* (セーフティ シート)、*seat belt* (シート ベルト)、および *seat belt buckle* (シート ベルトのバックル)

とグループ化されます。詳細は、 p. 130 [内包関係のコンセプト](#) を参照してください。

**共起:** この手法では、テキスト内の共起関係のコンセプトからカテゴリを作成します。ドキュメントおよびレコードでコンセプトまたはコンセプト パターンがいっしょに出現することが多いとき、共起関係のコンセプトはおそらくカテゴリ定義の値のものである基底となる関連を反映します。単語が頻繁に共起する場合、共起規則が作成され、新しいサブカテゴリのカテゴリの記述子として使用できます。たとえば、多くのレコードに単語 price (価格) および availability (有効性) が含まれている場合 (ただし、一方を含み、もう一方を含まないレコードはほとんどない)、これらのコンセプトを共起規則 (price & available) にグループ化し、たとえばカテゴリ price のサブカテゴリに割り当てることができます。詳細は、 p. 132 [共起規則](#) を参照してください。

- **最小レコード数:** 共起関係のコンセプトの重要性を判断できるようにするため、カテゴリの記述子として使用されるよう、指定の共起関係のコンセプトを含む必要のあるレコードの最小数を定義します。

**最大検索距離:** カテゴリ作成前の手法による検索の距離を選択します。値が小さいほど、取得する結果は少なくなります。ただし、これらの結果はノイズが少なく、またリンクや関連性が大きくなります。値が大きいほど、取得する結果は多くなります。ただし、これらの結果の信頼性または関連性が弱くなります。このオプションはすべての手法にグローバルに適用されますが、共起およびセマンティック ネットワークに対する効果は最も大きくなります。

**特定のコンセプトがグループ化されないようにする:** 出力の 2 つのコンセプトがグループ化またはペアとならないように処理を停止します。コンセプト ペアを作成または管理するには、[\[ペアを管理\]](#) をクリックします。詳細は、 p. 126 [例外ペアのリンクの管理](#) を参照してください。

**可能な場合ワイルドカードを使用して一般化:** アスタリスク (\*) のワイルドカードを使用して、製品が一般的な規則をカテゴリに生成することができますようになります。たとえば、[\[アップル タルト + .\]](#) や [\[アップル ソース + .\]](#) などの複数の記述子を作成する代わりに、[\[アップル \\* + .\]](#) のようにワイルドカードを使用します。ワイルドカードを使用して一般化すると、以前と同じように、ちょうど同じ数のレコードを取得する場合があります。ただし、このオプションには、数の縮小やカテゴリの記述子の簡略化という利点があります。また、このオプションを使用すると、新しいテキスト データ (例: 長期的/周期的研究) にこれらのカテゴリを使用してより多くのレコードをカテゴリ化する機能を拡大します。

### カテゴリを作成するその他のオプション

適用するグループ化手法を選択するほか、次のように、その他の作成オプションを編集することができます。

**最大数のカテゴリがさくせいされました。**このオプションを使用して、[カテゴリを作成] をクリックすると作成できるカテゴリ数を制限します。この値を高く設定し、関心の低いカテゴリを削除すると、よりよい結果が生成される場合があります。

**カテゴリあたりの記述子数またはサブカテゴリ数の最小値:** カテゴリが作成するために含む必要のある記述子数およびサブカテゴリ数の最小値を定義します。多くのレコードをキャプチャしないカテゴリの作成が制限されます。

**同じ記述子が複数のカテゴリに含まれることを許可する:** このオプションを選択すると、記述子を次の作成される複数のカテゴリに使用できるようにします。項目が一般的にまたは「自然に」2 つ以上のカテゴリになり、より良い品質のカテゴリを作成するため、このオプションが一般的に選択されます。このオプションを選択しない場合、複数のカテゴリのレコードの重複が少なくなり、データのタイプによっては、これが望ましい設定となります。ただし、多くのデータタイプでは、記述子を1つのカテゴリに制限すると、品質またはカテゴリの範囲が損なわれます。たとえば、car seat manufacturer というコンセプトがあったとします。このオプションを指定すると、このコンセプトは、テキスト car seat に基づいてあるカテゴリに、また manufacturer というテキストに基づいて別のカテゴリに使用されます。ただし、このオプションが選択されていない場合、2つのカテゴリを取得できますが、コンセプト car seat manufacturer は、car seat および manufacturer がそれぞれ出現するレコード数など、いくつかの要素に基づいて、最も一致するカテゴリにのみ、記述子として使用されます。

**次の方法で重複するカテゴリ名を解決:** 名前が既存のカテゴリと同じ新規カテゴリまたはサブカテゴリの処理方法を選択します。新規カテゴリ（およびその記述子）と同じ名前を持つ既存カテゴリとを結合できます。あるいは、既存のカテゴリに重複する名前があった場合、カテゴリの作成をスキップすることもできます。

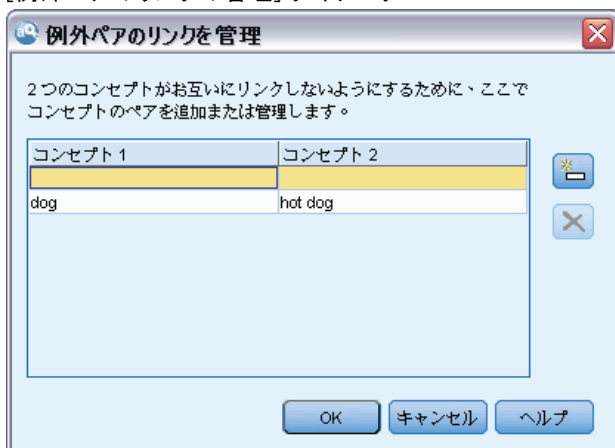
## 例外ペアのリンクの管理

カテゴリ作成およびコンセプト マッピングの間、既知の関連性によって内部アルゴリズムが語をグループ化します。2つのコンセプトが対応しないように、またはお互いにリンクしないようにするためには、[カテゴリ作成詳細設定] ダイアログ、および [コンセプト マップ インデックス設定] ダイアログでこの機能をオンにして、[ペアを管理] ボタンをクリックします。

表示される [リンクの例外を管理] ダイアログでコンセプト ペアを追加、編集または削除できます。1行につき1つのペアを入力します。ここでペアを入力すると、カテゴリ作成または拡張時のグループ化、コンセプト マッピングが行われなくなります。必要に応じて、単語を正確に入力します。たとえば、単語のアクセント付きバージョンがアクセントのないバージョンとは同じではありません。

たとえば、hot dog および dog がグループ化されていないことを確認したい場合、ペアをテーブル内の各行に追加できます。

図 6-11  
[例外ペアのリンクの管理] ダイアログ



## 言語学的手法について

カテゴリを作成または拡張する場合、「派生関係のコンセプトの語幹」、「内包関係のコンセプト」、「セマンティック ネットワーク」（英語テキストのみ）、および「共起規則」など、さまざまな詳細言語カテゴリ作成手法から選択することができます。これらの手法は個別に、またはそれぞれを組み合わせてカテゴリを作成することができます。

これらの設定について詳しく知らなくても、使用することができます。デフォルトでは、最も一般的で平均的な設定がすでに選択されています。必要に応じて、高度な設定のダイアログを省略して、カテゴリをすぐに作成または拡張することができます。同様に、ここで変更を行うと、最後に使用した設定が記憶されているため、設定ごとに設定ダイアログに戻る必要はありません。

ただし、すべてのデータセットが一意であるため、手法の数やそれらを適用する順序は、時間によって変わる場合がありますので、注意してください。テキスト マイニングの目標が、データセットによって異なる場合があるため、それぞれの手法を検証して、指定したテキスト データにとってどの手法が最良の結果を生み出すかを確認する必要があります。自動的手法では、データを完全にカテゴリ化できません。そのため、データに合った 1 つまたは複数の自動的手法を見つけ、適用することをお勧めします。

カテゴリを作成する、主な自動化言語法は、次のとおりです。

- **派生関係のコンセプトの語幹:** コンセプト コンポーネントが形態的に関連するかどうかを分析するとき、コンセプトを取得し、そのコンセプトに関連するその他のコンセプトを検索することによって、カテゴリ

を作成します。詳細は、 p.128 [派生関係のコンセプトの語幹](#) を参照してください。

- **ないほう関係のコンセプト:** コンセプトを取得し、そのコンセプトを含むその他のコンセプトを見つけることによって、カテゴリを作成します。詳細は、 p.130 [内包関係のコンセプト](#) を参照してください。
- **セマンティック ネットワーク:** 各コンセプトの考えられうる意味を、単語の関係の拡張インデックスから特定することによって開始し、関連するコンセプトをグループ化することによってカテゴリを作成します。詳細は、 p.131 [セマンティック ネットワーク](#) を参照してください。このオプションは、英語テキストにのみ使用できます。
- **共起:** 新しいカテゴリを作成、カテゴリを拡張するために、または別のカテゴリ手法の入力として使用できる共起規則を作成します。詳細は、 p.132 [共起規則](#) を参照してください。

## 派生関係のコンセプトの語幹

派生関係のコンセプトの語幹による手法は、コンセプト コンポーネントが形態的に関連するかどうかを分析するとき、コンセプトを取得し、そのコンセプトに関連するその他のコンセプトを検索することによって、カテゴリを作成します。コンポーネントは単語です。コンセプトの各コンポーネントの末尾（接尾辞）を確認し、それらから派生するその他のコンセプトを見つけることによって、コンセプトのグループ化を試みます。単語がお互いに派生している場合、共有するか、意味の上で近い傾向にあります。末尾を特定するために、内部の言語固有の規則が使用されます。たとえば、コンセプト *opportunities to advance* は、コンセプト *opportunity for advancement* および *advancement opportunity* とグループ化されます。

いかなる種類のテキストにも派生関係のコンセプトの語幹を使用できます。それ自体によって作成されるカテゴリはごくわずかであり、各カテゴリに含まれるコンセプトも少数です。各カテゴリのコンセプトは、類義語または状況的に関連しています。手動でカテゴリを作成する場合でも、このアルゴリズムを作成すると役立ちます。見つかった類義語は、特に関心のあるコンセプトの類義語である場合があります。

注:それらを明示的に指定して、コンセプトがグループ化されないようにすることができます。詳細は、 p.126 [例外ペアのリンクの管理](#) を参照してください。

## キーワードのコンポーネント化および活用の無効化

派生関係のコンセプトの語幹による手法または内包関係のコンセプトの手法が適用されると、キーワードはまずコンポーネント（単語）に分割され、コンポーネントの活用が無効化されます。手法が適用されると、コンセプトおよびそれらに関連したキーワードが読み込まれ、スペース、ハイフン、アポストロフィなどの区切り文字に基づいて、コンポーネントに分

割されます。たとえば、system administrator というキーワードは、{administrator, system} のように、コンポーネントに分割されます。

ただし、元のキーワードの一部は使用できず、ストップワードとして参照されます。英語の場合、こうした無視できるコンポーネントには、a、and、as、by、for、from、in、of、on、or、the、to、および with などがあります。

たとえば、キーワード examination of the data のコンポーネントは {data, examination} のようになり、of および the は無視できるとみなされます。また、コンポーネント セットでは、コンポーネントの順序は意味を持ちません。これにより、次の 3 つのキーワードは同等とすることができます。cough relief for child、child relief from a cough、および relief of child cough。これらはすべて、{child, cough, relief} という同じコンポーネント セットです。キーワードのペアが同等のものとして特定されるごとに、対応するコンセプトを結合して、すべてのキーワードを参照する新しいコンセプトを形成します。

また、キーワードのコンポーネントは活用している場合があるため、言語固有の規則が内部的に適用され、複数形など、活用した変異形にかかわらず、同等のキーワードを特定します。このようにして、活用のない単数形が level であるため、キーワード level of support および support levels を同等のものとして特定することができます。

### 派生関係のコンセプトの語幹の機能

キーワードがコンポーネント化され、活用がなくなった場合（前セクションを参照）、派生関係のコンセプトの語幹アルゴリズムがコンポーネントの末尾、または接尾辞を分析し、コンポーネントの語幹を検索して、そのコンセプトを、同じ、または類似したルートを持つ他のコンセプトとグループ化します。末尾は、テキスト言語に特有の言語派生規則を使用して特定されます。たとえば、接尾辞 ical で終わるコンセプト コンポーネントは、同じ語源を持ち、接尾辞 ic で終わるコンセプトから派生するという、英語のテキストの派生規則があるとします。この規則（および活用の無効化）を使用すると、アルゴリズムはコンセプト epidemiologic study および epidemiological studies をグループ化できます。

キーワードはすでにコンポーネント化され、(in や of などの) 無視できるコンポーネントが特定されているため、派生関係のコンセプトの語幹アルゴリズムは、コンセプト studies in epidemiology を epidemiological studies とグループ化することもできます。

一連のコンポーネント派生関係の規則は、このアルゴリズムでグループ化されるコンセプトの大半が同義語となるように選ばれました。たとえば、epidemiologic studies、epidemiological studies、studies in epidemiology という 3 つのコンセプトはすべて同義のキーワードです。完全性を高くするために、一部の派生規則を使用すると、アルゴリズムによって、状況的に関連するコンセプトをグループ化できます。たと

例えば、アルゴリズムは `empire builder` や `empire building` などのコンセプトをグループ化できます。

## 内包関係のコンセプト

内包関係のコンセプトの手法は、コンセプトを取得し、語彙系列のアルゴリズムを使用してカテゴリを作成し、その他のコンセプトに含まれるコンセプトを特定します。コンセプト内の単語が別のコンセプトの部分集合である場合、規定となるセマンティックの関係を反映します。内包関係のコンセプトは、どんな種類のテキストにも使用できる強力な手法です。

この方法は、セマンティック ネットワークと組み合わせるとうまく動作しますが、個別に使用することもできます。レコードにドメイン固有の用語または専門用語が多く含まれている場合、内包関係のコンセプトを使用するとより良い結果が出ます。これは事前に類義語辞書を使用して、特別なキーワードが適切に抽出・グループ化されるように調整してある場合に、特にいい結果が得られます。

## 内包関係のコンセプトの機能

内包関係のコンセプト アルゴリズムを適用する前に、キーワードはコンポーネント化され、活用がなくなります。詳細は、p. 128 [派生関係のコンセプトの語幹](#) を参照してください。次に、内包関係のコンセプト アルゴリズムはコンポーネント セットを分析します。各コンポーネント セットについて、アルゴリズムは最初のコンポーネント セットの部分集合である別のコンポーネント セットを検索します。

たとえば、コンポーネント セットが `{breakfast, continental}` のコンセプト `continental breakfast` があり、コンポーネント セットが `{breakfast}` のコンセプト `breakfast` がある場合、アルゴリズムは、`continental breakfast` は `breakfast` の一種であると結論付け、これらとともにグループ化します。

より大きな例では、[抽出結果] パネルにコンセプト `seat` が表示され、このアルゴリズムを適用する場合、`safety seat`、`leather seat`、`seat belt`、`seat belt buckle`、`infant seat carrier`、および `car seat laws` のようなコンセプトは該当するカテゴリにグループ化されます。

キーワードはすでにコンポーネント化され、(in や of などの) 無視できるコンポーネントが特定されているため、内包関係のコンセプト アルゴリズムは、コンセプト `advanced spanish course` にコンセプト `course in spanish` が含まれると認識します。

注: それらを明示的に指定して、コンセプトがグループ化されないようにすることができます。詳細は、p. 126 [例外ペアのリンクの管理](#) を参照してください。



## セマンティック ネットワーク

このリリースでは、セマンティック ネットワーク手法は、英語テキストにのみ使用できます。

この手法では、単語の関係の組み込みネットワークを使用してカテゴリを作成します。このため、キーワードが具体的で、あまりあいまいでなければ、この手法を使用すると、非常に良い結果を生成することができます。ただし、この手法が非常に技術的/専門的なコンセプト間に多くのつながりを見つけることを期待することができません。こうしたコンセプトを処理する場合、内包関係のコンセプトおよび派生関係のコンセプトの語幹による手法がより有用な場合があります。

### セマンティック ネットワークの機能

セマンティック ネットワーク手法は、既知の単語の関係を利用して、類義語または下位語のカテゴリを作成します。**下位語**は、1 つのコンセプトがある種の 2 番目のコンセプトである場合、階層の関係性があり、ISA リレーションシップとも呼ばれます。たとえば、animal がコンセプトである場合、動物の種類である cat、kangaroo は animal の下位語となります。

類義語および下位語の関係性のほか、セマンティック ネットワークの手法では、<Location タイプからコンセプト間の部分的なリンクおよび全体のリンクを検証します。たとえば、この手法ではコンセプト normandy、provence、および france を、ノルマンディおよびプロバンスは、フランスの一部であるため、1 つのカテゴリにグループ化します。

セマンティック ネットワークは、セマンティック ネットワークの各コンセプトの考えられる意味を特定することから始めます。コンセプトが類義語または下位語として特定されると、1 つのカテゴリにグループ化されます。たとえば、この手法をつかうと、次の 3 つのコンセプトからなる 1 つのカテゴリを作成されます。**生食用リンゴ**、**デザート**のリンゴ、および**グラニー スミス**。なぜならば、セマンティック ネットワークには次のような情報が含まれるからです。1) **デザート**のリンゴは**生食用リンゴ**の類義語であり、2) **グラニー スミス**は**生食用リンゴ**の一種である（**生食用リンゴ**の下位語という意味で）。

個別にみると、多くのコンセプト、特にユニタームがあいまいです。たとえば、コンセプト buffet は食事の種類、あるいは家具を表す場合があります。一連のコンセプトに meal、furniture、および buffet がある場合、アルゴリズムは meal または furniture のいずれかによる buffet のグループ化を選択するよう強制します。アルゴリズムによる選択は、レコードのコンテキストにおいては適切でない場合があります。

セマンティック ネットワーク手法は、特定の種類のデータによる内包関係のコンセプトにおいて優れています。セマンティック ネットワークと内包関係のコンセプトでは、apple pie が pie の一種であることを認識しますが、tart も pie の一種であることを認識できるのはセマンティック ネットワークだけです。

セマンティック ネットワークは、他の手法を組み合わせることで機能します。たとえば、セマンティック ネットワーク手法と内包関係のコンセプトの手法を選択し、セマンティック ネットワークによりコンセプト teacher をコンセプト tutor とグループ化した (tutor は teacher の一種であるため) とします。内包アルゴリズムはコンセプトをgraduate tutor と tutor にグループ分けし、結果として、2つのアルゴリズムが共同してアウトプット カテゴリを作成します。アウトプットカテゴリには、tutor、graduate tutor、teacherが含まれます。

### セマンティック ネットワークのオプション

この手法では、さまざまな追加設定が重要である場合があります。

- **【最大検索距離】**を変更します。カテゴリ作成前に手法による検索の距離を選択します。値が小さいほど、作成される結果は少なくなります。ただし、これらの結果はノイズが少なく、またリンクや関連性が大きくなります。値が大きいほど、取得する結果は多くなります。ただし、これらの結果の信頼性または関連性が弱くなります。

たとえば、距離に応じて、Danish pastry から coffee roll (上位) まで、そして bun (祖父母) および bread まで上方に検索します。

作成されるカテゴリが大きすぎる、あるいはあまりに多くのものがグループ化されていると感じられる場合は、検索距離を短縮すれば、より小さなカテゴリを作成でき、作業がしやすくなります。

**重要!** 誤ったグループ化と行うと結果に大きな悪影響をおよぼす場合があります。この手法を手法する場合、オプション **語幹文字数が次の最小値以上のときにスペルを調整する** ([抽出] ダイアログ ボックスで定義) を適用せず、Fuzzy Grouping を行うことをお勧めします。

### 共起規則

共起規則を使用すると、レコードのセット内で強い関連を持つコンセプトを見つけ、グループ化することができます。レコードでコンセプトが共に頻繁に見つかる場合、共起はおそらくカテゴリ定義の値のものである基底となる関連を反映します。この手法はカテゴリを作成、カテゴリを拡張するために、または別のカテゴリ手法の入力として使用できる共起規則を作成します。レコードのあるセット内で頻繁に同時に現れ、他のレコードのいずれにも個別にあまり現れない場合、2つのコンセプトは強力に共起します。この手法を使用して、少なくとも数百のレコードを持つ大きなデータセットを使用して良い結果を作成することができます。

たとえば、多くのレコードに単語 price および availability が含まれている場合、これらのコンセプトを共起規則 (price & available) にグループ化できます。また、コンセプト peanut butter、jelly、sandwich

が個別で現れるより頻繁に同時に現れる場合、これらはコンセプト共起規則 (peanut butter&jelly & sandwich) にグループ化されます。

**重要!** 以前のリリースでは、共起規則および類義語規則は、大カッコで囲まれていました。このリリースの場合、大カッコはパターン結果を示します。その代わりに、共起規則や類義語規則は、(スピーカー システム | スピーカー) のように、カッコで囲まれます。

### 共起規則の機能

この手法では、レコードをスキャンし、同時に現れる傾向のある 2 つ以上のコンセプトを探します。レコードのあるセット内で頻繁に同時に現れ、他のレコードのいずれにも個別にあまり現れない場合、2 つ以上のコンセプトは強力に共起します。

共起するコンセプトが見つかった場合、カテゴリ規則が形成されます。これらの規則は、& ブール型演算子を使用してつながっている 2 つ以上のコンセプトで構成されています。これらの規則は、規則内のコンセプトのセットがレコードですべて共起する場合、自動的にレコードをカテゴリに自動的に分類するという、論理ステートメントです。

### 共起規則のオプション

共起規則の手法を使用している場合、作成される規則に影響を与えるいくつかの設定を調整できます。

- **【最大検索距離】** を変更します。カテゴリ作成前に手法による検索の距離を選択します。値が小さいほど、作成される結果は少なくなります-ただし、これらの結果はノイズが少なく、またリンクや関連性が大きくなります。値が大きいほど、取得する結果は多くなります。ただし、これらの結果の信頼性または関連性が弱くなります。共起について作業する際、検索距離のデフォルト値を使用すると非常に多数の共起が作成され、その多くはリンクも薄いため興味を引かないという結果になります。検索距離を短縮すると、弱い共起をフィルタでき、より有意な結果が得られます。
- **レコードの最小数:** 共起の重要性を確認できるようにするため、カテゴリの記述子として使用されるよう、指定の共起を含む必要のあるレコードの最小数を定義します。データ セットが小さい場合、このオプションの設定が低いほど、共起をより見つけやすくなります。

注: それらを明示的に指定して、コンセプトがグループ化されないようにすることができます。詳細は、p. 126 [例外ペアのリンクの管理](#) を参照してください。

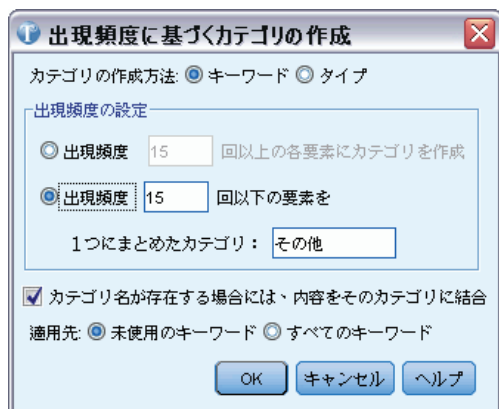
## 頻度の詳細設定

単純で機械的な出現頻度に基づく手法に基づいて、カテゴリを作成できます。この手法を使用して、指定されたレコードまたはの数を超えて見つかった項目（タイプ、コンセプト、またはパターン）ごとに 1 つずつカテゴリを作成できます。また、あまり頻繁に出現しない項目すべてに 1 つカテゴリを作成できます。度数ごとに、テキスト全体の出現数の合計に対し、対象の抽出されたコンセプト（およびその類義語）、タイプ、またはパターンを含むレコードの数を参照します。

一般的または重要な回答を示すため、頻繁に出現する項目をグループ化すると、関心のある結果を生成できます。他の手法が適用されると、この手法が未使用の抽出結果に非常に役立ちます。他にカテゴリが存在しない場合、別のアプリケーションが抽出直後にこの手法を実行し、結果を編集して、関心のないカテゴリを削除し、これらのカテゴリを拡張して、より多くのレコードに一致するようにします。詳細は、p.135 [カテゴリの拡張](#) を参照してください。

この手法を使用する代わりに、[抽出結果] パネルのレコード数の多い順にコンセプトまたはコンセプト パターンを並べ替え。上位のコンセプトまたはコンセプト パターンを [カテゴリ] パネルにドラッグ アンド ドロップして、対応するカテゴリを作成することができます。

図 6-12  
詳細設定:[度数分布表] ダイアログ ボックス



**カテゴリの記述子を生成:** 記述子の入力の種類を選択します。詳細は、p. 118 [カテゴリの作成](#) を参照してください。

- **コンセプトレベル:** このオプションを選択すると、コンセプトまたはコンセプト パターンの頻度が使用されます。タイプがカテゴリ作成の入力として選択されている場合はコンセプトが使用され、タイプ パターンが選択された場合はコンセプト パターンが使用されます。一般的には、この手法をコンセプト レベルに適用すると、コンセプトおよ

びコンセプト パターンがより低いレベルの尺度を示すため、より特定の結果を作成します。

- **タイプレベル** :このオプションを選択すると、タイプまたはタイプ パターンの頻度が使用されます。タイプがカテゴリ作成の入力として選択されている場合はタイプが使用され、タイプ パターンが選択された場合はタイプ パターンが使用されます。この手法をタイプ レベルに適用すると、指定された幅広い回答に関してすばやく表示することができます。

**独自のカテゴリを持つ項目の最小レコード数**:このオプションを使用すると、頻繁に出現する項目からカテゴリを作成できます。出力が、少なくとも X 個のレコードで出現する記述子を含むカテゴリのみに制限されます。X は、このオプションに入力する値を示します。

**すべての残りの項目を次のカテゴリにグループ化**: このオプションを使用すると、あまり頻繁に出現しないすべてのコンセプトまたはタイプを、選択した名前の付いた「キャッチオール」カテゴリにグループ化します。デフォルトでは、カテゴリ名はその他です。

**カテゴリ入力**: 手法を適用するグループを選択します。

- **未使用の抽出結果**:既存のカテゴリで使用されていない抽出結果からカテゴリを作成できます。レコードが、複数のカテゴリと合致する傾向が最も小さくなり、作成されるカテゴリの数が制限されます。
- **すべての抽出結果**:抽出結果のいずれを使用してもカテゴリを作成できます。カテゴリがないまたは少ない場合に最も役立ちます。

**次の方法で重複するカテゴリ名を解決**: 名前が既存のカテゴリと同じ新規カテゴリの処理方法を選択します。新規カテゴリ（およびその記述子）と同じ名前を持つ既存カテゴリとを結合できます。あるいは、既存のカテゴリに重複する名前があった場合、カテゴリの作成をスキップすることもできます。

## カテゴリの拡張

拡張は、記述子を自動的に追加または拡張して、既存のカテゴリを「拡大」するプロセスです。その目的は、本来カテゴリに割り当てられていなかった関連レコードをキャプチャするより良いカテゴリを作成することです。

選択した自動グループ化手法では、既存のカテゴリ記述子に関連するコンセプト、TLA パターン、およびカテゴリ規則を特定しようとします。これらの新しいコンセプト、パターン、カテゴリ規則が新しい記述子として追加されるか、既存の記述子に追加されます。拡張するためのグループ化手法には、「派生関係のコンセプトの語幹」、「内包関係のコンセプト」、「セマンティック ネットワーク」（英語テキストのみ）、および「共起規則」が含まれます。[カテゴリ名から生成された記述子を使用して空白のカテゴリを拡張する]

の手法を使用すると、カテゴリ名の単語を使用して記述子を生成します。そのため、カテゴリ名が記述的であるほど、結果が良いものとなります。

注:カテゴリを拡張する場合、出現頻度に基づく手法は使用できません。

拡張は、カテゴリをインタラクティブに改善する重要な方法です。次に、カテゴリを拡張する場合の例をいくつか示します。

- [カテゴリ] パネルでコンセプト パターンをドラッグ/ドロップしてカテゴリを作成した後
- 手動でカテゴリを作成し、簡単なカテゴリ規則および記述子を追加した後
- 非常に記述的な名前を持つ コード フレームをインポートした後
- プロジェクト作成時に選択した TAP に由来するカテゴリを修正した後

カテゴリを複数回使用できます。たとえば、非常に記述的な名前を持つ事前定義済みカテゴリ ファイルをインポートした場合、[カテゴリ名から生成された記述子を使用して空白のカテゴリを拡張する] オプションを使用して拡張子、記述子の最初のセットを取得して、これらのカテゴリを再度拡張します。ただし、複数回拡張すると、記述子が拡張されて幅広くなると、あまりに一般的なカテゴリが生成される場合があります。作成グループ化手法および拡張グループ化手法では類似した基底のアルゴリズムを使用するため、カテゴリの作成後に直接拡張すると、より関心の高い結果の作成は期待できません。

### ヒント

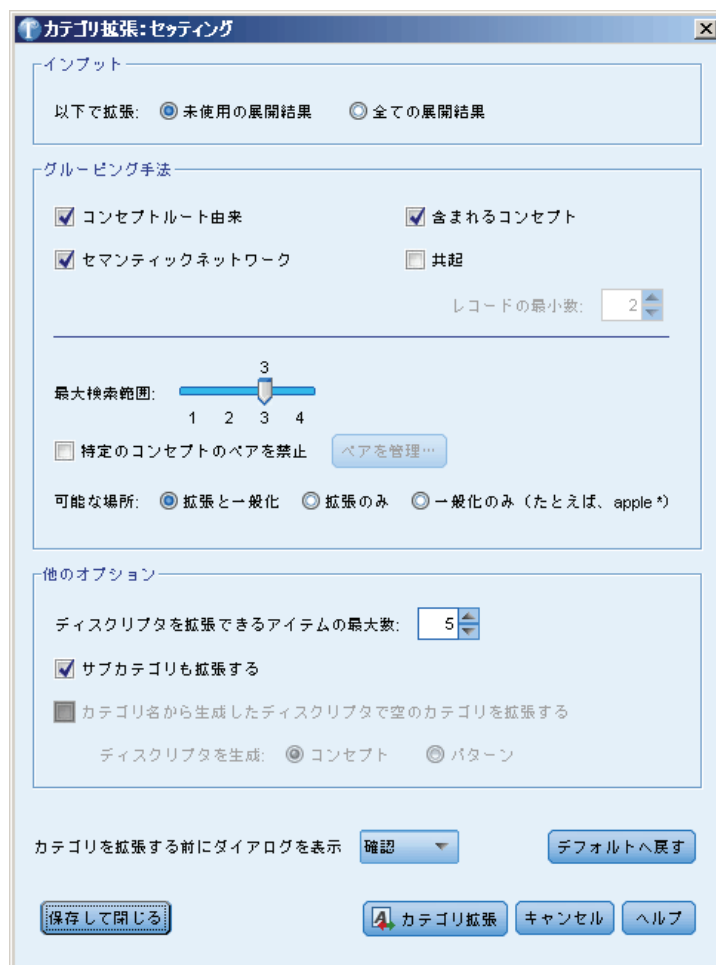
- 拡張を試みるが結果の使用は望まない場合、拡張を行った直後に操作をいつでも取り消す ([編集] > [取り消し]) ことができます。
- プロセス中、規則は個別に作成されるため、ドキュメントの同じセットに正確に一致するカテゴリのカテゴリ規則を 2 つ以上作成します。必要に応じて、カテゴリを確認し、カテゴリ記述子を手動で編集して重複を削除できます。詳細は、 p. 172 [カテゴリ記述子の編集](#) を参照してください。

### カテゴリを展開するには

- ▶ [カテゴリ] パネルで、展開するカテゴリを選択します。
- ▶ メニューから、[カテゴリ] > [カテゴリを展開] を選択します。プロンプトが表示されないようオプションを選択している場合、メッセージ ボックスが表示されます。
- ▶ 今すぐ作成するか、左記に設定を編集するかを選択します。
  - [今すぐ拡張] をクリックすると、現在の設定でカテゴリの拡張が開始されます。プロセスが開始し、進捗状況のダイアログが表示されます。
  - [編集] をクリックして、設定を確認し、変更します。

拡張しようとした後、新しい記述子が見つかったカテゴリには、[カテゴリ] パネルで「展開」という単語のフラグが立てられ、すばやくカテゴリを特定できます。「展開」というテキストは、再度展開するか、別の方法で編集、またはコンテキストメニューを使用してこれらを解除するまで残ったままです。

図 6-13  
[カテゴリを展開] ダイアログ ボックス



カテゴリの作成時または拡張時に使用できるそれぞれの手法は、特定の種類のデータおよび状況に適していますが、同じ分析で手法を組み合わせると役に立つ場合があります。カテゴリにグループ化されたコンセプトおよびタイプは、次にカテゴリを作成する場合も使用できます。つまり、複数のカテゴリのコンセプトを表示したり、重複するカテゴリを見つけることができます。

**カテゴリ入力。**カテゴリの展開に使用する入力を選択します。

- **未使用の抽出結果:** 既存のカテゴリで使用されていない抽出結果からカテゴリを作成できます。レコードが、複数のカテゴリと合致する傾向が最も小さくなり、作成されるカテゴリの数が制限されます。
- **すべての抽出結果:** 抽出結果のいずれを使用してもカテゴリを作成できません。カテゴリがないまたは少ない場合に最も役立ちます。

### グループ化手法

これらの手法の簡単な説明は、「[詳細言語設定](#)」( p.122 ) を参照してください。これらの手法には、次のものが含まれています。

- **派生関係のコンセプトの語幹**(日本語テキストには使用不可)
- **セマンティック ネットワーク** (英語テキストのみ)
- **内包関係のコンセプト**
- **共起 および 最小ドキュメント数のサブオプション**

これらのタイプは関連する結果を作成しないため、多くのタイプがセマンティック ネットワークから永続的に除外します。それらのタイプには、<Positive、<Negative、<IP、その他の非言語的タイプなどがあります。

**最大検索距離:** カテゴリ作成前の手法による検索の距離を選択します。値が小さいほど、取得する結果は少なくなります。ただし、これらの結果はノイズが少なく、またリンクや関連性が大きくなります。値が大きいほど、取得する結果は多くなります。ただし、これらの結果の信頼性または関連性が弱くなります。このオプションはすべての手法にグローバルに適用されますが、共起およびセマンティック ネットワークに対する効果は最も大きくなります。

**特定のコンセプトがグループ化されないようにする:** 出力の 2 つのコンセプトがグループ化またはペアとならないように処理を停止します。コンセプト ペアを作成または管理するには、[\[ペアを管理\]](#) をクリックします。詳細は、p.126 [例外ペアのリンクの管理](#) を参照してください。

**可能な場合、単純に拡張するか、ワイルドカードを使用して記述子を一般化するか、またはその両方を選択します。**

- **拡張および一般化:** このオプションは、選択したカテゴリを拡張し、記述子を一般化します。一般化を選択した場合、アスタリスク (\*) のワイルドカードを使用して、製品が一般的なカテゴリ規則をカテゴリに作成することができるようになります。たとえば、[\[アップル タルト + .\]](#) や [\[アップル ソース + .\]](#) などの複数の記述子を作成する代わりに、[\[アップル \\* + .\]](#) のようにワイルドカードを使用します。ワイルドカードを使用して一般化すると、以前と同じように、ちょうど同じ数のレコードを取得するケースが多くなります。ただし、このオプションには、数の縮小やカテゴリの記述子の簡略化という利点があります。また、このオプションを使用すると、新しいテキスト データ



(例: 長期的/周期的研究) にこれらのカテゴリを使用してより多くのレコードをカテゴリ化する機能を拡大します。

- **展開のみ:** 一般化せずにカテゴリを展開します。手動で作成したカテゴリには [展開のみ] オプションを選択し、[拡張および一般化] オプションを使用して同じカテゴリをもう一度展開すると便利です。
- **一般化のみ:** 別の方法でカテゴリを展開せずに、記述子を一般化します。

### カテゴリを拡張するその他のオプション

適用するグループ化手法を選択するほか、次のように、その他のオプションを編集することができます。

**記述子を拡張する場合の最大項目数:** 項目 (コンセプト、タイプおよびその他の式) で記述子を拡張する場合、単一の記述子に追加できる項目の最大数を定義します。この制限値を 10 に設定すると、最大 10 件の追加項目を既存の記述子に追加できます。10 件を超える項目を追加しようとする場合、10 番目の項目が追加されると、新しい項目の追加を停止します。そうすることにより、記述子のリストが短くなりますが、最も関心の高い項目が最初に使用されたことを保障するものではありません。[可能な場合ワイルドカードを使用して一般化] オプションを使用して、品質を落とすことなく拡張のサイズを縮小することが必要な場合があります。このオプションは、ブール値 & (AND) または ! (NOT) を含む記述子にのみ適用されます。

**サブカテゴリも展開:** 選択したカテゴリ下のサブカテゴリも展開します。

**カテゴリ名に基づき記述子を使用してカテゴリを拡張:** このオプションを選択すると、カテゴリ名を構成する単語に基づいて、各カテゴリの記述子を自動的に作成しようとします。カテゴリ名をスキャンして、名前の単語が抽出されたコンセプトに一致するかどうかを確認します。コンセプトが認識されると、そのコンセプトを使用して、合致するコンセプト パターンを検索し、コンセプトとパターンを使用してカテゴリの記述子を形成します。このオプションを選択すると、カテゴリ名が長く記述的である場合に、最良の結果を作成します。迅速にカテゴリの記述子を生成し、またカテゴリはこれらの記述子を含むレコードをキャプチャすることができます。別の場所からカテゴリをインポートしたり、長く記述的な名前を使用して手動でカテゴリを作成する場合に最も役立つオプションです。記述子が 0 件の、空のカテゴリにのみ適用されます。カテゴリにすでに記述子が含まれている場合、この方法では拡張されません。

**記述子を次の形式で生成:** このオプションは、先行のオプションがオンの場合のみ適用されます。

- **コンセプト:** 入力テキストから抽出されているかどうかに関係なく、記述子をコンセプトの形式で作成します。
- **パターン:** パターンが抽出されているかどうかに関係なく、記述子をパターンの形式で作成します。

## 手作業でのカテゴリの作成

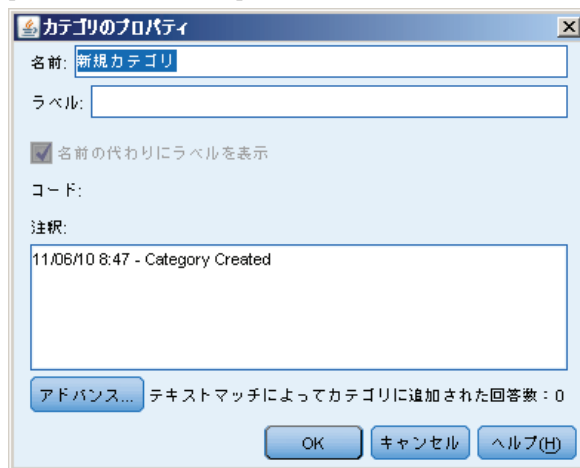
自動カテゴリ作成手法、コード フレーム マネージャおよび条件規則エディタを使用してカテゴリを作成するほか、カテゴリを手動で作成することもできます。手動で作成する方法は、次のとおりです。

- 要素を 1 つずつ追加する空のカテゴリを作成する。詳細は、p. 140 [カテゴリの新規作成または名前の変更](#) を参照してください。
- キーワード、タイプ、パターンを [カテゴリ] ウィンドウにドラッグする。詳細は、p. 141 [ドラッグ&ドロップによるカテゴリの作成](#) を参照してください。

## カテゴリの新規作成または名前の変更

空のカテゴリを作成して、コンセプトおよびタイプをカテゴリに追加できます。カテゴリの名前を変更することもできます。

図 6-14  
[カテゴリのプロパティ] ダイアログ



### 新規の空白カテゴリを作成するには

- ▶ [カテゴリ] パネルに移動します。
- ▶ メニューから、[カテゴリ] > [空白カテゴリを作成] を選択します。ダイアログボックスが開きます。
- ▶ [名前] フィールドでそのカテゴリの名前を入力します。

- ▶ [OK] をクリックすると名前が適用され、ダイアログ ボックスが閉じます。ダイアログ ボックスが閉じ、に新しいカテゴリ名がパネルに表示されます。

これでこのカテゴリに追加していくことができます。詳細は、 p. 171 [記述子のカテゴリへの追加](#) を参照してください。

### カテゴリの名前を変更するには

- ▶ カテゴリを選択して、[カテゴリ] > [カテゴリ名を変更]。ダイアログ ボックスが開きます。
- ▶ [名前] フィールドでそのカテゴリの新しい名前を入力します。
- ▶ [OK] をクリックすると名前が適用され、ダイアログ ボックスが閉じます。ダイアログ ボックスが閉じ、に新しいカテゴリ名がパネルに表示されます。

## ドラッグ & ドロップによるカテゴリの作成

ドラッグ & ドロップは手作業の手法であって、アルゴリズムに基づいたものではありません。次のように、[カテゴリ] ウィンドウでカテゴリを作成できます。

- コンセプト、タイプ、またはパターンを [抽出結果] パネルから [カテゴリ] パネルにドラッグ。
- 抽出されたコンセプトを [データ] パネルから [カテゴリ] パネルにドラッグ。
- 行全体を [データ] パネルから [カテゴリ] パネルにドラッグ。その行に含まれる抽出されたすべてのコンセプトおよびパターンで構成されたカテゴリを作成します。

注: [抽出結果] パネルでは、複数選択をサポートし、複数要素のドラッグ アンド ドロップを可能にします。

**重要!** テキストから抽出されていないコンセプトを [データ] パネルからドラッグ アンド ドロップできません。データで検出したコンセプトの抽出を強制する場合、そのコンセプトをタイプに追加する必要があります。その後、抽出を再度実行します。新しい抽出結果には、追加したばかりのコンセプトが含まれます。その結果をカテゴリに使用できます。詳細は、5 章 p. 95 [コンセプトのタイプへの追加](#) を参照してください。

### ドラッグ & ドロップでカテゴリを作成するには:

- ▶ [抽出結果] パネルまたは [データ] パネルから、1 つまたは複数のコンセプト、パターン、タイプ、レコード、または一部のレコードを選択します。

- ▶ マウス ボタンを押したまま、要素を既存のカテゴリまたはパネルの領域にドラッグして、新しいカテゴリを作成します。
- ▶ 要素をドロップする領域の上で、マウス ボタンを離します。要素が [カテゴリ] パネルに追加されます。変更が加えられたカテゴリは背景色が変わります。この色は、[カテゴリ フィードバック背景] と呼ばれます。詳細は、2 章 p. 19 [オプションの設定](#) を参照してください。

注: 結果として得られたカテゴリには、自動的に名前が付けられます。詳細は、7 章 p. 114 “[カテゴリのプロパティの編集](#)” を参照してください。詳細は、p. 169 [カテゴリのプロパティの編集](#) を参照してください。

どのレコードがカテゴリに割り当てられているかを確認する場合は、[カテゴリ] パネルで該当するカテゴリを選択します。[データ] パネルは自動的に更新され、そのカテゴリのすべてのレコードが表示されます。質問の回答全体のセットを表示するには、カテゴリ ツリー最上部の [すべてのレコード] ノードを選択します。

## 定義済みのインポートおよびエクスポート

独自のカテゴリを Microsoft Excel (\*.xls, \*.xlsx) ファイルに保存している場合、それらを IBM® SPSS® Text Analytics for Surveys にインポートできます。

また、使用中のプロジェクト内のカテゴリを Microsoft Excel (\*.xls, \*.xlsx) ファイルにエクスポートすることもできます。カテゴリをエクスポートすると、記述子やスコアなど、いくつかの追加情報を含めたり除外したりできます。詳細は、p. 152 [カテゴリのエクスポート](#) を参照してください。

重要: SPSS Text Analytics for Surveysバージョン4.0.1以降では、コード フレームの使用のほとんどが定義済みカテゴリと置き換えられています。たとえば、コード フレームのインポート ウィザードは定義済みカテゴリのインポート ウィザードに置き換えられています。しかしながら、この新しいウィザードでは既存のコード フレームをインポートすることができます。さらに、コード フレーム マネージャはサポートされていません。コード値を編集するには、メニューから **表示 > カテゴリコード** を選択して、カテゴリ パネルにコード列を表示し、必要に応じてコードを編集します。

定義済みカテゴリにコードがない場合、または新しいコードが必要な場合、メニューから **[カテゴリ] > [カテゴリを管理] > [コードを自動生成]** を選択して、[カテゴリ] パネルでカテゴリ セットの新しいコードのセットを自動的に生成できます。これにより、既存のコードがすべて削除され、自動的に番号が指定しなおされます。

## 定義済みカテゴリのインポート

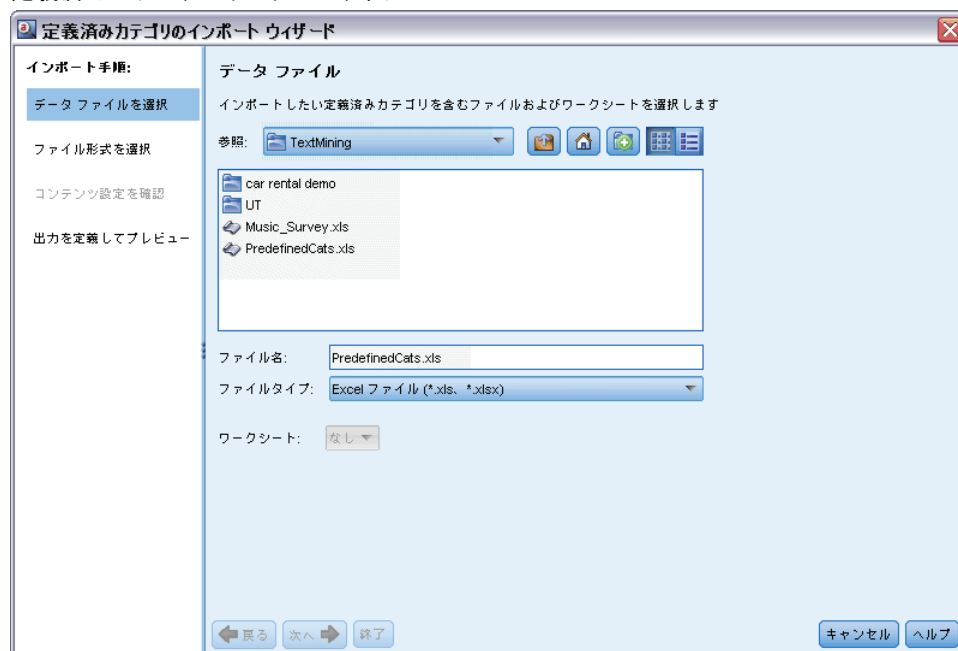
定義済みカテゴリを IBM® SPSS® Text Analytics for Surveys にインポートできます。インポートする前に、定義済みカテゴリ ファイルが Microsoft Excel (\*.xls, \*.xlsx) ファイルであり、サポート可能な形式のいずれかの構造であることを確認してください。形式を自動的に検知するよう選択することもできます。次の形式がサポートされています。

- フラットリスト形式： 詳細は、 p. 147 [フラット リスト形式](#) を参照してください。
- コンパクト形式： 詳細は、 p. 148 [コンパクト形式](#) を参照してください。
- インデント形式： 詳細は、 p. 150 [インデント形式](#) を参照してください。

### 定義済みカテゴリをインポートするには

- ▶ メニューから、[カテゴリ > カテゴリを管理 > 定義済みカテゴリのインポート](#) を選択します。定義済みカテゴリのインポート ウィザードが表示されます。

図 6-15  
定義済みカテゴリのインポート ウィザード

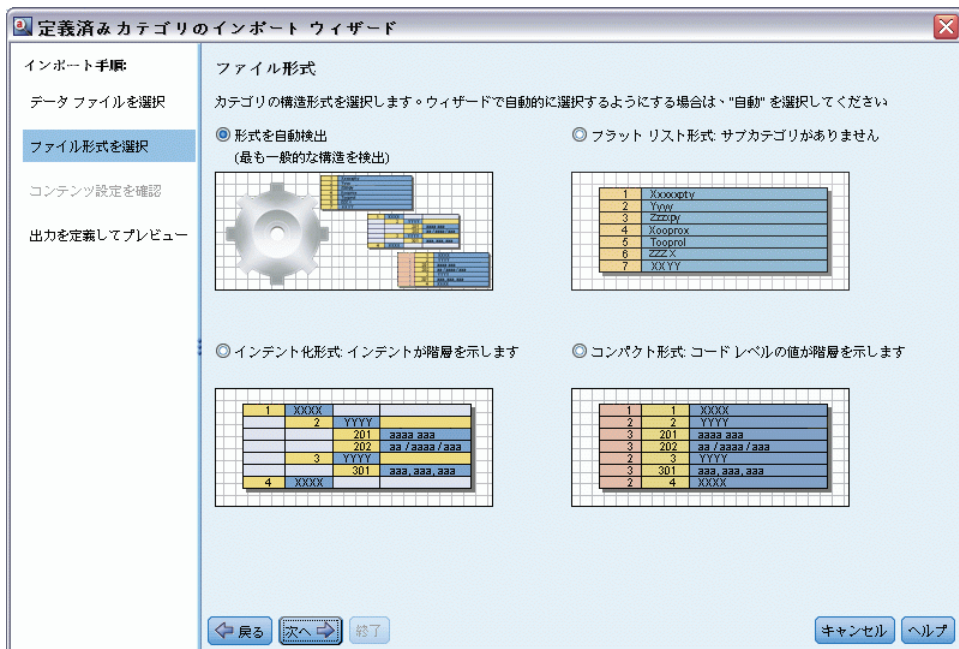


- ▶ [参照] ドロップダウン リストから、ファイルを投入するドライブとフォルダを選択します。
- ▶ リストからファイルを選択します。[ファイル名] テキスト ボックスにファイルの名前が表示されます。

- ▶ リストから、定義済みカテゴリを含むワークシートを選択します。ワークシート名が [ワークシート] フィールドに表示されます。
- ▶ [次へ] をクリックして、データ形式の選択を開始します。

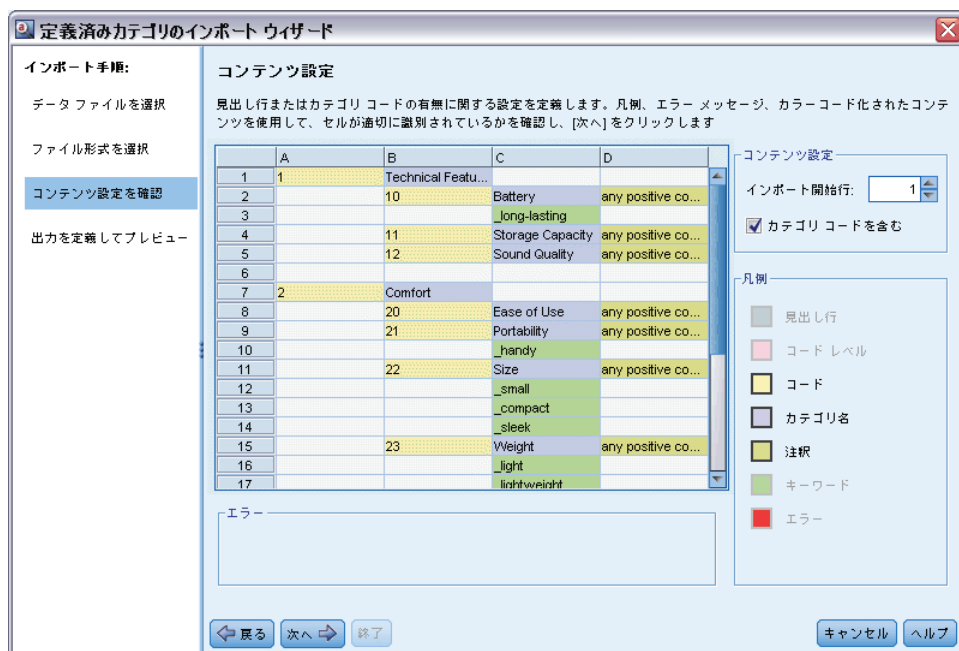
図 6-16

## [定義済みカテゴリのインポート] ダイアログボックス、データ形式のステップ



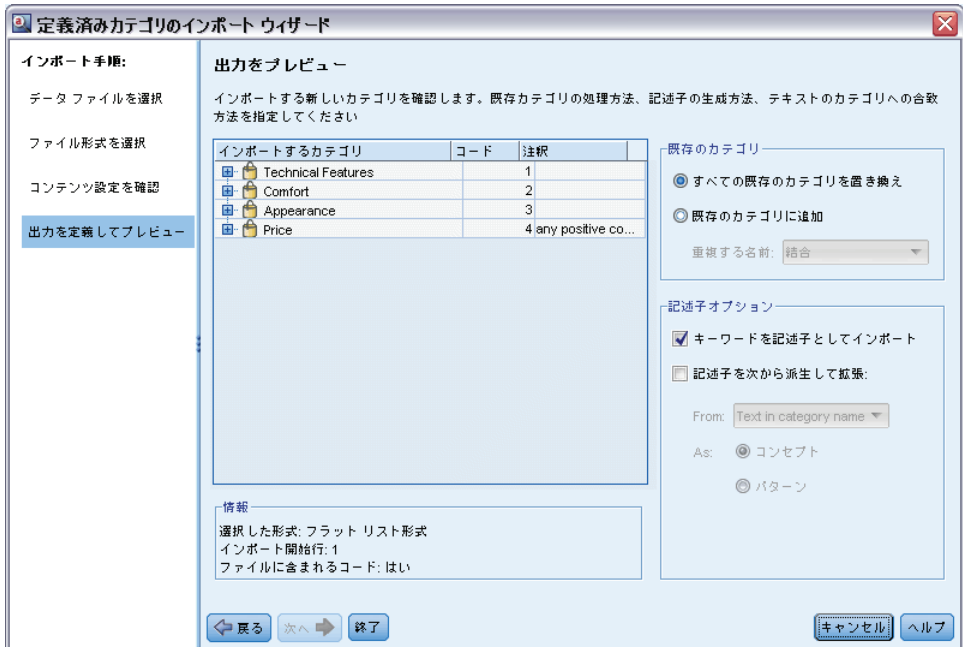
- ▶ ファイルの形式を選択するか、自動的に形式を検知しようとするオプションを選択します。自動検知は、最も一般的な形式に最も適しています。
  - フラットリスト形式: 詳細は、 p. 147 [フラットリスト形式](#) を参照してください。
  - コンパクト形式: 詳細は、 p. 148 [コンパクト形式](#) を参照してください。
  - インデント形式: 詳細は、 p. 150 [インデント形式](#) を参照してください。
- ▶ [次へ] をクリックして、追加のインポート オプションを定義します。形式の自動検知を選択した場合、最後の手順に進みます。

図 6-17  
定義済みカテゴリのインポート、インポート オプションの手順



- ▶ このワークシートの 1 行または複数行が列見出しまたはその他の外部情報である場合、[インポート開始行] オプションでインポート開始場所である行番号を選択します。たとえば、カテゴリ名が行 7 で始まる場合、ファイルを正しくインポートするためには、このオプションで行番号 7 を入力する必要があります。
- ▶ ファイルにカテゴリコードが含まれている場合、オプション [カテゴリコードを含む] を選択します。これにより、ウィザードがデータを正しく認識します。
- ▶ カラーコード化されたセルおよび凡例を確認し、データが正しく特定されるようにします。ファイルで検出されたエラーは赤で表示され、形式プレビュー テーブルの下に表示されます。不正な形式が選択された場合、戻って別の形式を選択します。ファイルを修正する必要がある場合、変更を行い、ファイルをもう一度選択してウィザードを再起動します。ウィザードを終了する前にすべてのエラーを修正してください。
- ▶ [次へ] をクリックして、インポートされる一連のカテゴリおよびサブカテゴリを確認し、これらのカテゴリの記述子の作成方法を定義します。

図 6-18  
[定義済みカテゴリのインポート] ダイアログボックス、プレビューのステップ



- ▶ テーブルでインポートされる一連のカテゴリを確認します。記述子として表示されるはずのキーワードが表示されない場合、インポート時に認識されなかったことが考えられます。正しく接頭辞が使用され、正しいセルに表示されていることを確認してください。
- ▶ プロジェクトの既存のカテゴリの処理方法を選択します。
  - **すべての既存のカテゴリを置き換え:** 既存のカテゴリすべてを削除し、新しくインポートされたカテゴリが代わりに使用されます。
  - **既存のカテゴリに追加カテゴリをインポートし、既存カテゴリと共通カテゴリを結合します。** 既存のカテゴリに追加する場合、重複カテゴリの処理方法を決定する必要があります。オプション [結合] を選択すると、カテゴリ名を共有する場合、インポートされたカテゴリは既存のカテゴリと結合されます。オプション [インポートから除外] は、同じ名前が存在する場合、カテゴリのインポートを禁止します。
- ▶ [キーワードを記述子としてインポート] は、関連するカテゴリの記述子としてデータで特定されるキーワードをインポートするオプションです。
- ▶ [記述子から派生してカテゴリを拡張] は、カテゴリ、またはサブカテゴリの名前を示す単語、および注釈を構成する単語空記述子を生成するオプションです。単語が抽出結果に合致する場合、記述子としてカテゴリに追加されます。このオプションを選択すると、カテゴリ名または注釈が長く記述的である場合に、最良の結果を作成します。迅速にカテゴリの記述



子を生成し、またカテゴリはこれらの記述子を含むレコードをキャプチャすることができます。

- [派生元] フィールドを使用して、記述子が派生するテキスト、名前またはカテゴリおよびサブカテゴリ、注釈内の単語から選択できます。
- [形式を指定] フィールドを使用して、これらの記述子をコンセプトまたは TLA パターンの形式で作成することを選択できます。TLA 抽出が行われない場合、ウィザードの [パターン] オプションが無効となります。

- ▶ [完了] をクリックすると、定義済みカテゴリが [カテゴリ] パネルにインポートされます。

## フラット リスト形式

このフラット リスト形式では、階層のない、上位レベルのカテゴリのみがあります。つまり、サブカテゴリやサブネットはありません。カテゴリ名は 1 つの列に表示されます。

図 6-19

フラット リスト形式の例

	A	B
1	1 Technical Features	
2	_reliable	
3	_durably constructed	
4	10 Technical Features/Battery	any positive comment about long battery life
5	_long-lasting	
6	11 Technical Features/Storage Capacity	any positive comment about the amount that can be stored or memory capacity
7	12 Technical Features/Sound Quality	any positive comment about sound, quality or music quality
8	20 Comfort/Ease of Use	any positive comment indicating that it is convenient, easy and user-friendly
9	21 Comfort/Portability	any positive comment about mobility or indicating that it is handy and easy to transport
10	22 Comfort/Size	any positive comment indicating that it is small or compact
11	23 Comfort/Weight	any positive comment indicating that it is lightweight
12	_light	
13	30 Appearance/Design	any positive comment about appealing style
14	_good-looking	
15	_stylish	
16	_sleek	
17	_well-designed	
18	31 Appearance/Color	any positive comment about color
19	4 Price	any positive comment about cost, such as cheap or affordable
20	50 Usage/Music	any positive comment about listening to songs
21	51 Usage/Video	
22	_movies	
23	52 Usage/Games	
24	53 Usage/Other (radio, internet, photo)	

この形式のファイルには、次の情報が含まれます。

- オプションの**コード**列には、各カテゴリを一意に特定する数値が入力されます。データ ファイルにコードを含まないよう指定 ([コンテンツ設定] で [カテゴリコードを含む] オプションを選択) した場合、カテゴリ名の左隣のセルに各カテゴリの一意のコードを含む列がなければなりません。デー

タにコードが含まれず、後でコードをいくつか作成したい場合、後でいつでもコードを生成できます（カテゴリ>カテゴリを管理>コードを自動生成）。表示>カテゴリコードを選択して、後からコードを編集できます。コードはカテゴリ パネルのコード列に表示され、手動で変更できます。

- 必須の**[カテゴリ名]** 列には、カテゴリのすべての名前が入力されています。この列は、この形式を使用してインポートする場合に必要です。
- カテゴリ名のすぐ右にあるオプションの**[注釈]** セル。この注釈は、カテゴリ/サブカテゴリを説明するテキストで構成されています。
- オプションの**[キーワード]** は、カテゴリの記述子としてインポートできます。認識できるようにするために、これらのキーワードは関連するカテゴリ/サブカテゴリ名のすぐ下にあるセルになければならず、「\_firearms, weapons / guns」のようにキーワードの前にアンダースコア（\_）を追加する必要があります。キーワード セルには、各カテゴリの説明に使用する 1 つまたは複数の単語を入力できます。これらの単語は、ウィザードの最後の手順の指定内容に応じて、記述子としてインポートしたり、無視されます。後で、記述子はテキストから抽出された結果と比較されます。合致が見つかった場合、レコードまたはドキュメントがこの記述子を含むカテゴリにスコアリングされます。

テーブル 6-3  
コード、キーワード、および注釈を含むフラットリスト形式

列 A	列 B	列 C
カテゴリ コード (オプション)	カテゴリ名	注釈
	_記述子/キーワード リスト (オプション)	

### コンパクト形式

コンパクト形式は、階層カテゴリで使用される点を除いて、フラット リスト形式と同じ構造です。そのため、各カテゴリおよびサブカテゴリの階層レベルを定義するには、コード レベル列が必要です。

図 6-20  
Microsoft Excel のコンパクト形式の定義済みカテゴリ ファイルの例

	A	B	C	D
1	1	1 Technical Features		
2		_reliable		
3		_durably constructed		
4	2	10 Technical Features/Battery	any positive comment about long battery life	
5		_long-lasting		
6	2	11 Technical Features/Storage Capacity	any positive comment about the amount that can be stored or memory capacity	
7	2	12 Technical Features/Sound Quality	any positive comment about sound, quality or music quality	
8	1	2 Comfort		
9	2	20 Comfort/Ease of Use	any positive comment indicating that it is convenient, easy and user-friendly	
10	2	21 Comfort/Portability	any positive comment about mobility or indicating that it is handy and easy to transport	
11	2	22 Comfort/Size	any positive comment indicating that it is small or compact	
12	2	23 Comfort/Weight	any positive comment indicating that it is lightweight	
13	1	3 Appearance		
14	2	30 Appearance/Design	any positive comment about appealing style	
15		_good-looking		
16		_stylish		
17		_sleek		
18		_well-designed		
19	2	31 Appearance/Color	any positive comment about color	
20	1	4 Price	any positive comment about cost, such as cheap or affordable	
21	1	5 Usage		
22	2	50 Usage/Music	any positive comment about listening to songs	
23	2	51 Usage/Video		
24		_movies		
25	2	52 Usage/Games		

この形式のファイルには、次の情報が含まれます。

- 必須の **[コード レベル]** 列には、その行の後続の情報の階層の位置を示す番号が入力されます。たとえば、値 1、2、3、が指定され、カテゴリおよびサブカテゴリの両方がある場合、1 はカテゴリ、2 はサブカテゴリ、3 はサブ-サブカテゴリを示します。カテゴリおよびサブカテゴリのみがある場合、1 はカテゴリを、2 はサブカテゴリを示します。カテゴリの深さの限り続きます。
- オプションの **コード** 列には、各カテゴリを一意に特定する値が入力されます。データ ファイルにコードを含まないように指定 ([コンテンツ設定] で **[カテゴリ コードを含む]** オプションを選択) した場合、カテゴリ名の左隣のセルに各カテゴリの一意のコードを含む列がなければなりません。データにコードが含まれず、後でコードをいくつか作成したい場合、後でいつでもコードを生成できます (カテゴリ > カテゴリを管理 > コードを自動生成)。表示 > カテゴリコードを選択して、後からコードを編集できます。コードはカテゴリ パネルのコード列に表示され、手動で変更できます。
- 必須の **[カテゴリ名]** 列には、カテゴリおよびサブカテゴリのすべての名前が入力されています。この列は、この形式を使用してインポートする場合に必要です。

- カテゴリ名のすぐ右にあるオプションの【注釈】セル。この注釈は、カテゴリ/サブカテゴリを説明するテキストで構成されています。
- オプションの【キーワード】は、カテゴリの記述子としてインポートできます。認識できるようにするために、これらのキーワードは関連するカテゴリ/サブカテゴリ名のすぐ下にあるセルになければならず、「\_firearms, weapons / guns」のようにキーワードの前にアンダースコア ( \_ ) を追加する必要があります。キーワードセルには、各カテゴリの説明に使用する 1 つまたは複数の単語を入力できます。これらの単語は、ウィザードの最後の手順の指定内容に応じて、記述子としてインポートしたり、無視されます。後で、記述子はテキストから抽出された結果と比較されます。合致が見つかった場合、レコードまたはドキュメントがこの記述子を含むカテゴリにスコアリングされます。

テーブル 6-4  
コードを含むコンパクト形式の例

列 A	列 B	列 C
階層コード レベル	カテゴリ コード (オプション)	カテゴリ名
階層コード レベル	サブカテゴリ コード (オプション)	サブカテゴリ名

テーブル 6-5  
コードを含まないコンパクト形式の例

列 A	列 B
階層コード レベル	カテゴリ名
階層コード レベル	サブカテゴリ名

## インデント形式

インデント ファイル形式の場合、コンテンツには階層があります。つまり、カテゴリと、1 レベルまたは複数レベルのサブカテゴリがあります。さらに、構造がこの階層を示すよう、インデント化されています。ファイルの各行にはカテゴリまたはサブカテゴリが含まれますが、サブカテゴリはカテゴリからインデント化され、サブ-サブカテゴリはサブカテゴリからインデント化されています。この構造を Microsoft Excel で手動で作成したり、別の製品からエクスポートし Microsoft Excel にインポートした構造を使用することもできます。

図 6-21  
Microsoft Excel のインデント構造のカテゴリの例

列 A	列 B	列 C
1 Technical Features		
2 _reliable		
3 _durably constructed		
4	10 Battery	any positive comment about long battery life
5	_long-lasting	
6	11 Storage Capacity	any positive comment about the amount that can be stored or memory capacity
7	12 Sound Quality	any positive comment about sound, quality or music quality
8		
9	2 Comfort	
10	20 Ease of Use	any positive comment indicating that it is convenient, easy and user-friendly
11	21 Portability	any positive comment about mobility or indicating that it is handy and easy to transport
12	22 Size	any positive comment indicating that it is small or compact
13	23 Weight	any positive comment indicating that it is lightweight
14	_light	
15	3 Appearance	
16	30 Design	any positive comment about appealing style
17	_good-looking	
18	_stylish	
19	_sleek	
20	_well-designed	
21	31 Color	any positive comment about color
22	4 Price	any positive comment about cost, such as cheap or affordable
23	5 Usage	
24	50 Music	any positive comment about listening to songs
25	51 Video	

- **最上位レベルのカテゴリコードおよびカテゴリ名**は、それぞれ列 A および列 B に表示されます。またはコードがない場合、カテゴリ名が列 A に表示されます。
- **サブカテゴリコードおよびサブカテゴリ名**は、それぞれ列 B および列 C に表示されます。またはコードがない場合、サブカテゴリ名が列 B に表示されます。サブカテゴリはカテゴリのメンバーです。上位レベルのカテゴリがない場合、サブカテゴリはありません。

テーブル 6-6  
コードを含むインデント構造

列 A	列 B	列 C	列 D
カテゴリコード (オプション)	カテゴリ名		
	サブカテゴリコード (オプション)	サブカテゴリ名	
		サブ-サブカテゴリコード (オプション)	サブ-サブカテゴリ名

テーブル 6-7  
コードのないインデント構造

列 A	列 B	列 C
カテゴリ名		
	サブカテゴリ名	
		サブ-サブカテゴリ名

この形式のファイルには、次の情報が含まれます。

- オプションの**コード**は、各カテゴリまたはサブカテゴリを一意に特定する数値でなければなりません。データ ファイルにコードを含まないよう指定（[コンテンツ設定] で [カテゴリコードを含む] オプションを選択）した場合、カテゴリ名の左隣のセルに各カテゴリまたはサブカテゴリの一意のコードがなければなりません。データにコードが含まれず、後でコードをいくつか作成したい場合、後でいつでもコードを生成できます（カテゴリ > カテゴリを管理 > コードを自動生成）。表示 > カテゴリコードを選択して、後からコードを編集できます。コードはカテゴリ パネルのコード列に表示され、手動で変更できます。
- 各カテゴリおよびサブカテゴリの必須の**[名前]**。サブカテゴリは、カテゴリから、各行の右側に 1 セルインデントされていなければなりません。
- カテゴリ名のすぐ右にあるオプションの**[注釈]**セル。この注釈は、カテゴリ/サブカテゴリを説明するテキストで構成されています。
- オプションの**[キーワード]**は、カテゴリの記述子としてインポートできます。認識できるようにするために、これらのキーワードは関連するカテゴリ/サブカテゴリ名のすぐ下にあるセルになければならず、「\_firearms, weapons / guns」のようにキーワードの前にアンダースコア（\_）を追加する必要があります。キーワードセルには、各カテゴリの説明に使用する 1 つまたは複数の単語を入力できます。これらの単語は、ウィザードの最後の手順の指定内容に応じて、記述子としてインポートしたり、無視されます。後で、記述子はテキストから抽出された結果と比較されます。合致が見つかった場合、レコードまたはドキュメントがこの記述子を含むカテゴリにスコアリングされます。

**重要!** 1 つのレベルでコードを使用する場合、各カテゴリおよびサブカテゴリにコードを含む必要があります。そうでない場合、インポート プロセスが失敗します。

## カテゴリのエクスポート

また、使用中のプロジェクト内のカテゴリを Microsoft Excel (\*.xls, \*.xlsx) ファイル形式にエクスポートすることもできます。エクスポートされるデータは、大部分がカテゴリ プロパティの [カテゴリ] パネルの現

在のコンテンツのデータです。そのため、ドキュメント スコア値もエクスポートする場合は、もう一度スコアリングを行うことをお勧めします。

#### 常にエクスポート...

- ある場合はカテゴリ コード
- カテゴリ (およびサブカテゴリ) 名
- ある場合はコード レベル (フラット/コンパクト形式)
- 列見出し (フラット/コンパクト形式)

#### オプションでエクスポート...

- ドキュメント スコア
- カテゴリの注釈
- 記述子名
- 記述子数

**重要!** 記述子をエクスポートする場合、それらはテキスト文字列とアンダースコアの接頭辞に変換されます。この製品に再度インポートする場合、パターン、条件規則、単純なコンセプトである記述子と区別する機能が失われます。これらのカテゴリの本製品で再利用する場合、テキスト分析パッケージ (TAP) を使用して、現在定義されているすべての記述子 (使用されているすべてのカテゴリ、コード、言語リソース) を保持することをお勧めします。TAP ファイルは IBM® SPSS® Text Analytics と IBM® SPSS® Text Analytics for Surveys の両方で使用できます。詳細は、3 章 p.44 [テキスト分析パッケージの使用](#) を参照してください。

#### 定義済みカテゴリをエクスポートするには

- ▶ メニューから、[カテゴリ > カテゴリを管理 > カテゴリのエクスポート](#) を選択します。カテゴリのエクスポート ウィザードが表示されます。

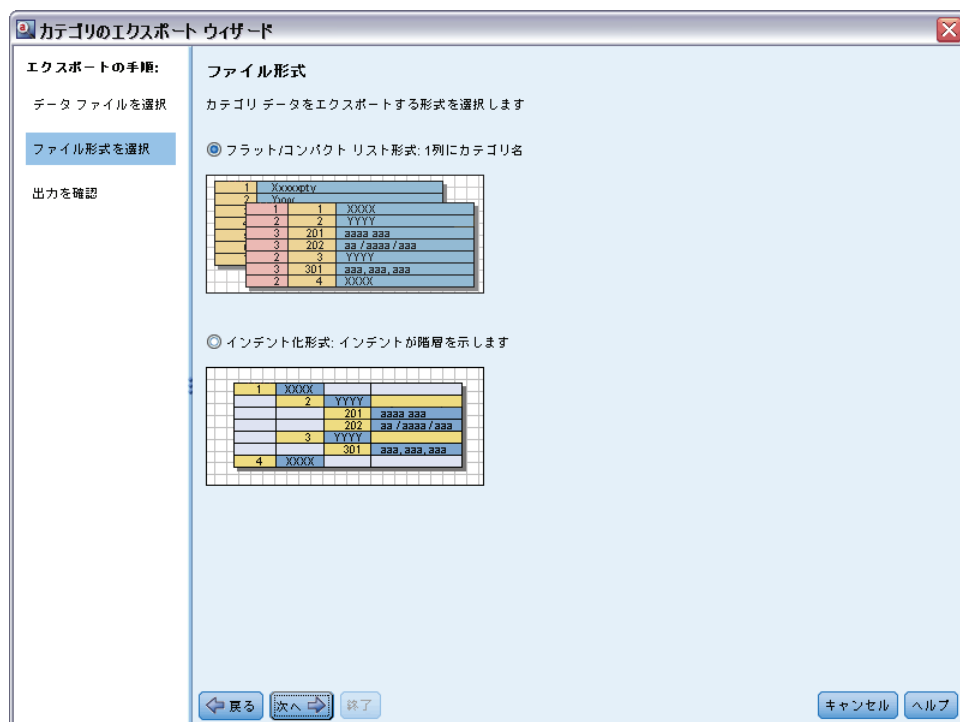
図 6-22  
カテゴリのエクスポート ウィザード、手順 1



- ▶ 場所を選択して、エクスポートするファイルの名前を入力します。
- ▶ [ファイル名] テキスト ボックスに出力ファイルの名前を入力します。
- ▶ [次へ] をクリックして、カテゴリ データをエクスポートする形式を選択します。



図 6-23  
カテゴリのエクスポート ウィザード、手順 2



- ▶ 次のいずれかの形式を選択します。
  - フラット リスト形式またはコンパクト リスト形式: 詳細は、 p. 147 [フラット リスト形式](#) を参照してください。フラット リスト形式には、サブカテゴリはありません。詳細は、 p. 148 [コンパクト形式](#) を参照してください。コンパクト リスト形式には階層カテゴリが含まれています。
  - インデント形式: 詳細は、 p. 150 [インデント形式](#) を参照してください。
- ▶ [次へ] をクリックして、エクスポートするコンテンツを選択し、提案されたデータを確認します。

図 6-24  
カテゴリのエクスポート ウィザード、手順 3



- ▶ エクスポート ファイルの内容を確認します。
- ▶ 注釈または記述子名など、エクスポートする追加内容を選択または選択解除します。
- ▶ [完了] をクリックすると、カテゴリがエクスポートされます。

## カテゴリ規則の使用

カテゴリを作成する方法はたくさんあります。それらの方法の 1 つには、キーワードを表すカテゴリ規則を定義することがあります。カテゴリ規則とは、抽出したコンセプト、タイプ、パターン、およびブール型演算子を使用した論理式に基づいて、レコードを自動的に分類するステートメントです。たとえば、「このカテゴリに、アルゼンチンではなく、大使館という抽出したコンセプトを含む」という内容を意味する式を作成することができます。

共起関係およびコンセプトの語幹 (カテゴリ> 設定を行う> 詳細設定: 言語) のような、グループ化手法を使用したカテゴリ作成時には、カテゴリ ルールの一部は自動的に作成されますが、データおよびコンテキストに対するカテゴリの理解に基づき、ール エディタで手動でカテゴリ ルールを作成する

こともできます。各規則は単一のカテゴリに関連付けられるため、規則を満たすレコードがそのカテゴリにスコアリングされます。

カテゴリ規則により、より大きな特異性を持つ回答をカテゴリ化できることで、テキストマイニングの結果およびより高度な数量分析の品質および生産性を向上させます。経験およびビジネス情報により、データおよびコンテキストに対する特定の理解を与える場合があります。このように理解することによって、情報をカテゴリに変換し、ブール型ロジックと抽出した要素を結合して、レコードをより効率的かつ正確にカテゴリ化できます。

これらの規則を作成する機能により、ビジネス情報を製品の抽出テクノロジーに重ねて、コード化の精度、効率性および生産性を拡張することができます。

注: 条件規則がどのようにテキストに合致するかについての例は、「[カテゴリ規則の例](#)」( p. 164 ) を参照してください。

## カテゴリ規則シNTAX

共起関係およびコンセプトの語幹 ( [カテゴリ > 設定を行う > 詳細設定: 言語](#) ) のような、グループ化手法を使用したカテゴリ作成時には、カテゴリ ルールの一部は自動的に作成されますが、データおよびコンテキストに対するカテゴリの理解に基づき、ルール エディタで手動でカテゴリ ルールを作成することもできます。各規則は単一のカテゴリの説明であるため、規則を満たすレコードが自動的にそのカテゴリにスコアリングされます。

注: 条件規則がどのようにテキストに合致するかについての例は、「[カテゴリ規則の例](#)」( p. 164 ) を参照してください。

規則を作成または編集する場合、規則を条件規則エディタで開く必要があります。コンセプト、タイプ、またはパターンを追加し、またワイルドカードを使用して一致を拡張することができます。抽出されたコンセプト、タイプ、パターンを使用すると、すべての関連コンセプトを検出することができます。

**重要!** 一般的なエラーを回避するために、コンセプトを [抽出結果] パネル、[テキスト リンク分析] パネルまたは [データ] パネルから直接条件規則エディタにドラッグアンドドロップしたり、使用できる場合はコンテキスト メニューを使用して追加することを推奨します。

コンセプト、タイプ、パターンが認識されると、テキストの隣にアイコンが表示されます。



抽出コンセプト



抽出タイプ



抽出パターン

## 条件規則シンタックスおよび演算子

次の表には、条件規則シンタックスを定義する文字を示しています。これらの文字をコンセプト、タイプ、パターンと共に使用して、規則を作成します。

テーブル 6-8  
サポートされるシンタックス

文字	説明
&	「and」ブール型演算子。たとえば、次のような $a \ \& \ b$ には、 $a$ および $b$ が含まれます。 <ul style="list-style-type: none"> <li>- 侵略 &amp; アメリカ合衆国</li> <li>- 2016 &amp; オリンピック</li> <li>- 良い &amp; リンゴ</li> </ul>
	「or」ブール型演算子は包含的演算子で、要素の一部またはすべてが見つかった場合に一致することを意味します。たとえば、次のような $a \   \ b$ には、 $a$ または $b$ が含まれます。 <ul style="list-style-type: none"> <li>- 攻撃   フランス</li> <li>- コンドミニウム   アパート</li> </ul>
!()	「not」ブール型演算子。たとえば、次のような $!(a)$ には、 $a$ が含まれません。 $!(\text{良い} \ \& \ \text{ホテル})$
*	使用方法に応じて、1 文字から単語全体にいたるまでの文字を示すワイルドカード 詳細は、p. 161 <a href="#">カテゴリ規則におけるワイルドカードの使用</a> を参照してください。
()	式の区切り文字。カッコ内の式を最初に評価します。
+	順序特有のパターンを形成するために使用するパターン コネクタ。この演算子がある場合は、大カッコを使用する必要があります 詳細は、p. 159 <a href="#">カテゴリ規則内の TLA パターンの使用</a> を参照してください。
[]	カテゴリ規則内部の抽出した TLA パターンに基づいて合致を検出する場合、パターン区切り文字が必要です。ブラケット内の内容は TLA パターンを参照し、単純な共起に基づいてコンセプトまたはタイプに合致しません。この TLA パターンを抽出していない場合、合致は発生しません。詳細は、p. 159 <a href="#">カテゴリ規則内の TLA パターンの使用</a> を参照してください。パターンではなくコンセプトとタイプの合致に着目している場合は、角カッコを使用しないでください。 注: 古いバージョンの場合、カテゴリ作成手法で生成された共起規則や類義語規則は、大カッコに囲まれていました。すべての新しいバージョンの場合、大カッコは TLA パターンの存在を示します。その代わりに、共起規則による手法や類義語規則による手法で作成された規則は、(スピーカー システム スピーカー) のように、カッコで囲まれます。

& 演算子および | 演算子は、 $a \ \& \ b = b \ \& \ a$  および  $a \ | \ b = b \ | \ a$  のように相互的です。

## バックスラッシュによる文字のエスケープ

シンタックス文字でもある文字がコンセプトに含まれている場合、その文字の前にバックスラッシュを追加して、規則が正しく解釈されるようにする必要があります。バックスラッシュ (¥) 文字を使用して、特別な意味を

持つ文字をエスケープします。エディタにドラッグ アンド ドロップすると、自動的にバックslashが追加されます。

次の条件規則シンタックス文字を条件規則シンタックスとしてでなく、そのまま扱う場合は、その文字の前にバックslashを追加する必要があります。

`&!|+<>()[]*`

たとえば、コンセプト `r&d` に「and」演算子 (&) が使用されているため、条件規則エディタで入力する場合は、`r\&d` となるようにバックslashが必要です。

## カテゴリ規則内の TLA パターンの使用

テキスト リンク分析パターンを、カテゴリ規則で明示的に定義し、より具体的で文脈上の結果を取得することができます。カテゴリ規則でパターンを定義すると、より単純な抽出結果を省略し、抽出されたテキスト リンク分析パターン結果に基づいたドキュメントおよびレコードのみを合致させます。

**大カッコで区切る:** カテゴリ規則ではなく TLA パターンを使用している場合、TLA パターンを大カッコ [ ] で囲む必要があります。抽出した TLA パターンに基づいて合致を検出する場合、パターン区切り文字が必要です。カテゴリ規則には、タイプ、コンセプト、またはパターンが含まれるため、カッコはカッコ内の内容が抽出された TLA パターンを参照するということを明確にします。この TLA パターンを抽出していない場合、合致は発生しません。[カテゴリ] パネルにリンゴ + 良い のようにカッコのないパターンがあった場合、パターンがカテゴリ規則エディタ外部でカテゴリに直接追加されたことを示します。たとえば、コンセプト パターンを抽出結果パネルからカテゴリに直接追加する場合、大カッコは表示されません。ただし、規則内でパターンを使用する場合、[バナナ + !(良い)] のようにカテゴリ規則内の大カッコで囲む必要があります。

**パターンで + 記号を使用:** IBM® SPSS® Text Analytics for Surveys では、2 つの部分に分かれたパターンを作成できます。順序が重要であることを示す場合、+ 記号を使用して、要素を接続します。順序が重要でない場合、代わりに & ブール型演算子を使用できます。broccoli is better” と “it’s better than broccoli” という 2 つのテキストでは、better という単語の位置が重要です。

たとえば、次のような日本語のテキストがあるとします。「私はパイナップルが好きだ」と「私はパイナップルが嫌いだが、イチゴが好きだ」。like & pineapple という式はこの両方のテキストに一致します。それはコンセプト式であり、テキスト リンク規則ではないからです。

(角カッコでくくられていない)。式 pineapple + like は “I like pineapple” とのみ一致します。なぜならば、2 番目のテキストでは、like という語は今度は strawberries 関連づけられるからです。

**パターンによりグループ化:** 独自のパターンを使用して規則を簡略化することができます。3 つの式、cayenne peppers + like、chili peppers + like、および peppers + like の式をキャプチャするとします。これらを、[\* peppers & like] のようにして、単一のカテゴリ規則にグループ化できます。hot peppers + good というもう 1 つの式がある場合、これら 4 つの式を [\* peppers + <Positive>] のような規則でグループ化することができます。

**パターン内の順序:** 出力をよりよく構成するために、製品とともにインストールされたテンプレートに提供されているテキストリンク分析規則が、文の語順に関係なく、同じ順序で基本パターンを出力しようとします。たとえば、テキスト「Good presentations.」を含むレコードおよび「the presentations were good」を含む別のレコードがある場合、いずれのテキストも同じ規則で合致し、出力の順序は、presentation + good や good + presentation ではなく、コンセプトパターン規則の presentation + good と同じ順序になります。また、例のパターンのような 2 スロットパターンで、意見ライブラリのタイプに割り当てられたコンセプトは、デフォルトでは apple + bad のように出力の最後に表示されます。

テーブル 6-9  
パターン シンタックスおよびブール型演算子の使用

式	レコードが一致する条件
[ ]	任意の TLA パターンを含む。抽出した TLA パターンに基づいて合致を検出する場合、カテゴリ規則内でパターン区切り文字が必要です。ブラケット内の内容は、単なるコンセプトおよびタイプではなく TLA パターンを参照します。この TLA パターンを抽出していない場合、合致は発生しません。そのため、パターンを含んでいない規則を作成する場合、!( [ ] ) を使用することができます。
[a]	パターン内の位置に関係なく、少なくとも 1 つの要素が a であるパターンを含む。たとえば、[deal] は、[deal + good] または [deal + .] に一致します。
[a + b]	コンセプト パターンを含む。たとえば、[deal + good] となります。注: 他の要素を追加せずにこのパターンをキャプチャする場合、パターンを使用して規則を作成するのではなく、パターンを直接カテゴリに追加することをお勧めします。
[<A> + <B>]	最初のスロットがタイプ <A>、2 番目のスロットがタイプ <B> のパターンを含み、ちょうど 2 つのパターンがある。+ 記号は、一致する要素の順序が重要であることを示します。たとえば、[<Budget> + <Negative>] となります。注: 他の要素を追加せずにこのパターンをキャプチャする場合、パターンを使用して規則を作成するのではなく、パターンを直接カテゴリに追加することをお勧めします。

式	レコードが一致する条件
[<A> & <B>]	タイプ <A> および <B> のタイプ パターンを含む。たとえば、[<Budget> & <Negative>] となります。この TLA パターンは抽出されませんが、そのように記述された場合、[<Budget> + <Negative>]   [<Negative> + <Budget>] のようになります。+ 記号は、一致する要素の順序が重要であることを示します。
[a + .]	a が唯一のコンセプトであるパターンを含み、そのパターンの他のスロットには何もない。たとえば、[deal + .] は、唯一の出力がコンセプト deal であるコンセプト パターンに一致します。コンセプト deal をカテゴリ記述子として追加した場合、deal を含むすべてのレコードを、deal に関して肯定的な記述を含むコンセプトとして取得します。ただし、[deal + .] を使用すると、deal を示すこれらのレコード パターン結果のみに合致し、他の関係性または意見は deal + fantastic と合致しません。 注：他の要素を追加せずにこのパターンをキャプチャする場合、パターンを使用して規則を作成するのではなく、パターンを直接カテゴリに追加することをお勧めします。
[<A> + <>]	<A> が唯一のタイプであるパターンが含まれている。たとえば、[<Budget> + <>] は、唯一の出力がタイプ <Budget> のコンセプトであるパターンに一致します。 注：[price + <>] ではなく、[<Budget> + <>] のように、タイプ パターンでパターンの + 記号の後に使用する場合にのみ、<> を使用して空のタイプを示すことができます。 注：他の要素を追加せずにこのパターンをキャプチャする場合、パターンを使用して規則を作成するのではなく、パターンを直接カテゴリに追加することをお勧めします。
[a + !(b)]	コンセプト a を含み、コンセプト b を含まないパターンを少なくとも 1 つ含む。少なくとも 1 つのパターンを含む必要があります。 たとえば、[price + !(high)] またはタイプの場合は [!(<Fruit> <Vegetable>) + <Positive>] となります。
!([<A> & <B>])	特定のパターンを含まない。たとえば、!([<Budget> & <Negative>]) となります。

注：条件規則がどのようにテキストに合致するかについての例は、「[カテゴリ規則の例](#)」（p. 164）を参照してください。

## カテゴリ規則におけるワイルドカードの使用

ワイルドカードを規則のコンセプトに追加して、マッチング機能を拡張することができます。アスタリスク \* ワイルドカードを単語の前および/または後に追加して、コンセプトがどのように一致するかを指定できます。ワイルドカードの使用方法には次の 2 種類があります。

- **ワイルドカードの接頭(尾)辞**: ワイルドカードが接頭辞または接尾辞として使用されます。文字列とアスタリスクの間にスペースはありません。たとえば、operat\* は、operat、operate、operates、operations、operational などと一致します。
- **単語ワイルドカード**: ワイルドカードがコンセプトの前または後ろに使用されます。文字列とアスタリスクの間にスペースがあります。たとえば、\* operation は、operation、surgical operation、post operation などと一致します。たとえば、\* operat\* \* のように、ワイルドカードが接頭(尾)辞とともに単語ワイルドカードを使用できます。この場合、operation、surgical operation、telephone operator、operatic aria などと一致します。最後の例のように、範囲があまりに幅広くなったり、不要なマッチをキャプチャしないよう、ワイルドカードを注意して使用することをお勧めします。

#### 例外:

- ワイルドカードは、単独で使用できません。たとえば、(apple | \* ) は無効です。
- ワイルドカードをタイプ名の一致に使用することはできません。<Negative\*> は、タイプ名に合致しません。
- 特定のタイプをワイルドカードで検索されたコンセプトに対する合致から除外することはできません。コンセプトが割り当てられるタイプは自動的に使用されます。
- ワイルドカードは、語の終わりであっても初めであっても (open\* account) 、または独立した要素であっても (open \* account) 語の連鎖の途中に使用することはできません。タイプ名にワイルドカードを使用することはできません。たとえば、apple\* recipe など、word\* word は「applesauce recipe」や他の言葉にも合致しません。ただし、apple\* \* は、applesauce recipe、apple pie、apple などの言葉に合致します。また、apple \* toast など、word\* word は 2 つの語の間にアスタリスクを使用しているため、apple cinnamon toast という語には合致しません。ただし、apple\* \* は、apple cinnamon toast、apple、apple pie などに合致します。

テーブル 6-10  
ワイルドカードの使用方法

式	レコードが一致する条件
*apple	文字で終了し、接頭辞として任意の数の文字を使用しているコンセプトを含む。例:*apple は、apple で終了し、次のように接頭辞を使用します。 - apple - pineapple - crabapple
apple*	文字で開始し、接尾辞として任意の数の文字を使用しているコンセプトを含む。例:*apple は、apple で開始し、次のように接尾辞を使用、または使用しません。 - apple



式	レコードが一致する条件
	<ul style="list-style-type: none"> <li>- applesauce</li> <li>- applejack</li> </ul> <p>たとえば、apple* &amp; !(pear*   quince) には、文字 apple で始まるコンセプトが含まれますが、文字 pear またはコンセプト quince で始まるコンセプトは含まれません。そのため、次のコンセプトとは一致しません。apple &amp; quince 次のコンセプトとは一致します。</p> <ul style="list-style-type: none"> <li>- applesauce</li> <li>- apple &amp; orange</li> </ul>
*product*	<p><b>product</b> という文字を含み、接頭辞または接尾辞、または両方として任意の数の文字を使用しているコンセプトを含む。 例:*product* は、次のコンセプトと一致します。</p> <ul style="list-style-type: none"> <li>- product</li> <li>- byproduct</li> <li>- unproductive</li> </ul>
* loan	<p>単語 loan を含み、単語の前に別の単語と組み合わせる場合があるコンセプトを含む。たとえば、* loan は次のコンセプトと一致します。</p> <ul style="list-style-type: none"> <li>- loan</li> <li>- car loan</li> <li>- home equity loan</li> </ul> <p>たとえば、[* delivery + &lt;Negative&gt;] は、前半が単語 delivery で終わり、後半にタイプ &lt;Negative&gt; を含むコンセプトを含み、次のコンセプト パターンと一致します。</p> <ul style="list-style-type: none"> <li>- package delivery + slow</li> <li>- overnight delivery + late</li> </ul>
event *	<p>単語 event を含み、単語の後に別の単語が続く場合があるコンセプトを含む。たとえば、event * は次のコンセプトと一致します。</p> <ul style="list-style-type: none"> <li>- event</li> <li>- event location</li> <li>- event planning committee</li> </ul>
* apple *	<p>任意の単語で始まり、次に apple で始まる単語が続き、別の単語が続く場合があるコンセプトを含む。* は 0 または n を意味するため、apple にも合致します。たとえば、* apple* は次のコンセプトと一致します。</p> <ul style="list-style-type: none"> <li>- gala applesauce</li> <li>- granny smith apple crumble</li> <li>- famous apple pie</li> <li>- apple</li> </ul> <p>たとえば、[* reservation* * + &lt;Positive&gt;] は、前半が単語 reservation のあるコンセプトを含み（コンセプト内の場所は関係ない）、後半にタイプ &lt;Positive&gt; を含み、次のコンセプトパターンと一致します。</p> <ul style="list-style-type: none"> <li>- reservation system + good</li> <li>- online reservation + good</li> </ul>

注: 条件規則がどのようにテキストに合致するかについての例は、「カテゴリ規則の例」 ( p.164 ) を参照してください。

## カテゴリ規則の例

それらを表現するために使用するシンタックスに基づき、規則がレコードにどのように合致するかを示すために、次の例について考えてみます。

### 例のレコード

次の 2 つのレコードがあるとします。

- レコード A: “when I checked my wallet, I saw I was missing 5 dollars.”
- レコード B: “\$5 was found at the picnic area, but the blanket was missing.”

次の 2 つの表には、コンセプト パターンおよびタイプ パターンのほかにコンセプトおよびタイプに抽出される内容を示しています。

### 例から抽出されるコンセプトとタイプ

テーブル 6-11  
コンセプトとタイプの抽出例

抽出コンセプト	コンセプトのタイプ
wallet	<不明>
missing	<否定的>
USD5	<通貨>
blanket	<不明>
picnic area	<不明>

### 例から抽出される TLA パターン

テーブル 6-12  
TLA パターン出力の抽出例

抽出されるコンセプトパターン	抽出されるタイプパターン	レコード
picnic area + .	<不明> + <>	レコード B
wallet + .	<不明> + <>	レコード A
blanket + missing	<不明> + <否定的>	レコード B
USD5 + .	<通貨> + <>	レコード B
USD5 + missing	<通貨> + <否定的>	レコード A

### カテゴリ規則の合致

次の表には、カテゴリ規則エディタに入力できるシンタックスをいくつか示しています。すべての規則が機能するわけではなく、またすべてが同じレコードに合致するわけではありません。異なるシンタックスが合致したレコードにどのように影響するかを確認してください。

テーブル 6-13  
条件規則のサンプル

条件規則シンタックス	結果
USD5 & missing	抽出コンセプト missing および抽出コンセプト USD5 の両方が含まれているため、レコード A および B のいずれにも合致します。これは以下と同じになります。 (USD5 & missing)
missing & USD5	抽出コンセプト missing および抽出コンセプト USD5 の両方が含まれているため、レコード A および B のいずれにも合致します。これは以下と同じになります。 (missing & USD5)
missing & <通貨>	抽出コンセプト missing およびタイプ <通貨> に合致するコンセプト両方が含まれているため、レコード A および B のいずれにも合致します。これは以下と同じになります。 (missing & <通貨>)
<通貨> & missing	抽出コンセプト missing およびタイプ <通貨> に合致するコンセプト両方が含まれているため、レコード A および B のいずれにも合致します。これは以下と同じになります。 (<通貨> & missing)
[USD5 + missing]	レコード B は USD5 + missing を含む TLA パターン出力を作成していないため、A とは合致しますが B とは合致しませんが B とは合致しません (前の表を参照)。これは次の TLA パターン出力と同じになります。 USD5 + missing
[missing + USD5]	抽出された TLA パターン (前の表参照) が、最初の位置に missing を使用して表現した順序に合致しないため、レコード A にも B にも合致しません。これは次の TLA パターン出力と同じになります。 USD5 + missing
[missing & USD5]	こうした TLA パターンがレコード B から抽出されているため、A とは合致しますが B とは合致しません。& 文字を使用すると、合致時の順序が重要でないことを示すため、この規則では [missing + USD5] または [USD5 + missing] のいずれかを検索します。レコード A の [USD5 + missing] のみに合致があります。
[missing + <通貨>]	抽出された TLA パターンがこの順序に合致していないため、レコード A にも B にも合致しません。TLA 出力はキーワード (USD5 + missing) またはタイプ (<通貨> + <Negative>) にのみ基づくため、同等のものはありませんが、コンセプトおよびタイプを組み合わせません。

条件規則シンタックス	結果
[<通貨> + <否定的>]	TLA パターンがレコード B から抽出されているため、レコード A に合致しますが B には合致しません。以下の TLA 出力と同じになります。 <通貨> + <否定的>
[<否定的> + <通貨>]	抽出された TLA パターンがこの順序に合致していないため、レコード A にも B にも合致しません。意見テンプレートの場合、デフォルトでは、トピックが意見とともに検出されると、トピック (<通貨>) は最初のスロットに位置し、意見 (<否定的>) は 2 番目のスロットに位置します。

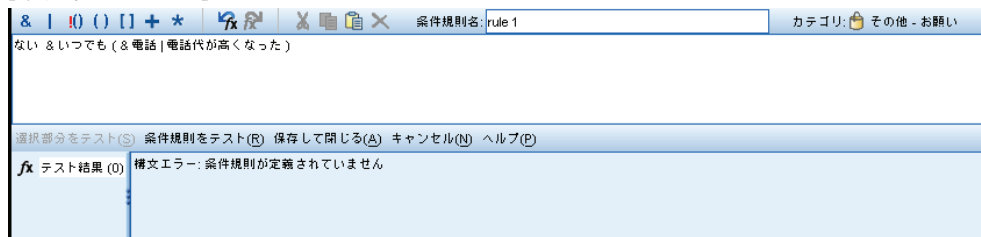
## カテゴリ規則の作成

規則を作成または編集する場合、規則を条件規則エディタで開く必要があります。コンセプト、タイプ、またはパターンを追加し、またワイルドカードを使用して一致を拡張することができます。認識されたコンセプト、タイプ、パターンを使用すると、すべての関連コンセプトを検出するため、便利です。たとえば、コンセプトを使用すると、そのすべての関連キーワード、複数形、および類義語も規則を満たします。同様に、タイプを使用すると、そのすべてのコンセプトも規則にキャプチャされます。

既存の規則を編集するか、カテゴリ名を右クリックして [条件規則を作成] を選択して、条件規則エディタを開くことができます。

図 6-25

[条件規則エディタ] パネル



コンテキストメニューを使用、ドラッグアンドドロップ、または手動でコンセプト、タイプおよびパターンをエディタに入力します。これらのブール型演算子 (&, !(), |) やブラケットを使用して、条件規則式を形成します。一般的なエラーを回避するために、コンセプトを [抽出結果] パネルまたは [データ] パネルから直接条件規則エディタにドラッグアンドドロップすることを推奨します。エラーを回避するため、規則のシンタックスには十分注意してください。詳細は、p. 157 [カテゴリ規則シンタックス](#) を参照してください。

注: 条件規則がどのようにテキストに合致するかについての例は、「[カテゴリ規則の例](#)」 ( p. 164 ) を参照してください。

### 規則を作成するには

- ▶ データがまだ抽出されていない、または抽出が過去のものである場合は、抽出してください。詳細は、5章 p.88 [データの抽出](#) を参照してください。
- ▶ [カテゴリ] パネルで、規則を追加するカテゴリを選択します。
- ▶ メニューから、[カテゴリ] → [条件規則の作成] を選択します。ウィンドウにエディタのカテゴリ条件規則エディタのパネルが開きます。
- ▶ [条件規則名] フィールドに、規則の名前を入力します。名前を入力しない場合、自動的に式を名前として使用します。規則の名前は、後で変更できます。
- ▶ より大きい式のテキスト フィールドで、次の作業を実行できます。
  - テキストをフィールドに直接入力するか、別のパネルからドラッグアンドドロップします。抽出されたコンセプト、タイプ、パターンのみを使用します。たとえば、cats という単語を入力しても単数形の cat のみが [抽出結果] パネルに表示される場合、エディタは cats を認識できなくなります。単数形には自動的に複数形が含まれる場合があり、そうでない場合はワイルドカードを使用することができます。詳細は、p. 157 [カテゴリ規則シンタックス](#) を参照してください。
  - 規則に追加するコンセプト、タイプ、またはパターンを選択してメニューを使用します。
  - ブール型演算子を規則のリンク要素に追加します。ツールバー ボタンを使用して、「and」のブール型演算子 &、「or」のブール型演算子 |、「not」のブール型演算子 !()、カッコ ()、パターンのブラケット [ ] を規則に追加します。
- ▶ [条件規則をテスト] ボタンをクリックして、規則が適格であることを確認します。詳細は、p. 157 [カテゴリ規則シンタックス](#) を参照してください。見つかったレコードの数は、テキスト [テスト結果] の隣のカッコの中に表示されます。このテキストの右側には、認識された条件規則内の要素か、エラーメッセージが表示されます。タイプ、パターン、またはコンセプトの隣のグラフィックに赤い疑問符が表示されている場合、要素が既知の抽出に一致しないことを示します。一致しない場合、規則によってレコードは検出されません。
- ▶ 規則の一部をテストするには、該当する部分を選択して、[選択部分をテスト] をクリックします。
- ▶ 問題が見つかった場合は、必要な変更を行い、規則を再度テストします。
- ▶ 終了したら、[保存して閉じる] をクリックして、規則をもう一度保存し、エディタを閉じます。新しい規則名がカテゴリに表示されます。

図 6-26  
[カテゴリ] パネルの規則

カテゴリ	#ディス...	#レコー...
music	26	76
child's music		1
music catalogue		1
play music		2
music choice		1
amounts of music		1
library of music		1
collection of music		1
bank of music		1
digital music hands		1
share music		1
music quality		1
music		49
selection of music		1
music device		1
music videos		1
variety of music		1
[<Budget> & (<Positive>   low   small)]		1
download music	3	3
software to download music		1

## 規則の編集および削除

ある条件規則を作成・保存したならば、この条件規則はいつでも編集できます。詳細は、p. 157 [カテゴリ規則シンタックス](#) を参照してください。

ある条件規則が不要になった場合、これを削除できます。

### 規則を編集するには

- ▶ [カテゴリ] パネルで、編集するカテゴリを選択します。
- ▶ メニューから、[カテゴリ] → [条件規則の編集] を選択するか、規則名をダブルクリックします。条件規則を表示したエディタが開きます。
- ▶ 既存の結果およびツールバー ボタンを使用して、変更を行います。
- ▶ 規則を再テストして、期待される結果が返されることを確認します。
- ▶ [保存して閉じる] をクリックして、規則をもう一度保存し、エディタを閉じます。

### 規則を削除するには

- ▶ [カテゴリ] パネルで、削除するカテゴリを選択します。

- ▶ メニューから [編集] → [削除] を選択します。この条件規則がカテゴリから削除されます。

## カテゴリの編集および調整

カテゴリを作成すると、それらを検証して、何らかの調整を行う必要が常にあります。言語リソースを調整するほか、定義を結合またはクリーンアップしたり、カテゴリ化されたレコードの一部をチェックするための方法を探すことによってカテゴリを確認する必要があります。また、ニュアンスや特徴が分かるようにカテゴリが定義されるよう、カテゴリのレコードの確認を行うこともできます。

ビルトインで、自動のカテゴリ作成手法を使用して、カテゴリを作成できますが、これらのカテゴリに何らかの調整が必要な場合がよくあります。1 つまたは複数の手法を使用した後、多くの新規カテゴリがウィンドウに表示されます。カテゴリ定義が適切なものとなるまで、カテゴリのデータを確認して調整を行うことができます。詳細は、p. 115 [カテゴリとは](#) を参照してください。

カテゴリ調整のオプションがいくつかあります。その多くについては次のページで説明します。

- カテゴリのプロパティの編集 (名前の変更、ラベルの追加、注釈の追加)
- 記述子のカテゴリへの追加
- カテゴリの編集
- カテゴリの移動
- 階層カテゴリのフラット化
- カテゴリの結合
- テキスト マッチの追加
- 回答のカテゴリへの投入
- カテゴリのコピーおよび再利用
- カテゴリの削除
- 言語リソースの変更および再抽出
- カテゴリをどのように組み合わせるかを視覚化して調整 詳細は、7 章 p. 181 [グラフの視覚化](#) を参照してください。

## カテゴリのプロパティの編集

IBM® SPSS® Text Analytics for Surveys の他の多くの要素と同様に、カテゴリの以下のようなプロパティを編集できます。名前、ラベル、注釈、および高度なテキスト マッチ エントリ。詳細は、p. 116 [カテゴリのプロパティ](#) を参照してください。編集できるプロパティのほか、カテゴリ定義に

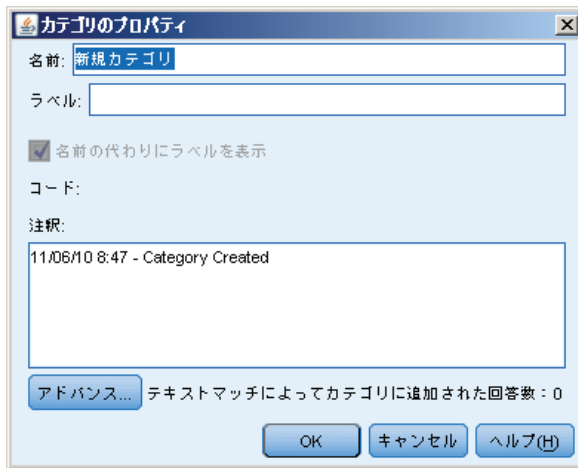
含まれる項目数、つまりカテゴリを構成するキーワード、タイプ、TLA パターン、またはカテゴリ規則の数を表示することもできます。またコードフレームマネージャ内にあるコード値に対応するコード番号も表示できます。

### カテゴリのプロパティを編集するには

- ▶ メニューから、[カテゴリ]>[カテゴリのプロパティ] を選択します。[カテゴリのプロパティ] ダイアログボックスが開きます。

図 6-27

[カテゴリのプロパティ] ダイアログ



- ▶ 必要に応じて、[名前] フィールドに新しい名前を入力してカテゴリの名前を変更します。
- ▶ カテゴリ名またはラベルを変更します。
- ▶ [カテゴリ] パネルなどのインターフェイスで、カテゴリ名の代わりにラベルを使用するには、[名前代わりにラベルを表示] を選択します。
- ▶ 必要に応じて、[注釈] フィールドに注釈を入力します。
- ▶ 単語または句をカテゴリ定義に投入するには、[詳細] をクリックして、テーブルにテキスト マッチを入力します。詳細は、p. 176 [カテゴリ内のテキストマッチ](#) を参照してください。
- ▶ [OK] をクリックし、変更を適用します。



## 記述子のカテゴリへの追加

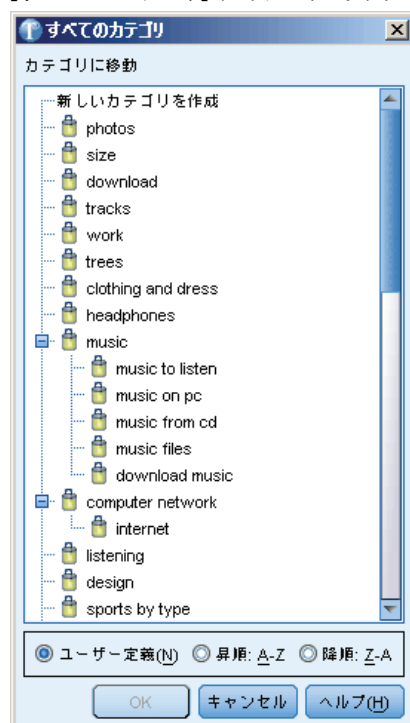
自動的手法を使用した後に、カテゴリ定義のいずれにも使用されなかった抽出結果が作成される場合があります。結果のリストを [抽出結果] ウィンドウで確認する必要があります。カテゴリに含ませたい要素があった場合、これを既存のあるいは新規のカテゴリに追加します。

### コンセプトまたはタイプをカテゴリに追加するには

- ▶ [抽出結果] パネルおよび [データ] パネルから、新規または既存のカテゴリに追加する要素を選択します。
- ▶ メニューから、[カテゴリ] → [カテゴリに追加] を選択します。[すべてのカテゴリ] ダイアログボックスにカテゴリのセットが表示されます。選択した要素を追加したいカテゴリを選択します。要素を新規カテゴリに追加する場合、[新規カテゴリ] を選択します。最初に選択した要素を使用した新しいカテゴリが、[カテゴリ] パネルに表示されます。

図 6-28

[すべてのカテゴリ] ダイアログ ボックス



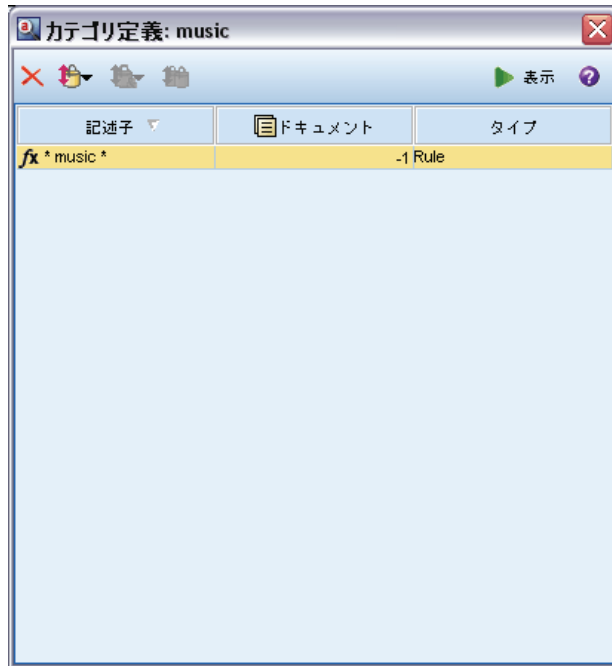
## カテゴリ記述子の編集

いくつかのカテゴリを作成すると、各カテゴリを開いてその定義を構成するすべての記述子を表示できます。[カテゴリ定義] ダイアログ ボックスで、カテゴリ記述子にあらゆる編集を行うことができます。また、カテゴリがカテゴリ ツリーに表示されている場合、ここで処理することもできます。

### カテゴリを編集するには

- ▶ [カテゴリ] パネルで編集したいカテゴリを選択します。
- ▶ メニューから [表示] → [カテゴリ定義] を選択します。[カテゴリ定義] ダイアログ ボックスが開きます。






図 6-29  
[カテゴリ定義] ダイアログ ボックス



- ▶ 編集したい記述子を選択し、該当するツールバー ボタンをクリックします。

次の表で、カテゴリ定義を編集できるツールバーボタンについて説明します。

テーブル 6-14  
ツールバー ボタンおよび説明

アイコン	説明
	選択した記述子をカテゴリから削除します。
	選択した記述子を新規または既存のカテゴリに移動します。
	選択した記述子を & カテゴリ規則の形式でカテゴリに移動します。詳細は、 p. 156 <a href="#">カテゴリ規則の使用</a> を参照してください。
	選択した各記述子を、独自の新規カテゴリとして移動します。
 表示	選択した記述子に従って、[データ] パネルおよび [視覚化] パネルの表示内容を更新します。

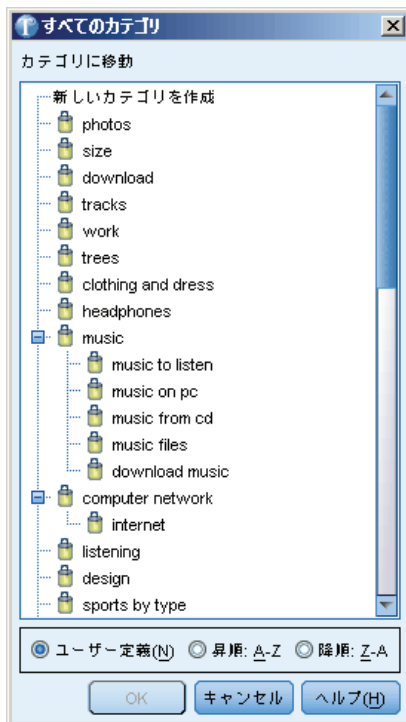
## カテゴリの移動

カテゴリを別の既存カテゴリに投入または記述子を別のカテゴリに移動したい場合、カテゴリを移動できます。

### カテゴリを移動するには

- ▶ [カテゴリ] パネルで、別のカテゴリに移動したいカテゴリまたは記述子を選択します。
- ▶ メニューから、[カテゴリ] → [カテゴリに移動] を選択します。メニューに一連のカテゴリが表示され、リストの上部には最近作成されたカテゴリが表示されます。選択したコンセプトを移動したいカテゴリ名を選択します。
  - 探している名前が表示されたら、その名前を選択します。選択した要素がそのカテゴリに追加されます。
  - 名前が表示されない場合、[もっと表示] を選択すると [すべてのカテゴリ] ダイアログ ボックスが表示され、リストからカテゴリを選択します。

図 6-30  
[すべてのカテゴリ] ダイアログ ボックス



## カテゴリの結合・組み合わせ

2 つ以上の既存カテゴリを 1 つの新規カテゴリに結合したい場合、それらを結合できます。カテゴリを結合する場合、一般名の付いた新規カテゴリが作成されます。カテゴリ記述子に使用されているすべてのコンセプト、タイプ、およびパターンがこの新規カテゴリに移動します。このカテゴリ名は、カテゴリのプロパティを編集することで後から変更できます。詳細は、p. 169 [カテゴリのプロパティの編集](#) を参照してください。

### カテゴリまたはカテゴリの一部を結合するには

- ▶ [カテゴリ] パネルで、結合したい要素を選択します。
- ▶ メニューから、[カテゴリ] → [カテゴリの結合] を選択します。カテゴリ プロパティ ダイアログボックスが表示されるので、新しく作成したカテゴリの名前を入力します。選択したカテゴリがサブカテゴリとして新しいカテゴリに結合されます。

## 回答のカテゴリへの強制投入

回答をカテゴリに投入またはカテゴリから除外すると、実際のカテゴリ定義を変更することなく、自動カテゴリ作成手法で作成されたカテゴリに定義を上書きできます。カテゴリは複数の回答を含むものですが、この回答のうちある特定の回答は、このカテゴリに含めるべきではない、と判断することもあります。この場合、カテゴリ定義からキーワードを削除することなく、回答をそのカテゴリから除外することができます。

この強制機能は、回答はそのカテゴリに適合していない（あるいは適合している）にも関わらず、そのカテゴリに割り当てられてしまっている（あるいは割り当てられていない）という特別な場合に使用されます。典型的な例としては、「このピザはすごくよかったです。焦げて、冷めたピザはみんな大好きなはずです。」。ここで、Pos: [<食べ物> + <Positive>] というカテゴリで、レストランのサービスなどの食べ物に関する好意的な意見を取り出していると仮定します。この回答はおそらくこのカテゴリに割り当てられるでしょう。回答をカテゴリに強制投入あるいはカテゴリから強制排除するには：

### 回答をカテゴリに投入またはカテゴリから除外するには

- ▶ [データ] パネルで、特定のカテゴリに投入またはカテゴリから除外したい回答を選択します。
- ▶ メニューから、[カテゴリ] → [回答をカテゴリに強制投入] または [カテゴリ] → [回答をカテゴリに強制排除] を選択します。サブメニューにカテゴリのリストが表示されます。
- ▶ この回答を強制投入または強制排除したいカテゴリを選択します。カテゴリがたくさん作成されている場合、サブメニュー内に表示されていないカテゴリもあるかもしれません。
  - この場合、サブメニューの一番下にある [もっと表示] を選択します。[すべてのカテゴリ] ダイアログ ボックスが開き、カテゴリを選択して [OK] をクリックすると変更が適用されます。
  - 回答を新規カテゴリに強制投入したい場合、[新規の空白カテゴリ] を選択します。カテゴリツリー内にデフォルト名である「新規カテゴリ」が表示されます。

カテゴリに1つまたは複数の強制回答が含まれる場合、強制投入または強制排除と呼ばれる疑カテゴリがツリーのカテゴリ名の下に表示されます。また、[データ] パネルで [強制投入] 列および [強制排除] 列を表示することによって、カテゴリに強制投入または強制排除された回答が分かります。詳細は、p. 104 [データ] パネル を参照してください。

### 強制された回答の状態を消去するには

- ▶ [データ] パネルで、カテゴリに強制投入またはカテゴリから強制排除したくない回答を選択します。
- ▶ メニューから、強制投入するには [カテゴリ] → [回答をカテゴリに強制投入] または強制排除するには [カテゴリ] → [回答をカテゴリに強制排除] を選択します。回答が強制されているカテゴリの前にはチェックマークが付いています。

図 6-31

[データ] ウィンドウ枠からの回答の強制

	Id	回答	カテゴリ
1	1	平日に時間があまりないときや休日のんびりしたい時	とき 時
2	2	休日の朝、昼	昼 休日
3	3	食事をご馳走したくなかったとき。	なる 休日 食事 とき
4	4		
5	5		なる 時 面倒くさい 外
6	6	行きたいお店を ど	とき 時 行く

メニュー項目	サブメニュー項目
カテゴリに追加	
条件規則に追加	Ctrl+U
強制投入	新規の空白カテゴリ...
強制排除	<input type="checkbox"/> 季節
回答にフラグを付ける	<input type="checkbox"/> 場
ソート	<input checked="" type="checkbox"/> 家
すべて選択	Ctrl+A
コピー	Ctrl+C
検索	Ctrl+F
列を表示	<input type="checkbox"/> 昼
	<input type="checkbox"/> ストレス
	<input type="checkbox"/> 会う
	<input type="checkbox"/> 面倒くさい
	<input type="checkbox"/> 地域
	<input type="checkbox"/> 話
	<input type="checkbox"/> 入学
	<input type="checkbox"/> 日

- ▶ サブメニューで、オンになっており強制を解除したいカテゴリを選択します。チェックマークが除去され、回答の強制が解除されます。

### 強制された回答の状態を消去するには

すべての強制された回答の状態を消去する手順は、次のとおりです。

- ▶ メニューから、[カテゴリ] → [すべて消去] → [強制投入] または [カテゴリ] → [すべて消去] → [強制排除] を選択します。回答の強制状態が解除されます。

## カテゴリ内のテキストマッチ

言語リソースを介して単語または句の抽出を強制しようとし、他の言語処理規則によって抽出されていない場合、テキスト マッチ エントリを作成して、抽出結果を使用せずに、そのテキストを含むカテゴリを特定のカテゴリに直接割り当てることができます。

テキスト マッチ エントリ（単語または句）をカテゴリに追加する場合、IBM® SPSS® Text Analytics for Surveys はその単語または句を含む回答を自動的にこのカテゴリに割り当てます。テキスト マッチングは、この定義を最も有効に利用するため、すでにこの単語を言語リソースに追加しようとしている場合にのみ使用する必要があります。詳細は、10 章 p.244 **キーワードの強制** を参照してください。

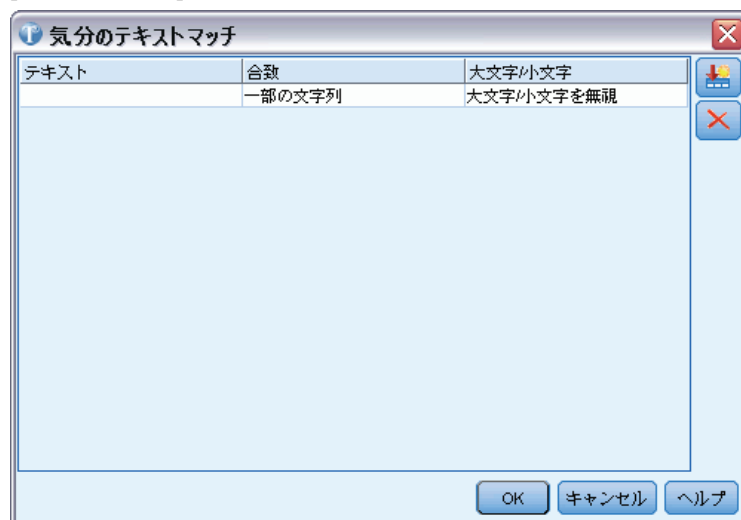
テキストマッチがカテゴリ定義に追加される場合は、テキストマッチと呼ばれる疑カテゴリが、カテゴリ ツリー内のカテゴリの下に表示されません。詳細は、p.104 **[データ] パネル** を参照してください。

また、削除したい行を選択して **[削除]** ボタンをクリックすると、このテーブルからエントリを削除することもできます。

### テキスト マッチ エントリをカテゴリ定義に追加するには

- ▶ **[データ]** パネルで、カテゴリ定義に投入したい単語または句を特定します。
- ▶ **[カテゴリ]** パネルで、この単語または句を強制投入するカテゴリを選択します。
- ▶ メニューから、**[カテゴリ]** > **[カテゴリのプロパティ]** を選択します。**[カテゴリのプロパティ]** ダイアログボックスが開きます。
- ▶ **[詳細]** をクリックします。**[テキスト マッチ]** ダイアログボックスが開きます。

図 6-32  
[テキスト マッチ] ダイアログ ボックス



- ▶ テーブルで、**[テキスト]** 列の最初のセルに単語または句を入力します。

- ▶ この単語または句が回答にあるテキストにどのように一致させるかを選択します。入力したものと完全に同じ語や語句をマッチさせるには、[語全体]を選択します。単語または句をより長い句と一致させるには、[一部の文字列]を選択します。
- ▶ 入力した単語または文字列が大文字と小文字を区別する場合、[大文字/小文字]列で[大文字/小文字を区別]を選択します。
- ▶ [OK] をクリックすると変更が保存され、ダイアログ ボックスが閉じます。テキストマッチ項目を使用してこのカテゴリに割り当てられた回答の個数が更新され、ダイアログボックス内に表示されます。
- ▶ [OK] をクリックし、変更を適用します。

## カテゴリパネルのコピー

同一のあるいは複数のアンケート内で、全く同じあるいは類似した質問を使用する場合、カテゴリの定義を再利用すると、時間をかなり節約できます。同一プロジェクト内である質問から他の質問にカテゴリをコピーできます。カテゴリを再利用する際には、回答データにカテゴリをマッチさせるために再度抽出を行う必要があります。再抽出する前は、[カテゴリ]パネルにカテゴリが頻度を示す疑問符 (?) と共に表示されます。

注:別のプロジェクトのカテゴリを再利用するには、プロジェクトのカテゴリおよびリソースを含むテキスト分析パッケージを作成し、ウィザードで新しいプロジェクトを作成する場合にこのテキスト分析パッケージ (TAP) を使用することをお勧めします。詳細は、3 章 p.44 [テキスト分析パッケージの使用](#) を参照してください。

### 同じプロジェクトのカテゴリをコピーするには

- ▶ カテゴリをコピーしたい質問に移動します。
- ▶ [カテゴリ] パネルのツリーで、すべてのカテゴリを選択します。
- ▶ メニューから、[編集] → [コピー] を選択してカテゴリをコピーします。
- ▶ これらのカテゴリを貼り付ける質問に移動します ([表示] → [質問])。
- ▶ メニューから、[編集] → [貼り付け] を選択してカテゴリを貼り付けます。カテゴリがパネルに追加されます。再抽出していないため、頻度は不明です。このため、度数は疑問符 (?) で表示されます。

## カテゴリの印刷

[カテゴリ] パネルで印刷したいツリー ビューを選択します。



### カテゴリ ツリー ビューを印刷するには

- ▶ [カテゴリ] パネルのツリーで、印刷したい内容に従って、ツリーの要素を展開、縮小または並べ替えます。
- ▶ メニューから、[ファイル] → [印刷] → [カテゴリを印刷] を選択します。[印刷プレビュー] ダイアログ ボックスが開きます。

図 6-33  
[印刷プレビュー] ダイアログ ボックス



- ▶ [印刷] ボタンをクリックすると、ダイアログ ボックスに表示されている通りに印刷されます。

## カテゴリの削除

カテゴリを保持しない場合、削除することができます。カテゴリを削除すると、別のカテゴリに使用されていないコンセプトは [抽出結果] パネルの [未使用の抽出] タブに表示されます。

### カテゴリを削除するには

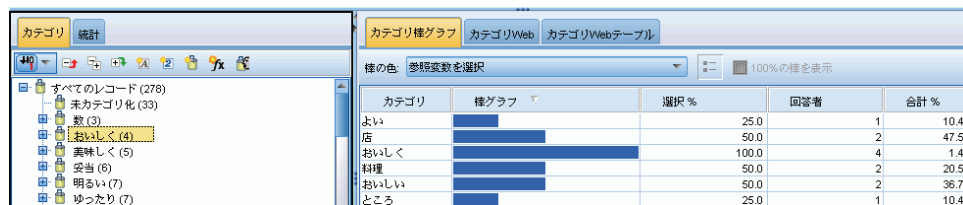
- ▶ [カテゴリ] パネルで、削除したいカテゴリを選択します。
- ▶ メニューから [編集] → [削除] を選択します。

# グラフの視覚化

カテゴリを作成する場合、時間をかけてカテゴリ定義、含まれる回答、およびカテゴリの重複を確認することが重要になります。[視覚化] パネルには、カテゴリに関するいくつかの視点が表示されています。[視覚化] パネルは、質問ビューの右上隅に表示されます。表示されない場合、[表示] メニュー（[表示] → [パネル] → [視覚化]）からこのパネルにアクセスできます。

このビューの [視覚化] パネルには、回答のカテゴリ化における共通性について 3 つの視点が表示されています。このパネルの図表やグラフを使用して、カテゴリ化の結果を分析したり、カテゴリまたはレポートの調整を行うことができます。カテゴリを調整する場合、このパネルを使用して、カテゴリ定義を確認し、あまりに類似している（回答の 75% 以上を共有しているなど）またはあまりに異なるカテゴリを明らかにできます。2 つのカテゴリがあまりに似ている場合、2 つのカテゴリの結合することができます。また、一方のカテゴリから特定の記述子を削除して、カテゴリ定義の調整することもできます。分析でまたはレポート用に使用するために、このパネルの結果をコピー、貼り付けおよび印刷することができます。

図 7-1  
[カテゴリ] パネルおよび [視覚化] パネル



[抽出結果] パネルまたは [カテゴリ] パネルで選択した内容に応じて、このパネルの各タブで、回答とカテゴリの間の該当する交互作用を表示できます。それぞれは、同様の情報を、異なる方法で、またはさまざまなレベルの詳細情報とともに表示されます。必要に応じて、[オプション] ダイアログ ボックスで、これらのグラフおよび図表で使用されている色をカスタマイズできます。詳細は、2 章 p. 21 オプション: [表示] タブ を参照してください。

注: 上位 5 つのカテゴリの棒グラフなど、集計グラフを生成することもできます。これらのグラフを HTML にエクスポートして、プレゼンテーションに使用できます。詳細は、4 章 p. 62 棒グラフのエクスポート を参照してください。

[視覚化] パネルには、次のようなグラフおよび図表が表示されます。

- **カテゴリ棒グラフ:** テーブルおよび棒グラフを使用して、選択に該当する回答と関連するカテゴリとの間の重なりを示します。また、棒グラフは、カテゴリ内の回答数の、回答数の合計に対する比率も示します。参照変数をインポートしている場合は、それを選択して、各カテゴリのレコードの参照変数値を比較することもできます。詳細は、 p.182 [カテゴリ棒グラフ](#) を参照してください。
- **カテゴリ Web グラフ:** このグラフは、その他のパネルの選択部分に従って、回答が属するカテゴリの回答の重なりを示します。詳細は、 p.183 [カテゴリ Web グラフ](#) を参照してください。
- **カテゴリ Web テーブル:** このテーブルは、[カテゴリ Web グラフ] タブと同じ情報をテーブル形式で表示します。このテーブルには 3 つの列があり、列の見出しをクリックするとソートできます。詳細は、 p.184 [カテゴリ Web テーブル](#) を参照してください。

詳細は、 6 章 p.100 [テキスト データの分類](#) を参照してください。

## カテゴリ棒グラフ

このタブには、選択に該当する回答と関連するカテゴリとの間の重なりを示すテーブルおよび棒グラフが表示されます。また、棒グラフは、カテゴリ内の回答数の、回答数の合計に対する比率も示します。このグラフのレイアウトは編集できません。

この棒グラフのコンテキスト メニューを使用して、列をソート、グラフの色を変更、グラフの内容を選択、内容をコピー、そして凡例を表示または隠すことができます。

テーブルには、次の列が表示されます。

- **カテゴリ:** 選択したカテゴリの名前が表示されます。デフォルトでは、選択した中で最も一般的なカテゴリが最初に表示されます。
- **棒グラフ:** 指定されたカテゴリのレコード数の、レコード数の合計に対する比率を視覚的に表示します。
- **選択 %:** カテゴリのレコード数の、選択部分に表示されたレコード数の合計に対する比率に基づいたパーセンテージを示します。
- **回答:** 指定したカテゴリの選択部分の回答の数を示します。
- **合計 %:** 指定したカテゴリのレコード数の合計の、この質問の（選択部分ではない）レコード数の合計に対する比率に基づいたパーセンテージを示します。

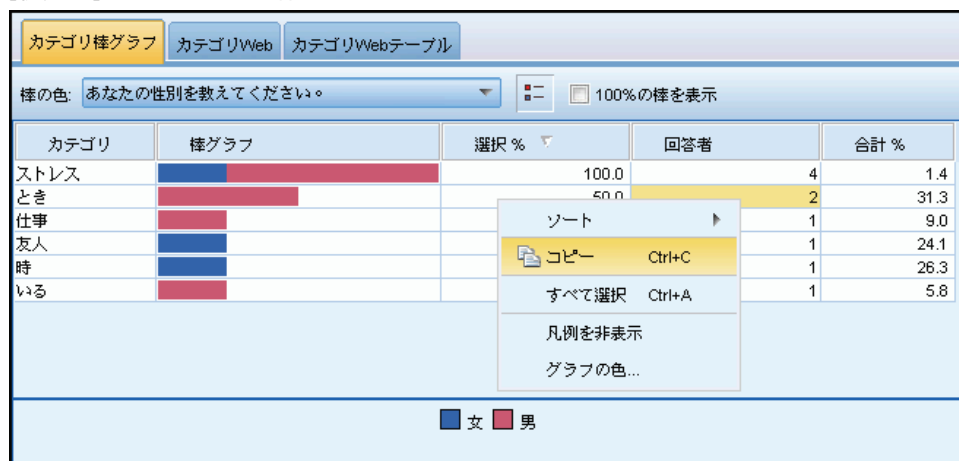
ドロップダウン リストから使用できる参照変数を選択して、これらの値を比較することもできます。参照変数を選択すると、テーブル内の棒グラフが、参照変数の値に従って分割され、カラー コード化されます。棒グラフのカラー コード化された各参照変数をクリックすると、[データ] ウィンドウが更新され、参照変数値に従って、回答のサブセクション

が表示されます。参照変数値の凡例を表示するには、[凡例] ツールバー ボタンをクリックします。

図 7-2  
[凡例] ツールバー ボタン



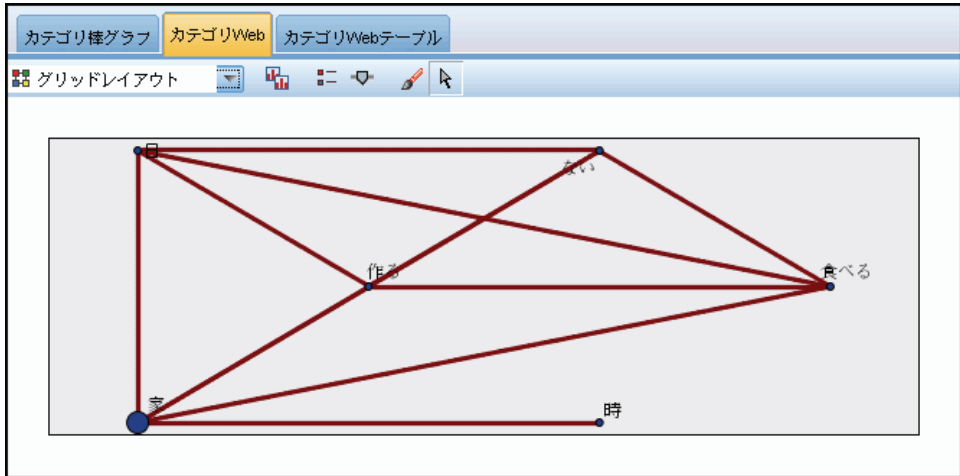
図 7-3  
[視覚化] パネル:カテゴリ棒グラフ



## カテゴリ Web グラフ

このタブには、カテゴリ Web グラフが表示されます。Web グラフは、その他のパネルの選択部分に従って、回答が属するカテゴリの回答の重なりを示します。カテゴリ ラベルがある場合は、グラフにこれらのラベルが表示されます。このパネルのツールバー ボタンを使用して、グラフのレイアウト（ネットワーク、サークル、有向、またはグリッド）を選択できます。

図 7-4  
カテゴリ Web グラフのグリッド レイアウト



Web グラフで、各ノードはカテゴリを示します。マウスを使用し、パネル内のノードを選択して移動できます。ノードのサイズは、選択部分のカテゴリのレコードの数に基づいた相対的なサイズを示します。カテゴリ間の線の太さと色は、含まれている共通のレコードの数を示します。探索的分析モードでノードの上にマウス ポインタを停止させると、ヒントにカテゴリの名前（またはラベル）およびカテゴリ内のレコードの全体数が表示されます。

注: デフォルトでは、ノードを移動できるグラフの探索的分析モードが有効化されています。ただし、編集モードに切り替えて、色、フォント、凡例など、グラフのレイアウトを編集できます。詳細は、p. 185 [グラフのツールバーおよびパレットの使用](#) を参照してください。

## カテゴリ Web テーブル

このタブには、[カテゴリ Web グラフ] タブと同じ情報をテーブル形式で表示されます。このテーブルには次の 3 つの列があり、列の見出しをクリックするとソートできます。

- **度数:** 2 つのカテゴリで共有している、または共通のレコードの数を表示します。
- **カテゴリ 1:** 最初のカテゴリの名前、そして含まれるレコード数の合計がカッコ内に表示されます。
- **カテゴリ 2:** 2 番目のカテゴリの名前、そして含まれるレコード数の合計がカッコ内に表示されます。

図 7-5  
[視覚化] パネル:カテゴリ Web テーブル

度数	カテゴリ 1	カテゴリ 2
2	メニュー(2)	作る(14)
2	食べる(12)	メニュー(2)
2	食べる(12)	作る(14)
1	ずる(27)	とき(87)
1	ずる(27)	家(41)
1	ずる(27)	食事(26)
1	ずる(27)	作る(14)
1	とき(87)	家(41)
1	とき(87)	食事(26)
1	とき(87)	作る(14)



## グラフのツールバーおよびパレットの使用




カテゴリ Web グラフにツールバーがあり、グラフに行うさまざまな操作を実行できるいくつかの共通パレットをすぐに使用できます。探索的分析ビュー モードまたは編集ビュー モードから選択できます。

探索モードでは、視覚化によって表現されたデータや値を分析的に検討することができます。一方、編集モードでは、視覚化のレイアウトや外観を変更することができます。たとえば、フォントや色を自分の組織のスタイル ガイドに合わせて変更することが可能です。このモードを選択するには、メニューから [表示] → [視覚化パネル] → [編集モード] を選ぶか、ツールバーにあるアイコンをクリックします。

編集モードには、視覚化のレイアウトのさまざまな要素に影響を与えるいくつかのツールバーがあります。使用しないものがある場合は、そのツールバーを非表示にしてダイアログ ボックスにおけるグラフの表示領域を増やすことができます。ツールバーの選択または選択解除を行うには、[表示] メニューで目的のツールバーまたはパレットの名前をクリックします。

テーブル 7-1  
Text Analytics のツールバー ボタン

ボタン/リスト	説明
	編集モードを有効化します。編集モードに切り替えて、フォントの拡大、会社のスタイル ガイドに合った色への変更、またはラベルや凡例の削除など、グラフの外観を変更できます。詳細は、p. 186 <a href="#">視覚化の編集</a> を参照してください。
	探索的分析モードを有効化します。デフォルトでは、探索的分析モードがオンになっており、グラフの周囲でノードを移動およびドラッグ、そしてグラフ オブジェクトの上でマウス ポインタを定義させて、ヒントの詳細情報を表示できます。

ボタン/リスト	説明
	<p>でグラフの Web 表示の種類を選択します。</p> <ul style="list-style-type: none"> <li>■ <b>サークルレイアウト:</b> どのようなグラフにも適用できる一般的なレイアウト。リンクに方向がないことを想定してグラフをレイアウトし、すべてのノードを同様に扱います。ノードは円の周囲にのみ配置されます。</li> <li>■ <b>ネットワークレイアウト:</b> どのようなグラフにも適用できる一般的なレイアウト。リンクに方向がないことを想定してグラフをレイアウトし、すべてのノードを同様に扱います。ノードは、レイアウト内に自由に配置されます。</li> <li>■ <b>有向レイアウト:</b> 方向のあるグラフにのみ使用できるレイアウト。このレイアウトは、ルート ノードから葉ノードへのツリー上の構造を作成し、色別に構成します。</li> <li>■ <b>グリッドレイアウト:</b> どのようなグラフにも適用できる一般的なレイアウト。リンクに方向がないことを想定してグラフをレイアウトし、すべてのノードを同様に扱います。ノードは領域内のグリッド ポイントにのみ配置されます。</li> </ul>
	<p>凡例を表示する切り替えボタン。ボタンを押さない場合、判例は表示されません。</p>
	<p>グラフの下にリンク スライダーを表示する切り替えボタン。矢印をスライドして、結果を絞り込むことができます。</p>

## 視覚化の編集

編集モードには、視覚化を編集するためのオプションがいくつか用意されています。以下を行うことができます。

- テキストの編集および書式設定。
- 塗りつぶしの色、透明度、および枠やグラフ要素のパターンの変更。
- 枠線および線の色と破線化の変更。
- 点要素の縦横比および形の回転および変更。
- 棒や点などのグラフ要素のサイズの変更。
- 余白とパディングを使用した項目の周りのスペースの調節。
- 番号の書式の指定。
- 軸および尺度設定の変更。
- カテゴリ軸のカテゴリの並び替え、除外、および集約。
- パネルの方向の設定。
- 座標システムの変換。
- 統計量、グラフ要素の種類、および衝突変更子の変更。
- 判例の位置の変更。
- 視覚化スタイルシートの適用。



以下のトピックでは、これらの作業の実行方法について説明します。グラフの編集の一般的な規則も読んでおくことをお勧めします。

### 編集モードに切り替える方法

- ▶ メニューから次の項目を選択します。  
表示 > 編集モード

## 視覚化編集の一般的な規則

### 編集モード

すべての編集は編集モードで行われます。編集モードを有効にするには、メニューから次の項目を選択します。

表示 > 編集モード

### 選択部分

編集で使用できるオプションは選択に依存します。選択内容に応じて、有効になるツールバーやプロパティ パレットのオプションは異なります。現在の選択には、有効になっている項目のみが適用されます。たとえば、軸が選択されると、[尺度]、[大分割の目盛り]、および [小分割の目盛り] タブがプロパティ パレットで有効になります。

視覚化で項目を選択する際のヒントをいくつか示します。

- 項目をクリックして選択します。
- グラフ要素（散布図の点や棒グラフの棒）を 1 回だけクリックして選択します。最初の選択の後で、もう一度クリックしてグラフ要素のグループまたは単一のグラフ要素に選択を絞り込みます。
- すべての選択を解除するには Esc キーを押します。

### パレット

視覚化で項目が選択されている場合、さまざまなパレットが選択状態を反映するために更新されます。パレットには、選択に対する編集を行うコントロールが含まれます。パレットはツールバーの場合と制御用の UI とタブを持ったパネルの場合があります。パレットは非表示にすることができますので、編集時に必要なパレットが表示されているか確認してください。[表示] メニューで、現在表示されているパレットを確認してください。

ツールバー パレットまたは他のパレットの左側で空いている場所をクリックしてからドラッグして、パレットを再配置することができます。視覚的な反応により、パレットを移動することができる場所が示されます。ツールバー以外のパレットの場合は、閉じるボタンをクリックして

パレットを非表示にしたり、取り外しボタンをクリックしてパレットを別のウィンドウに表示することもできます。ヘルプボタンをクリックすると、そのパレットのヘルプが表示されます。

### 自動設定

いくつかの設定では、[-自動-] オプションが提供されています。これは、値が自動的に適用されることを意味します。どの自動設定が使用されるかは、それぞれの視覚化およびデータ値により異なります。値を入力して、自動設定を上書きすることができます。自動設定を復元したい場合は、現在の値を削除して Enter キーを押します。この設定により、[-自動-] が再び表示されます。

### 項目の除外/非表示

視覚化では、さまざまな項目の除外/非表示を行うことができます。たとえば、凡例や軸ラベルを非表示にすることができます。項目を削除するには、選択して Delete キーを押します。項目の削除が認められていない場合は、何も起こりません。項目を誤って削除してしまった場合は、Ctrl + Z キーを押して削除を取り消します。

### 状態

いくつかのツールバーでは、現在の選択の状態が反映されますが、反映されないものもあります。プロパティ パレットは常に状態を反映します。ツールバーが状態を反映しない場合は、ツールバーについて説明しているトピックでそのことが言及されます。

## テキストの編集および書式設定

テキストを適切に編集して、テキスト ブロック全体の書式を変更することができます。データ値に直接リンクされているテキストを編集することはできません。たとえば、目盛りラベルの内容はその基になっているデータに由来しているので編集することはできません。しかし、視覚化ではすべてのテキストの書式設定が可能です。

### テキストの適切な編集方法

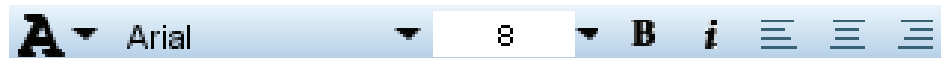
- ▶ テキスト ブロックをダブルクリックします。この操作により、すべてのテキストが選択されます。テキストを編集している間は視覚化のほかの部分を変更することはできないので、すべてのツールバーが無効になります。

- ▶ 既存のテキストに代わるテキストを入力します。テキストをもう一度クリックしてカーソルを表示することもできます。カーソルを任意の場所に置いて、追加のテキストを入力します。

### テキストの書式を設定する方法

- ▶ テキストを含む枠を選択します。テキストをダブルクリックしないでください。
- ▶ フォント ツールバーを使ってテキストの書式を設定します。ツールバーが有効になっていない場合は、テキストを含んでいる枠が選択されているかどうか確認してください。テキスト自体が選択されていると、ツールバーは無効になります。

図 7-6  
フォント ツールバー



フォントを変更することができます。

- 色
- ファミリー (Arial や Verdana など)
- サイズ (pc などのほかの単位を指定しない場合、単位は pt になります)
- 太さ (W)
- テキスト枠に適した位置合わせ

書式設定は枠内のテキストすべてに適用されます。特定のテキストのブロックにおける個々の文字や単語の書式を変更することはできません。

## 色、パターン、破線化、および透明度の変更

視覚化では、さまざまな項目が、塗りつぶしと枠線を持っています。最も分かりやすい例は、棒グラフにおける棒です。棒の色は塗りつぶしの色です。また、棒は黒い実線の枠線で囲まれていることがあります。

視覚化では、そのほかにも塗りつぶし色を持った項目があります。塗りつぶし色が透明の場合は、塗りつぶしが存在していることに気づかないかもしれません。たとえば、軸ラベルのテキストについて考えてみてください。これらのテキストは“浮動”テキストのようにも思われますが、実際には透明の塗りつぶし色を持った枠に現れています。軸ラベルを選択することで、枠を確認することができます。

視覚化全体を囲んでいる枠を含めて、視覚化におけるすべての枠は、塗りつぶし色と枠線のスタイルを持つことができます。また、すべての塗りつぶしには、関連する調節可能な不透明/透明レベルがあります。

### 色、パターン、破線化、および透明度の変更する方法

- ▶ 書式設定を行いたいテキストを選択します。たとえば、棒グラフの棒やテキストを含む枠を選択します。視覚化がカテゴリ変数やフィールドで分割されている場合、個別のカテゴリに対応するグループを選択することもできます。これにより、そのグループに割り当てられたデフォルトの外観を変更することができます。たとえば、積み上げ棒グラフで 1 つの積み上げグループの色を変更することができます。
- ▶ 塗りつぶしや枠線の色、あるいは塗りつぶしパターンを変更するには、色ツールバーを使用してください。

図 7-7  
色ツールバー

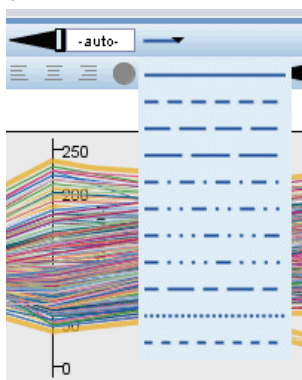


注: このツールバーには現在の選択状態は反映されません。

色や塗りつぶしを変更するために、ボタンをクリックして表示されているオプションを選択したり、ドロップダウンの矢印をクリックして別のオプションを選択したりすることができます。色の場合は、赤色の斜線が付いている白色があることに注意してください。これは透明色です。これを使用すると、たとえば、ヒストグラムで棒の枠線を非表示にすることができます。

- 最初のボタンは塗りつぶし色を制御します。
  - 2 つ目のボタンは枠線の色を制御します。
  - 3 つ目のボタンは塗りつぶしパターンを制御します。枠線の色が塗りつぶしパターンとして使用されます。このため、塗りつぶしパターンは可視の枠線がある場合のみ見ることができます。
  - 4 つ目のコントロールは、塗りつぶし色とパターンの非透明度を制御するスライダーとテキストボックスです。このパーセントが低いと非透明度が低く、透明度が高いこととなります。100% の場合は完全に非透明です。
- ▶ 枠線や線の破線化を変更するには、線ツールバーを使用します。

図 7-8  
線ツールバー



注: このツールバーには現在の選択状態は反映されません。

他のツールバーと同様に、ボタンをクリックして表示されているオプションを選択したり、ドロップダウンの矢印をクリックして別のオプションを選択したりすることができます。

## 点要素の形および縦横比の回転および変更

点要素を回転したり、別の定義済みの形を割り当てたり、あるいは縦横比を変更したりすることができます。

### 点要素を変更する方法

- ▶ 点要素を選択します。個々の点要素の縦横比および形の回転および変更を行うことはできません。
- ▶ 点を変更するには、シンボル ツールバーを使用します。

図 7-9  
シンボル ツールバー



- 最初のボタンを使用すると、点の形を変更することができます。ドロップダウンの矢印をクリックして、定義済みの形を選択します。
- 2 つ目のボタンを使用すると、点を特定のコンパス位置に回転することができます。ドロップダウンの矢印をクリックして、針を目的の位置にドラッグします。
- 3 つ目のボタンを使用すると、縦横比を変更することができます。ドロップダウンの矢印をクリックして、表示される長方形をクリックしてドラッグします。この長方形の形は縦横比を示します。

## グラフ要素のサイズの変更

視覚化では、グラフ要素のサイズを変更することができます。これらには、棒、線、および点などが含まれます。グラフ要素のサイズが変数やフィールドで決められている場合、指定されたサイズが最小サイズになります。

### グラフ要素のサイズを変更する方法

- ▶ サイズを変更したいグラフ要素を選択します。
- ▶ シンボル ツールバーで利用できるオプションに対して、スライダを使用するか、特定のサイズを入力してサイズを指定します。異なる単位を指定しない限り、単位はピクセルになります（単位の略語の完全なリストについては下記を参照してください）。(30% などのように) パーセントで指定することもできます。これは、グラフ要素が利用可能な領域中の指定されたパーセントを使用することを意味します。利用可能な領域は、グラフ要素の種類や個々の視覚化によって異なります。

テーブル 7-2  
有効な単位の略語

略語	単位
センチ	センチメートル
対象	インチ
mm	ミリメートル
pc	pica
pt	ポイント
px	ピクセル

図 7-10  
シンボル ツールバーでのサイズ制御



## 余白とパディングの指定

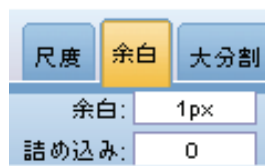
視覚化で枠の周りや内部に空白がありすぎる場合や、逆になさ過ぎる場合、余白とパディングの設定を変更することができます。マージンは枠とその周りにある他の項目との間にある空白の量です。パディングは、枠線と枠の中身の間にある空白の量です。

### 余白とパディングを指定する方法

- ▶ 余白とパディングを指定したい枠を選択します。これは、テキスト枠、凡例の周りの枠、または棒や点などのグラフ要素を表示しているデータ枠に対して行えます。

- ▶ プロパティ パレットの [余白] タブで設定を指定します。cm や in などの別の単位を指定しない限り、すべてのサイズはピクセルになります。

図 7-11  
[余白] タブ



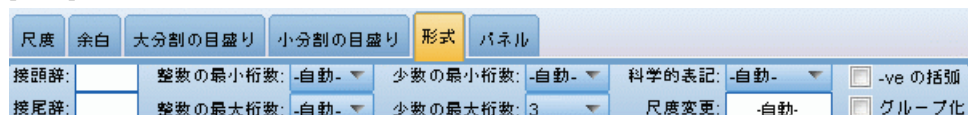
## 数値の書式設定

連続軸の目盛りラベルや、数値を表示するデータ値ラベルの数値の書式を指定できます。たとえば、目盛りラベルに表示される数値が 1000 単位になるように設定することができます。

### 数値書式を指定する方法

- ▶ 数値を含む連続軸の目盛りラベル、またはデータ値ラベルを選択します。
- ▶ プロパティ パレットで [書式] タブをクリックします。

図 7-12  
[書式] タブ



- ▶ 目的の数値書式設定オプションを選択します。

**接頭辞。** 数値の先頭に表示する文字。たとえば、数値が U.S. ドルを通貨とする給与を表している場合は、ドル記号 (\$) を入力します。

**接尾辞。** 数値の最後に表示する文字。たとえば、数値がパーセントを表している場合は、パーセント記号 (%) を入力します。

**整数の最小桁数。** 少数を含む数字の表記の整数部分として表示される最小桁数。実際の値に最小桁数分の桁がない場合、値の整数部分には 0 が追加されます。

**整数の最大桁数。** 少数を含む数字の表記の整数部分として表示される最大桁数。実際の値が最大桁数より大きい桁の場合、値の整数部分はアスタリスク (\*) により置き換えられます。

**少数の最小桁数。** 少数を含む数字の表記または科学的表記の少数部分として表示される最小桁数。実際の値に最小桁数分の桁がない場合、値の少数部分には 0 が追加されます。

**少数の最大桁数。**少数を含む数字の表記または科学的表記の少数部分として表示される最大桁数。実際の値が最大桁数より大きい桁の場合、値の少数は適切な桁数に丸められます。

**科学的。**数値を科学的表記法で表示するかどうか。科学的表記は、非常に大きなまたは小さな数値に便利です。**-自動-**を使用すると、アプリケーションが科学的表記が適切かどうか判定します。

**尺度変更。**元の値を分割する数値である、目盛りの単位です。数値が大きく桁数が多いため、目盛りラベルが長くなるのを避けたいときに、目盛りの単位を変更できます。目盛りラベルの数値の書式を変更する場合、軸の表題を編集して、数値をどのように解釈すべきかを説明してください。たとえば、スケール軸に給与を表示するケースで、ラベルが 30,000、50,000、および 70,000 にあるとします。目盛りの単位として「1000」を入力すると、30、50、70 と表示されます。その際、スケール軸の表題を編集し、「1000 ドル単位」というテキストを含めるようにします。

**-ve の括弧。**マイナスの値を表示するときに括弧で囲むかどうか。

**グループ化。**数値の桁区切り文字を表示するかどうか。ユーザーのコンピュータの現在のロケールにより、数値のグループ化に使用される文字が決定されます。

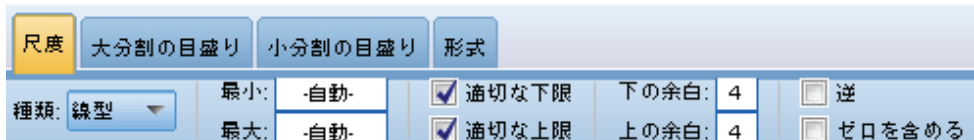
## 軸および尺度設定の変更

軸および尺度の変更には、いくつかのオプションがあります。

### 軸および尺度設定を変更する方法

- ▶ 軸ラベルや目盛りラベルなど、軸の任意の部分を選択します。
- ▶ プロパティ パレットの [尺度]、[大分割の目盛り]、および [小分割の目盛り] タブを使って、軸および尺度の設定を変更します。

図 7-13  
プロパティ パレット



### [尺度] タブ

**タイプ:** 尺度が線形なのか、あるいは変換されているのかを指定します。尺度変換は、データを理解したり、統計的推論に必要な仮定を行うのに役立ちます。散布図において、独立変数と従属変数またはフィールドの関係が非線型である場合は、変換されたスケールを使用します。また尺度変換



は、歪んだヒストグラムを対称形に近づけて正規分布に見えるようにするために使用できます。その際、変換されるのはデータを表示する尺度だけです。実際のデータは変換されません。

- **線型。** 線型の、変換されていない尺度を指定します。
- **対数。** 底が 10 の対数を使用して変換された尺度を指定します。0 および負の値を含めるために、この変換では変更されたバージョンの対数関数を使用されています。この「安全対数」関数は  $\text{sign}(x) * \log(1 + \text{abs}(x))$  として定義されます。つまり `safeLog(-99)` は次と等しくなります。

$$\text{sign}(-99) * \log(1 + \text{abs}(-99)) = -1 * \log(1 + 99) = -1 * 2 = -2$$

- **べき乗。** 0.5 のべき乗指数を使って、べき乗変換尺度を指定します。負の値を含めるために、この変換では変更されたバージョンのべき乗関数を使用されています。この「安全対数」関数は  $\text{sign}(x) * \text{pow}(\text{abs}(x), 0.5)$  として定義されます。つまり `safePower(-100)` は次と等しくなります。

$$\text{sign}(-100) * \text{pow}(\text{abs}(-100), 0.5) = -1 * \text{pow}(100, 0.5) = -1 * 10 = -10$$

**最小値/最大値/適切な下限/適切な上限。** 尺度の範囲を指定します。[適切な下限] と [適切な上限] を指定すると、アプリケーションはデータに基づいた適切な尺度を選択します。最大値と最小値は、通常、それぞれデータの最大数より大きい整数と最少数より小さい整数なので “適切” です。たとえば、データの範囲が 4 から 92 の場合、尺度の適切な下限と上限は、データの実際の最大値と最小値を使用するよりむしろ 0 と 100 のようにします。小さすぎる範囲を設定して重要な項目が非表示になってしまわないように注意してください。また、[ゼロを含める] オプションが選択されている場合、明示的な最小値と最大値を設定できない点に注意してください。

**下の余白/上の余白。** 軸の上端および下端に余白を作成します。余白は、選択されている軸に対して垂直方向に確保されます。cm や in などの別の単位を指定しない限り、すべての単位はピクセルになります。たとえば、縦軸の [上の余白] を 5 に設定する場合、5 px の水平の余白がデータ枠の上端に置かれます。

**逆。** 尺度が逆なのかどうかを指定します。

**ゼロを含める。** 尺度に 0 が含まれる必要があることを示します。このオプションは一般的に、一番短い棒の最上部付近から値が始まるのではなく、0 から棒が始まるように棒グラフで使用されます。このオプションが選択されている場合、尺度の範囲としてユーザー指定の最小値と最大値を指定することはできないので、[最小値] と [最大値] は無効になります。

### [大分割の目盛り]/[小分割の目盛り] タブ

目盛りまたは目盛りマークは軸に表示される線です。これらは特定の区間またはカテゴリにおける値を示します。[大分割の目盛り] はラベルの付いた目盛りマークです。これらは他の目盛りマークよりも長いです。[小分割の目盛り] は大分割の目盛りの間にある目盛りマークです。いくつかのオ

プシオンは特定の目盛りの種類のためのものですが、大分割および小分割の目盛りではほとんどのオプションが利用できます。

**目盛りを表示。** 大分割または小分割の目盛りをグラフに表示するかどうか指定します。

**グリッド線を表示。** 大分割または小分割の目盛りにグリッド線を表示するかどうかを指定します。**グリッド線**は、軸から軸までグラフ全体に引かれる線です。

**位置。** 軸に対応する目盛りマークの位置を指定します。

**長さ。** 目盛りマークの長さを指定します。cm や in などの別の単位を指定しない限り、すべての単位はピクセルになります。

**基本。** 大分割の目盛りにのみ適用されます。大分割の目盛りが最初に表示される値を指定します。

**デルタ。** 大分割の目盛りにのみ適用されます。大分割の目盛りの間の差異を指定します。つまり、大分割の目盛りは  $n$  番目の値ごとに表示され、 $n$  がデルタ値になります。

**分割。** 小分割の目盛りにのみ適用されます。大分割の目盛りの間の小分割の目盛りによる区分の数を指定します。小分割の目盛りの実際の数は、区分の数より 1 つ少ない数となります。たとえば、大分割の目盛りが 0 と 100 に存在しているとします。「2」を小分割の目盛りによる区分の数として指定すると、実際には小分割の目盛りが 50 に 1 つ追加され、0-100 の範囲を 2 つの区分に分割します。

## カテゴリの編集

カテゴリ軸のカテゴリをいくつかの方法で編集することができます。

- カテゴリの表示の並び順を変更する。
- 特定のカテゴリを除外する。
- データセットに出現しないカテゴリを追加する。
- 小さなカテゴリを 1 つのカテゴリに集約または結合する。

### カテゴリの並び順を変更する方法

- ▶ カテゴリ軸を選択します。[カテゴリ] パレットにより、軸上にカテゴリが表示されます。

注:パレットが見えない場合は、有効になっているか確認してください。

- ▶ [カテゴリ] パレットで、ドロップダウン リストから並び替えオプションを選択します。

**ユーザー指定。** パレットに表示される順番に基づいてカテゴリを並び替えます。カテゴリをリストの一番上、あるいは一番下まで動かすには、矢印ボタンを使用します。

**データ。** データセット内での出現順に基づいてカテゴリを並び替えます。

**名前:** パレットに表示されている名前を使って、カテゴリをアルファベット順に並び替えます。

**値。** パレットにある括弧内に表示された値を使用して、基礎となるデータの値順にカテゴリを並び替えます。メタデータのあるデータ ソース (IBM® SPSS® Statistics データ ファイルなど) のみがこのオプションをサポートします。

**統計量。** カテゴリごとに計算された統計量に基づいて、カテゴリを並び替えます。使用できる統計量は度数、パーセント、平均値などです。このオプションはグラフで統計量が使用されている場合のみ使用できます。

### カテゴリを追加する方法

デフォルトでは、データセットに出現するカテゴリのみが使用できます。必要に応じて、視覚化にカテゴリを追加することができます。

- ▶ カテゴリ軸を選択します。[カテゴリ] パレットにより、軸上にカテゴリが表示されます。

注:パレットが見えない場合は、有効になっているか確認してください。

- ▶ [カテゴリ] パレットで、[カテゴリの追加] ボタンをクリックします。

図 7-14  
[カテゴリの追加] ボタン



- ▶ [新しいカテゴリの追加] ダイアログ ボックスで、カテゴリ名を入力します。
- ▶ [OK] をクリックします。

### 特定のカテゴリを除外する方法

- ▶ カテゴリ軸を選択します。[カテゴリ] パレットにより、軸上にカテゴリが表示されます。

注:パレットが見えない場合は、有効になっているか確認してください。

- ▶ [カテゴリ] パレットにて、[含める] リストでカテゴリ名を選択して [X] ボタンをクリックします。除外したカテゴリを元に戻すには、[除外する] リストでカテゴリ名を選択し、リストの右にある矢印をクリックします。

### 小さなカテゴリを集約または結合する方法

小さすぎるために個別に表示する必要のないカテゴリを結合することができます。たとえば、円グラフにカテゴリがたくさんあるのであれば、10パーセント未満のカテゴリを集約することを検討してみてください。集約できるのは、相加的な統計のみです。例えば、平均値は相加的でないため、平均値同士を合計することはできません。したがって、平均値を使用してカテゴリを結合または集約することはできません。

- ▶ カテゴリ軸を選択します。[カテゴリ] パレットにより、軸上にカテゴリが表示されます。

注:パレットが見えない場合は、有効になっているか確認してください。

- ▶ [カテゴリ] パレットで、[集約] を選択してパーセントを指定します。合計のパーセントが指定された数値よりも小さいカテゴリは結合されて1つのカテゴリになります。パーセントは、図表に表示されている統計量に基づきます。度数に基づく統計量および合計の統計量の場合のみ、集約は利用できます。

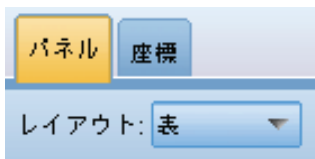
## 方向パネルの変更

視覚化でパネルを使用している場合、パネルの方向を変更することができます。

### パネルの方向を変更する方法

- ▶ 視覚化の任意の一部を選択します。
- ▶ プロパティ パレットで [パネル] タブをクリックします。

図 7-15  
[パネル] タブ



- ▶ [レイアウト] からオプションを選択します。

**テーブル。** 行または列がすべての個別の値に割り当てられた表のようにパネルを配置します。

**置き換え。** パネルを表のように配置しますが、元の行と列は入れ替えられます。このオプションはグラフ自体を置き換えるものとは異なります。このオプションを選択する際は、x 軸と y 軸は変化しないことに注意してください。

**リスト。** 各セルが値の組み合わせを表しているリストのように、パネルを配置します。列と行は個別の値には割り当てられません。このオプションを使用すると、必要に応じてパネルを折り返すことができます。

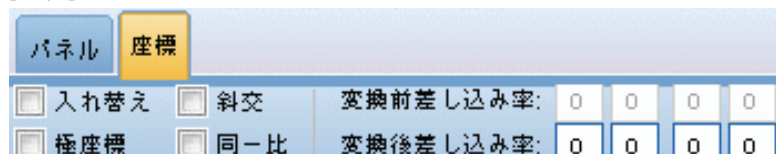
## 座標システムの変換

多くの視覚化は平面、長方形、または座標システムで表示されます。座標システムは、必要に応じて変換することができます。たとえば、極座標変換を座標システムに適用して、斜めの影付き効果を追加して、軸を置き換えることができます。これらの変換が現在の視覚化に既に適用されている場合は、変換を取り消すこともできます。たとえば、円グラフが極座標システムに描かれているとします。必要に応じて、極変換を取り消して、円グラフを単一の積み上げ棒グラフとして長方形の座標システムに表示することができます。

### 座標システムを変換する方法

- ▶ 変換する座標システムを選択します。個別のグラフの周りの枠を選択して、座標システムを選択します。
- ▶ プロパティ パレットで [座標] タブをクリックします。

図 7-16  
[座標] タブ



- ▶ 座標システムに適用したい変換を選択します。変換を取り消すために、変換の選択を解除することもできます。

**置き換え。** 軸の方向を変更することを**置き換え**と呼びます。これは 2-D の視覚化で縦と横の軸を入れ替えるのと似ています。

**極座標。** 極座標変換は特定の角度およびグラフの中心部からの特定の距離で、グラフ要素を描きます。円グラフは、個々の棒を特定の角度で描いた極変換による 1-D の視覚化です。レーダー図は、特定の角度およびグラフの中心部からの特定の距離でグラフ要素を描いた、極変換による 2-D の視覚化です。3-D の視覚化には、奥行き表す次元が追加されます。

**斜交。** 斜交変換は 3-D 効果をグラフ要素に追加します。この変換によりグラフ要素に奥行きが追加されますが、この奥行きは純粋に装飾的なものです。この奥行きが特定のデータ値の影響を受けることはありません。

**同一比。**同じ比率を適用すると、それぞれの尺度上での同じ距離がデータ値における同じ差異を指定するようになります。たとえば、両尺度における 2 cm はデータ値での 1000 の差をあらわします。

**変換前差し込み率。**変換後に軸がクリップしている場合は、変換前に差し込みをグラフに追加することができます。座標システムに変換が適用される前に、差し込みによって各次元が一定のパーセント縮みます。最小 x、最大 x、最小 y、および最大 y の順番で、各次元を制御することができます。

**変換後差し込み率。**グラフの縦横比を変更する場合は、変換の適用後にグラフを差し込みを追加することができます。座標システムに変換が適用された後に、差し込みによって各次元が一定のパーセント縮みます。グラフが変換されていない場合でも、これらの差し込みを適用することができます。最小 x、最大 x、最小 y、および最大 y の順番で、各次元を制御することができます。

## 統計量とグラフ要素の変更

を別の種類に変換したり、グラフ要素の描画に使用される統計量を変更したり、グラフ要素が重なった際に何が起こるのかを決定する衝突変更子を指定したりすることができます。

### グラフ要素を変換する方法

- ▶ 変換したいグラフ要素を選択します。
- ▶ プロパティ パレットで [要素] タブをクリックします。

図 7-17  
[要素] タブ



- ▶ [種類] リストから新しいグラフ要素の種類を選択します。

グラフ要素の種類	説明
点	特定のデータ ポイントを示すマーカー。ポイント要素は、散布図やその他関連する視覚化で使用されます。
間隔	特定のデータ値で描画され、始点と別のデータ値との間の領域を埋める四角形の領域。区間要素は、棒グラフおよびヒストグラムで使用されます。
線(L)	データ値を接続する線。
パス	データセットに表示された順にデータ値を接続する線。

グラフ要素の種類	説明
面	線と始点の間の領域を満たし、データ要素を接続する線。
多角形	データ領域を囲む多角形。多角形要素は、分割された散布図またはマップで使用できます。
スキーマ	外れ値を示すひげとマーカを持つボックスで構成された要素。スキーマ要素はボックスプロットに使用されます。

### 統計量を変更する方法

- ▶ 統計量を変更したいグラフ要素を選択します。
- ▶ プロパティ パレットで [要素] タブをクリックします。
- ▶ [要約] ドロップダウン リストから新しい統計量を選択します。統計量を選択するとデータが集計されることに注意してください。集計されていないデータを視覚化で表示するには、[要約] リストから[[統計なし]]を選択します。

### 連続型フィールドから計算された要約統計

- **平均値.** 中心傾向の測定値。観測値の合計をケース数で割った算術平均。
- **中央値.** ケースの中央付近にある値です。50 パーセンタイルです。ケース数が偶数の場合、中央値は、昇順または降順に保存されたときの 2 つのまん中のケースの平均になります。中央値は、外れ値に対して敏感でない、中心化傾向の測定値です。それに対して平均値は、いくつかの極端に大きい、または小さい値に影響されます。
- **モード.** 最も多く出現する値。複数の値が最高の頻度で出現する場合は、それぞれが最頻値となります。
- **最小.** 数値型変数にとる最も小さい値。
- **最大.** 数値型変数の最大値。
- **範囲.** 最小値と最大値の差異。
- **中間域** : 範囲の中間、つまり最小値との差が最大値との差と等しい値です。
- **合計.** 欠損値のないすべてのケースに対する変数の値の合計または全体。
- **累積合計** : 値の累積合計。各グラフィック要素は、サブグループの合計と、以前のグループすべての総合計を示します。
- **パーセント合計** : 合計した変数に基づいたサブグループ内の、全グループの合計に対するパーセンテージ。

- **累積パーセント合計** : 合計したフィールドに基づいたサブグループ内の、全グループの合計に対する累積パーセンテージ。各グラフィック要素は、サブグループのパーセンテージと、以前のグループすべての総パーセンテージを示します。
- **分散**. 平均値のまわりの値の散らばりの程度。平均値からの偏差の平方和を、有効観測値の合計数から 1 を引いたもので割って求めます。分散の単位はその変数の単位の 2 乗です。
- **標準偏差**. 平均の周辺のばらつき度。正規分布ではデータの 68% が平均 - SD と平均 + SD のなかに含まれ、データの 95% が平均 -2 SD と平均 +2 SD のなかに含まれます。たとえば、平均が 45 で、標準偏差が 10 である場合、正規分布ではデータの 95% が 25 と 65 の間に含まれます。
- **標準誤差**. サンプル間で検定統計量の値がどの程度ばらついているかを測ったもの。ある統計量のサンプル分布の標準偏差です。たとえば、平均値の標準誤差はサンプル平均の標準偏差です。
- **尖度**. 観測値が中心の周りに群がる度合いの測定値。正規分布の場合、尖度統計値は 0 です。正の尖度は、正規分布に対して、観測が分布の中心あたりによりクラスタ化されており、分布の極値まで両裾が薄くなることを示します。急尖的分布の両裾は、正規分布に対して厚くなります。負の尖度は、正規分布に対して、観測のクラスタがより小さくなり、分布の極値まで両裾が厚くなることを示します。急尖的分布の両裾は、正規分布に対して厚くなります。
- **歪度**. 分布の非対称の測定値。正規分布は対称で、歪度は 0 となります。有意な正の歪度を持つ分布では、右の裾が長くなります。有意な負の歪度を持つ分布では、左の裾が長くなります。一般に、歪度がその標準誤差の 2 倍より大きい場合は、正規分布から逸脱していると考えられます。

次の領域の統計で、サブグループごとに複数のグラフィック要素が作成される場合があります。区間、領域、または辺のグラフィック要素を使用する場合、領域統計では、範囲を示すグラフィック要素が作成されます。他のすべてのグラフィック要素は、2 つの要素を生成します。一方は範囲の始点を示し、もう一方は範囲の終点を示します。

- **領域 : 範囲** : 最小値と最大値の間の値の範囲。
- **領域 : 平均値の 95 % の信頼区間** : 母集団の平均値を含む 95 % の確率をもつ値の範囲。
- **領域 : 個別の 95 % の信頼区間** : このケースが指定された予測値を含む 95 % の確率をもつ値の範囲。
- **領域 : 平均値の上下 1 標準偏差** : 平均値の上下 1 標準偏差間の値の範囲。
- **領域 : 平均値の上下 1 標準誤差** : 平均値の上下 1 標準誤差間の値の範囲。



### 度数ベースの要約統計

- **度数** :行/ケースの数。
- **累積度数** :行/ケースの累積数。各グラフィック要素は、サブグループの度数と、以前のグループすべての度数合計を示します。
- **度数のパーセント** :行/ケースの総合系に対する、各サブグループ内の行/ケースのパーセンテージ。
- **度数の累積パーセント** :行/ケースの総合系に対する、各サブグループ内の行/ケースの累積パーセンテージ。各グラフィック要素は、サブグループのパーセンテージと、以前のグループすべての総パーセンテージを示します。

### 衝突変更子を指定する方法

衝突変更子により、グラフ要素同士が重なった際に何が起こるのかが決定されます。

- ▶ 衝突変更子を指定したいグラフ要素を選択します。
- ▶ プロパティ パレットで [要素] タブをクリックします。
- ▶ [変更子] ドロップダウン リストから、衝突変更子を選択します。-自動-を使用すると、アプリケーションがグラフ要素の種類と統計量に対して適切な衝突変更子を判定します。

**オーバーレイ**。複数のグラフ要素が同じ値の場合に、それぞれの上部に描画します。

**積み上げ**。同じデータ値を持った通常は重ね描きされる複数のグラフ要素を積み上げます。

**ドッジ**。重ね描きする代わりに、グラフ要素を同じ値に表示されている他のグラフ要素の隣に移動します。グラフ要素は左右対称に並べられます。つまり、グラフ要素はそれぞれ中心点の両側に移動されます。避ける機能は、クラスタ化とよく似ています。

**積み重ね**。重ね描きする代わりに、グラフ要素を同じ値に表示されている他のグラフ要素の隣に移動します。グラフ要素は左右非対称に並べられます。つまり、一番下のグラフ要素が尺度上の特定の値に位置した状態で、グラフ要素はそれぞれの上に積み上げられます。

**JitterNormal (結合)**。同じ値のグラフ要素を、正規分布を使用してランダムに再配置します。

**JitterUniform (結合)**。同じ値のグラフ要素を、一様分布を使用してランダムに再配置します。

## 凡例の位置の変更

グラフに凡例が含まれる場合、凡例は通常グラフの右側に表示されます。この位置は必要に応じて変更することができます。

### 凡例の位置を変更する方法

- ▶ 凡例を選択します。
- ▶ プロパティ パレットで [凡例] タブをクリックします。

図 7-18  
[凡例] タブ



- ▶ 位置を選択します。

## 視覚化および視覚化データのコピー

[全般] パレットには、視覚化とそのデータをコピーするためのボタンが含まれます。

図 7-19  
[視覚化のコピー] ボタン



**視覚化のコピー。** この操作により、視覚化がクリップボードにイメージとしてコピーされます。複数のイメージ フォーマットが利用可能です。イメージを他のアプリケーションに貼り付ける際は、[形式を選択して貼り付け] オプションを選択して、貼り付けで利用可能なイメージ フォーマットから任意のものを選択することができます。

図 7-20  
[視覚化データのコピー] ボタン



**視覚化データのコピー。** この操作により、視覚化を描画するときを使用される基のデータがコピーされます。データはクリップボードにテキスト形式または HTML テキスト形式でコピーされます。データを他のアプリケー

ションに貼り付ける際は、[形式を選択して貼り付け] オプションを選択して、これらのフォーマットのうちのどちらかを選択することができます。

## キーボードのショートカット

テーブル 7-3  
キーボードのショートカット

ショートカット キー	関数
Ctrl + スペース	探索モードと編集モードの切り替え
削除	視覚化の項目の削除
Ctrl + Z	元に戻す
Ctrl+Y	やり直す
F2	グラフにおける項目の選択のためのアウトラインの表示

# パート III: リソース エディタ

# テンプレートとリソース

IBM® SPSS® Text Analytics for Surveys は、主要キーワードをテキストデータから迅速にかつ正確にキャプチャします。この抽出プロセスは、テキストデータからの情報抽出を管理する言語リソースに大きく依存しています。詳細は、1 章 p.4 [抽出の方法](#) を参照してください。リソース エディタ ウィンドウで、これらのリソースを調整できます。

ソフトウェアをインストールすると、専門的なリソースも取得します。これらの付属リソースは、特定の言語と特定の応用分野で、数年にわたる調査と調整の結果得られたもので、ユーザーはその恩恵を受けることができます。ただし、これらの特別なリソースは、使用するデータの文脈には、完全に一致してはいまないので、ユーザー側の組織のデータ向けに、これらのリソーステンプレートを編集したり、独自に調整したカスタムライブラリを作成して使用することができるようになっています。これらのリソースの形式は多岐にわたり、それぞれプロジェクトで使用できます。リソースは、次の中にあります。

- **リソース テンプレート:**テンプレートは、製品への意見といったように、ある特定の領域や文脈に特化したリソースをまとめた形で、ライブラリのセット、タイプ、およびアドバンズ リソースで構成されています。
- **テキスト分析パッケージ (TAP):**テンプレートに保存されているリソースに加え、リソースをもとに作成した専門的カテゴリセットをまとめたテキスト分析パッケージ(TAP)は、カテゴリとリソースをいっしょに保存して再利用することを可能にします。詳細は、3 章 p.44 [テキスト分析パッケージの使用](#) を参照してください。
- **ライブラリ:**ライブラリは、TAP およびテンプレートの構成要素として使用されます。それらは、プロジェクトのリソースに個別に追加されます。各ライブラリはいくつかの辞書で構成され、タイプのリスト、類義語リスト、不要語リストを定義、管理するために使用されます。ライブラリは個別に提供されていますが、テンプレートおよびTAP でパッケージ化されています。詳細は、9 章 p.221 [ライブラリの使用](#) を参照してください。

注:抽出時、いくつかのコンパイル済み内部辞書も使用されます。これらのコンパイル済み辞書には、コア ライブラリのタイプを補完する多くの定義が含まれています。これらのコンパイル済み辞書は編集できません。

リソース エディタ を用いることで、抽出結果(キーワード、タイプ、およびパターン) を出力する際に適用されるリソースセットへのアクセスが可能となります。リソース エディタ で実行するタスクには、次のような数多くのものがあります。

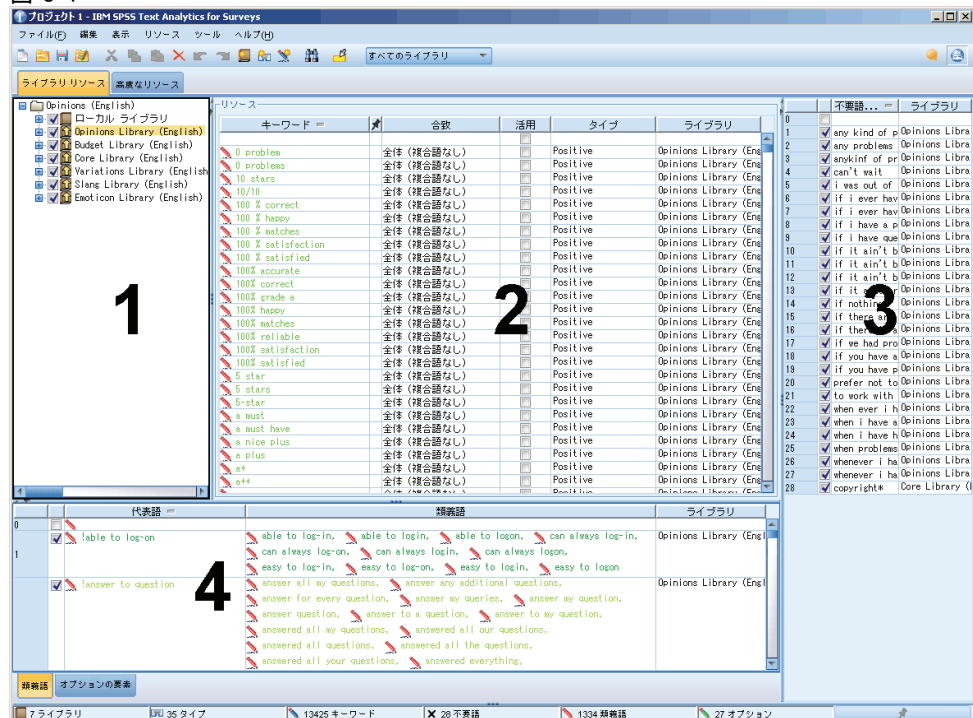
- **ライブラリの使用**。詳細は、9 章 p.221 [ライブラリの使用](#) を参照してください。
- **キーワード辞書の作成**。詳細は、10 章 p.238 [タイプの作成](#) を参照してください。
- **キーワードを辞書に追加**。詳細は、10 章 p.240 [キーワードを追加](#) を参照してください。
- **類義語の作成**。詳細は、10 章 p.249 [類義語の定義](#) を参照してください。
- **TAP のリソースの更新**。詳細は、3 章 p.46 [テキスト分析パッケージの更新](#) を参照してください。
- **テンプレートの作成**。詳細は、p.211 [テンプレートの作成および更新](#) を参照してください。
- **テンプレートのインポートおよびエクスポート**。詳細は、p.215 [テンプレートのインポートおよびエクスポート](#) を参照してください。
- **ライブラリの公開**。詳細は、9 章 p.232 [ライブラリの公開](#) を参照してください。

## エディタのインターフェイス

リソース エディタ で実行する操作は、言語リソースの管理および調整を中心に展開しています。これらのリソースは、テンプレートおよびライブラリの形で保存されています。詳細は、10 章 p.236 [キーワード辞書](#) を参照してください。

## [ライブラリ リソース] タブ

図 8-1



インターフェイスは、次のような 4 つの部分で構成されています。

**1. [ライブラリ ツリー] パネル:** 左上部分にあり、ここにはツリーが表示されます。このツリー内のライブラリを使用可あるいは使用不可にしたり、ツリー内でライブラリを選択することで、表示内容を制約することもできます。このツリーでは、右クリックをしてポップアップメニューを使って、さまざまな操作を実行できます。ツリー内のライブラリを展開すると、そこに含まれるタイプが表示されます。特定のライブラリのみにも焦点を当てたい場合、[表示] メニューからこのリストをフィルタリングすることもできます。

**2. [キーワード辞書のキーワード リスト] パネル:** ライブラリ ツリーの右側にあるこのパネルには、ツリーで選択されたライブラリのキーワード辞書のキーワード リストが表示されます。キーワード辞書は、ラベル (タイプ) ごとにまとめられた語を集めたものです。抽出エンジンがテキストデータを読み込むとき、テキスト内の単語をキーワード辞書内で定義されている語と比較します。抽出されたキーワードがキーワード辞書内にある場合、このタイプ名が割り当てられます。キーワード辞書は、(タイプという) 共通点でキーワードをまとめた個別の辞書と考えていいでしょう。たとえば、コア ライブラリの <Location> タイプには、new orleans、great britain、および new york などの概念が含まれます。これらのキーワードはすべて、地名を示します。ライブラリには、1 つま

たは複数のキーワード辞書が含まれます。詳細は、10 章 p.236 [キーワード辞書](#) を参照してください。

**3. [不要語辞書] パネル:** 右側にあるこのパネルには、最終的な抽出結果から除外されるキーワードの集合が表示されます。ここに登録されたキーワードは [抽出結果] パネルには表示されません。不要語キーワードは選択するライブラリに保存できます。ただし、[不要語辞書] パネルには、ライブラリ ツリーに表示されるすべてのライブラリに不要語登録されたすべてのキーワードが表示されます。詳細は、10 章 p.253 [不要語辞書](#) を参照してください。

**4. [類義語辞書] パネル:** 左下にあるこのパネルには、類義語およびオプションの要素がそれぞれのタブに表示されます。類義語およびオプションの要素を使用すると、最終的な抽出結果の代表語に基づいて類似したキーワードをグループ化できます。この辞書には既知の類義語やユーザー定義の類義語および要素、そして一般的なスペルミスと正しいスペルのペアが含まれています。類義語の定義およびオプションの要素は、選択するライブラリに保存できます。ただし、[類義語辞書] パネルには、ライブラリ ツリーに表示されるすべてのライブラリのすべての内容が表示されます。このパネルにはすべてのライブラリのすべての類義語またはオプションの要素が表示されますが、ツリーのすべてのライブラリの類義語は、このパネルでいっしょに表示されます。1つのライブラリに含まれる類義語辞書は1つだけです。詳細は、10 章 p.247 [類義語辞書](#) を参照してください。

注:

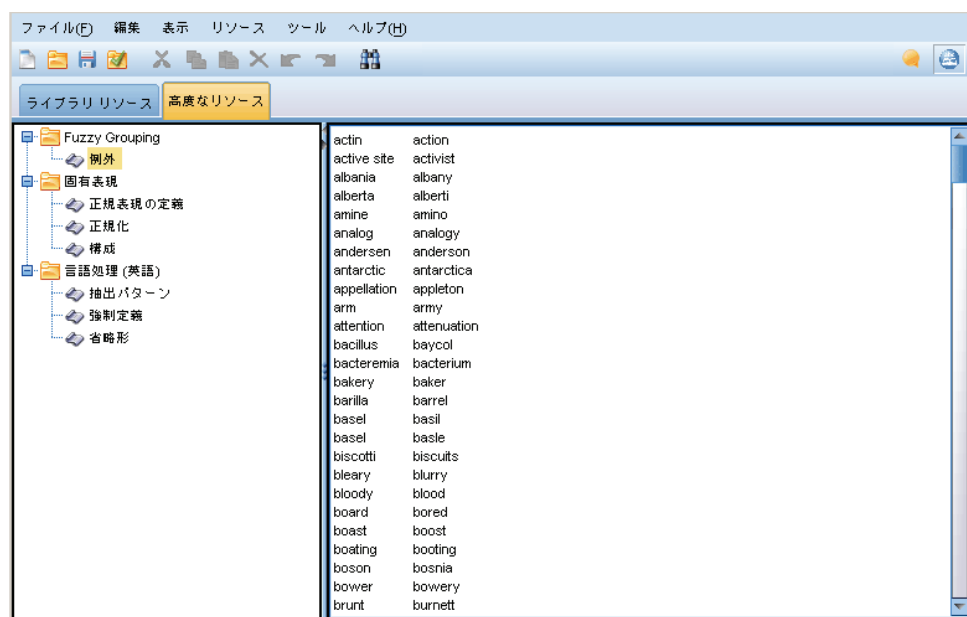
- 1つのライブラリに関する情報のみ表示されるようフィルタリングしたい場合、ツールバーのドロップダウン リストを使用して、ライブラリビューを変更できます。[すべてのライブラリ] という上位レベルのエントリおよび各ライブラリの追加エントリが含まれます。詳細は、9 章 p.225 [ライブラリの表示](#) を参照してください。

### [高度なりソース] タブ

エディタ ビューの 2 番目のタブで高度なりソースを使用できるようになりました。このタブで高度なりソースを確認および編集することができます。詳細は、11 章 p.256 [アドバンス リソースについて](#) を参照してください。



図 8-2  
アドバンス リソース

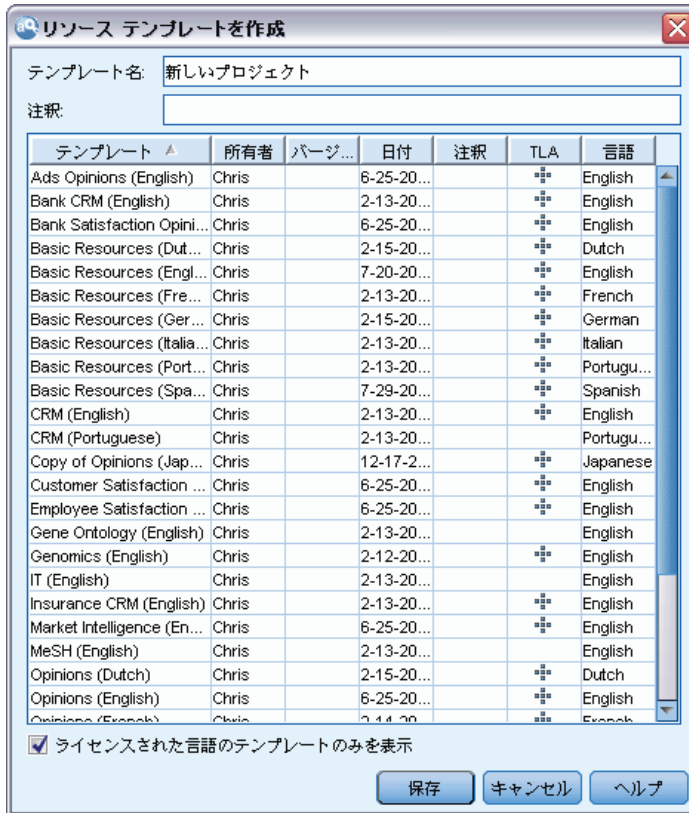


## テンプレートの作成および更新

リソースに変更を行い、今後それらを再利用したい場合、リソースをテンプレートとして保存することができます。保存する場合は、既存のテンプレート名を使用して保存するのか、新しい名前を付けるのかを選択できます。その後、テンプレートを読み込むと、同じリソースを取得することができます。

注: ライブラリを公開して共有することもできます。詳細は、9 章 p. 230 [ライブラリの共有](#) を参照してください。

図 8-3  
[リソース テンプレートを作成] ダイアログ ボックス



### テンプレートを作成 (または更新) するには

- ▶ リソース エディタ ビューのメニューで、[リソース] > [リソース テンプレートを作成] を選択します。[リソース テンプレートを作成] ダイアログ ボックスが開きます。
- ▶ 新しいテンプレートを作成する場合は、[テンプレート名] フィールドに新しい名前を入力します。既存のテンプレートを現在読み込まれたリソースで上書きする場合は、テーブルでテンプレートを選択します。
- ▶ [保存] をクリックして、テンプレートを作成します。

## リソース テンプレートの切り替え

現在読み込まれたリソースを別のテンプレートからのコピーに置き換えたい場合、これらのリソースに切り替えることができます。これによって、現在セッション内に読み込まれている。

リソース エディタ に内容をコピーしたいテンプレートを選択し、[OK] をクリックします。これにより、このプロジェクトのリソースが置き換えられます。

図 8-4  
[リソースを切り替え] ダイアログ ボックス



### リソースを切り替えるには

- ▶ リソース エディタ ビューのメニューで、[リソース]>[リソース テンプレートの切り替え] を選択します。[テンプレートの切り替え] ダイアログ ボックスが開きます。
- ▶ テーブルに表示されたテンプレートから、使用したいテンプレートを選択します。
- ▶ [OK] をクリックして、現在読み込まれているこれらのリソースを中止し、代わりに選択したテンプレートのリソースのコピーを読み込みます。リソースに変更を行い、今後使用するためにライブラリを保存したい場合、切り替える前にそれらを公開、更新、共有することができます。詳細は、9 章 p. 230 [ライブラリの共有](#) を参照してください。

## テンプレートの管理

テンプレート名の変更、テンプレートのインポートおよびエクスポート、または古いテンプレートの削除など、テンプレートを扱う上で基本的な管理方法がいくつかあります。これらの管理は [テンプレートを管理] ダイアログ ボックスで実行されます。テンプレートをインポートおよびエクスポートすると、テンプレートを他のユーザーと共有できます。詳細は、p. 215 [テンプレートのインポートおよびエクスポート](#) を参照してください。

注:この製品とともにインストールされた(付属の)テンプレートの名前を変更したり、削除することはできません。その代わりとして、名前を変更したい場合には、インストールしたテンプレートを開き、新しい名前でテンプレートを作成します。ユーザー定義のテンプレートは削除することができますが、付属のテンプレートを削除しようとする、このライブラリがインストールされたときの一番最初のバージョンで置き換えられます。

図 8-5  
[テンプレートを管理] ダイアログ ボックス



### テンプレートの名前を変更するには

- ▶ メニューの [リソース] → [リソース テンプレートの管理] を選択します。[テンプレートを管理] ダイアログ ボックスが開きます。
- ▶ 名前を変更したいテンプレートを選択し、[名前を変更] をクリックします。表の名前のセルが編集可能なフィールドとなります。
- ▶ 新しい名前を入力し、Enter キーを押します。確認のダイアログ ボックスが開きます。
- ▶ 名前を変更する場合は、[はい] をクリックします。そうでない場合は、[いいえ] をクリックします。

### テンプレートを削除するには

- ▶ メニューの [リソース] → [リソース テンプレートの管理] を選択します。[テンプレートを管理] ダイアログ ボックスが開きます。
- ▶ [テンプレートを管理] ダイアログ ボックスで、削除したいテンプレートを選択します。
- ▶ [削除] をクリックします。確認のダイアログ ボックスが開きます。

- ▶ [はい] をクリックすると削除され、[いいえ] をクリックすると操作はキャンセルされます。[はい] をクリックすると、テンプレートが削除されます。

## テンプレートのインポートおよびエクスポート

テンプレートをインポートおよびエクスポートすることによって、テンプレートを他のユーザーまたはマシンと共有できます。テンプレートは内部データベースに格納されますが、ハードドライブに \*.lrt ファイルとしてエクスポートできます。リソース エディタ の [テンプレートを管理] ダイアログ ボックスでテンプレートをインポートおよびエクスポートできます。

### テンプレートをインポートするには

- ▶ ダイアログ ボックスの [インポート] をクリックします。[テンプレートをインポート] ダイアログ ボックスが開きます。

図 8-6  
[テンプレートをインポート] ダイアログ ボックス

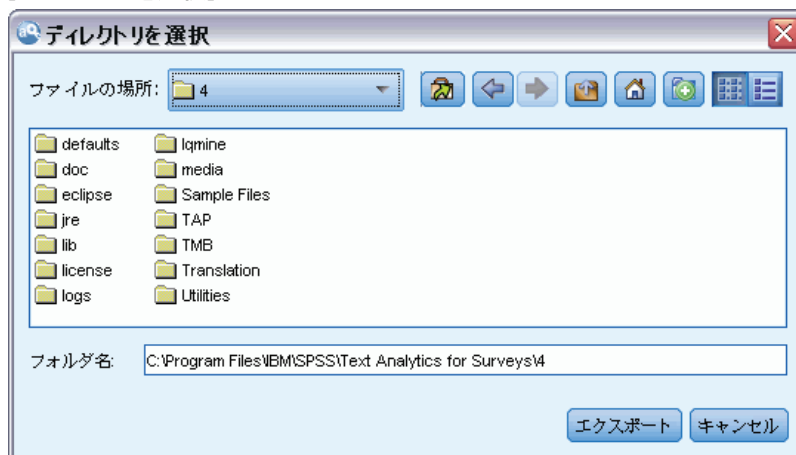


- ▶ インポートするリソース テンプレート ファイル (\*.lrt) を選択して、[インポート] をクリックします。インポートしているテンプレートを別の名前で保存するか、既存のテンプレートを上書きできます。ダイアログ ボックスが閉じ、表にテンプレートが表示されます。

### テンプレートをエクスポートするには

- ▶ ダイアログ ボックスで、エクスポートしたいテンプレートを選択し、[エクスポート] をクリックします。[ディレクトリを選択] ダイアログ ボックスが開きます。

図 8-7  
[ディレクトリを選択] ダイアログ ボックス



- ▶ エクスポート先のディレクトリを選択し、[エクスポート] をクリックします。ダイアログ ボックスが閉じ、テンプレートがエクスポートされ、ファイルの拡張子(\*.lrt) がつきます。

## リソースのバックアップ

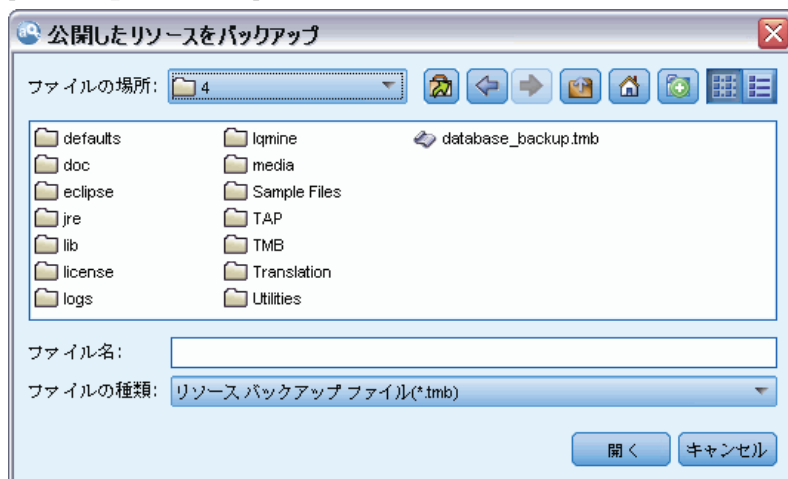
セキュリティ上の観点から、リソースのバックアップが必要な場合があります。

**重要!** 復元を実行すると、データベースの内容はすべて完全に消去され、バックアップファイルの内容しか使用できなくなります。これには処理中の作業も含まれます。

### リソースをバックアップするには

- ▶ メニューの [リソース] → [バックアップ ツール] → [リソースをバックアップ] を選択します。[バックアップ] ダイアログ ボックスが開きます。

図 8-8  
[リソースをバックアップ] ダイアログ ボックス

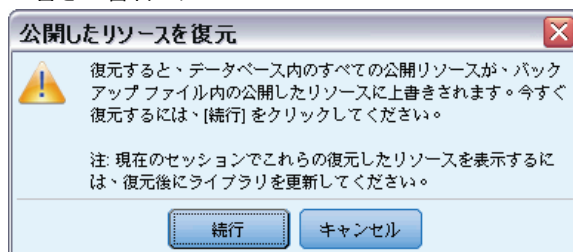


- ▶ バックアップ ファイルの名前を入力して、[保存] をクリックします。ダイアログ ボックスが閉じ、バックアップ ファイルが作成されます。

### リソースを復元するには

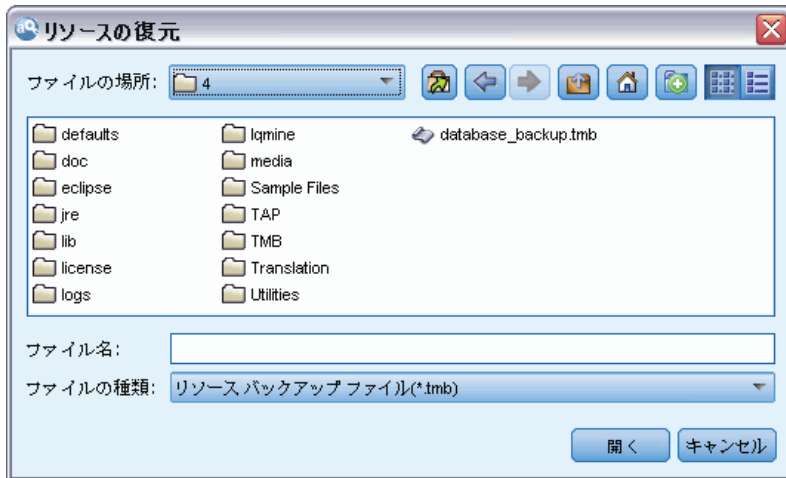
- ▶ メニューの [リソース] → [バックアップ ツール] → [リソースを復元] を選択します。復元するとデータベースの現在の内容が上書きされることの警告が表示されます。

図 8-9  
上書きの警告メッセージ



- ▶ [はい] をクリックして、先に進みます。プロジェクトが開いている場合、メモリ内にあるためプロジェクトは保持されます。ただし、新しく復元したデータベースに保持するにはもう一度保存する必要があります。ダイアログ ボックスが開きます。

図 8-10  
[リソースを復元] ダイアログ ボックス



- ▶ 復元したいバックアップ ファイルを選択し、[開く] をクリックします。ダイアログ ボックスが閉じ、リソースが復元されます。

**重要!** 復元を実行すると、データベースの内容はすべて完全に消去され、バックアップファイルの内容しか使用できなくなります。これには処理中の作業も含まれます。

## リソース ファイルのインポート

この製品以外のところでリソースファイルに直接変更を加えた場合、そのライブラリを選択してインポートの手順を踏むことにより、それらのファイルを選択したライブラリにインポートできます。ディレクトリごとインポートする場合、対象となるファイルすべてを、特定の使用中のライブラリにインポートすることもできます。インポートできるのは、\*.txt ファイルのみです。

インポートしようとするファイルには、各行に項目1つを記載し、その内容は、以下の構造となります。

- 語または語句のリスト（各行に1つずつ）。ファイルはキーワード辞書のキーワード リストとしてインポートされ、キーワード 辞書の名前はファイル名から拡張子を除いたものとなります。
- term1<TAB>term2 のようなエントリのリストとして構成されている場合、類義語のリストとしてインポートされます。term1 は基本キーワードで term2 は代表語です。



## 1 つのリソース ファイルをインポートするには

- ▶ メニューの [リソース] → [ファイルをインポート] → [単一ファイルをインポート] を選択します。[ファイルをインポート] ダイアログ ボックスが開きます。

図 8-11  
[ファイルをインポート] ダイアログ ボックス

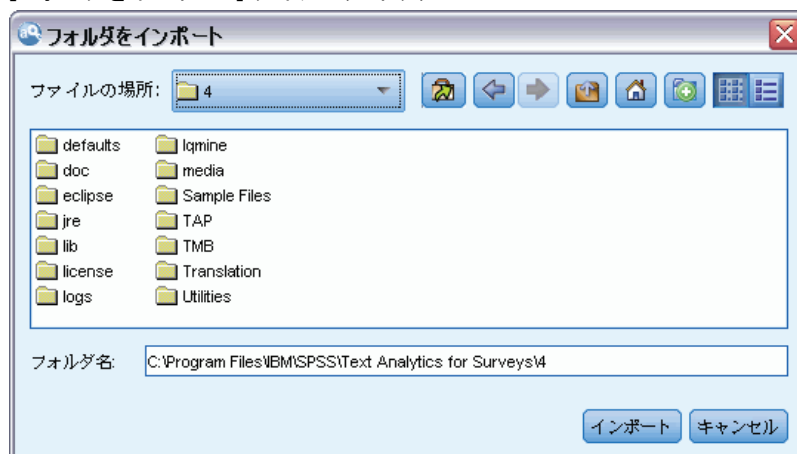


- ▶ インポートしたいファイルを選択し、[インポート] をクリックします。ファイルの内容は内部形式に変換され、ライブラリに追加されます。

## ディレクトリ内のすべてのファイルをインポートするには

- ▶ メニューの [リソース] → [ファイルをインポート] → [フォルダ全体をインポート] を選択します。[フォルダをインポート] ダイアログ ボックスが開きます。

図 8-12  
[フォルダをインポート] ダイアログ ボックス



- ▶ [インポート] リストから、インポートしたいすべてのリソースについて、ライブラリを選択します。[デフォルト] オプションを選択すると、ディレクトリの名前で、新しいライブラリが作成されます。
- ▶ ファイルをインポートするディレクトリを選択します。サブディレクトリは読み込まれません。
- ▶ [インポート] をクリックします。ダイアログ ボックスが閉じ、インポートされたリソース ファイルの内容が辞書およびアドバンス リソース ファイルの形式でエディタに表示されます。

# ライブラリの使用

テキスト データからキーワードを抽出してグループ化するために抽出エンジンで使用するリソースには、常に 1 つ以上のライブラリが含まれています。リソース エディタ の左上部分にあるライブラリ ツリーに一連のライブラリが表示されます。ライブラリは、キーワード、類義語、および不要語の3種類の辞書で構成されています。詳細は、10 章 p.236 [ライブラリ辞書について](#) を参照してください。

リソース テンプレートまたは選択したTAP のリソースには、複数のライブラリが含まれており、テキスト データからすぐにキーワード抽出を開始できるようになっています。しかしユーザーは、独自のライブラリを作成して、それらを再利用できるよう公開することもできます。詳細は、p.232 [ライブラリの公開](#) を参照してください。

たとえば、自動車産業に関連するテキスト データを頻繁に扱っているとします。データを分析した後、カスタム化された言語リソースを作成し、業界特有の用語や隠語を扱えるようにします。リソース エディタ を使用して、新しいテンプレートを作成し、テンプレート内にライブラリを作成して自動車に関するキーワードを抽出し、グループ化します。このライブラリの情報が再び必要になるため、ライブラリを中央リポジトリに公開し、**[ライブラリを管理]** ダイアログ ボックスで使用できるようにします。また異なるプロジェクトで独立して再利用できるようになります。

また、その業界のさらに各分野(電子機器、エンジン、冷却装置、あるいは場合によっては特定の製造業者や市場)などに特有のキーワードをグループ化する必要性が出てくる場合もあります。グループごとにライブラリを作成して公開することで、テキスト データでこれを使用できます。こうすれば、自分のテキスト データのコンテキストの状況に最も適切なライブラリを追加できます。

注:追加リソースは、**[アドバンス リソース]** タブで設定および管理できます。一部のリソースはすべてのライブラリに適用され、固有表現、Fuzzy Grouping の例外などを管理します。詳細は、11 章 p.256 [アドバンス リソースについて](#) を参照してください。

## 付属ライブラリ

デフォルトでは、複数のライブラリが IBM® SPSS® Text Analytics for Surveys と共にインストールされます。これらの事前に形式設定されたライブラリを使用して、さまざまなタイプのほか、多くの事前定義されたキーワードや類義語を使用できます。これらの付属ライブラリは、複数の異なるドメイン向けに調整されます。また、複数の言語で使用できます。

多くのライブラリがありますが、最も一般的に使用されているのは次のとおりです。

- **ローカル ライブラリ:** ユーザー定義の辞書の格納に使用します。デフォルトではすべてのリソースに追加される空のライブラリです。空白のキーワード辞書も含まれます。テキスト分析ウィンドウのリソースに直接変更または調整を行う（単語をタイプに追加するなど）場合に役立ちます。この場合、これらの変更および調整は、リソース エディタ のライブラリ ツリーに表示された最初のライブラリに自動的に保存されます。デフォルトでは、これがローカル ライブラリとなります。このライブラリはプロジェクト データ固有であるため、このライブラリは公開できません。この内容を公開したい場合は、まずライブラリの名前を変更する必要があります。
- **コア ライブラリ:** 人名、地名、組織名、商品名、そして不明を示す 5 つの基本的なビルトインのタイプで構成されているため、多くのケースで使用されます。キーワード辞書の 1 つに記載されているのは少数のキーワードですが、コア ライブラリに記載されているタイプは、テキストマイニング製品に付属する内部のコンパイル済み辞書の頑健なタイプを補います。これらの内部コンパイル済み辞書には、各タイプの多くのキーワードが含まれています。このため、キーワード辞書のキーワードリストにキーワードは表示されませんが、コア タイプで抽出およびタイプ指定できます。つまり、Johnのみがコア ライブラリの <Person> キーワード辞書で出現する場合に、George をタイプ <Person> として抽出できます。同様に、コア ライブラリがない場合でも、それらのタイプを含むコンパイル済み辞書が抽出エンジンで使用されるため、抽出結果にそれらのタイプが表示される場合があります。
- **意見ライブラリ:** テキスト データの意見パターンを抽出する場合、最も一般的に使用されます。このライブラリには、嗜好、識別子、優先順位を示す単語が数多く含まれています。これらは-他のキーワードと連携して使用された場合に、-主題についての意見を示すものです。このライブラリには、多くのビルトインのタイプ、類義語および不要語が含まれています。また、テキスト リンク分析に使用されるパターン規則の大きなセットも含まれています。
- **予算ライブラリ:** コストを参照するキーワードを抽出するために使用します。このライブラリには、価格または品質に関する形容詞、識別子、意見を示す多くの単語および句が含まれています。
- **バリエーション ライブラリ:** 特定の言語バリエーションが適切にグループ化するために類義語定義が必要なケースを追加するために使用します。ライブラリには、類義語定義のみが含まれます。

テンプレート外の付属ライブラリの一部はいくつかのテンプレートの内容に似ていますが、テンプレートは特定のアプリケーション向けに調整され、追加のアドバンス リソースを含んでいます。一般的なテンプレートに個別のライブラリを追加するのではなく、処理しているテキスト デー

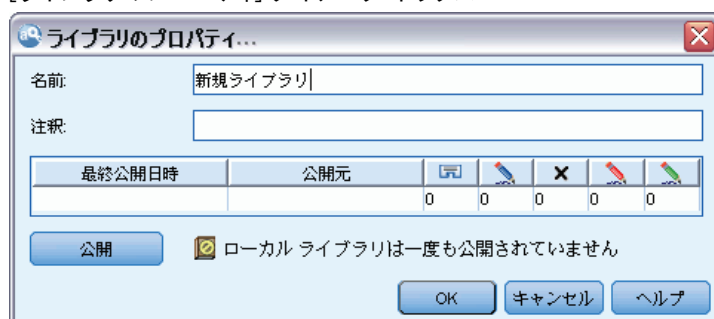
タの種類向けに作成されたテンプレートを使用し、これらのリソースに変更を行うことをお勧めします。

コンパイル済み辞書も、SPSS Text Analytics for Surveys に付属しています。コンパイル済み辞書は抽出プロセスで常に使用され、デフォルトライブラリのビルトインのキーワード辞書に対する多くの補足的定義が含まれています。これらのリソースはコンパイルされていないため、表示も編集もできません。ただし、これらのコンパイル済み辞書にタイプが割り当てられたキーワードを他の辞書に強制投入することができます。詳細は、10 章 p.244 [キーワードの強制](#) を参照してください。

## ライブラリの作成

ライブラリはいくつでもできます。新しいライブラリを作成した後、このライブラリ内で辞書を作成し、キーワード、類義語、不要語を入力できます。

図 9-1  
[ライブラリのプロパティ] ダイアログ ボックス



### ライブラリを作成するには

- ▶ メニューの [リソース] → [新規ライブラリ] を選択します。[ライブラリをプロジェクトに追加] ダイアログが開きます。
- ▶ [名前] テキスト ボックスにライブラリの名前を入力します。
- ▶ 必要に応じて、[注釈] テキスト ボックスにコメントを入力します。
- ▶ ライブラリに入力する前にこのライブラリを公開したい場合は、[公開] をクリックします。詳細は、 p. 230 [ライブラリの共有](#) を参照してください。後でいつでも公開することができます。
- ▶ [OK] をクリックしてライブラリを作成します。ダイアログ ボックスが閉じ、ツリー ビュー内にライブラリが表示されます。ツリー内のこのライブラリを展開すると、空白のキーワード辞書が自動的に含まれていることがわかります。いますぐここにキーワードを追加することもできます。詳細は、10 章 p.240 [キーワードを追加](#) を参照してください。

## パブリック ライブラリを追加

別のプロジェクト データからライブラリを再利用したい場合、パブリックライブラリであれば、ライブラリを現在のリソースに追加することができます。パブリック ライブラリとは、公開されているライブラリです。詳細は、 p. 232 [ライブラリの公開](#) を参照してください。

パブリック ライブラリを追加すると、ローカル コピーがプロジェクトデータに埋め込まれます。このライブラリに変更することはできますが、これらの変更を共有したい場合には、このライブラリのパブリックバージョンを再度公開する必要があります。

パブリック ライブラリを追加すると、このライブラリと他のライブラリ間でキーワードのタイプが同じでない場合、[競合を解決してください] ダイアログ ボックスが表示されます。これらの競合を自分で解決するか、あるいはそこに提示された解決方法を承認して、この操作を完了する必要があります。詳細は、 p. 234 [競合の解決](#) を参照してください。

注: プロジェクトを開くまたはプロジェクトを閉じて公開するときにライブラリを更新する場合、ライブラリが同期しない可能性が低くなります。詳細は、 p. 230 [ライブラリの共有](#) を参照してください。

図 9-2  
[ライブラリを追加] ダイアログ ボックス



### ライブラリを追加するには

- ▶ メニューの [リソース] → [ライブラリを追加] を選択します。[ライブラリを追加] ダイアログ ボックスが開きます。
- ▶ リストのライブラリを選択します。
- ▶ [追加] をクリックします。新しく追加されたライブラリと既存のライブラリとの間に競合箇所がある場合には、この競合を解決するか、ライブラリを変更しないと次に進めません。詳細は、 p. 234 [競合の解決](#) を参照してください。

## キーワードおよびタイプの検索

エディタ内の [検索] 機能を使用して、様々な場所を検索できます。エディタのメニューから [編集] → [検索] を選択すると、検索ツールバーが表示されます。このツールバーを使用して、一回に 1 つの出現ずつ検索できます。もう一度 [検索] をクリックすると、次に出現しているキーワードが検索できます。

検索するとき、エディタは検索ツールバーのドロップダウン リストに表示されている1つあるいは複数のライブラリのみを検索します。[すべてのライブラリ] が選択されている場合、エディタ内のすべてが検索されます。

検索を開始すると、対象となっている部分から検索を始めます。検索はセッションごとに行われ、開始位置であるアクティブなセルに戻るまでループします。矢印を使うことで、検索の順番を逆にできます。検索で大文字と小文字を区別するかどうかを選択することもできます。

### ウィンドウ内の文字列を検索するには

- ▶ メニューから [編集] → [検索] を選択します。検索ツールバーが表示されます。
- ▶ 検索したい文字列を入力します。
- ▶ [検索] ボタンをクリックして検索を開始します。該当するキーワードまたはタイプの次の出現が強調表示されます。
- ▶ ボタンをもう一度クリックして次の出現しているものに移動します。

## ライブラリの表示

ある特定のライブラリまたはすべてのライブラリの内容を表示できます。これは、ライブラリがたくさんあったり、あるいは、ある特定のライブラリの内容を公開前に確認する際に便利です。ビューを変更しても、この [ライブラリ リソース] タブの表示内容が変わるだけで、これによって抽出のプロセスでライブラリが使用されなくなる、というものではありません。詳細は、 p. 227 **ローカル ライブラリを使用不可に** を参照してください。

デフォルトのビューは [すべてのライブラリ] で、これはツリー内にすべてのライブラリを、また他のウィンドウにその内容が表示するものです。ツールバーのドロップダウン リストまたはメニューの選択 ([表示] → [ライブラリ]) によってこの選択範囲を変更できます。1 つのライブラリが表示されている場合、他のライブラリのすべての項目がビューから表示されなくなります。抽出時に読み取ることができます。

### ライブラリビューを変更するには

- ▶ [ライブラリ リソース] タブのメニューで [表示] → [ライブラリ] を選択します。すべてのローカル ライブラリを含んだメニューが開きます。
- ▶ 1つのライブラリを選択するか、あるいは [すべてのライブラリ] を選択して、その内容を表示させます。ウィンドウの内容はこの選択によって変わってきます。

## ローカル ライブラリの管理

パブリック ライブラリに対し、ローカル ライブラリはプロジェクト内またはテンプレート内のライブラリです。詳細は、 p. 228 [パブリック ライブラリの管理](#) を参照してください。ローカルライブラリの基本的な管理方法としては次のようなものがあります。ローカル ライブラリの名前の変更、無効化、削除。

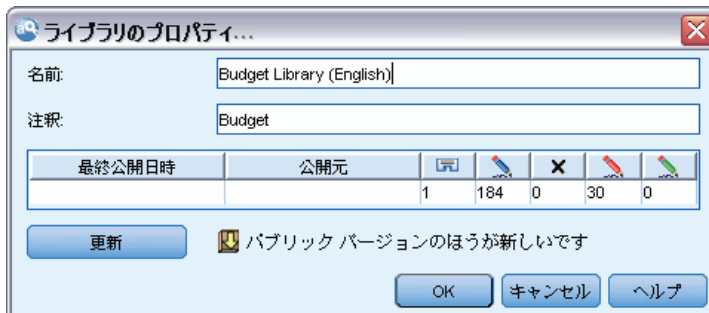
### ローカル ライブラリの名前の変更

ローカル ライブラリの名前は変更できます。ローカルライブラリの名前を変更した場合、これと同じパブリックバージョンがあった場合、それらの関係性はなくなってしまうます。つまり、それ以降の変更はこのパブリックライブラリと共有されることはないということです。ローカルライブラリを新しい名前のもので再度公開することはできます。この場合にも、このローカルライブラリにおける変更は、元の名前のパブリックライブラリに対して反映されません。

注:パブリック ライブラリの名前は変更できません。

- ▶ メニューから [編集] → [ライブラリのプロパティ] を選択します。[ライブラリのプロパティ] ダイアログ ボックスが開きます。

図 9-3  
[ライブラリのプロパティ] ダイアログ ボックス





### ローカル ライブラリの名前を変更するには

- ▶ ツリー ビュー内で、名前を変更したいライブラリを選択します。
- ▶ [名前] テキスト ボックスにライブラリの新しい名前を入力します。
- ▶ [OK] をクリックし、ライブラリの新しい名前を確定します。ダイアログ ボックスが閉じ、ツリービュー内にあるライブラリの名前が更新されます。

## ローカル ライブラリを使用不可に

抽出プロセスからライブラリを一時的に除外したい場合には、ツリービュー内で、このライブラリの名前の左側にあるチェックボックスをオフにします。これによって、このライブラリはプロジェクト内に保持されますが、その内容は、競合のチェックならびに抽出のプロセスでは無視されるようになります。

### ライブラリを無効化するには

- ▶ ライブラリ ツリー パネルで、使用しないライブラリを選択します。
- ▶ スペースバーをクリックします。名前の左側にあるチェック ボックスがオフになります。

## ローカル ライブラリの削除

パブリック バージョンのライブラリを削除せずにライブラリを削除することができます。その逆も可能です。ローカル ライブラリを削除すると、プロジェクトのみのライブラリおよびすべての内容が削除されます。ローカル バージョンのライブラリを削除しても、他のプロジェクトまたはパブリック バージョンのライブラリは削除されません。詳細は、p. 228 [パブリック ライブラリの管理](#) を参照してください。

### ローカル ライブラリを削除するには

- ▶ ツリー ビューで、削除したいライブラリを選択します。
- ▶ ライブラリを削除するには、メニューから、[編集] → [削除] を選択します。ライブラリは削除されます。
- ▶ このライブラリを公開したことがない場合には、このライブラリを削除するか保存するかを尋ねるメッセージが表示されます。[削除] をクリックして次に進むか、[保持] をクリックし、このライブラリを保持します。

注:1 つのライブラリは必ず保持する必要があります。

## パブリック ライブラリの管理

ローカル ライブラリを再利用するために、ローカル ライブラリを公開して処理し、[ライブラリを管理] ダイアログ ボックスに表示することができます ([リソース] → [ライブラリを管理])。詳細は、 p. 230 [ライブラリの共有](#) を参照してください。パブリックライブラリの基本的な管理方法としては、パブリック ライブラリのインポート、エクスポート、または削除があります。パブリック ライブラリの名前は変更できません。

図 9-4  
[ライブラリを管理] ダイアログ ボックス



### パブリック ライブラリのインポート

- ▶ [ライブラリを管理] ダイアログ ボックスの [インポート...] をクリックします。[ライブラリをインポート] ダイアログ ボックスが開きます。

図 9-5  
[ライブラリをインポート] ダイアログ ボックス



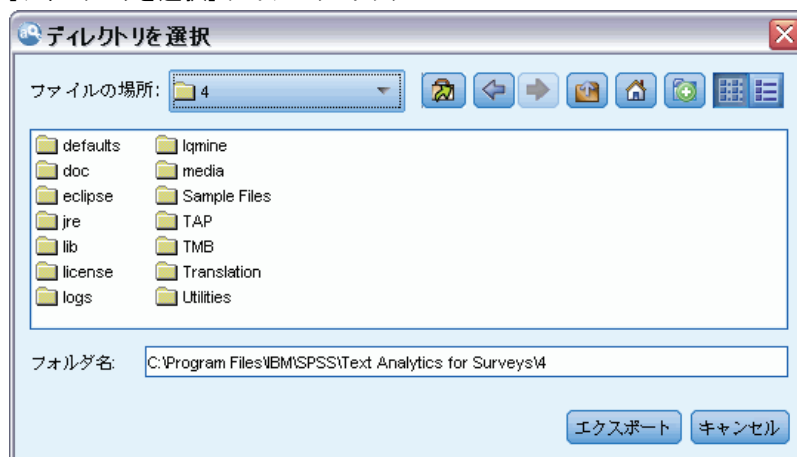
- ▶ インポートしたいライブラリ (\*.lib) を選択し、このライブラリをローカルに追加したい場合、[現在のプロジェクトにライブラリを追加] のチェックボックスをオンにします。
- ▶ [インポート] をクリックします。ダイアログ ボックスが閉じます。同じ名前のパブリック ライブラリがすでに存在する場合、インポートしようとしているライブラリの名前を変更するか、あるいは現在のパブリックライブラリを上書きするのかをたずねるメッセージが表示されます。

### パブリック ライブラリのエクスポート

パブリック ライブラリを .lib 形式でエクスポートすると、ライブラリを共有できるようになります。

- ▶ [ライブラリを管理] ダイアログ ボックスで、リストからエクスポートしたいライブラリを選択します。
- ▶ [エクスポート] をクリックします。[ディレクトリを選択] ダイアログ ボックスが開きます。

図 9-6  
[ディレクトリを選択] ダイアログ ボックス



- ▶ エクスポート先のディレクトリを選択し、[エクスポート] をクリックします。ダイアログ ボックスが閉じ、ライブラリ ファイル (\*.lib) がエクスポートされます。

### パブリック ライブラリの削除

パブリック バージョンのライブラリを削除せずにローカル ライブラリを削除することができます。その逆も可能です。ただし、ライブラリがこのダイアログ ボックスから削除されると、ローカル バージョンがもう一度公開されるまでプロジェクトに追加できなくなります。

製品とともにインストールされたライブラリを削除すると、最初にインストールされていたバージョンが復元されます。

- ▶ [ライブラリを管理] ダイアログ ボックスで、削除したいライブラリを選択します。該当する見出しをクリックして、リストをソートすることができます。
- ▶ [削除] をクリックしてライブラリを削除します。IBM® SPSS® Text Analytics for Surveys が、ローカル バージョンのライブラリがパブリック ライブラリと同じかどうかを検証します。同じであった場合には、警告なくこのライブラリが削除されます。しかしライブラリのバージョンが異なっている場合、パブリックバージョンを保持するのか、あるいは削除するのかをたずねる警告が表示されます。

## ライブラリの共有

ライブラリを使用して、複数のプロジェクト間で共有しやすい方法でリソースを扱うことができます。ライブラリには2つの状態、すなわち2つのバージョン（版）があります。特定のプロジェクトと関連するライブラリは**ローカル ライブラリ**と呼ばれます。プロジェクトで作業している間、たとえば 野菜ライブラリに多くの変更を加えることができます。変更が他のデータでも役立つ場合、この野菜というライブラリの**パブリック ライブラリ** 版を作成することで、これらのリソースを他でも使用できるようになります。パブリック ライブラリは、他のプロジェクトに使用可能です。






[ライブラリを管理] ダイアログ ボックスにパブリック ライブラリが表示されます。このようなパブリックバージョンのライブラリは、他の文脈のリソースに追加できます。こうすることで、ユーザーが作成したカスタムの言語リソースを他でも活用できます。

付属ライブラリ（インストール時に含まれるライブラリ）は、最初はパブリックライブラリです。これらのライブラリ内のリソースを編集してから、これを新しいパブリックバージョンとすることも可能です。これらの新しいバージョンを、他の新しいプロジェクトで使用できるようになります。

自分のライブラリを使って作業をし、これに変更を加えていった場合、このライブラリと他のバージョンのライブラリが同期しなくなってきました。場合によっては、ローカルバージョンのほうがパブリックバージョンより最新であったり、また反対に、パブリックバージョンのほうがローカルバージョンよりも最新であったりします。他のプロジェクト内からパブリック バージョンが更新された場合、一方のライブラリに含まれていない変更がパブリック バージョンおよびローカル バージョンに含まれている場合もあります。ライブラリの個々のバージョンの同期がなくなった場合、これらを再度同期させることができます。ライブラリバージョンの同期は、ローカルライブラリの再公開や更新によって行います。

、プロジェクトを開閉する際にはいつでも、アップデートや再発行が必要なライブラリの同期化を試みる必要があります。またローカルライブラリの同期状況は、ツリービュー内のライブラリ名の隣にあるアイコンや、[ライブラリのプロパティ] ダイアログボックスを表示することによって簡単にわかります。また、同期はメニューからいつでも行うことができます。次の表は、あり得る5つの状態とそれに対応するアイコンです。

テーブル 9-1  
ローカル ライブラリの同期の状態

アイコン	ローカル ライブラリの状態の説明
	未公開-ローカル ライブラリは公開されていません。
	同期-ローカル ライブラリのバージョンおよびパブリック ライブラリのバージョンが同じです。ローカル ライブラリにも適用されます。ローカル ライブラリはプロジェクト固有のリソースのみを含むとされているため、公開できません。
	古い-パブリック ライブラリ バージョンの方がローカル バージョンに比べて新しいものです。ローカルバージョンを更新して、これらの変更を反映させます。
	新しい-ローカル ライブラリ バージョンの方がパブリック バージョンに比べて新しいものです。このローカルバージョンをパブリックバージョンとして再公開できます。
	非同期-ローカル ライブラリおよびパブリック ライブラリに、一方には含まれていない変更があります。この場合、ローカルライブラリを更新するのか、あるいはこれを公開するのかを決めなくてはなりません。更新を選んだ場合、最後に更新あるいは公開した以降の変更はすべて失われます。一方、公開を選んだ場合、パブリックバージョンに加えられた変更は上書きされて失われます。

注:プロジェクトを開くまたはプロジェクトを閉じて公開するときにライブラリを更新する場合、ライブラリが同期しない可能性が低くなります。

ライブラリに変更するとこのライブラリを含む他のプロジェクトに役立つと考えた場合はいつでも、ライブラリを再公開できます。変更すると他のプロジェクトに役立つ場合は、それらのプロジェクトのローカルバージョンを更新できます。このように、新しいライブラリを作成または多くのパブリック ライブラリをリソースに追加することによってデータに適用される各コンテキストまたはドメインのプロジェクトを作成できます。

あるパブリックライブラリが共有されている場合、ローカルバージョンとパブリックバージョンの間で差異が出てくる可能性は高くなります。プロジェクトを閉じるときに開くまたは公開する場合、[ライブラリを管理] ダイアログ ボックスでライブラリと同期していないバージョンのライブラリを公開または更新できるメッセージが表示されます。パブリックライブラリのバージョンがこのローカルバージョンよりも新しい場合、更新するかどうかをたずねるダイアログボックスが開きます。パブリックバージョンで更新する代わりに、現在のローカルバージョンを保持するのか、あるいは更新を現在のローカルライブラリに反映させるのかを選択できます。

## ライブラリの公開

ライブラリをこれまで公開していなかった場合、これを公開するとデータベース内に、ローカルライブラリのパブリックコピーが作成されます。ライブラリを再度公開した場合、ローカルライブラリの内容が、既存のパブリックバージョンの内容を置き換えます。再公開した後、他のプロジェクトのこのライブラリを更新し、ローカルバージョンがパブリックバージョンと同期するようになります。ライブラリを公開できる場合でも、ローカルバージョンは常にプロジェクトに格納されます。

**重要!** ローカルライブラリが変更されており、またこれに対応するパブリックライブラリも変更されている場合、これは同期していない（非同期）と見なされます。このような場合、まず変更されたパブリックバージョンでローカルバージョンを更新し、次にローカルバージョンを再度公開することで、両方のバージョンを全く同じにすることを推奨します。変更が加えられたローカルバージョンを先に公開してしまうと、パブリックバージョンの変更が上書きされてしまいます。

図 9-7  
[ライブラリを公開] ダイアログ ボックス



### ローカル ライブラリをデータベースに公開するには

- ▶ メニューの [リソース] → [ライブラリを公開] を選択します。[ライブラリを公開] ダイアログ ボックスが開き、デフォルトでは、公開の必要があるライブラリがすべて選択されています。
- ▶ 公開または再公開したい各ライブラリの左側にあるチェック ボックスをオンにします。
- ▶ [公開] をクリックし、このライブラリを [ライブラリを管理] データベースに公開します。

## ライブラリの更新

プロジェクトを開く、またはプロジェクトを閉じるときに公開する場合、パブリックバージョンと同期しないライブラリを更新または公開できます。パブリックライブラリのバージョンがローカルバージョンよりも新しい場合、このライブラリを更新するかどうかをたずねるダイアログボックスが開きます。パブリックバージョンで更新せずに現在のローカルバージョンを保持するのか、あるいは現在のプロジェクト内のローカルバージョンをパブリックのもので置き換えるのかを選択できます。パブリックバージョンがローカルバージョンよりも新しい場合、ローカルバージョンを更新して、パブリックバージョンの内容と同期させることができます。更新とは、パブリックバージョン内の変更を、ローカルバージョンに適用することです。

注:プロジェクトを開くまたはプロジェクトを閉じて公開するときライブラリを更新する場合、ライブラリが同期しない可能性が低くなります。詳細は、p. 230 [ライブラリの共有](#) を参照してください。

図 9-8  
[ライブラリを更新] ダイアログボックス



### ローカル ライブラリを更新するには

- ▶ メニューの [リソース] → [ライブラリを更新] を選択します。。 [ライブラリを更新] ダイアログボックスが開き、デフォルトでは、更新の必要のあるライブラリがすべて選択されています。
- ▶ 公開または再公開したい各ライブラリの左側にあるチェック ボックスをオンにします。
- ▶ [更新] をクリックし、ローカル ライブラリを更新します。

## 競合の解決

### ローカル ライブラリとパブリック ライブラリの競合

プロジェクトを開くと、IBM® SPSS® Text Analytics for Surveys が、ローカル ライブラリと [ライブラリを管理] ダイアログ ボックスに表示されたライブラリとの比較を行います。プロジェクトのローカル ライブラリがパブリック バージョンと同期していない場合、[ライブラリの同期] ダイアログ ボックスが開きます。ここで使用したいライブラリのバージョンを選択する際には、次のようないくつかの方法があります。

- **すべてをローカル ライブラリで:** このオプションでは、ローカルライブラリをすべてそのまま保持します。いつでもこれらを再公開あるいは更新できます。
- **すべてをこのマシン上の公開済みライブラリで:** このオプションでは、表示されたローカルライブラリをデータベース内のバージョンで置き換えます。
- **すべてを最新のライブラリで:** このオプションでは、表示されたローカルライブラリをデータベース内のバージョンで置き換えます。
- **その他:** このオプションでは、使用したいバージョンをユーザーが表から選びます。

### 強制キーワードの競合

パブリックライブラリを追加したり、あるいはローカルライブラリを更新した場合、リソース内で、当該ライブラリと他のライブラリとの間で、キーワードやタイプの競合や重複が発見されることがあります。この場合、提示された競合の解決方法から選択します。あるいは競合や重複を変更するための、[強制キーワードを編集] ダイアログボックスが表示されます。詳細は、10 章 p. 244 [キーワードの強制](#) を参照してください。

図 9-9  
[強制キーワードを編集] ダイアログ ボックス





[強制キーワードを編集] ダイアログ ボックスには、競合するキーワードまたはタイプのペアが表示されます。各ペアを見やすくするために、異なる背景色が使われています。これらの色は、[オプション] ダイアログ ボックスで変更できます。詳細は、2 章 p.21 [オプション: \[表示\] タブ](#) を参照してください。[強制キーワードを編集] ダイアログ ボックスには、次の 2 つのタブがあります。

- **重複:** このタブには、このライブラリ内に含まれる重複するキーワードが表示されます。キーワード部分にある押しピンのアイコンは、キーワードのこの出現が強制されていることを表わします。黒いXのアイコンがある場合、他のところで強制されているため、キーワードのこの出現は抽出時に無視されることを示します。
- **ユーザー定義:** このタブには、競合ではなく、キーワード辞書のキーワードパネルにおいて手作業で強制されたキーワードのリストが含まれています。

注: ライブラリを追加・更新すると、[強制キーワードを編集] ダイアログ ボックスが開きます。このダイアログ ボックスをキャンセルしても、ライブラリの更新または追加はキャンセルされません。

### 競合を解決するには

- ▶ [強制キーワードを編集] ダイアログ ボックスで、強制したいキーワードの [使用] 列で選択します。
- ▶ 完了したら、[OK] をクリックして強制キーワードを適用し、ダイアログ ボックスを閉じます。[キャンセル] をクリックすると、このダイアログ ボックスで行った変更がキャンセルされます。

# ライブラリ辞書について

テキスト データの抽出に使用されるこれらのリソースは、テンプレートおよびライブラリの形式で保存されています。ライブラリは、3 つの辞書で構成されています。

- **キーワード辞書**には、1 つのラベル、またはタイプ名に基づいてグループ化されたキーワードの集合が含まれています。抽出エンジンがテキスト データを読み取る場合、テキストの単語を、キーワード辞書で定義したキーワードと比較します。抽出時、タイプのキーワードおよび類義語の活用形が、コンセプトという代表語にグループ化されます。抽出されたコンセプトは、キーワードとして出現するキーワード辞書に割り当てられます。エディタの左上のウィンドウと中央のパネル（ライブラリツリーとキーワードのパネル）でキーワード辞書を管理できます。詳細は、 p. 236 [キーワード辞書](#) を参照してください。
- **類義語辞書**には、最終抽出結果で、類義語、またはコンセプトという 1 つの代表語の下で類似したキーワードをグループ化するために使用する類義語またはオプションの要素として定義される単語の集合が含まれています。[類義語] タブおよび [オプション] タブを使用してエディタの左下のパネルで、類義語辞書を管理できます。詳細は、 p. 247 [類義語辞書](#) を参照してください。
- **不要語辞書**には、最終抽出結果から削除されるキーワードおよびタイプの集合が含まれています。エディタの一番右側のパネルで不要語辞書を管理できます。 詳細は、 p. 253 [不要語辞書](#) を参照してください。

詳細は、 9 章 p. 221 [ライブラリの使用](#) を参照してください。

## キーワード辞書

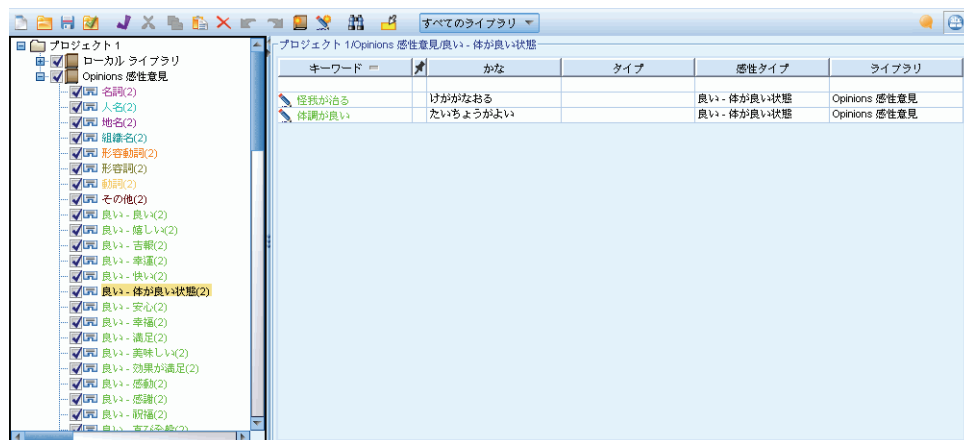
**キーワード辞書**は、1 つのタイプ名、またはラベル、およびキーワードのリストで構成されています。キーワード辞書は、エディタの [ライブラリ リソース] タブの左上のパネルおよび中央のパネルで管理されます。  
ビュー>リソース エディタ (メニュー)。

抽出エンジンがテキスト データを読み取る場合、テキストの単語を、キーワード辞書で定義したキーワードと比較します。キーワードは言語リソースのキーワード辞書にある語および句です。

単語がキーワードに一致した場合、そのキーワードにタイプ名が割り当てられます。抽出時にリソースが読み込まれた場合、テキスト内で見つかったキーワードはいくつかの処理手順を経て、[抽出結果] パネルでコンセプトとなります。同じキーワード辞書に含まれる複数のキーワードが抽出エンジンによって類義語と判断される場合、最も頻繁に出現するキー

ワードに基づいてグループ化され、[抽出結果] パネルで**コンセプト**として表示されます。たとえば、キーワード question および query は、最終的にコンセプト名 question で表示されます。

図 10-1  
ライブラリ ツリーおよびキーワードパネル



キーワード辞書のリストが、左側のライブラリ ツリー パネルに表示されます。各キーワード辞書の内容は、中央のパネルに表示されます。キーワード辞書には、キーワードのリスト以上のものが含まれています。テキスト データの語および語句がキーワード辞書で定義されたキーワードに一致する方法は、定義されたマッチ オプションによって決まります。**マッチ オプション**は、キーワードがテキストデータの候補語または候補句に関してどのように固定されているかを指定します。詳細は、p. 240 **キーワードを追加** を参照してください。

また、キーワードの活用形を自動的に生成して辞書に追加するかどうかを指定して、キーワード辞書のキーワードを拡張することができます。活用形を生成して、単数形の複数形、複数形の単数形、および形容詞をキーワードに自動的に追加することができます。詳細は、p. 240 **キーワードを追加** を参照してください。

注: キーワード辞書がなく、テキストから抽出されたコンセプトは、自動的に <Unknown> のタイプとなります。

## ビルトインのタイプ

IBM® SPSS® Text Analytics for Surveys は付属ライブラリおよびコンパイル済み辞書の形式で一連の言語リソースに付属しています。付属ライブラリには、<地名>、<組織名>、<人名>、および <商品名> を含むビルトインのキーワード辞書が含まれています。

これらのキーワード辞書は、抽出エンジンによって使用され、タイプ <Location> をコンセプト paris に割り当てるように、タイプを抽出したコンセプトに割り当てます。多くのキーワードがビルトインのキーワード辞書で定義されていますが、すべての可能性をカバーしているわけではありません。そのため、辞書に追加するか独自に作成することができます。特定の付属キーワード辞書の内容の詳細は、[タイプのプロパティ] ダイアログ ボックスの注釈を参照してください。ツリーでタイプを選択し、コンテキスト メニューから [編集] → [プロパティ] をクリックします。

注: 付属ライブラリのほか、(抽出エンジンにも使用される) コンパイル済み辞書には、ビルトインのキーワード辞書に対する定義の多くの補足が含まれていますが、その内容は製品では表示されません。ただし、これらのコンパイル済み辞書にタイプが割り当てられたキーワードを他の辞書に強制投入することができます。詳細は、p. 244 [キーワードの強制](#) を参照してください。

## タイプの作成

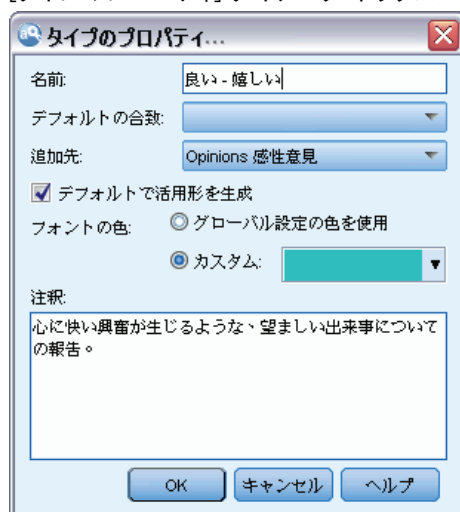
キーワード辞書を作成して、類似したキーワードをグループ化できます。この辞書に出現するキーワードが抽出プロセスで見つかった場合、キーワードはそのタイプ名に割り当てられ、コンセプト名で抽出されます。ライブラリを作成すると、すぐにキーワードを入力できるような空白のキーワード辞書が含まれます。

食べ物に関するテキストを分析しており、野菜に関連するキーワードをグループ化する場合、独自の <Vegetables> キーワード辞書を作成できます。テキストに出現する重要なキーワードと感じた場合は、carrot、broccoli、および spinach などのキーワードを追加できます。抽出時、これらのキーワードのいずれかが見つかった場合、コンセプトとして抽出され、<Vegetables> タイプに割り当てられます。

キーワードの活用形を生成できるため、単語または表現のすべての形式を定義する必要はありません。このオプションを選択すると、抽出エンジンは、他の形式からキーワードの単数形または複数形を、このタイプに割り当てられているものとして自動的に認識します。このオプションは、動詞または形容詞の活用形が必要な場合が少ないため、タイプには主に名詞が含まれている場合に役立ちます。

**重要!** 意見ライブラリに新しいタイプを作成しないことをお勧めします。新しいタイプを作成しても、処理時に考慮に入れられることはありません。意見ライブラリの内容は、パターンの作成に使用されるため、その他のライブラリとは異なった方法で処理されます。代わりに、ライブラリ内に存在するタイプを処理するか、新しいタイプをプロジェクトの別のライブラリに追加します。

図 10-2  
[タイプのプロパティ] ダイアログ ボックス



**名前:** 作成しているキーワード辞書に指定する名前。複数のタイプ名が同じ単語で始まる場合は特に、タイプ名にスペースを使用しないことをお勧めします。

**デフォルトの合致:** デフォルトの合致属性は、抽出エンジンにこのキーワードがテキスト データにどのように合致するかを指示します。キーワードをキーワード辞書に追加すると、これが自動的キーワードに割り当てられる合致属性となります。キーワード リストで合致の選択を手動でいつでも選択することができます。オプションには、[キーワード全体]、[語頭]、[語末]、[どこでも]、[語頭あるいは語末]、[完全かつ語頭]、[完全かつ語末]、[完全かつ (語頭あるいは語末)]、および [全体 (複合語なし)] があります。詳細は、 p. 240 [キーワードを追加](#) を参照してください。

**追加先:** 新しいキーワード辞書を作成するライブラリを指定します。

**デフォルトで活用形を生成:** 抽出エンジンに、文法的形態論を使用して、キーワードの単数形または複数形など、この辞書に追加するキーワードの類似した形式をキャプチャおよびグループ化するように指示します。このオプションは、タイプにほとんど名詞が含まれている場合に特に役立ちます。このオプションを選択すると、このタイプに追加されたすべての新しいキーワードには自動的にこのオプションが与えられますが、リストで手動で変更できます。

**フォントの色:** このフィールドを指定すると、インターフェイスでこのタイプの結果と他のタイプの結果とを区別できるようになります。[グローバル設定の色を使用] を選択すると、タイプのデフォルト色がこのキーワードに使用されます。このデフォルト色は、[オプション] ダイアログ ボックスで設定されます。詳細は、 2 章 p. 21 [オプション: \[表示\] タブ](#) を

参照してください。[カスタム] を選択すると、ドロップダウン リストから色を選択できます。

**注釈:** このフィールドはオプションで、任意のコメントまたは説明に使用できます。

### キーワード辞書を作成するには

- ▶ 新しいキーワード辞書を作成するライブラリを選択します。
- ▶ メニューの [ツール] → [新規タイプ] を選択します。[タイプのプロパティ] ダイアログ ボックスが開きます。
- ▶ [名前] テキスト ボックスにキーワード辞書の名前を入力し、必要なオプションを選択します。
- ▶ [OK] をクリックしてキーワード辞書を作成します。新しいタイプがライブラリ ツリー パネルに表示され、中央パネルに表示されます。すぐにキーワードを追加できます。詳細は、[キーワードを追加](#)を参照してください。

注:ここでは、リソース エディタ ビューでどのように変更を行うかについて説明します。[抽出結果] パネルまたは[データ] パネルで直接、この種類の調整を行うこともできます。詳細は、5 章 p.91 [抽出結果の調整](#) を参照してください。

## キーワードを追加

ライブラリ ツリー パネルにはライブラリが表示され、ツリーを展開すると、ツリーに含まれているキーワード辞書が表示されます。ツリーの選択によって、中央のパネルのキーワードリストに選択したライブラリまたはキーワード辞書のキーワードが表示されます。

図 10-3  
[キーワード] パネル

キーワード	かな	タイプ	感性タイプ	ライブラリ
記述	記述	名詞		Basic Resources 基本リソ
記述	記述	形容動詞		Basic Resources 基本リソ
test	test	人名		Basic Resources 基本リソ
test	test	組織名		Basic Resources 基本リソ
test	test	形容動詞		Basic Resources 基本リソ
gfg	gfg		良い - 嬉しい	Basic Resources 基本リソ
ffg	ffg		良い - 嬉しい	Basic Resources 基本リソ
dfvdf	dfvdf		良い - 嬉しい	Basic Resources 基本リソ

リソース エディタ では、[キーワード] パネルで直接または [新しいキーワードを追加] ダイアログ ボックスを使用して、キーワードをキーワード辞書に追加できます。追加するキーワードは、単語でも複合語でもか

ありません。リストの一番上に空白の行があり、そこに新しいキーワードを追加できます。

注:ここでは、リソース エディタ ビューでどのように変更を行うかについて説明します。[抽出結果] パネルまたは[データ] パネルで直接、この種類の調整を行うこともできます。詳細は、5 章 p.91 [抽出結果の調整](#) を参照してください。

### [キーワード] 列

この列のセルに、単語または複合語を入力します。キーワードが表示される色は、キーワードが保存または強制投入されるタイプの色によって異なります。[タイプのプロパティ] ダイアログ ボックスでタイプの色を変更できます。詳細は、p.238 [タイプの作成](#) を参照してください。

### [強制] 列

このセルで押しピンのアイコンを押すと、抽出エンジンは、他のライブラリのこの同じキーワードの他の出現を無視します。詳細は、p.244 [キーワードの強制](#) を参照してください。

### [合致] 列

この列ではマッチ オプションを選択して、抽出エンジンにこのキーワードがテキスト データにどのように合致するかを指示します。例については表を参照してください。タイプのプロパティを編集して、デフォルト値を変更できます。詳細は、p.238 [タイプの作成](#) を参照してください。メニューから[編集]→[合致方法を変更]を選択します。次に示すのは基本的なマッチ オプションで、これらを組み合わせて使用することもできます。


- **語頭**:辞書のキーワードがテキストから抽出したコンセプトの初めの部分に合致する場合、このタイプが割り当てられます。たとえば、apple と入力すると、apple tart が合致します。
- **語末**:辞書のキーワードがテキストから抽出したコンセプトの終わりの部分に合致する場合、このタイプが割り当てられます。たとえば、apple と入力すると、cider apple が合致します。
- **どこでも**:辞書のキーワードがテキストから抽出したコンセプトの一部に合致する場合、このタイプが割り当てられます。たとえば、apple と入力すると、[どこでも] のタイプが、apple tart、cider apple、および cider apple tart に割り当てられます。
- **キーワード全体**:テキストから抽出したコンセプト全体が辞書キーワードにそのまま合致する場合、このタイプが割り当てられます。キーワードを[キーワード全体]として追加すると、[完全かつ語頭]、[完全かつ語末]、[完全かつどこか]、または[全体(複合語なし)]がキーワードの抽出を強制します。

さらに、<Person> タイプは、edith piaf や mohandas gandhi のように名前の 2 つの部分抽出するため、姓について記載されていないときに名を抽出する場合、名をこのキーワード辞書に明示的に追加したい場合があります。たとえば、edith のすべてのインスタンスを名前として取得したい場合、[キーワード全体] または [完全かつ語頭] を使用して、edith を <Person> タイプに追加する必要があります。

- [全体 (複合語なし)]: テキストから抽出したコンセプト全体が辞書キーワードにそのまま合致する場合、このタイプが割り当てられ、キーワードが長い複合語に合致しないよう、抽出が停止します。たとえば、apple と入力すると、[全体 (複合語なし)] オプションは、apple にタイプを割り当て、別に強制されていない限り複合語 apple sauce は抽出されません。

ZV, i·¥, Å, ÍAfL[f[fh apple がキーワード辞書にあると想定します。マッチオプションによって、この表にはテキスト内で見つかった場合に抽出およびタイプが割り当てられるコンセプトが表示されます。

テーブル 10-1  
合致の例

マッチ オプション - キーワード:  apple	抽出コンセプト			
	apple	apple tart	ripe apple	homemade apple tart
キーワード全体	✓			
語頭		✓		
語末			✓	
語頭あるいは語末		✓	✓	
完全かつ語頭	✓	✓		
完全かつ語末	✓		✓	
完全かつ (語頭あるいは語末)	✓	✓	✓	
どこでも		✓	✓	✓
完全かつどこか	✓	✓	✓	✓
全体 (複合語なし)	✓	抽出なし	抽出なし	抽出なし

### [活用] 列

この列では、抽出エンジンが抽出時にこのキーワードの活用形を生成し、すべてをグループ化するかどうかを選択します。この列のデフォルト値は [タイプのプロパティ] で定義されますが、場合によっては、列で直接このオプションを変更できます。メニューから [編集] → [活用形を変更] を選択します。



### [タイプ] 列

この列で、ドロップダウン リストからキーワード辞書を選択します。タイプのリストは、ライブラリ ツリー パネルでの選択に応じてフィルタリングされます。リストの最初のタイプは、ライブラリ ツリー パネルで選択されたデフォルト タイプです。メニューから [編集] → [タイプを変更] を選択します。

### [ライブラリ] 列

キーワードが格納されているライブラリが表示されます。ライブラリ ツリー パネルでキーワードを別のタイプにドラッグ アンド ドロップして、そのライブラリを変更できます。

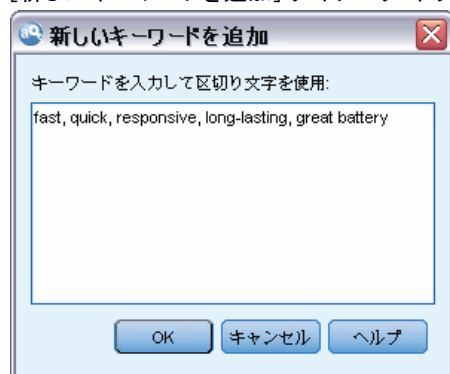
#### キーワード辞書に1つのキーワードを追加するには

- ▶ ライブラリ ツリー パネルで、キーワードを追加したいキーワード辞書を選択します。
- ▶ 中央のパネルのキーワード リストで、使用できる最初の空白セルにキーワードを入力し、このキーワードに必要なオプションを設定します。

#### キーワード辞書に複数のキーワードを追加するには

- ▶ ライブラリ ツリー パネルで、キーワードを追加したいキーワード辞書を選択します。
- ▶ メニューの [ツール] → [新規キーワード] を選択します。[新しいキーワードを追加] ダイアログ ボックスが開きます。

図 10-4  
[新しいキーワードを追加] ダイアログ ボックス



- ▶ キーワードを入力するか、キーワードのセットをコピーして貼り付けて、選択したキーワード辞書に追加したいキーワードを入力します。複数のキーワードを入力する場合、[オプション] ダイアログで定義された区切り文字を使用してキーワードを区切るか、各キーワードを新しい行に追

加するつようがあります。詳細は、2 章 p.19 [オプションの設定](#) を参照してください。

- ▶ [OK] をクリックすると、キーワードが辞書に追加されます。マッチ オプションは、このキーワード ライブラリのデフォルトのオプションに自動的に設定されます。ダイアログ ボックスが閉じ、辞書に新しいキーワードが表示されます。

## キーワードの強制

キーワードを特定のタイプに割り当てる場合、対応するキーワード辞書に追加することができます。ただし、同じ名前を持つ複数のキーワードがある場合、抽出エンジンはどのタイプを使用するかを認識する必要があります。そのため、使用するタイプを選択するよう要求するメッセージが表示されます。この操作をタイプへのキーワードの**強制**といいます。このオプションは、コンパイル済み（内部、編集不可能）辞書からのタイプの割り当てを上書きする場合に役立ちます。通常、キーワードが重複しないようにすることをお勧めします。

強制しても、このキーワードの他の出現を「削除」するわけではありません。抽出エンジンによって無視されます。キーワードを強制または強制を解除することによって、使用する出現を後で変更することができます。また、パブリック ライブラリを追加またはパブリック ライブラリを更新する場合、キーワードをキーワード辞書に投入することが必要な場合もあります。

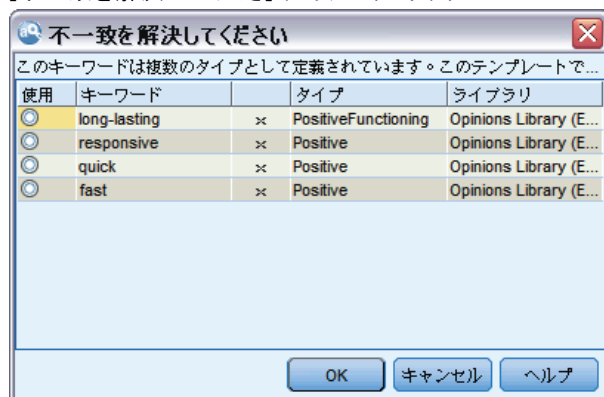
図 10-5  
強制の状態のアイコン

キーワード	かな	タイプ	感性タイプ	ライブラリ
魚は新鮮だ	さかなはしんせんた		良い - 良い	Opinions 感性意見
馬鹿にされた	ばかにされた		悪い - 怒り	Opinions 感性意見
肌えに苦しむ	うえにくるしむ		悪い - 苦しい	Opinions 感性意見
風邪を引いた	かぜをひいた		悪い - 体が悪い状態	Opinions 感性意見
頭が痛い	あたまがいたい		悪い - 体が悪い状態	Opinions 感性意見
頑張ってください	がんばってください		その他 - 激励	Opinions 感性意見
韓流ブームです	はんりゅうぶームです		良い - 好評・人気	Opinions 感性意見
面白味を感じる	おもしろみをかんじる		良い - 楽しみ全般	Opinions 感性意見
間違が多い	まちがいがおおい		悪い - 悪い	Opinions 感性意見
配慮が行き届いていた	はいりよがいきとどいてい		良い - 対応が親切	Opinions 感性意見

キーワード パネルの 2 番目の列、[強制] 列で、度のキーワードが強制されているか、または無視されているかを確認できます。押しピンのアイコンが表示されている場合、キーワードのこの出現が強制されていることを示します。黒い X のアイコンがキーワードの後に表示されている場合、他の場所で強制されているため、キーワードのこの出現は抽出時に無視されることを示します。また、キーワードを強制すると、強制されたタイプの色で表示されます。**タイプ 1** および **タイプ 2** のキーワードが **タイプ 1** に強制投入されると、**タイプ 1** に定義されたフォントの色で表示されます。

アイコンをダブルクリックして、状態を変更できます。キーワードが他の場所で表示されている場合、[不一致を解決してください] ダイアログボックスが開き、使用する出現を選択できます。

図 10-6  
[不一致を解決してください] ダイアログ ボックス



## タイプの名前の変更

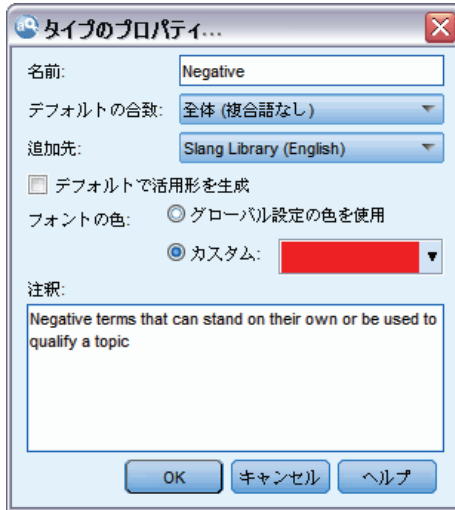
[タイプのプロパティ] を編集して、キーワード辞書の名前を変更したり、その他の辞書の設定を変更することができます。

**重要!** 複数のタイプ名が同じ単語で始まる場合は特に、タイプ名にスペースを使用しないことをお勧めします。コア ライブラリまたは意見ライブラリでタイプの名前を変更したり、デフォルトの合致属性を変更しないことをお勧めします。

### タイプの名前を変更するには

- ▶ ライブラリ ツリー パネルで、名前を変更したいキーワード辞書を選択します。
- ▶ マウスを右クリックし、コンテキスト メニューから [タイプのプロパティ] をクリックします。[タイプのプロパティ] ダイアログ ボックスが開きます。

図 10-7  
[タイプのプロパティ] ダイアログ ボックス



- ▶ [名前] テキスト ボックスにキーワード辞書の新しい名前を入力します。
- ▶ [OK] をクリックして、新しい名前を承認します。新しいタイプ名がライブラリ ツリー パネルに表示されます。

## タイプの移動

キーワード辞書をライブラリ内の別の場所、またはツリー内の別のライブラリにドラッグすることができます。

注: ビルトインのタイプは移動しないことをお勧めします。

### ライブラリ内のタイプの順序を変更するには

- ▶ ライブラリ ツリー パネルで、移動したいキーワード辞書を選択します。
- ▶ メニューで [編集] → [1つ上に移動] を選択すると、ライブラリ ツリー パネルでキーワード辞書が 1 つ上の位置に移動します。[編集] → [1つ下に移動] を選択すると 1 つ下の位置に移動します。

### タイプを別のライブラリに移動するには

- ▶ ライブラリ ツリー パネルで、移動したいキーワード辞書を選択します。
- ▶ マウスを右クリックし、コンテキスト メニューから [タイプのプロパティ] をクリックします。[タイプのプロパティ] ダイアログ ボックスが開きます。(タイプを別のライブラリにドラッグ アンド ドロップすることができます)。

- ▶ [追加先] リスト ボックスで、キーワード辞書を移動したいライブラリを選択します。
- ▶ [OK] をクリックします。ダイアログ ボックスが閉じ、タイプが選択したライブラリに移動します。

## タイプの無効化および削除

キーワード辞書を一時的に削除したい場合、ライブラリ ツリー パネルの辞書名の左側にあるチェック ボックスをオフにして無効化することができます。これは、ライブラリ内に辞書を保存したいが、競合を検証する場合および抽出プロセス時に内容を無視することを示します。

ライブラリからキーワード辞書を永続的に削除することもできます。

注: コア ライブラリまたは意見ライブラリのビルトインのタイプは削除しないことをお勧めします。代わりに、無効化することをお勧めします。

### キーワード辞書を無効化するには

- ▶ ライブラリ ツリー パネルで、無効化したいキーワード辞書を選択します。
- ▶ スペースバーをクリックします。タイプ名前の左側にあるチェック ボックスがオフになります。

### キーワード辞書を削除するには

- ▶ ライブラリ ツリー パネルで、削除したいキーワード辞書を選択します。
- ▶ メニューから、[編集] → [削除] を選択すると、キーワード辞書が削除されます。

## 類義語辞書

類義語辞書は、1 つの代表語に基づいて、類似したキーワードをグループ化できるキーワードの集合です。類義語辞書は、[ライブラリ リソース] タブの一番下のパネルで管理されます。 **ビュー>リソースエディタ** (メニュー)。この辞書では 2 つの形式の類義語を定義できます: **類義語**および**オプションの要素**。このパネルでタブをクリックして切り替えることができます。

テキスト データの抽出を実行した後、他のコンセプトの類義語または活用形であるコンセプトをいくつか見つけることができます。オプションの要素および類義語を特定して、抽出エンジンがこれらを 1 つの代表語にマップされるよう強制できます。

類義語やオプションの要素を使用すると、より頻度の高い数のより重要で代表的なコンセプトに結合することによって、[抽出結果] パネルのコンセプト数を削減することができます。

図 10-8  
[類義語辞書] パネル

	代表語	類義語	ライブラリ
0	<input checked="" type="checkbox"/> answer to problem	actions to correct the errors, address the problem, address the problem(s), address your problem, address the problem, address your problems, answer to problem, handling of this issue, help with problems, problems addresses	Customer Satisfaction Library
1	<input checked="" type="checkbox"/> ideal	ideal	Customer Satisfaction Library
2	<input checked="" type="checkbox"/> in-store service	customer service at the store, customer service at the stores, customer service in the store, customer service reps in the store, everyone in the store, experience at the store, girl at the store, guys in the store, in store customer service, in-store tech representative, in-store technical representative, lady at the store, man at the store, people at the store, people in the store, people that work there, representative at the store, reps in the store, reps to assist customer, retail service person, service at the store, service center,	Customer Satisfaction Library Customer Satisfaction Library
4			

類義語 オプションの要素

## 類義語

類義語とは、同じ意味を持つ複数の語を関連付けたものです。また、類義語はキーワードを略語とグループ化したり、一般的にスペルミスのある単語と正しいスペルの単語とをグループ化したりするために使用できます。[類義語] タブでこれらの類義語を定義できます。

類義語定義は、2 つの部分で構成されています。1 つ目は代表語で、抽出エンジンがすべての類義語キーワードをグループ化する基準となるキーワードです。この代表語が別の代表語の類義語として使用されていない限り、または不要語として登録されていない限り、[抽出結果] パネルに表示されるコンセプトとなります。2 つめは、代表語の下にグループ化される類義語のリストです。

たとえば、automobile という単語を vehicle という単語に置き換えたい場合、automobile が類義語となり、vehicle が代表コンセプトとなります。

[類義語] 列にはどんな言葉も入力できますが、その語が抽出時に見つからず、キーワードのマッチ オプションが [キーワード全体] である場合、類義語は出現しません。ただし、類義語をこのキーワードの下にグループ化するために、代表語を抽出する必要がありません。

図 10-9  
類義語辞書、[類義語] タブ

	代表語	類義語	ライブラリ
0	<input checked="" type="checkbox"/> busy	busy, busy	CRM Library (English)
1	<input checked="" type="checkbox"/> lessee	lessee	CRM Library (English)
2	<input checked="" type="checkbox"/> inmessage left	am im tcm, call left message, call im, call intc, call intcb, called and left message, called and im, im, im am, im ans mach, im call, im pager, im pg, im to cfb, im to call, imess, imsg, leave message on her answer machine, left a brief message, left a message, left a message on answer machine, left a message on her answer machine, left a message on his answer machine, left a message on his answer machine, left a message on his answer phone, left a message to callback on answer phone, left another message, left her a message on her answer phone, left her a message on her answerphone, left mes, left mess, left message,	CRM Library (English)

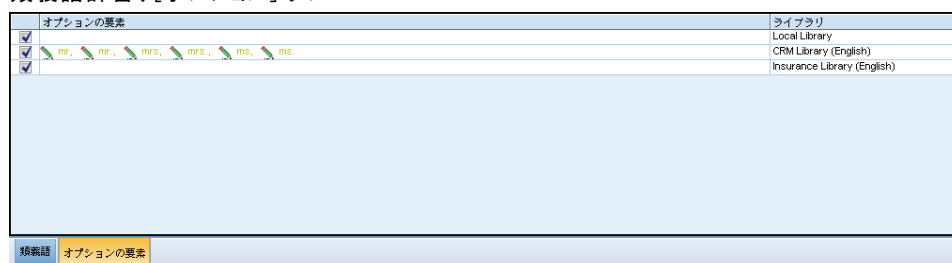
類義語 オプションの要素

## オプションの要素

オプションの要素は、テキスト内で若干異なったように出現する場合でも類義語を保持するために抽出時に無視できる結合キーワード内のにんいの語を示します。オプションの要素は単語で、結合語から削除された場合、別のキーワードと合致を作成できます。これらの単語は、結合語内のどこでも（語頭、語中、語末）出現します。[オプション] タブでオプションの要素を定義できます。

たとえば、ibm および ibm corp のキーワードをグループ化する場合、corp をオプションの要素として扱うよう宣言する必要があります。また別の例を示すと、キーワード access をオプションの要素として指定し、抽出時に internet access speed および internet speed が見つかった場合、最も頻繁に出現するキーワードに基づいてグループ化されます。

図 10-10  
類義語辞書、[オプション] タブ



## 類義語の定義

[類義語] タブで、テーブルの一番上の空白行に類義語定義を入力できます。まず代表語とその類義語を定義します。この定義を格納するライブラリも選択できます。抽出時、類義語のすべての出現を、最終的な抽出で代表語に基づいてグループ化します。詳細は、p. 240 [キーワードを追加](#) を参照してください。

たとえば、テキスト データに多数の電話通信関連の情報が含まれている場合、cellular phone、wireless phone、mobile phone という 3 つのキーワードが取得されるとします。この場合、cellular および mobile を wireless の類義語として定義する必要があります。これらの類義語を定義すると、cellular phone および mobile phone の抽出されたすべての出現は、wireless phone と同じキーワードとして扱われ、キーワード リストにいっしょに表示されます。

キーワード辞書を作成している場合、キーワードを入力し、そのキーワードに対して 3 つまたは 4 つの類義語を考えることができます。この場合、類義語辞書にすべてのキーワードと代表語を入力し、類義語をドラッグすることができます。

類義語は、類義語の活用形（複数形など）にも適用されます。コンテキストに応じて、キーワードをどのように代用するかについて制限が必要な場合があります。特定の文字を使用して、類義語の処理の程度に制限を加えることができます。

- **感嘆符 (!):** `!synonym` のように類義語の前に直接感嘆符を追加すると、`!` の意義のの活用形は代表語によって代用されないことを示します。ただし、`!target-term` のように代表語の前に感嘆符を追加すると、複合代表語の一部または変異形がさらる類義語を受け入れないようにすることを示します。
- **アスタリスク (\*):** `*synonym*` のようにアスタリスクを類義語の直後に置くと、この語を代表語に置き換えることを示します。たとえば、`manage*` を類義語に、`management` を代表語に定義すると、`associate managers` が代表語 `associate management` に置き換えられます。また、`internet *` のようにごとアスタリスクの間にスペースを追加することもできます (`*synonym *`)。代表語に `internet`、類義語に `internet * *` および `web *` を定義すると、`internet access card` および `web portal` は `internet` に置き換えられます。この辞書では、語または文字列の頭をアスタリスクにすることはできません。
- **カレット (^):** `^synonym` のように類義語の前にカレットとスペースを追加すると、キーワードが類義語から始まる場合にのみ、類義語のグループ化が適用されることを示します。たとえば、`^ wage` を類義語に、`income` を代表語に定義して両方が抽出される場合、キーワード `income` に基づいてグループ化されます。ただし、`minimum wage` と `income` が抽出されると、`minimum wage` が `wage` で始まっていないため、それらはいっしょにグループ化されません。この記号と類義語の間にスペースを追加する必要があります。
- **ドル記号 (\$):** `$synonym $` のように類義語の後にスペースドル記号を追加すると、キーワードが類義語で終わる場合にのみ、類義語のグループ化が適用されることを示します。たとえば、`cash $` を類義語に、`money` を代表語に定義して両方が抽出される場合、キーワード `money` に基づいてグループ化されます。ただし、`cash cow` と `money` が抽出されると、`cash cow` が `cash` で終わっていないため、それらはいっしょにグループ化されません。この記号と類義語の間にスペースを追加する必要があります。
- **カレット (^) およびドル記号 (\$):** カレットとドル記号を `^synonym $` のように同時に使用すると、完全一致の場合にのみキーワードが類義語と合致します。つまり、類義語のグループ化が行われるよう、抽出されたキーワードの類義語の前後に語があってはいけないことを意味します。たとえば、`^ van $` を類義語に、`truck` を代表語に定義すると、`van` は `truck` とグループ化されますが、`marie van guerin` がそのままになります。また、カレットとドル記号を使用して類義語を定義して、この語がソース テキスト内のどこにでも出現する場合、類義語は自動的に抽出されます。



図 10-11  
類義語辞書、[類義語] タブの例

	代表語	類義語	ライブラリ
0			
1	busy	busyu, busyy	CRM Library (English)
2	message	lcses	CRM Library (English)
	message left	adm lntcmb, call left message, call lm, call lntc, call lntcb, called and left message, called and lm, lm, lm alm, lm ans mach, lm call, lm pager, lm pg, lm to ctb, lm to call, lmsg, lmsg, leave message on her answer machine, left a brief message, left a message, left a message on answer machine, left a message on her answer machine, left a message on his answer machine, left a message on his answer machine, left a message on ins answer phone, left a message to callback on answer phone, left another message, left her a message on her answer phone, left her a message on her answerphone, left mes, left mess, left message.	CRM Library (English)

### 類義語エントリを追加するには

- ▶ 類義語パネルを表示し、左下の [類義語] タブをクリックします。
- ▶ テーブルの一番上の空白行で、[代表語] 列に代表語を入力します。入力した代表語が色つきで表示されます。この色は、キーワードが表示されるまたは強制されるタイプを示します。キーワードが黒で表示されている場合、キーワード辞書にはこのキーワードがないことを意味します。
- ▶ 代表語の右側の 2 番目のセルをクリックして、類義語のセットを入力します。[オプション] ダイアログ ボックスで定義したグローバル区切り文字を使用して、各エントリを区切ります。詳細は、2 章 p.19 [オプションの設定](#) を参照してください。入力したキーワードが色つきで表示されます。この色は、キーワードが出現するタイプを示します。キーワードが黒で表示されている場合、キーワード辞書にはこのキーワードがないことを意味します。
- ▶ 最後のセルをクリックして、この類義語定義を格納するライブラリを選択します。

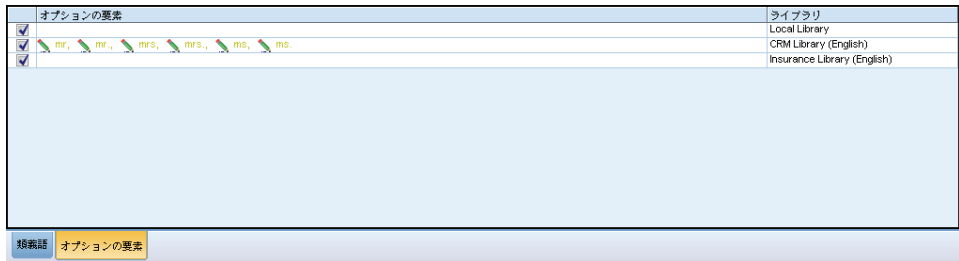
注:ここでは、リソース エディタ ビューでどのように変更を行うかについて説明します。[抽出結果] パネルまたは [データ] パネルで直接、この種類の調整を行うこともできます。詳細は、5 章 p.91 [抽出結果の調整](#) を参照してください。

## オプションの要素の定義

[オプション] タブで、必要なライブラリのオプションの要素を定義します。これらのエントリは、各ライブラリでグループ化されます。ライブラリがライブラリ ツリー パネルに追加されると、空白のオプションの要素行が [オプション] タブに追加されます。

すべてのエントリが、自動的に小文字の語に変換されます。抽出エンジンは、エントリをテキスト内の小文字および大文字の語に合致させます。

図 10-12  
類義語辞書、[オプション] タブ



注:[オプション] ダイアログ ボックスで定義した区切り文字を使用して、キーワードを区切ります。詳細は、2 章 p.19 [オプションの設定](#) を参照してください。入力しているオプションの要素にキーワードの一部と同じ区切り文字が含まれている場合、その前にバックスラッシュを追加する必要があります。

### エントリを追加するには

- ▶ 類義語パネルを表示し、エディタの左下にある [オプション] タブをクリックします。
- ▶ このエントリを追加するライブラリの [オプションの要素] 列のセルをクリックします。
- ▶ オプションの要素を入力します。[オプション] ダイアログ ボックスで定義したグローバル区切り文字を使用して、各エントリを区切ります。詳細は、2 章 p.19 [オプションの設定](#) を参照してください。

## 類義語の無効化および削除

辞書でエントリを無効化して、一時的に削除することができます。エントリを無効化すると、そのエントリは抽出時に無視されます。

類義語辞書の古いエントリを削除することもできます。

### エントリを無効化するには

- ▶ 辞書で、無効化したいエントリを選択します。

- ▶ スペースバーをクリックします。エントリの左側にあるチェックボックスがオフになります。

注: エントリの左側にあるチェックボックスをオフにして、無効化することもできます。

#### 類義語エントリを削除するには

- ▶ 辞書で、削除したいエントリを選択します。
- ▶ メニューから、[編集] → [削除] を選択または Del キーを押すと、キーワード辞書が削除されます。エントリは辞書内から除外されます。

#### オプションの要素エントリを削除するには

- ▶ 辞書で、削除したいエントリをダブルクリックします。
- ▶ キーワードを手動で削除します。
- ▶ Enter を押して変更を適用します。

## 不要語辞書

不要語辞書は、語、句、一部の文字列のリストです。不要語辞書のエントリに合致またはエントリを含むキーワードは無視されるか、抽出から除外されます。不要語辞書は、エディタの右側のパネルで管理されます。通常、このリストに追加するキーワードは、穴埋めのための単語または句で、一貫性を維持するためにテキストで使用されテキストにとって重要でなく、抽出結果を混乱させる場合があります、こういったキーワードを不要語辞書に追加することで、全く抽出されないようにできます。

不要語辞書は、エディタの [ライブラリ リソース] タブの右上のパネルで管理されます。 **ビュー > リソース エディタ (メニュー)**。

図 10-13  
[不要語辞書] パネル

	不要語リスト	ライブラリ
0	<input type="checkbox"/>	
1	<input checked="" type="checkbox"/> problem description	CRM Library (English)
2	<input checked="" type="checkbox"/> in respect of	Insurance Library (English)
3	<input checked="" type="checkbox"/> n2	Insurance Library (English)
4	<input checked="" type="checkbox"/> see event log	Insurance Library (English)

不要語辞書で、テーブルの一番上の空白行に語、句、または一部の文字列を入力できます。文字列を不要語辞書に 1 つまたは複数の語として、またはアスタリスクをワイルドカードとして使用し、単語の一部として追加することができます。不要語辞書で宣言されたエントリを使用して、コンセ

プトを抽出から除外します。エントリが、キーワード辞書などインターフェイスの別の場所でも宣言されている場合、他の辞書では、現在除外されていることを示す取り消し線が表示されます。この文字列は、テキストデータに出現する必要はなく、また適用されるキーワード辞書の一部として宣言する必要はありません。

注: 類義語エントリで代表語としても機能するコンセプトを不要語辞書に追加すると、代表語およびそのすべての類義語も不要語として登録されます。詳細は、 p. 249 [類義語の定義](#) を参照してください。

### ワイルドカード (\*) の使用

不要語のエントリを一部の文字列として扱うことを示すためにアスタリスクのワイルドカードを使用します。抽出エンジンで見つかった、不要語辞書で入力した文字列で始まるまたは終わる語を含むキーワードは、最終的な抽出からは除外されます。ただし、ワイルドカードの使用が認められない場合が 2 つあります。

- \*- のようにアスタリスクの前にダッシュ (-) が追加されている場合
- \*' s のようにアスタリスクの前にアポストロフィー ( ' ) が追加されている場合

テーブル 10-2  
不要語エントリの例

エントリ	例	結果
語	next	next という語が含まれている場合、コンセプト (またはそのキーワード) は抽出されません。
句	for example	for example という句が含まれている場合、コンセプト (またはそのキーワード) は抽出されません。
一部	copyright*	copyrighted、copyrighting、copyrights、または copyright 2010 のように、copyright の変異形に合致または変異形を含むコンセプト (またはそのキーワード) は不要語として登録されます。
一部	*ware	freeware、shareware、software、hardware、beware、または silverware のように、ware の変異形に合致または変異形を含むコンセプト (またはそのキーワード) は不要語として登録されます。

### エントリを追加するには

- ▶ テーブルの一番上の空白行で、キーワードを入力します。入力したキーワードが色つきで表示されます。この色は、キーワードが出現するタイプを示します。キーワードが黒で表示されている場合、キーワード辞書にはこのキーワードがないことを意味します。

### エント리를無効化するには

不要語辞書でエント리를無効化して、エント리를一時的に削除することができます。エント리를無効化すると、そのエント리는抽出時に無視されます。

- ▶ 不要語辞書で、無効化したいエント리를選択します。
- ▶ スペースバーをクリックします。エント리의左側にあるチェック ボックスがオフになります。

注:エント리의左側にあるチェック ボックスをオフにして、無効化することもできます。

### エント리를削除するには

不要語辞書の不要なエント리를削除することもできます。

- ▶ 不要語辞書で、削除したいエント리를選択します。
- ▶ メニューから [編集] → [削除] を選択します。エント리는辞書内から除外されます。

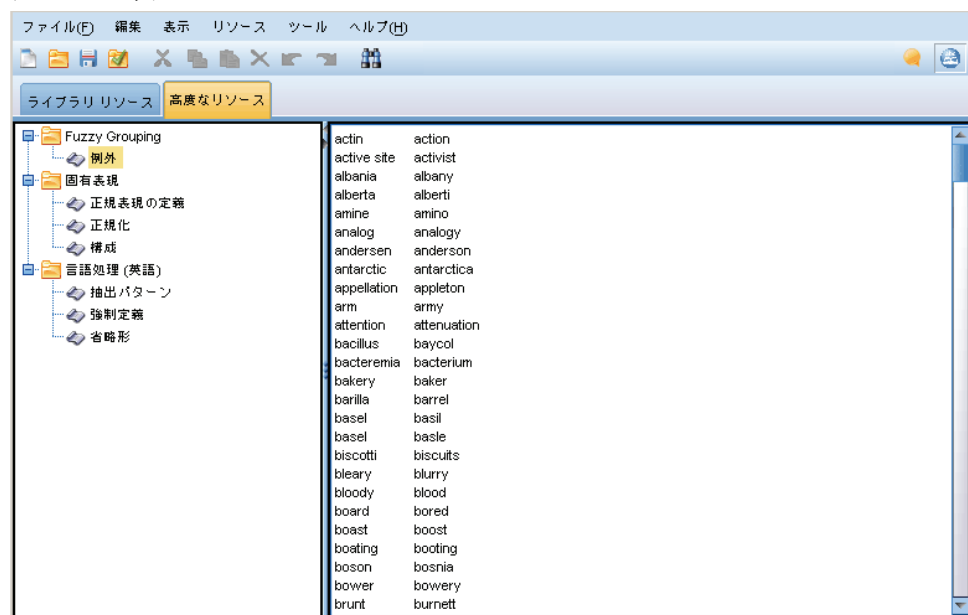
# アドバンス リソースについて

キーワード辞書、不要語辞書および類義語辞書のほか、Fuzzy Grouping 設定や固有表現キーワード辞書など、さまざまなアドバンス リソースの設定を使用することもできます。これらのリソースは、リソース エディタビューの [高度なリソース] タブで処理することができます。すべてのプロジェクトのデフォルトとして変更を保存することも、元の内容に戻すこともできます。

[アドバンス リソース] タブに移動すると、次の情報を編集できます。

- **Fuzzy Grouping (例外):** Fuzzy Grouping (スペル修正) アルゴリズムから単語のペアを除外するために使用します。詳細は、p. 258 [Fuzzy Grouping](#) を参照してください。
- **固有表現:** 抽出時に適用される正規表現や正規化規則のほか、どの固有表現を抽出できるかを有効化および無効化するために使用します。詳細は、p. 260 [固有表現](#) を参照してください。
- **言語処理:** 文を構築する (抽出パターンおよび強制定義) そして選択した言語の略語を使用する特別な方法を宣言するために使用します。詳細は、p. 265 [言語処理](#) を参照してください。

図 11-1  
アドバンス リソース



注: 検索/置換ツールバーを使用して、情報を迅速に検索したり、一様の変更をセクションに行ったりすることができます。詳細は、 p. 257 [置換](#) を参照してください。

### アドバンスリソースを編集するには

- ▶ 編集するリソース セクションを検索および選択します。内容が右側のパネルに表示されます。
- ▶ メニューまたはツールバーのボタンを使用して、必要に応じて内容を切り取り、コピー、または貼り付けることができます。
- ▶ このセクションで、書式設定規則を使用して、変更したいファイルを編集します。編集を行うとすぐに、変更が保存されます。ツールバーの取り消しまたはやり直しの矢印を使用して、以前の変更に戻ります。

## 検索

特定のセクションで、情報の迅速な検索が必要な場合があります。検索機能を使用して、特定の規則をすぐに見つけることができます。セクション内の情報を検索するために、検索ツールバーを使用できます。

図 11-2  
検索ツールバー



### 検索機能を使用するには

- ▶ 検索するリソース セクションを検索および選択します。内容がエディタの右側のパネルに表示されます。
- ▶ メニューから [編集] → [検索] を選択します。[アドバンス リソースを編集] ダイアログ ボックスの右上に、検索ツールバーが表示されます。
- ▶ テキスト ボックスに検索したい文字列を入力します。ツールバー ボタンを使用して、大文字と小文字の区別、部分一致、検索の方向を制御します。
- ▶ [検索] を選択して、検索を開始します。一致が見つかった場合は、テキストがウィンドウで強調表示されます。
- ▶ [検索] もう一度クリックすると、次の一致を検索します。

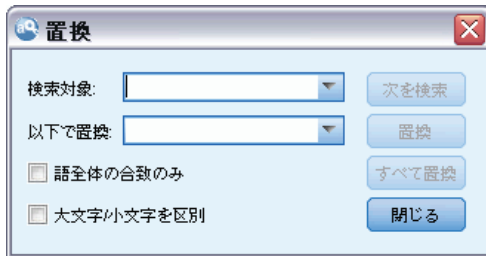
## 置換

アドバンス リソースへより広い更新が必要な場合があります。置換機能を使用すると、内容に一様の更新を行うことができます。

### 置換機能を使用するには

- ▶ 検索および置換するリソース セクションを検索および選択します。内容がエディタの右側のパネルに表示されます。
- ▶ メニューから [編集] → [置換] を選択します。[置換] ダイアログ ボックスが開きます。

図 11-3  
[置換] ダイアログ ボックス



- ▶ [検索対象] テキスト ボックスで、検索する文字列を入力します。
- ▶ [以下で置換] テキスト ボックスで、見つかったテキストの代わりに使用する文字列を入力します。
- ▶ 完全な語のみを検索または置換する場合、[語全体の合致のみ] を選択します。
- ▶ 大文字と小文字が完全に一致する語のみを検索または置換する場合、[大文字/小文字を区別] を選択します。
- ▶ [次を検索] を選択して、一致を検索します。一致が見つかった場合は、テキストがウィンドウで強調表示されます。この一致を置換したくない場合は、置換したい一致が見つかるまで [次を検索] をクリックします。
- ▶ [置換] を選択して、選択した一致を置換します。
- ▶ [置換] を選択して、セクション内のすべての一致を置換します。置換が行われた数を示したメッセージが表示されます。
- ▶ 置換が終了したら、[閉じる] をクリックします。ダイアログ ボックスが閉じます。

注: 誤って置換を行った場合、ダイアログ ボックスを閉じ、メニューの [編集] → [取り消し] を選択して、置換を取り消すことができます。取り消したい変更ごとに 1 回ずつ実行する必要があります。

## Fuzzy Grouping

[抽出設定] ダイアログで、[語幹文字数が次の最小値以上のときにスペルを調整する] を選択している場合、Fuzzy Grouping アルゴリズムが有効化されます。



Fuzzy Grouping を使用すると、抽出語から最初の母音を除くすべての母音および二重および三重の子音を一時的に取り除いて比較し、残りが同じかどうかを確認して、一般的にスペルミスのある語やスペルの近い語をグループ化することができます。抽出プロセス時、Fuzzy Grouping 機能は、抽出キーワードに適用され、結果を比較して、一致があるかどうかを判断します。一致があった場合、元のキーワードは最終的な抽出リストにいっしょにグループ化されます。それらのキーワードは、データ内で最も頻繁に出現するキーワードに基づいてグループ化されます。

注:比較されている 2 つのキーワードが <Unknown> タイプを除く異なるタイプに割り当てられている場合、このペアに Fuzzy Grouping 手法は適用されません。つまり、この手法を適用するには、キーワードが同じタイプまたは <Unknown> タイプに割り当てられている必要があります。

この機能を有効にして、類似したスペルの 2 つのキーワードが不正にグループ化されていることがわかった場合、それらを Fuzzy Grouping から除外する必要があります。不正に一致したペアを [アドバンス リソース] タブの [例外] セクションに入力することによって、Fuzzy Grouping から除外することができます。詳細は、p. 256 [アドバンス リソースについて](#) を参照してください。

次の例は、Fuzzy Grouping がどのように実行されるかについて説明しています。Fuzzy Grouping が有効な場合、これらの語が同じであると示され、以下のように一致します。

color -> colr	mountain -> montn
colour -> colr	montana -> montn
modeling -> modlng	furniture -> furntr
modelling -> modlng	furnature -> furntr

前者の例では、mountain および montana をグループ化から除外することが考えられます。次のように [例外] セクションに入力します。

```
mountain      montana
```

**重要!** Fuzzy Grouping の例外では、特定の類義語規則が適用されているため、2 つの単語がグループ化される場合があります。この場合、感嘆符のワイルドカード (!) を使用して類義語を入力し、単語が出力で類義語とならないようにする必要があります。詳細は、10 章 p. 249 [類義語の定義](#) を参照してください。

### Fuzzy Grouping の例外の書式規則

- 1 行につき 1 つの例外ペアのみを定義する。
- 単語または複合語を使用する。
- 語には小文字の実を使用する。大文字は無視されます。
- ペア内の各語を区切るには TAB 文字を使用する。

## 固有表現

特定の種類のデータを扱っている場合、日付、社会保障番号、またはその他の固有表現を抽出したいと考える場合があります。これらのエンティティは、エンティティを有効化または無効化できる構成ファイルで明示的に宣言されています。詳細は、p. 264 [構成](#) を参照してください。抽出エンジンから出力を最適化するために、非言語処理からの入力を正規化し、事前に定義された形式に従って同様のエンティティをグループ化します。詳細は、p. 263 [正規化](#) を参照してください。

注:抽出設定で、固有表現の抽出を有効にしたり無効にしたりできます。

### 使用できる固有表現

次の表の固有表現を抽出できます。カッコ内はタイプ名です。

住所 (<Address>)	組織名 (<Organization>)
アミノ酸 (<Aminoacid>)	割合 (<Percent>)
通貨 (<Currency>)	商品名 (<Product>)
日付 (<Date>)	プロテイン (<Gene>)
遅延 (<Delay>)	電話番号 (<PhoneNumber>)
数値 (<Digit>)	時間 (<Time>)
電子メール アドレス (<email>)	米国社会保障番号 (<SocialSecurityNumber>)
HTTP/URL アドレス (<url>)	計量 (<Weights-Measures>)
IP アドレス (<IP>)	

### 処理するテキストの削除

固有表現の抽出を行う前に、入力テキストが削除されます。この手順で、次の一時的な変更が行われ、固有表現が以下のようなものとして特定および抽出されます。

- 複数のスペースの行列は、1 つのスペースに置き換えられます。
- 表形式はスペースに置き換えられます。
- 単一の行末文字または配列文字はスペースの置き換えられ、複数の行末文字の配列はパラグラフの最後としてマークされます。行末は、復帰改行 (CR) および改行 (LF) またはそれらの両方で示されます。
- HTML タグおよび XML タグは一時的に除外および無視されます。

## 正規表現の定義

固有表現抽出時、正規表現の特定に使用される正規表現定義の編集またはそれへの追加が必要な場合があります。これは [アドバンス リソース] タブの [正規表現の定義] セクションで行われます。詳細は、p. 256 [アドバンス リソースについて](#) を参照してください。

ファイルはいくつかのセクションに分割されます。最初のセクションは [macros] です。このセクションのほか、固有表現ごとにセクションが存在する場合があります。このファイルにセクションを追加できます。各セクション内で、規則には番号が付けられています (regexpl、regexp2、など)。これらの規則は 1 ~ n の順番に番号を付けなければなりません。番号が欠けていると、このファイルの処理が一度に中断します。

エンティティが言語に依存する場合があります。構成ファイルの言語パラメータの値が 0 以外の場合、エンティティは言語に依存すると見なされます。詳細は、p. 264 [構成](#) を参照してください。エンティティが言語に依存する場合、[english/PhoneNumber] のようにセクション名の前に言語を示す必要があります。PhoneNumber エンティティの言語の値に 2 が指定されている場合、このセクションには英語の電話番号にのみ適用される規則が含まれます。

**重要!** エディタでこのファイルや他のファイルに変更を行い、抽出エンジンが必要に応じて機能しない場合、ツールバーの [元のものにリセット] オプションを使用して、ファイルを付属の下の内容に戻します。このファイルは、正規表現に対する特定のレベルの親密度が必要です。この領域においてさらに支援が必要な場合、IBM Corp. にご連絡ください。

### 特殊文字: []()¥\*+?|^\$

以下の特殊文字を除くすべての文字はそれ自身に一致します。これらの特殊文字は、表現内の特別な目的に使用されます。.[{}()\\*\+?|^\$これらの文字をその文字として使用するには、定義で文字の前にバックスラッシュ (\) を追加する必要があります。

たとえば、Web アドレスを抽出しようとしている場合、終止符はエンティティにとって非常に重要な文字であるため、次のようにバックスラッシュを追加する必要があります。

```
www\.[a-z]+\.[a-z]+
```

### 繰り返し演算子および識別子: ? + \* []

定義をより柔軟性のあるものにするために、正規表現に標準的なワイルドカードをいくつか使用できます。使用できるのは \* ? +

- アスタリスク \* は、0 以上の先行文字列があることを示します。たとえば、ab\*c は、「ac」、「abc」、「abbbc」などに一致します。
- プラス記号 + は、1 つ以上の先行文字列があることを示します。

たとえば、`ab+c` は、「`abc`」、「`abbc`」、「`abbbc`」などに一致しますが、「`ac`」には一致しません。

- 疑問符 `+` は、0 または 1 つの先行文字列があることを示します。たとえば、`modell?ing` は、「`modeling`」、「および「`modeling`」のどちらにも一致します。
- 繰り返し制限を示す中カッコ `{}` は、繰り返しの境界を示します。たとえば、
  - ▶ `[0-9]{n}` は、ちょうど `n` 回繰り返された数値に一致します。たとえば、`[0-9]{4}` は“1998”に一致し、“33”および“19983”には一致しません。
  - ▶ `[0-9]{n,}` は、`n` 回以上繰り返された数値に一致します。たとえば、`[0-9]{3,}` は“199”または“1998”に一致しますが“19”には一致しません。
  - ▶ `[0-9]{n,m}` は、`n ~ m` 回繰り返された数値に一致します。たとえば、`[0-9]{3,5}` は“199”、“1998”または“19983”に一致しますが、“19”および“199835”には一致しません。

### 任意のスペースおよびハイフン

定義内に任意のスペースを追加したい場合があります。たとえば、「`uruguayan pesos`」、「`uruguayan peso`」、「`uruguay pesos`」、「`uruguay peso`」、「`pesos`」または「`peso`」などの通貨を抽出したい場合、スペースで区切られた 2 つの単語があるという事実に対処する必要があります。この場合、この定義は `(uruguayan |uruguay )?pesos?` として記述されます。`pesos/peso` と共に使用する場合、`uruguayan` または `uruguay` の後にはスペースが続くため、任意のスペースは、任意の行列 (`uruguayan |uruguay`) 内で定義される必要があります。`(uruguayan|uruguay)? pesos?` のように任意の行列内にスペースがない場合、スペースが必要であるため「`pesos`」または「`peso`」とは一致しません。

リスト内にハイフンを含む一連の文字を探している場合、ハイフンを最後に定義する必要があります。たとえば、カンマ (,) またはハイフン (-) を検索する場合、`[,-]` を使用し、`[-,]` は決して使用しません。

### リストおよびマクロの文字列の順序

短い行列の前に最も長い行列を定義する必要があります。短い行列に一致があるため、最も長い行列が読み取られなくなるためです。たとえば、「`billion`」または「`bill`」という文字列を検索している場合、「`billion`」を「`bill`」の前に定義する必要があります。つまり、`(billion|bill)` ではなく、`(bill|billion)` となります。マクロは一連の文字列であるため、これはマクロにも適用されます。

## 定義セクションの規則の順序

1 行ごとに 1 つの規則を定義します。各セクション内で、規則には番号が付けられています (regexpl、regexp2、など)。これらの規則は 1 ~ n の順番に番号を付けなければなりません。番号が欠けていると、このファイルの処理が一度に中断します。エントリを無効にするには、正規表現の定義に使用する各行の初めに # 記号を追加します。エントリを有効にするには、行の前の # 文字を削除します。

各セクションで、確実に処理するために、最も具体的な規則を最も一般的な規則の前に定義する必要があります。たとえば、「month year」の形式および「month」の形式で日付を検索したい場合、「month year」の規則を「month」の規則の前に定義する必要があります。次に、その定義を示します。

```
#@# January 1932
regexpl=${MONTH},?[0-9]{4}

#@# January
regexp2=${MONTH}
```

そして、次のようには定義してはいけません。

```
#@# January
regexpl=${MONTH}

#@# January 1932
regexp2=${MONTH},?[0-9]{4}
```

## 規則でのマクロの使用

いくつかの規則で特定の行列を使用している場合、マクロを使用できません。この行列の定義を変更する必要がある場合、該当する箇所を一度だけ変更する必要があり、それを参照するすべての規則を変更する必要はありません。たとえば、次のようなマクロがあるとします。

```
MONTH=((january|february|march|april|june|july|august|september|october|
november|december)|(jan|feb|mar|apr|may|jun|jul|aug|sep|oct|nov|dec)(\.)?)
```

マクロの名前を参照する場合、必ず `$()` で囲みます。例: `regexpl=${MONTH}`  
すべてのマクロは `[macros]` で定義する必要があります。

## 正規化

固有表現を抽出する場合、出現したエンティティは正規化され、事前定義された形式に従って同様のエンティティとグループ化されます。たとえば、語内の通貨記号および同格の文字は同じように扱われます。正規化エントリは、[アドバンス リソース] タブの [正規化] セクションで行われます。詳細は、p. 256 [アドバンス リソースについて](#) を参照してください。ファイルはいくつかのセクションに分割されます。

**重要!** このファイルはアドバンス ユーザーのみが使用できます。このファイルの変更が必要な可能性はほとんどありません。この領域においてさらに支援が必要な場合、IBM Corp. にご連絡ください。

### 正規化の書式規則

- 1 行につき正規化エントリを 1 つだけ追加する。
- このファイルのセクションを厳密に重視する。新しいセクションを追加することはできません。
- エントリを無効にするには、その行の初めに # 記号を追加する。エントリを有効にするには、行の前の # 文字を削除します。

## 構成

固有表現構成ファイルに抽出したい固有表現タイプを有効化および無効化することができます。必要のないエンティティを無効にすることによって、必要な処理時間を短縮することができます。これは [アドバンス リソース] タブの [構成] セクションで行われます。詳細は、p.256 [アドバンス リソースについて](#) を参照してください。非言語学的抽出が有効になると、抽出エンジン抽出時にこの構成ファイルを読み取り、固有表現タイプを抽出するかどうかを判断します。

このファイルのシンタックスは次のようになります。

```
#name<TAB>言語<TAB>Code
```

テーブル 11-1  
構成ファイルのシンタックス

列ラベル	説明
#name	固有表現の語は、固有表現抽出の他の 2 つの必須ファイルで参照されます。ここで使用される名前は大文字と小文字が区別されます。
言語	レコードの言語。特定の言語を選択することが最適ですが、[任意] オプションもあります。指定できるオプションは次のとおりです。0 = IP/URL/電子メール アドレスなど、正規表現が言語特有でなく、さまざまな言語のいくつかのテンプレートで使用できる任意のオプション、1 = フランス語、2 = 英語、4 = ドイツ語、5 = スペイン語、6 = オランダ語、8 = ポルトガル語、10 = イタリア語です。
Code	品詞コード。多くのエンティティは、一部の場合を除いて “s” の値をとります。指定できる値は次のとおりです。s = ストップワード、a = 形容詞、n = 名詞です。有効にすると、固有表現が最初に抽出され、抽出パターンを適用し、より大きいコンテキストでその役割を特定します。たとえば、割合は “a” の値が与えられます。固有表現として 30% が抽出されるとします。それは形容詞として特定されます。テキストに「30% salary increase」という文字列が含まれていた場合、“30%” の固有表現は “ann” (形容詞、名詞、名詞) の品詞パターンに適合します。

## エンティティ定義の順序

このファイルでエンティティが宣言される順序は重要であり、それらがどのように抽出されるかに影響を与えます。表示された順序に適用されません。順序を変更すると、結果も変わります。最も具体的な固有表現は、最も一般的な固有表現の前に定義する必要があります。

たとえば、固有表現 “Aminoacid” は次のように定義されます。

```
regex1=$(AA)-?(NUM))
```

\$ (AA) は、特定のアミノ酸に対応する固有の 3 文字の行列である、“(ala|arg|asn|asp|cys|gln|glu|gly|his|ile|leu|lys|met|phe|pro|ser)” に対応します。

一方、固有表現 Gene はより一般的であり、次のように定義されます。

```
regex1=p[0-9]{2,3}
regex2=[a-z]{2,4}-?[0-9]{1,3}-?[r]
regex3=[a-z]{2,4}-?[0-9]{1,3}-?p?
```

[構成] セクションで「Gene」を「Aminoacid」の前に定義すると、「Gene」の「regex3」が最初に一致するため、「Aminoacid」は一致しなくなります。

## こうせいの書式規則

- 列内の各語を区切るには TAB 文字を使用する。
- 行を削除しない。
- 前述の表示されたシンタックスを重視する。
- エントリを無効にするには、その行の初めに # 記号を追加する。エンティティを有効にするには、行の前の # 文字を削除します。

## 言語処理

現在使用される各言語には、キーワードを表現、文を構築、省略形を使用する特別な方法があります。[言語処理] セクションでは、抽出パターンを編集、これらのパターンの定義を強制、[言語] ドロップダウン リストで選択した言語の省略形を宣言することができます。

- 抽出パターン
- 強制定義
- 省略形

## 抽出パターン

レコードから情報を抽出する場合、抽出エンジンは、品詞抽出パターンのセットをテキストの単語の「積み重ね」に適用して、抽出の候補のキーワード（単語および句）を特定します。抽出パターンを追加または変更できます。

品詞には、名詞、形容詞、過去分詞、決定詞、前置詞、人の名、イニシャル、助詞など、文法的な要素が含まれます。これら一連の要素が、品詞の抽出パターンを構成しています。IBM Corp. のテキストマイニング製品では、各品詞が 1 つの文字で表され、パターンが定義しやすくなります。たとえば、形容詞は小文字の「a」で表されます。デフォルトでは、サポートされているコードのセットがデフォルトの抽出パターンの上位に、パターンのセットと、各パターンの例と共に表示され、使用される各コードについて理解しやすくなります。

### 抽出パターンの書式設定規則

- 1 行ごとに 1 つのパターン。
- 行頭に # を使用してパターンを無効化する。

単語の指定された順序は抽出エンジンによって一度だけ読み込まれ、エンジンが合致を検出した最初の抽出パターンに割り当てるため、抽出パターンの表示順が非常に重要になります。

## 強制定義

レコードから情報を抽出するとき、抽出エンジンはテキストをスキャンし、出現するすべての単語の品詞を特定します。状況によっては、単語がいくつかの異なる役割に適合する場合があります。単語が特定の品詞の役割を取得するよう強制する場合、または処理からその単語を完全に除外する場合、[アドバンス リソース] タブの [強制定義] セクションで指定できます。詳細は、p. 256 [アドバンス リソースについて](#) を参照してください。

指定した単語の品詞を強制するには、次のシンタックスを使用して、このセクションに 1 行追加します。

```
term:code
```

テーブル 11-2  
シンタックスの説明

エントリ	説明
term	キーワード名。
code	品詞の役割を示す 1 文字のコード。ユニタームあたり最大 6 つの品詞コードを表示できます。また、additional:s のように小文字のコード s を使用して、単語が複合語/句に抽出されないようにできます。



### 強制定義の書式設定規則

- 単語ごとに 1 行。
- キーワードにコロンの使用不可。
- 単語を抽出しないようにするには小文字の `s` を品詞コードに使用する。
- 1 行あたり最大 6 つの品詞コードを使用する。サポートされている品詞コードは、[抽出パターン] セクションに表示されます。詳細は、p.266 [抽出パターン](#) を参照してください。
- 部分一致の場合、文字列の最後にアスタリスク (\*) をワイルドカードとして使用する。たとえば、`add*:s` と入力すると、`add`、`additional`、`additionally`、`addendum`、および `additive` のような単語はキーワードとして、あるいは複合語キーワードの一部として抽出されなくなります。ただし、単語の合致がコンパイル済み辞書または強制定義でキーワードとして明示的に宣言されている場合は、抽出されます。たとえば、`add*:s` および `addendum:n` を入力すると、`addendum` がテキスト内で見つかった場合に抽出されます。

## 省略形

抽出エンジンがテキストを処理しているとき、検出されたピリオドは、文の終了を示す指標として認識されます。これは通常、正常な処理ですが、省略形がテキスト内にある場合、ピリオド文字のこうした処理は適用されません。

テキストからキーワードを抽出し、特定の省略形の処理が不適切であったことがわかった場合、このセクションで省略形を明示的に宣言する必要があります。

注:省略形が類義語定義に出現、またはキーワード辞書でキーワードとして定義されている場合、ここで省略形の投入を追加する必要はありません。

### 省略形の書式設定規則

- 1 行ごとに 1 つの省略形を定義する。

# 日本語テキストの例外

日本語テキストは IBM® SPSS® Text Analytics for Surveys の他のサポート言語と同じような方法で処理およびマイニングされますが、多くの相違点があります。小さな相違点については、本マニュアルのほかのすべての言語の指示とともに記載されますが、大きな相違点については、この付録の章で説明します。

## 日本語テキストの抽出およびカテゴリ化

日本語テキストをマイニングする場合、サポートされている他の言語と同じような処理となります。詳細は、1 章 p.3 [テキストマイニングについて](#) を参照してください。ただし、日本語の処理では、次のような異なる部分があります。

### 抽出の方法

回答の主要キーワードの抽出時、IBM® SPSS® Text Analytics for Surveys は言語学に基づくテキスト分析に依存します。このアプローチを用いると統計に基づくシステムがもたらすようなスピードと費用対効果が得られます。また人の手を介することがほとんどないので、極めて高い精度が得られます。言語学に基づくテキスト分析は、自然言語処理とも、計量言語学とも呼ばれる研究の分野に基づきます。

日本語テキストにおける抽出プロセス時の統計的アプローチと言語学的アプローチとの差を、例として**沈む**という単語を使用して説明することができます。この単語を使用した、**日が沈む**や、**気分が沈む**という表現があります。統計的アプローチのみを使用すると、**日**、**気分**、および**沈む**という単語がそれぞれ個別に抽出されます。ただし、言語学的手法に基づく感性分析を使用すると、**日**、**気分**、および**沈む**が抽出されるだけでなく、**気分が沈む**も抽出され、タイプ <悪い - 悲しみ全般> に割り当てられます。感性分析で言語学に基づいた手法を使用すると、より意味のある表現を抽出できます。感性の分析およびキャプチャを行い、定義というより信頼できるアプローチによって、言語学的テキストマイニングを実行してテキストの曖昧さに切り込みます。

抽出プロセスがどうなっているのかを理解しておく、言語リソース（ライブラリ、タイプ、類義語など）を微調整する際に役に立ちます。抽出プロセスのステップには以下のものがあります。

- ソース データの標準フォーマットへの変換
- 候補のキーワードの特定

- 類義語の等価クラスおよび統合の特定
- タイプの割り当て
- 二次分析によるインデックスの付与、および必要に応じてパターン マッチ

### 手順 1: ソース データの標準フォーマットへの変換

最初のステップでは、後続の分析に利用できるように、インポートしたデータを決まった形式に変換します。この変換は内部的に実行され、元のデータは変更されません。

### 手順 2: 候補のキーワードの特定

言語学的抽出において、候補となるキーワードを特定する際の言語リソースの役割を理解しておくのは大切なことです。言語リソースは、抽出が実行されるごとに使用されます。言語リソースは、テンプレート、ライブラリ、およびコンパイル済み辞書の形式で保存されています。ライブラリには、語のリスト、関係性、また抽出の実行や調整に使用される情報が含まれています。基幹辞書は表示・編集ができません。ただし、残りのリソースをリソース エディタ で編集できます。

コンパイル済み辞書は、SPSS Text Analytics for Surveys の抽出エンジンの主要な、内部コンポーネントです。これらのリソースには、品詞コード（名詞、動詞、形容詞など）を含む基本形のリストを収めた一般辞書が含まれています。また、リソースには、<地名>、<組織>、または<人名>に多くの抽出されたキーワードを割り当てるために使用する、予約済みのビルトインのタイプも含まれています。詳細は、 p. 278 [日本語テキストで使用できるタイプ](#) を参照してください。

これらコンパイル済み辞書のほか、製品にはいくつかのライブラリが付属し、それらを使用して、コンパイル済み辞書のタイプ定義およびコンセプト定義を補い、また類義語を提供することができます。これらのライブラリ、および作成したユーザー指定のライブラリは、いくつかの辞書で構成されています。これらには、キーワード辞書、類義語辞書、および不要語辞書が含まれています。詳細は、 p. 274 [日本語テキストのリソースの編集](#) を参照してください。

データがインポートおよび変換されると、抽出エンジンは抽出の候補のキーワードの特定を開始します。候補となるキーワードとは、テキスト内の概念を特定するのに使用される語や、語の集まりのことです。テキストを処理しているとき、単語（**ユニターム**）および複合語（**マルチターム**）は、品詞パターン抽出を使用して特定されます。たとえば、品詞パターンが「<地名> + <名詞>」のマルチターム **青森りんご** は、2 つの部分に分けられます。そして、候補の感性キーワードは、感性テキスト リンク分析を使用して特定されます。

たとえば、次のような日本語のテキストがあるとします。写真が新鮮で良かった。この場合、抽出エンジンは、感性テキスト リンク規則のいずれかを使用して (品物) + が + 良い を合致させた後、感性タイプ 良い - 褒め・賞賛 を割り当てます。

注: 前述のコンパイル済み一般辞書にあるキーワードは、ユニタームとして重要でないまたは言語学的にあいまいであるすべての単語を示します。これらの単語は、ユニタームを特定するときに不要語に追加されます、ただし、それらは、品詞を決定またはより長い候補の複合語 (マルチターム) を参照している場合に再評価されます。

### 手順 3: 類義語の等価クラスおよび統合の特定

候補のユニタームおよびマルチタームが特定された後、正規化辞書を使用して、等価クラスを特定します。等価クラスは、ある語句の基本形、すなわち同じ語句の2つの表現を1つの形で表わしたものです。句を等価クラスに割り当てる目的は、たとえば、side effect と 副作用 を別のコンセプトとして扱わないようにすることです。等価クラスのどのコンセプトを使用するか、つまり、side effect または副作用のどちらを主要キーワードとして使用するかを判断するために、抽出エンジンは、次の規則を順に適用します。

- ライブラリのユーザー指定の形式。
- コンパイル済みリソースで定義されている最も頻度の高い形式。

### 手順 4: タイプの割り当て

次に、抽出されたコンセプトにタイプを割り当てます。タイプは、コンセプトの意味上のグループ化です。基幹辞書ならびにライブラリの両方がこのステップで使用されます。タイプには、上位レベルのコンセプト、肯定的な単語および否定的な単語、人名、地名、組織名などが含まれます。詳細は、10 章 p.236 キーワード辞書を参照してください。

日本語リソースには、タイプの異なるセットがあります。詳細は、p.278 日本語テキストで使用できるタイプを参照してください。

### 手順 5: イベント抽出によるインデックスの付与およびパターン マッチ

レコードのセット全体に、テキストの位置と各等価クラスの代表キーワードの間に ポインタを確定してインデックスを付けます。候補のコンセプトの活用形インスタンスはすべて、候補の基本型としてインデックスが付けられます。基本形ごとに全体の出現頻度が計算されます。

SPSS Text Analytics for Surveys は、タイプやコンセプトだけでなく、それらの関係性も見つけることができます。この製品ではいくつかのアルゴリズムおよびライブラリを使用でき、またタイプおよびコンセプトの間のテキスト リンク分析の関係性パターンを抽出する機能が用意されています。特定の意見 (製品の反応など) を検出しようとする場合、特に役立ちます。

## 二次抽出の手順

日本語テキストで抽出を実行する場合、基本キーワードおよび人名、地名、組織名、名詞、形容詞、動詞、形容動詞、およびその他の 8 つの基本タイプからコンセプトを自動的に取得します。ただし、日本語テキスト向けに用意されたデフォルトのリソースを活用するために、[感性分析] または [係り受け分析] の二次分析機能のいずれかを選択する必要があります。

二次分析を選択すると、テキスト リンク分析パターンを抽出して、テキスト内のキーワード間の関係性を明らかにすることができます。

**二次分析:** 抽出が開始したとき、基本キーワード抽出が、タイプのデフォルト セットを使用して行われます。詳細は、 p. 278 [日本語テキストで使用できるタイプ](#) を参照してください。ただし、二次分析機能を選択すると、抽出機能にコンセプトの一部として助詞や助動詞が含まれているため、より多くの詳細なコンセプトを取得することができます。たとえば、「肩の荷が下りた」という文があるとします。この例では、基本キーワード抽出によって各コンセプトを、**肩** (肩)、**荷** (荷)、**下りる** (下)のように個別に抽出できますが、これらのワードの関係は抽出できません。しかし、感性分析を適用すると、コンセプト = **肩の荷が下りた** というより高次の意味をもつコンセプトが抽出され、<<良い-安心>> という感性タイプとして割り当てられます。なお感性分析については、このほかに多くの感性タイプがあります。さらに、二次分析機能を選択すると、テキストリンク分析結果も生成できます。

注: 二次分析を呼び出すと、抽出プロセスにより時間がかかります。詳細は、 p. 271 [二次抽出の手順](#) を参照してください。

- **係り受け解析:** このオプションを選択すると、キーワード、およびキーワードに助詞等を加えた語を、基本の品詞タイプのコンセプトとして抽出します。また係り受けテキスト リンク分析 (TLA) に基づいたパターン結果を出力します。
- **感性分析:** この分析機能を選択すると、追加の抽出コンセプト、および可能な場合は、TLA パターン結果の抽出が行われます。基本タイプのほか、嬉しい、吉報、幸運、安心、幸福など 80 を超える感性タイプも利用できます。これらのタイプを使用して、感情、感性、意見の表現のテキストでコンセプトおよびパターンを検出します。感性分析のフォーカスを管理するオプションは3つあります: **すべての感性**、**代表的感性のみ**、**結論のみ**。

## 感性分析のオプション

日本語テキストを処理するとき、感性分析機能を使用して、追加のコンセプトおよびタイプを抽出できます。この分析機能には、意見、間隔、感性をテキスト データから抽出できる、80 を超える追加タイプが含まれます。また、二次分析として **感性分析** を選択した場合、抽出エンジンに抽出する感性を指示する次のオプションからいずれかを選択する必要があります。

- すべての感性
- 代表的感性のみ
- 結論のみ

抽出時、感性分析機能はレコードを、それぞれに述語が含まれる節に分割します。たとえば、「4月になったがまだ寒い。」というテキストは、文末記号「。」を1つだけ使用しているにもかかわらず、分析機能によって2つの節があると解釈されます。各節が抽出エンジンによって検証され、選択したサブオプションに適しているかどうかを確認します。

サンプル テキスト「案内してくれた仲居さんは無愛想だったが、部屋は広くて申し分なかった。夕食も満足。」を使用して、3つのオプションを検証してみましょう。“抽出時、元のテキストは次の節に分割されます。

- 案内してくれた仲居さんは無愛想だったが、
- 部屋は広くて申し分なかった。
- 夕食も満足。

### すべての感性

このオプションを指定すると、リソースおよび感性テキスト リンク規則に一致するすべての感性、意見および感情が抽出されます。この例では、サンプル テキストから次のコンセプトが抽出されます。

テーブル A-1  
[すべての感性] オプションを使用した場合の出力の例

Concept	タイプ
仲居さんは無愛想だった	<悪いー対応が不親切>
部屋は広くて	<良いー満足>
申し分なかった	<良いー満足>
満足	<良いー満足>

注: 前述の表の2番目および3番目の行で、どのようにして同じ節から2つのコンセプトを取得するかを示しています。

### 代表的感性のみ

このオプションを指定すると、各節で表現されているより代表的な意見または感性のみを抽出します。テキストに複数の意見または感性がある場合、アルゴリズムが適用されます。このアルゴリズムは、節で見つかった感性および単語の位置の重要度を判断しようとします。重要度が同じ感性キーワードが2つ見つかった場合、節の最初の感性キーワードではなく後の方の場所にある感性キーワードが抽出されます。

内部アルゴリズムおよび単語の位置が適用された場合、この節の 2 番目の申し分なかったより重要であると認識されるため、部屋は広くては適用されません。

テーブル A-2

[代表的感性のみ] オプションを使用した場合のテキストの出力の例

Concept	タイプ
仲居さんは無愛想だった	<悪いー対応が不親切>
申し分なかった	<満足>
満足	<満足>

### 結論のみ

このオプションを指定すると、感性キーワードをレコード全体の結論を示すものとして特定および抽出されるよう強制します。すべてのテキストに結論があるわけではなく、このオプションを指定してもテキストの指定した部分から何も抽出されない場合もあります。また、レコードが長いほど、分析機能で主な結論を特定することが困難になります。めったにありませんが、複数の結論が抽出される場合があります。

満足 の場合、これが表現された感性の重要な結論であると見なされます。

テーブル A-3

[結論のみ] オプションを使用した場合のテキストの出力の例

Concept	タイプ
満足	<満足>

## カテゴリ化の方法

いくつかの手法から選択して、カテゴリを作成できます。すべてのデータセットが一意であるため、手法の数やそれらを適用する順序は、変わる場合があります。結果の解釈が、他の人とは異なる場合があるため、テキスト データにとってどの手法が最良の結果を生み出すか、それぞれの手法を検証する必要があります。

このガイドの場合、**カテゴリの作成**は、カテゴリ定義の生成および、1 つまたは複数のビルトインの手法を使用した分類を指し、また**カテゴリ化**は、スコアリング、またはラベル付け、一意の識別子（名前/ID/値）を各レコードのカテゴリ定義に割り当てるプロセスのことを指します。

カテゴリ作成時、抽出されたコンセプトおよびタイプはカテゴリの構築ブロックとして使用されます。カテゴリを作成すると、カテゴリ定義の要素に一致するテキストが含まれる場合、レコードが自動的にカテゴリに割り当てられます。

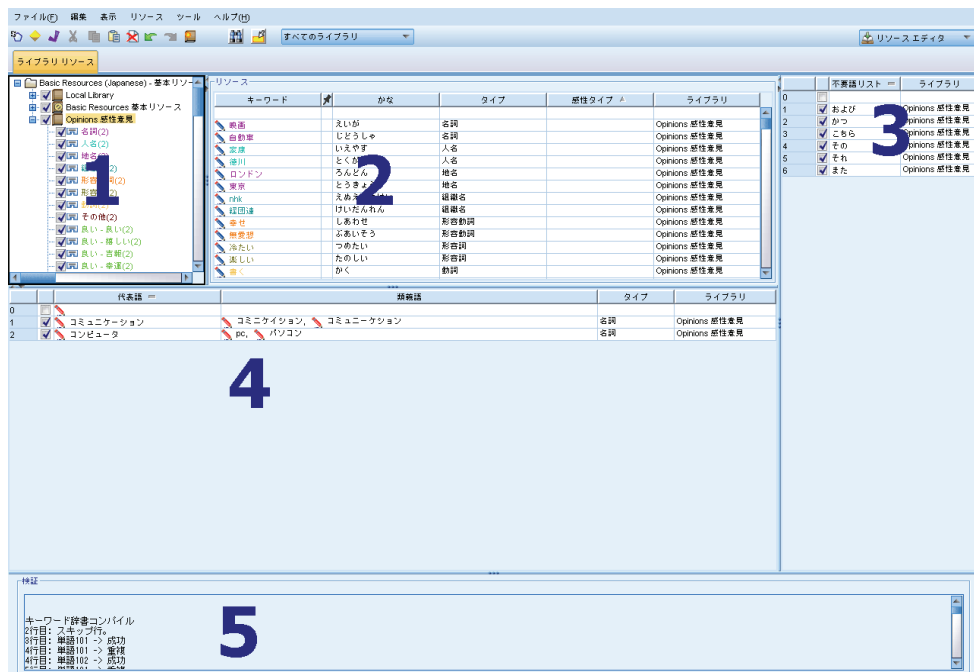
IBM® SPSS® Text Analytics for Surveys には、自動カテゴリ作成手法がいくつか用意されており、レコードを迅速にカテゴリ化することができます。使用できるそれぞれの手法は、特定の種類のデータおよび状況に適していますが、同じ分析で手法を組み合わせ、全範囲のレコードをキャプチャすると役に立つ場合があります。複数のカテゴリのコンセプトを表示したり、重複するカテゴリを見つけることができます。

## 日本語テキストのリソースの編集

IBM® SPSS® Text Analytics for Surveysバージョン 4以降では、新しいテンプレートと、新しいテキスト分析パッケージ (TAP) が日本語で使用できます。キーワードを追加および編集してリソースに変更を加え、データに合わせてカスタマイズすることができます。テキスト分析パッケージには、肯定的な感性、否定的な感性、コンテキスト/一般的な感性を示すカテゴリで構成されたカテゴリ セットも含まれています。

リソース エディタ でリソースの作業ができます。エディタはすべてのテキスト言語で同様の作業をしますが、日本語テキストでの作業については、次のような重要な違いがあります。

図 A-1  
日本語テキストのリソース エディタ ウィンドウ





日本語テキストのリソースを編集する場合の主な相違点は次のとおりです。[ライブラリ リソース] タブの 4 つのメイン パネルの詳細は、[エディタのインターフェイス](#) p. 208 を参照してください。

**1. [ライブラリ] パネル:** 左上のこの領域は、他の言語と同じように機能します。ただし、新しいタイプを作成できない、タイプの名前を変更できないなど、いくつか異なる点があります。詳細は、9 章 p. 221 [ライブラリの使用](#) を参照してください。

**2. キーワード辞書の [キーワード] パネル** ライブラリ ツリーの右側にあるこのパネルは、日本語テキストの場合、大きく異なります。キーワード名があるほか、かな名を追加でき、またキーワードに関連付けることができる 1 つまたは 2 つのタイプを選択できます。ただし、日本語以外の言語できるように、キーワードの活用形あるいは日本語キーワードのマッチ オプションを割り当てることはできません。詳細は、p. 275 [日本語のライブラリ ツリー、タイプ、キーワードのパネル](#) を参照してください。

**3. [類義語辞書] パネル:** 日本語テキスト リソースには、リソースのすべての類義語を定義できる [類義語] タブがあります。[類義語] タブにもう 1 つ [タイプ] 列があり、そこで入力した類義語のタイプを指定する必要があります。詳細は、p. 283 [日本語テキストの類義語辞書の使用](#) を参照してください。注: また、[オプションの要素] タブは日本語テキストの言語リソースには適用されないため、使用できません。

**4. [不要語辞書] パネル:** ワイルドカードの使用がサポートされていない点を除き、日本語テキスト リソースのこのパネルに異なる点はありません。

**5. [検証] パネル:** 日本語テキストの場合、抽出前にリソースを確認するための [検証] パネルがあります。日本語テキストから抽出する場合、抽出プロセスが始まる前に変更が検出されると、抽出エンジンは自動的にリソースを再コンパイルします。実行時のエラーを回避するため、抽出前にリソースを再コンパイルして検証し、発生したエラーを修正できます。詳細は、p. 285 [日本語リソースの検証およびコンパイル](#) を参照してください。

注: 日本語テキスト向けに編集可能なアドバンス リソースまたはテキストリンク規則がないため、これらのタブは使用できません。

## 日本語のライブラリ ツリー、タイプ、キーワードのパネル

日本語辞書のライブラリおよびタイプの使用方法は、他の言語の場合と非常に似ています。詳細は、10 章 p. 236 [キーワード辞書](#) を参照してください。

ただし、次のように異なる点があります。

- 日本語テキスト リソースにタイプの異なるセットがある。詳細は、p. 278 [日本語テキストで使用できるタイプ](#) を参照してください。

- タイプの作成、または名前の変更はできないが、プロパティは編集できる。詳細は、p. 282 日本語のタイプのプロパティの編集を参照してください。
- キーワードのかな名の指定、1 つのタイプまたは 2 番目の感性タイプへの割り当てなど、キーワードの追加および編集ができる。詳細は、p. 275 日本語のライブラリ ツリー、タイプ、キーワードのパネルを参照してください。

ライブラリ ツリー パネルには、キーワード辞書のほか、ライブラリが表示されます。左側のライブラリまたはタイプを選択すると、右側のキーワード パネルに、選択したライブラリまたはキーワード辞書のキーワードが表示されます。[キーワード] パネルで直接または [キーワードを追加] ダイアログ ボックスを使用して、キーワードをキーワード辞書に追加できます。追加するキーワードは、単語でも複合語でもかまいません。リストの一番上に空白の行があり、そこに新しいキーワードを追加できます。

キーワード辞書のキーワードを定義すると、デフォルトでは名詞と見なされ、自動的にタイプ <名詞> に割り当てられます。ただし、タイプを <動詞>、<形容詞>、<地名> など別の基本タイプに変更できます。抽出エンジンが、このキーワードが [タイプ] 列で割り当てられているタイプと同じ品詞として機能していることを認識した場合、キーワードはそのタイプに割り当てられ、抽出されます。また、キーワードを [感性タイプ] 列のいずれかの感性タイプに割り当てることができます。感性二次分析を使用すると、テキストには 2 回目の処理が行われ、キーワードを検索して感性タイプに割り当てようとします。さらに、感性タイプおよび基本タイプの 2 つを定義し、二次感性分析時に抽出エンジンでこのキーワードがいずれのタイプにも合致することが認識された場合、感性タイプが優先され、抽出結果パネルおよびテキスト リンク分析結果に表示されます。たとえば、動詞が <動詞> タイプとして抽出され、「愛されている」のような肯定的な種類のタイプとしても抽出された場合、感性のキャプチャが単なる品詞よりより重要である場合が多いため、このキーワードはインターフェイス上では肯定的なタイプに割り当てられているものとして表示されます。

図 A-2  
日本語リソースの [ライブラリ] パネルおよび [キーワード] パネル

キーワード	かな	タイプ	感性タイプ	ライブラリ
映画	えいが	名詞		Opinions 感性意見
自動車	じどうしゃ	名詞		Opinions 感性意見
家康	いえやす	人名		Opinions 感性意見
徳川	とくがわ	人名		Opinions 感性意見
ロンドン	ろんどん	地名		Opinions 感性意見
東京	とうきょう	地名		Opinions 感性意見
HP	えぬあいちけい	組織名		Opinions 感性意見
経団連	けいだんれん	組織名		Opinions 感性意見
幸せ	しあわせ	形容動詞		Opinions 感性意見
無愛想	むあいそう	形容動詞		Opinions 感性意見
冷たい	つめたい	形容詞		Opinions 感性意見
楽しい	たのしい	形容詞		Opinions 感性意見
書く	かく	動詞		Opinions 感性意見
見る	みる	動詞		Opinions 感性意見
かなり	かなり	その他		Opinions 感性意見
いかなる	いかなる	その他		Opinions 感性意見
かなり	かなり	その他		Opinions 感性意見
お願してもいいかしら	おねがいてもいいかしら		その他 - お願	Opinions 感性意見
取材させて下さい	しゅざいさせてください		その他 - お願	Opinions 感性意見
一緒に行きませんか	いっしょにいきませんか		その他 - 動詞	Opinions 感性意見
今度お会いに行きましょう	こんどのお会いに行きましょう		その他 - 動詞	Opinions 感性意見

テーブル A-4  
[キーワード] パネルの列の説明

列名	列の説明
キーワード	単語または複合語を入力します。キーワードが表示される色は、キーワードが保存または強制投入されるタイプの色によって異なります。[タイプのプロパティ] ダイアログ ボックスでタイプの色を変更できます。詳細は、 p. 282 <a href="#">日本語のタイプのプロパティの編集</a> を参照してください。通常、キーワードは漢字で書かれますが、かなを組み合わせる場合もあります。 <b>重要!</b> カタカナを使用した動詞の入力はサポートされていません。
強制	このセルに押しピンのアイコンをクリックして投入すると、抽出エンジンは、他のライブラリのこの同じキーワードの他の出現を無視します。詳細は、 10 章 p. 244 <a href="#">キーワードの強制</a> を参照してください。すべての言語で同じように機能します。
かな	漢字のキーワード名の読みがなを入力します。
タイプ	キーワードを割り当てる基本タイプ名を選択します。詳細は、 p. 278 <a href="#">日本語テキストで使用できるタイプ</a> を参照してください。
感性タイプ	2 番目の分析が実行されると、キーワードを割り当てる感性タイプ名を選択します。詳細は、 p. 278 <a href="#">日本語テキストで使用できるタイプ</a> を参照してください。
ライブラリ	キーワードが格納されているライブラリを選択します。ライブラリ ツリー パネルでキーワードを別のタイプにドラッグ アンド ドロップして、そのライブラリを変更できます。

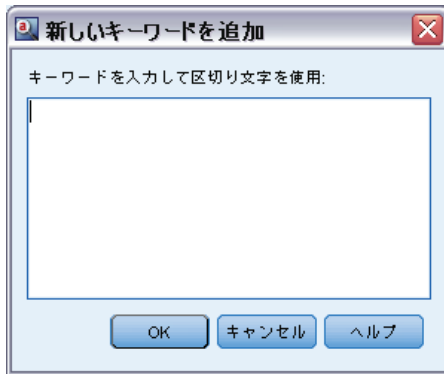
#### キーワード辞書に 1 つのキーワードを追加するには

- ▶ ライブラリ ツリー パネルで、キーワードを追加したいキーワード辞書を選択します。
- ▶ 中央のパネルのキーワード リストで、使用できる最初の空白セルにキーワードを入力し、このキーワードに必要なオプションを設定します。

#### キーワード辞書に複数のキーワードを追加するには

- ▶ ライブラリ ツリー パネルで、キーワードを追加したいキーワード辞書を選択します。
- ▶ メニューの [ツール] → [新規キーワード] を選択します。[新しいキーワードを追加] ダイアログ ボックスが開きます。

図 A-3  
[新しいキーワードを追加] ダイアログ ボックス



- ▶ キーワードを入力するか、キーワードのセットを貼り付けて、選択したキーワード辞書に追加したいキーワードを入力します。複数のキーワードを入力する場合、[オプション] ダイアログで定義された区切り文字を使用してキーワードを区切るか、各キーワードを新しい行に追加するつようがあります。詳細は、2 章 p.19 [オプションの設定](#) を参照してください。
- ▶ [OK] をクリックすると、キーワードが辞書に追加されます。ダイアログボックスが閉じ、辞書に新しいキーワードが表示されます。

## 日本語テキストで使用できるタイプ

日本語リソースに新しいタイプを追加することはできませんが、日本語リソースからキーワードを追加および削除することができます。次の表に、現在使用できるタイプの一覧を示します。

### 基本抽出のタイプ

抽出が開始されると、次のタイプが使用されます。

テーブル A-5  
基本抽出のタイプ

タイプ	説明
名詞	「車」や「映画」など、物を示す単語人名、地名、組織名は別にカテゴリ化されます。
人名	「徳川」や「家康」など、特定の人物の名前に該当する名詞。「徳川家康」のように、名と姓を組み合わせても人名となります。
地名	「東京」やロンドンなど、特定の場所を示す名詞。
組織名	「経団連」など、特定の会社や組織を示す名詞。
形容動詞	物事の特徴または状況を説明する「静か」のような単語。「静かでない」や「静かなこと」のように句として使用することができます。

タイプ	説明
形容詞	物事の特徴または状況を説明する「楽しい」のような単語。「楽しくなる」や「楽しいこと」のように句として使用することができます。
動詞	動きまたは動作を説明する、タイプ I (子音語幹) 動詞、タイプ II (母音語幹) 動詞、不規則 (サ行変格活用、カ行変格活用) 動詞などの単語。
その他	「非常に」、「いかなる」、「そして」、「ありがとう」など、助動詞、連体詞、接続詞、感嘆詞などの語。

### 感性分析のタイプ

感性抽出に二次分析機能を選択した場合、8 つの基本タイプのほか、多くのタイプを取得します。

テーブル A-6  
感性分析のタイプ

タイプ	説明
良い - 良い	「良い」と分類できる、一般的に肯定的なものの表現。
良い - 嬉しい	喜びの感動を生み出す望ましいイベントを説明。
良い - 吉報	相当な努力によってのみ可能な喜びのイベントを説明。
良い - 幸運	たまたま、またはまったくの偶然によってのみ可能な喜びのイベントを説明。
良い - 快い	喜びの心理的な感覚に切り替える状況または環境であることを示す表現。
良い - 体が良い状態	病気、怪我、疲労のない体の状態、または身体的な状況が改善している状態を説明。
良い - 安心	落ち着いて、傷ついたり被害を受けるリスクがないことを説明。
良い - 幸福	特に好ましい状況、またはある人の動作から愛情を、または誕生の状況にあることを説明。
良い - 満足	心が落ち着く望ましいイベントを説明。
良い - 美味しい	食べ物の味が良いことを示す。
良い - 効果が満足	あるものが、期待された効果をあげたことを示す。
良い - 感動	何かの重要性、意味または価値が特別に良いことを示す。
良い - 感謝	相手の動作を肯定的に認識することを示す。
良い - 祝福	ある人物の状況が望ましいという (話者にとってある程度受け入れられる) 気持ちを表現。
良い - 喜び全般	話者にとってあまり関連がない肯定的なイベント。
良い - 楽しい	交友、娯楽、レクリエーションなどの活動を示すまたは予期する。
良い - 可笑しい	楽しい刺激を与えるユーモラスな性質があることを示す。
良い - 笑い	良いまたはユーモラスなものによって起こる笑いを示す。

## A 付録

タイプ	説明
良い - 期待	良いイベントが将来起こることを示す。
良い - 楽しみ全般	話者にとってあまり関連がない楽しいイベントまたは肯定的な活動/動作。
良い - 金額への賞賛	購入者の立場から、望ましい金銭的価値があることを示す。
良い - 対応が早い	サービスがタイムリーに提供または完了することを示す。
良い - 対応が親切	サービス提供者の態度または動作に配慮が行き届いていることを示す。
良い - 説明が良い	情報の種類/または品質、およびその提供の方法が適切であるということを示す。
良い - 対応への賞賛	サービスの提供者を褒める、上記以外の考え。
良い - 褒め・賞賛	あるものの性質、機能、動作を褒める、上記以外の考え。
良い - 好き	あるものを所有したいまたは近くにいたいという欲求を示す。
良い - 入会希望	あるグループの一員になりたいまたは一員のままでいたいという欲求を示す。
良い - 買いたい	あるものを得るためにお金を使いたいまたは使う予定があることを示す。
良い - 好評・人気	あるものを必要としているまたは褒めている人の数が目標を超えていることを示す。
良い - 売れた	あるものを購入した人々の有無または購入数または金額が目標を超えていることを示す。
悪い - 悪い	「悪い」と分類できる、一般的に否定的なものの表現。
悪い - 怒り	予定通りに物事が起こらなかった場合に感じる怒りの感情。
悪い - 批判	他者が適切な選択をしなかったことに対する考えを表す。
悪い - お叱り	他者をある人物の意図に従うよう強制させる言葉または動作。
悪い - 誹謗・中傷	他者のあまりに低い評価を示すために使用する言葉。
悪い - 軽蔑	他者の性格、能力、または性質が著しく書けているということを示す。
悪い - 恨み	他者によってもたらされる不利な点に対する報復または憤りを示す。
悪い - 嫌がらせ	コミュニケーションを抑制する目的で使用する言葉。
悪い - 不満	希望する者または状態を取得できなかったことによる、不快な感情。
悪い - 不味い	食べ物の味が悪いことを示す。
悪い - 効果が不満	あるものが、期待された効果をあげていないことを示す。
悪い - 金額が不満	購入者の立場から、あるものの金銭的価値が望ましいものでないことを示す。

タイプ	説明
悪い - 対応への不満	サービス提供者に過失があることを示す。
悪い - 対応が遅い	サービスが不適切な時に行われたこと、またはサービスがまだ行われていないことを示す。
悪い - 対応が不親切	サービス提供者の態度または動作によってもたらされる不快な感情を示す。
悪い - 説明が悪い	情報の種類/または品質、およびその提供の方法が不適切であるということを示す。
悪い - 返答なし	状況が満足いく場合であっても、サービス提供者が適切な回答を提供しなかったことを示す。
悪い - 不快	否定的な心理的感情に切り替える状況または環境であることを示す。
悪い - 怒り全般	上記以外の怒りの感情。話者の組織または会社によってもたらされる一般的な怒り、または言及された怒りによってもたらされた出来事の説明。
悪い - 悲しい	あるものを失ったまたは得られなかった場合の不快な感情。
悪い - 凶報	十分な努力にかかわらず、ある目的を達成できなかったことを示す。
悪い - 不運	自分の過失ではなく、不幸な偶然または運によってもたらされた否定的な結果を示す。
悪い - ショック	予測できない、否定的なものまたは発生によってショックを受けたこと、または適切な回答が見つからないことを示す。
悪い - 残念	期待されることがそのようにできなかった場合にもたらされる悲しい感情。
悪い - 落胆	悲しく落胆した感情によって支配されている状態。
悪い - 諦め	話者または他者によってもたらされた否定的な状態が改善できないことを示す。
悪い - 後悔	代わりのものがあったとしても、過去に適切な選択をできなかったことに対する考えを示す。
悪い - 謝罪	他者を機づつ行けたことに対する話者の認識を示す。
悪い - 淋しい	他者との関係が十分でないまたは、関係する人の数が少ないことを示す。
悪い - 哀れみ	他者の状況が話者よりとても悪いことに対する気持ちを示す。
悪い - 悩み	選択をする必要があるが、選択肢から選ぶことができないことを示す。
悪い - 困っている	行動が必要な状況に対応する効果的な方法がないことに対する気持ちを表す。
悪い - 苦しい	外的な要因または自分の誤りまたは間違いによって正常に行動できない、不快な心理的状况を示す。
悪い - 体が悪い状態	病気、怪我、疲労のある体の状態、または身体的な状況が改善していない状態を説明。
悪い - 不安	あることが望ましい状態で継続できないまたは期待通りにいかないということに対する感情を示す。
悪い - 恐怖	あるものによって、傷つく恐れがある状態を示す。

## A 付録

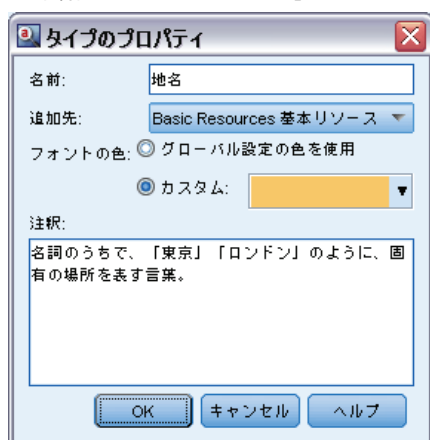
タイプ	説明
悪い - 悲しみ全般	指定されていない者に対する一般的な悲しみなど、上記以外の悲しみの感情。
悪い - 嫌い	あるものを遠ざけたいまたは、離れたいということを示す。
悪い - 退会希望	あるグループの一員であることを辞めたいまたは参加したくないという欲求を示す。
悪い - 買いたくない	あるものが欲しくないまたは言及したものに対して支払う予定がないことを示す。
悪い - 不評・不人気	あるものを好きな人々の数がある目標に達しなかったことまたは、そのものに対する否定的な感情を持つ人々が多くいることを示す。
悪い - 売れていない	あるものを購入した人々がいないまたは購入数または金額が目標に達していないことを示す。
その他 - 疑問	他人のより詳細な検討または検証が必要な情報を必要としていることを示す。
その他 - 問い合わせ	他人が持っている情報を要求していることを示す。
その他 - 要望	(他人に直接責任がある、または話者より順位が低い場合) 他者に問題を解決するよう命じることを示す。
その他 - 提案・忠告	(他人に直接責任がある、または話者より順位が低い場合) より良い行動を行うよう命じることを示す。
その他 - お願い	(他人に直接責任がない、または話者より順位が低い場合) 他者に何か実行するよう要望する表現。
その他 - 激励	他者を勇気づけることを示す表現、または動作を促す説明。
その他 - 勧誘	他者に、話者と一緒に何かを行うことを要求する表現。
その他 - 驚き	イベントが突然であることまたは尺度が、論理的な判断/理解を超えていることを示す表現。
評価なし - 評価なし	評価に値する表現がない。

## 日本語のタイプのプロパティの編集

日本語リソースでタイプを作成することはできませんが、タイプのプロパティを編集して表示することができます。また、マッチ オプションや活用形などのオプションは日本語テキストに適用されません。



図 A-4  
日本語テキストリソースの [タイプのプロパティ] ダイアログ ボックス



**名前:** キーワード辞書の名前。

**追加先:** 新しいキーワード辞書を作成するライブラリを指定します。

**フォントの色:** このフィールドを指定すると、インターフェイスでこのタイプの結果と他のタイプの結果とを区別できるようになります。[グローバル設定の色を使用] を選択すると、タイプのデフォルト色がこのキーワードに使用されます。このデフォルト色は、[オプション] ダイアログ ボックスで設定されます。詳細は、2 章 p.21 [オプション: \[表示\] タブ](#) を参照してください。[カスタム] を選択すると、ドロップダウン リストから色を選択できます。

**注釈:** このフィールドはオプションで、任意のコメントまたは説明に使用できます。

#### タイプのプロパティを表示または編集するには

- ▶ プロパティを表示したいタイプを選択します。
- ▶ マウスを右クリックし、コンテキスト メニューから [タイプのプロパティ] をクリックします。[タイプのプロパティ] ダイアログ ボックスが開きます。
- ▶ 必要に応じて変更してください。
- ▶ [OK] をクリックすると、キーワード辞書への変更が保存されます。

## 日本語テキストの類義語辞書の使用

日本語テキストの場合、類義語辞書には、類義語を管理するタブ、[類義語] タブが 1 つだけあります。類義語とは、同じ意味を持つ複数の語を関連付けたものです。また、類義語はキーワードを略語とグループ化した

り、一般的にスペルミスのある単語と正しいスペルの単語とをグループ化したりするために使用できます。

図 A-5  
日本語テキストの類義語エントリ

	代表語	類義語	タイプ	ライブラリ
0				
1	コミュニケーション	コミュニケーション、コミュニケーション	名詞	Opinions 感性差覺
2	コンピュータ	pc、パソコン	名詞	Opinions 感性差覺

類義語定義は、2 つの部分で構成されています。代表語は、抽出エンジンがすべての類義語キーワードをグループ化する基準となるキーワードです。この代表語が別の代表語の類義語として使用されていない限り、または不要語として登録されていない限り、[抽出結果] パネルに表示されるコンセプトとなります。類義語のリストは、代表語の下にグループ化されるキーワードです。

[類義語] タブで、テーブルの一番上の空白行に類義語定義を入力できます。まず代表語とその類義語を定義します。この定義を格納するライブラリも選択できます。抽出時、類義語のすべての出現を、最終的な抽出で代表語に基づいてグループ化します。詳細は、10 章 p. 240 [キーワードを追加](#) を参照してください。

キーワード辞書を作成している場合、キーワードを入力し、そのキーワードに対して 3 つまたは 4 つの類義語が考えられます。この場合、類義語辞書にすべてのキーワードと代表語を入力し、類義語をドラッグすることができます。

**重要!** ワイルドカードおよび特殊文字は、日本語テキストの類義語ではサポートされていません。

### 類義語エントリを追加するには

- ▶ [類義語] パネルの [類義語] タブに表示されたテーブルの一番上の空白行で、[代表語] 列に代表語を入力します。入力した代表語が色つきで表示されます。この色は、キーワードが表示されるまたは強制されるタイプを示します。キーワードが黒で表示されている場合、キーワード辞書にはこのキーワードがないことを意味します。
- ▶ 代表語の右側の 2 番目のセルをクリックして、類義語のセットを入力します。[オプション] ダイアログ ボックスで定義したグローバル区切り文字を使用して、各エントリを区切ります。入力されたすべての類義語は、同じタイプでなければなりません。詳細は、2 章 p. 19 [オプションの設定](#) を参照してください。入力したキーワードが色つきで表示されます。この色は、キーワードが出現するタイプを示します。キーワードが黒で表示されている場合、キーワード辞書にはこのキーワードがないことを意味します。

- ▶ 3 番目の列、[タイプ] 列で、これらの類義語のタイプを指定します。ただし、代表語は、抽出時にタイプが割り当てられます。代表語がコンセプトとして抽出されない場合、抽出結果でこの列に表示されたタイプが代表語に割り当てられます。
- ▶ 最後のセルをクリックして、この類義語定義を格納するライブラリを選択します。

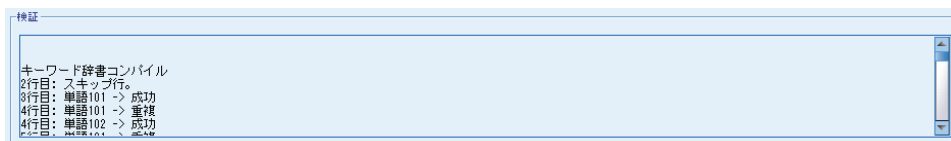
注:ここでは、リソース エディタ ビューでどのように変更を行うかについて説明します。[抽出結果] パネルまたは[データ] パネルで直接、この種類の調整を行うこともできます。詳細は、5 章 p.91 [抽出結果の調整](#) を参照してください。

## 日本語リソースの検証およびコンパイル

日本語テキストの場合、抽出前にリソースを確認するための [検証] パネルがあります。日本語テキストの抽出プロセスが開始される前に変更が検出されると、抽出エンジンは自動的にリソースを再コンパイルします。抽出時にエラーが見つかったら、処理を適切に完了できない場合があります。

コンパイル エラーを回避するために、リソース エディタ で変更を行った後、リソースを検証してからコンパイルすることをお勧めします。エラー メッセージが表示された場合、修正して検証をもう一度行うことができます。

図 A-6  
日本語テキストの [検証] パネル



### リソースを検証するには

- ▶ メニューの [ツール] → [リソースを検証] を選択します。[検証] パネルが開き、コンパイルおよびエラーのメッセージが表示されます。

## 日本語についてのその他の例外

### ユーザー定義リソースを上書きする内部リソース

日本語テキストの場合、デフォルトのリソースには、コンパイル済みの内部基本リソースがいくつか含まれています。これらの内部リソースは編集できません。このため、リソース エディタ を使用して、変更または調整を行うことができます。ほぼすべての場合、リソースで定義したキーワード、類義語および不要語リストのエントリは、コンパイル済み内

部リソースより優先されます。ただし、次の例の一部で示されているように、例外がいくつかあります。

- このようなキーワードを追加しても特定のタイプが抽出結果に影響を与えない場合があります。このようなケースは、データにいくつかの形態的要素、区切り記号、または記号を使用している長い文が含まれる場合に最も発生します。また、日本語テキスト リソースには、事前にコンパイルされている多くの一般的キーワードが含まれているため、特定の言語学的定義に常に強制投入される共通単語がいくつかあります。
- 抽出エンジンは、**ある**、**いる**、または**なる**の抽出を常に強制するため、これらのキーワードを除外することができない場合があります。
- キーワード**東京**のタイプを <地名> から <名詞> に変更することはできますが、キーワードのタイプをキーワード辞書を使用して <地名> から <動詞> または <形容詞> に変更しようとしても、抽出エンジンに無視されます。
- リソース エディタ または で行った変更がある文からの抽出結果に影響を与える場合がありますが、抽出プロセスの最後に各文の共起語を参照するため、別の文には影響を与えません。

### 半角カタカナの表示問題

半角カタカナの文字は、抽出時、全角カタカナに内部的に変換されますが、ユーザー インターフェイスの。半角カタカナ文字は [データ] パネルで強調表示できません。この問題を回避するには、処理前にレコード全体を全角カタカナに変換します。

### 大文字および小文字の使用

アプリケーションに読み込む場合、アルファベットの大文字は小文字に一時的に変換されます。ただし、[データ] パネルには、元のテキストと同じ文字でテキストが表示されます。小文字および大文字は、この製品では同じように扱われます。

# 注意

本情報は全世界で提供する製品およびサービスについて作成したものです。

本書に記載の製品、サービス、または機能がその他の国においては提供されていない場合があります。お住まいの国で利用可能な製品、サービス、および機能については、現地 IBM の営業担当員にお尋ねください。本書で IBM 製品、プログラム、またはサービスに言及していても、その IBM 製品、プログラム、またはサービスのみが使用可能であることを意味するものではありません。これらに代えて、IBM の知的所有権を侵害することのない、機能的に同等の製品、プログラム、またはサービスを使用することができます。ただし、IBM 以外の製品とプログラムの操作またはサービスの評価および検証は、お客様の責任で行っていただきます。

IBM は、本書に記載されている内容に関して特許権（特許出願中のものを含む）を保有している場合があります。本書の提供は、お客様にこれらの特許権について実施権を許諾することを意味するものではありません。実施権についてのお問い合わせは、書面にて下記宛先にお送りください。

IBM Director of Licensing, IBM Corporation, North Castle Drive,  
Armonk, NY 10504-1785, U. S. A.

2 バイト文字セット (DBCS) 情報についてのライセンスに関するお問い合わせは、お住まいの国の IBM Intellectual Property Department に連絡するか、書面にて下記宛先にお送りください。

神奈川県大和市下鶴間1623番14号 日本アイ・ビー・エム株式会社 法務・知的財産 知的財産権ライセンス渉外

**以下の条項は、イギリスまたはこのような条項が法律に反する他の国では適用されません。** International Business Machines は、明示的または黙示的に関わらず、第三者の権利の侵害しない、商品性または特定の目的に対する適合性の暗黙の保証を含むがこれに限定されない、いかなる保証なく、本出版物を「そのまま」提供します一部の州では、特定の取引の明示的または暗示的な保証の免責を許可していないため、この文が適用されない場合があります。

この情報には、技術的に不適切な記述や誤植を含む場合があります。情報については変更が定期的に行われます。これらの変更は本書の新版に追加されます。IBM は、本書に記載されている製品およびプログラムについて、事前の告知なくいつでも改善および変更を行う場合があります。

IBM 以外の Web サイトに対するこの情報内のすべての参照は、便宜上提供されているものであり、決してそれらの Web サイトを推奨するものではありません。これらの Web サイトの資料はこの IBM 製品の資料に含

まれるものではなく、これらの Web サイトの使用はお客様の責任によるものとします。

IBM は、お客様が提供するいかなる情報も、お客様に対してなんら義務も負うことのない、自ら適切と信ずる方法で、使用もしくは配布することができるものとします。

本プログラムのライセンス保持者で、(i) 独自に作成したプログラムとその他のプログラム（本プログラムを含む）との間での情報交換、および (ii) 交換された情報の相互利用を可能にすることを目的として、本プログラムに関する情報を必要とする方は、下記に連絡してください。

IBM Software Group, Attention: Licensing, 233 S. Wacker Dr., Chicago, IL 60606, USA.

本プログラムに関する上記の情報は、適切な使用条件の下で使用することができますが、有償の場合もあります。

本書で説明されているライセンス・プログラムまたはその他のライセンス資料は、IBM 所定のプログラム契約の契約条項、IBM プログラムのご使用条件、またはそれと同等の条項に基づいて、IBM より提供されます。

本書に記述された性能は、管理された環境下で判定されたものです。従って、異なる使用環境で取得された結果は大きく異なる可能性があります。一部の測定は開発段階のシステムで行われた場合もあり、これらの測定結果が一般的に入手可能なシステムと同じであることは保証されません。また、一部の測定値は外挿法により推定されています。実際の結果は異なる場合もあります。本書のユーザーは個々の環境で入手したデータを検証する必要があります。

IBM 以外の製品に関する情報は、その製品の供給者、出版物、もしくはその他の公に利用可能なソースから入手したものです。IBM は、それらの製品のテストは行っておりません。したがって、他社製品に関する実行性、互換性、またはその他の要求については確認できません。IBM 以外の製品の性能に関する質問は、それらの製品の供給者にお願いします。

この情報には、日常の業務処理で用いられるデータや報告書の例が含まれています。できる限り詳細に説明するため、例には、個人、企業、ブランド、製品などの名前が使用されています。これらの名称はすべて架空のものであり、実際の企業で使用される名称および住所とは一切関係ありません。

電子的に複製されたこの情報を表示する場合には、写真やカラーのイラストは表示されない場合があります。

## 商標

IBM、IBM ロゴ、および [ibm.com](http://www.ibm.com)、SPSS は、世界の多くの国で登録された IBM Corporation の商標です。IBM の商標の現在のリストは、<http://www.ibm.com/legal/copytrade.shtml> を参照してください。

Adobe、Adobe のロゴ、PostScript、および、PostScript のロゴは、Adobe Systems Incorporated のアメリカ合衆国、その他の国家、または両方における登録商標、または、商標です。

Intel、Intel ロゴ、Intel Inside、Intel Inside ロゴ、Intel Centrino、Intel Centrino ロゴ、Celeron、Intel Xeon、Intel SpeedStep、Itanium、Pentium は 米国およびその他の国における Intel Corporation または、その子会社の商標または登録商標です。

Linux は、Linus Torvalds 氏のアメリカ合衆国およびその他の国における登録商標です。

Microsoft、Windows、Windows NT、および Windows のロゴは、米国およびその他の国における Microsoft 社の商標です。

UNIX は、米国およびその他の国における The Open Group の登録商標です。

Java およびすべての Java ベースの商標およびロゴは、米国およびその他の国の Sun Microsystems, Inc. の商標です。

Microsoft 製品の画面コピーは、Microsoft Corporation の承認を得て掲載しています。

他の製品またはサービスの名称は、IBM または、他の会社の商標である場合があります。

# 索引

- 固有表現有効化, 264
- 共起規則手法, 8, 122, 125, 127-128, 132, 135, 138
- 最大検索距離, 125, 132, 138
- 言語学的手法, 3, 7, 9, 124
- 不要語辞書, 221, 253-255
- 衝突変更子, 200
- 類義語辞書, 221, 247, 251-253
- 二次分析
  - 感性分析, 271
  - 係り受け解析, 271
- 参照変数, 27-28, 36, 71
- 固有表現, 90
  - 米国社会保障番号, 260
  - 電話番号, 260
  - 住所, 260
  - 数値, 260
  - 日付, 260
  - 時間, 260
  - 計量, 260
  - 通貨, 260
  - HTTP アドレス/URL, 260
  - IP アドレス, 260
  - 正規化、NonLingNorm.ini, 263
  - 正規表現、RegExp.ini, 261
  - アミノ酸, 260
  - 有効化および無効化, 264
  - パーセント, 260
  - プロテイン, 260
  - 電子メール アドレス, 260
- 強制定義, 265-266
- 感性分析, 271
  - オプション, 272
- 法的注意, 287
- 追跡回答, 79
- 代表語, 251
- 作成順, 54
- 信頼性, 9
- 優先度, 19, 21, 23
- 再利用
  - 分類, 178
- 新機能, 1
- 日本語, 268, 274
  - 種類, 278, 285
  - タイプのプロパティ, 282
  - リソース エディタ, 274
- 極座標, 199
- 正規化, 263
- 活用形, 128, 236, 238-240, 282
- 無効化
  - 不要語辞書, 255
  - 類義語辞書, 252, 258
- 固有表現, 264
- キーワード辞書, 247
- ステータス バー, 80
- ライブラリ, 227
- 省略形, 265, 267
- 統計量
  - 説明, 201
  - 視覚化における編集, 201
- 複数形, 239
- 視覚化
  - 数値書式, 193
  - 破線化, 189
  - 透明度, 189
  - 余白, 192
  - 分類, 196
  - 尺度, 194
  - 編集, 186
  - 軸, 194
  - text, 188
  - 置き換え, 196, 198-199
  - コピー, 204
  - 座標システムの変換, 199
  - 色とパターン, 189
  - 凡例の位置, 204
  - 点の縦横比, 191
  - 点の回転, 191
  - 点の形, 191
  - パディング, 192
  - パネル, 196, 198
- 記述子, 102
  - 分類, 110, 116
  - カテゴリの編集, 172
  - 最適の選択, 111
- 類義語, 91, 247
  - 代表語, 249, 283
  - 追加, 93, 249, 283
  - 色, 251
  - ! ^ \* \$ 記号, 250
  - Fuzzy Grouping の例外, 89, 258
  - エントリの削除, 253
  - 日本語テキストの場合, 283
  - の定義, 248
- 作成
  - 類義語, 91, 93, 249
  - 分類, 108, 118, 141
  - 種類, 95
  - オプションの要素, 251
  - カテゴリ規則, 156-157, 166
  - キーワード辞書, 238, 282
  - 日本語の類義語, 283
  - 規則のあるカテゴリ, 157
  - 不要語辞書のエントリ, 254



- プロジェクト, 30
- ライブラリ, 223
- リソースからテンプレート, 211
- 保存
  - 抽出結果, 91
  - テンプレートとしてのリソース, 211
  - プロジェクト, 55
  - プロジェクトの自動保存, 19
  - リソース, 216
- 修正
  - 句読点エラー, 89
  - スペルミス, 89
- 倒置, 90
- 公開, 55, 232
  - パブリックライブラリの追加, 224
  - ライブラリ, 230
- 凡例
  - 位置, 204
- 分類, 7, 27, 100, 102, 115, 169, 273
  - 言語学的手法, 118, 135
  - 共起規則, 122, 127, 132
  - 手動作成, 140
  - 記述子, 110-111, 116
  - 追加先, 171
  - 関連性, 105-106
  - 作成, 108, 134, 141
  - 削除, 180
  - 印刷, 178
  - 名前, 116, 169
  - 手動, 140
  - 拡張, 127, 135
  - 方法, 108
  - 方略, 109
  - 構築, 7, 118, 122, 124, 127, 137
  - 注釈, 116, 169
  - 移動, 173
  - 結合, 174
  - 編集, 169, 172
  - Web グラフ, 181
  - エクスポート, 57
  - 空白カテゴリを新規作成, 140
  - グループ手法の使用, 122
  - コピー, 178
  - スコアリング, 103
  - セマンティック ネットワーク, 122, 127, 131
  - テキスト分析パッケージ, 44, 46
  - 出現頻度に基づく手法, 134
  - 単語の強制, 176
  - 名前の変更, 140
  - 回答の投入, 175
  - 手法の使用, 127
  - 結果の調整, 10, 169
  - 共通性のグラフ, 181
  - 内包関係のコンセプト, 122, 127, 130
  - 派生関係のコンセプトの語幹, 122, 127-128
  - プロパティ, 116, 169
  - ラベル, 116, 169
  - 削除
    - 類義語, 253
    - 分類, 180
    - オプションの要素, 253
    - カテゴリ規則, 168
    - キーワード辞書, 247
    - 除外されたエントリ, 255
    - ライブラリ, 227, 229
    - ライブラリの無効化, 227
    - リソース テンプレート, 213
  - 名前
    - 分類, 116, 169
    - キーワード辞書, 245
    - ライブラリ, 226
  - 品詞, 266
  - 商標, 288
  - 回答, 104
    - カテゴリへの投入, 175
    - 完了としてマーク, 79
    - フラグ, 79
  - 図表, 181
  - 変数
    - 参照変数, 27-28, 36, 71
    - 合致, 74
    - ID変数, 27-28, 36, 71
    - インポート, 32, 34, 67, 70
    - エクスポート, 57
    - テキスト変数, 27-28, 36, 71
    - データ ソースの変更, 65
    - プロパティの編集, 54
    - リフレッシュ, 77
  - 変更
    - テンプレート, 212
    - データソース, 65
  - 定義, 110, 116
  - 強制
    - 強制投入, 104
    - 強制排除, 104
    - 回答, 175
    - 用語, 244
    - コンセプト抽出, 98
    - 列の表示, 104
    - 単語をカテゴリに, 176
  - 手法
    - 共起規則, 122, 127, 132, 135
    - 頻度, 134
    - セマンティック ネットワーク, 122, 127, 131, 135
    - ドラッグ アンド ドロップ, 141
    - 内包関係のコンセプト, 122, 127, 130, 135

## 索引

- 派生関係のコンセプトの語幹, 122, 127-128, 135
- 抽出, 3-4, 83, 88-89, 221, 236, 268
  - 固有表現, 90
  - 結果, 27
  - コンセプト、タイプ、パターン, 83
  - 単語の強制, 98
  - 結果の保存, 91
  - 結果の調整, 10, 91
  - ユニターム, 5, 90, 269
- 注釈
  - カテゴリ, 116, 169
  - プロジェクトのために, 52
- 用語
  - 活用形, 236
  - 色, 239, 283
  - エディタの検索, 225
  - キーワードの強制, 244
  - タイプへの追加, 240
  - 日本語タイプへの追加, 275
  - 不要語辞書に追加, 254
  - マッチ オプション, 236
  - 単語をカテゴリに強制投入, 176
- 移動
  - 分類, 173
  - キーワード辞書, 246
- 種類, 236
  - 日本語, 278, 282, 285
  - 作成, 238, 282
  - 抽出, 83
  - 辞書, 221
  - エディタの検索, 225
  - コンセプトの追加, 91
  - タイプの頻度, 134
  - デフォルトの色, 21, 239, 283
  - ビルトインのタイプ, 237
- 管理
  - 分類, 169
  - パブリック ライブラリ, 228
  - ローカル ライブラリ, 226
- 編集
  - 分類, 169, 172
  - カテゴリ規則, 168
  - 抽出結果の調整, 91
  - プロパティ, 169
- 翻訳, 24, 77
  - 翻訳精度設定, 39, 76, 78
  - [翻訳設定] ダイアログ, 77
  - オプション ダイアログ、翻訳タブ, 24
  - 英語への翻訳, 38, 75, 77
- 表示
  - 分類, 181
  - データ, 53
  - ライブラリ, 225
- 設定, 19, 21, 23
- 辞書, 16, 236
  - 不要語, 221, 236, 253
  - 類義語, 221, 236, 247
  - 種類, 221, 236
- 追加
  - 記述子, 111
  - 類義語, 93, 249
  - 種類, 95
  - オプションの要素, 251
  - カテゴリにコンセプト, 171
  - キーワード辞書へのキーワード, 240
  - 日本語キーワード辞書へのキーワード, 275
  - キーワードを不要語辞書へ, 254
  - サウンド, 21, 23
  - 日本語の類義語, 283
  - パブリック ライブラリ, 224
- 除外
  - fuzzy の除外, 258
  - 不要語エントリの無効化, 255
  - カテゴリ リンク, 126
  - 抽出からコンセプト, 97
  - 辞書の無効化, 247, 252
  - ライブラリの無効化, 227
- 頻度, 134
- 色
  - 不要語辞書, 254
  - 類義語, 251
  - 色オプションの設定, 21
  - 棒グラフのバー, 65
  - タイプおよびキーワード, 239, 283
- 社会保障番号 (固有表現), 260
- 電話番号 (固有表現), 260
- 住所 (固有表現), 260
- 数値 (固有表現), 260
- 日付 (固有表現), 260
- 時間 (固有表現), 260
- 計量 (固有表現), 260
- 通貨 (固有表現), 260
- 感嘆符 (!), 250
- 付属 (デフォルト) ライブラリ, 221
- 自動- 設定, 188
- & | !() 演算子, 167
- ! 類義語の ^ \* \$ 記号, 250
- [視覚化] パネル
  - カテゴリ Web グラフ, 181
  - グラフの更新, 181
- [表示] ボタン, 103
- AND 演算子, 167
- Budget キーワード辞書, 237
- delimiter, 19
- Fuzzy Grouping の例外, 89, 256, 258

- HTTP/URL (固有表現), 260  
 IBM SPSS Data Collection, 36, 71  
   エクスポート, 58  
   データ ソースの変更, 65  
 IBM SPSS Statistics.sav ファイル, 32, 67  
   アウトプット フォーマット, 57  
   エクスポート, 58  
   データ ソースの変更, 65  
 ID変数, 27-28, 36, 71  
 IP アドレス (固有表現), 260  
 Language Weaver, 24, 38, 75  
 \*.lib, 228  
 Location キーワード辞書, 237  
 Microsoft Excel.xls / .xlsx ファイル, 33, 60, 69  
   アウトプット フォーマット, 57  
   データ ソースの変更, 65  
   定義済みカテゴリのインポート, 142-143  
   定義済みカテゴリのエクスポート, 152  
 Negative キーワード辞書, 237  
 NOT 演算子, 167  
 ODBC データソース, 34-35, 65, 70-71  
 OR 演算子, 167  
 Organization キーワード辞書, 237  
 Person キーワード辞書, 237  
 Positive キーワード辞書, 237  
 Product キーワード辞書, 237  
 \*.tap テキスト分析パッケージ, 40, 44-46, 49  
 TLA, 212  
 Uncertain キーワード辞書, 237  
 Unknown キーワード辞書, 237  
 Web グラフ, 181  
 Web テーブル, 181  
 .xls / .xlsx ファイル, 57, 60  
  
 アスタリスク (\*)  
   不要語辞書, 254  
   類義語, 250  
 アドバンス リソース, 256  
   エディタの検索と置換, 257  
 アミノ酸 (固有表現), 260  
 アンチリンク, 126  
  
 インデント形式, 150  
 インポート  
   ODBC, 34, 70  
   テンプレート, 215  
   データの更新, 77  
   データの準備, 28  
   データのインプット, 36, 71  
   データの再インポート, 65  
   パブリック ライブラリ, 228  
   定義済みカテゴリ, 143  
  
 エクスポート, 57  
   IBM SPSS Data Collectionのために, 58  
   IBM SPSS Statistics.sav ファイルとして, 58  
   Microsoft Excel.xls / .xlsx  
   ファイルとして, 60  
   アウトプット フォーマット, 57  
   棒グラフ, 62  
   テンプレート, 215  
   結果のカテゴリライズ化, 57-58, 60  
   パブリック ライブラリ, 229  
   定義済みカテゴリ, 152  
 句読点エラー, 89  
  
 オプション, 19  
   翻訳, 24  
   表示, 21  
   サウンド, 23  
   システム, 19  
 オプションの要素, 247  
   対象, 251  
   追加, 251  
   エントリの削除, 253  
   の定義, 249  
 回答およびカテゴリの関連性, 105-106  
  
 カテゴリ結合, 174  
 カテゴリ規則, 156-157, 164, 166-168  
   共起規則, 122, 127, 135  
   例, 164  
   類義語から, 7, 122-124, 127-128, 135, 138  
   コンセプト共起から, 7, 123-124, 128, 132, 138  
   シンタックス, 157  
 カテゴリ名, 102  
 新規カテゴリ, 140  
 未カテゴリ化, 102  
 カテゴリ Web グラフ/テーブル, 181, 183-184  
 [カテゴリ] パネル, 102  
 [カテゴリ] パネルの列の表示, 102  
 カテゴリ棒グラフ, 181-182  
 カテゴリの作成, 7, 118-119, 121-124, 126-135, 137, 140, 273  
   共起規則手法, 7, 124, 138  
   同格/横の関係をもとにグループ化, 137  
   セマンティック ネットワーク手法, 7, 124, 138  
   手法の使用, 7, 124  
   内包関係のコンセプトの手法, 138

## 索引

- 派生関係のコンセプトの語幹による手法, 7, 124, 138
- 分類リンクの例外, 126
- カテゴリの印刷, 178
- カテゴリの拡張, 135
- カテゴリの結合, 174
- カテゴリのラベル, 116, 169
- カレット記号 (^), 250
  
- キーワード辞書, 221
  - 無効化, 247
  - 類義語, 236
  - 削除, 247
  - 移動, 246
  - オプションの要素, 236
  - キーワードの強制, 244
  - キーワードの追加, 240
  - タイプの作成, 238, 282
  - 名前の変更, 245
  - 日本語のキーワードの追加, 275
  - ビルトインのタイプ, 237
- キーワードおよびタイプの検索, 225
- キーワードのコンポーネント化, 128
  
- クラスタ, 203
- グラフ
  - 編集, 185
  - カテゴリ Web グラフ, 181
  - 棒グラフのエクスポート, 62
  - グラフ要素のサイズ, 192
  - 探索的分析モード, 185
  - リフレッシュ, 181
- グラフ要素
  - 衝突変更子, 203
  - 変換, 200
  - 変更, 200
  - 種類, 200
- 棒グラフ, 62, 181
- グラフの更新, 181
- グラフの編集
  - グラフ要素のサイズ, 192
- グローバルな区切り文字, 19
  
- コア ライブラリ, 237
- コピー
  - 分類, 178
- コンセプト
  - 抽出, 83
  - カテゴリ内, 110, 116
  - カテゴリに追加, 110, 116, 171
  - 抽出からの除外, 97
  - タイプの作成, 91
  - タイプへの追加, 95
  - 最適な記述子, 111
  - 抽出への強制投入, 98
- コンセプトの無視, 97
- コンパクト形式, 148
- コンポーネント化, 128
- コード フレーム, 142-143
  
- サウンド オプション, 23
- サウンドのミュート, 23
- 復元されたファイル, 19
  
- 座標システム
  - 変換, 199
- 抽出した棒グラフのタイトル, 64
- ジッター, 203
  
- [スコア] ボタン, 103
- スコアリング, 103
- ステータス バー, 80
- すべての抽出結果:, 91
- すべてのドキュメント, 102
- スペル ミス, 28, 89, 258
- 更新する
  - グラフ, 181
  - テンプレート, 211
  - ライブラリ, 230, 233
- 生成するカテゴリ数の最大値, 126
  
- 言語処理セクション, 256, 265
  - 強制定義, 265-266
  - 省略形, 265, 267
  - 抽出パターン, 265-266
- セマンティック ネットワーク手法, 7, 122, 124, 127-128, 131, 135, 138
  
- タイプの頻度, 134
  
- テキスト分析, 3, 8-9, 12
- テキスト変数, 27-28, 36, 71
- テキスト マイニング, 3
- テキスト マッチ, 116, 169, 176
- テキスト分析パッケージ, 40, 44-46
- テキスト区切り文字, 19
- デフォルトのライブラリ, 221
- テンプレート, 5, 207, 269
  - 削除, 213
  - 復元, 216
  - TLA, 212
  - インポートとエクスポート, 215
  - テンプレートの切り替え, 212
  - 名前の変更, 213

- バックアップ, 216
- 更新または名前を付けて保存, 211
- リソースから作成, 211
- データ
  - 分類, 100, 118, 140
  - 抽出, 88
  - 表示, 53
  - IBM SPSS Data Collection, 36, 58, 71
  - IBM SPSS Statistics.sav ファイル, 32, 57-58, 67
  - Microsoft Excel.xls / .xlsx ファイル, 33, 57, 60, 69
  - ODBC, 35, 71
  - エクスポート, 57
  - カテゴリの作成, 7, 122, 124, 127, 135
  - ソート, 54
  - データ ソースの変更, 65
  - データソースの選択, 31, 66
  - 結果の調整, 91
  - 変数のプロパティの編集, 54
  - リフレッシュ, 77
- [データ] パネル, 104
- [表示] ボタン, 103
- [データ] パネルの列の表示, 104
- データと変数のソート, 54
- データの準備, 28
- データの調査, 2, 27-28, 36, 71
- データの再インポート, 65
  
- 検索と置換 (アドバンス リソース), 257
- ドキュメント列, 102-103
- ドッジ, 203
- ドラッグ アンド ドロップ, 141
- ドル記号 (\$), 250
  
- 固有表現の有効化, 264
- 固有表現の無効化, 264
- 自由記述式の質問, 2
- 選択形式の質問, 2
- 活用形の生成, 236, 238-240, 282
- 視覚化の編集, 186
  - 数値書式, 193
  - 自動設定, 187
  - 破線化, 189
  - 透明度, 189
  - 余白, 192
  - 分類, 196
  - 尺度, 194
  - 選択, 187
  - 軸, 194
  - 3-D 効果の追加, 199
  - text, 188
- カテゴリの結合, 196
- カテゴリの除外, 196
- カテゴリの並び替え, 196
- カテゴリを閉じる, 196
- 置き換え, 198-199
- 座標システムの変換, 199
- 色とパターン, 189
- 凡例の位置, 204
- 点の縦横比, 191
- 点の回転, 191
- 点の形, 191
- パディング, 192
- パネル, 198
- ルール, 187
- 名前の変更
  - 分類, 140, 169
  - キーワード辞書, 245
  - プロジェクト, 55
  - ライブラリ, 226
  - リソース テンプレート, 213
- 空白の回答, 28
- 結果の調整
  - 抽出結果, 10, 91
  - 分類, 10, 169
  - コンセプトの除外, 97
  - コンセプトのタイプへの追加, 95
  - コンセプト抽出を強制, 98
  - タイプの作成, 95
  - 類義語の追加, 93
- 規則の演算子 & | !() , 167
- 視覚化のコピー, 204
- 内包関係のコンセプトの手法, 8, 122, 124-125, 127-128, 130, 135
- 派生関係のコンセプトの語幹による手法, 122, 127-128, 135, 138
- 回答のマーク, 79
- 列の折り返し, 21
  
- パターン, 83, 86
- 抽出パターン, 265-266
- パレット
  - 非表示, 187
  - 移動, 187
  - 表示する, 187
- パーセント (固有表現), 260
  
- 質問ビュー, 12-13
- ビューア
  - テキスト分析ウィンドウ :, 12
  - 質問ビュー, 13
  - プロジェクト全体, 15
  - リソース エディタ ウィンドウ, 16

## 索引

- ファイルの復元, 19
- フォントの色, 239, 283
- フラグ回答, 79
- 完了フラグ, 79
- 重要フラグ, 79
- フラット リスト形式, 147
- プロジェクト, 27, 50
  - 作成, 30
  - 保存, 55
  - 共有, 79
  - 翻訳, 38, 75
  - カテゴリおよびリソースの選択, 40
  - カテゴリの再利用, 178
  - 開く, 50
  - ステータス バー, 80
  - テキスト分析パッケージ, 40
  - データソース, 31, 66
  - 名前の変更, 55
  - 変数の選択, 36, 71
  - プロパティ, 52
  - ライブラリのオプション, 19
- プロジェクトの自動保存, 19
- プロジェクトの共有, 79
- プロジェクト全体ビュー, 12, 15
- プロジェクトを開く, 50
- プロテイン (固有表現), 260
- プロパティ
  - 分類, 116, 169
  - 変数, 54
  - 日本語のタイプ, 282
  - プロジェクト, 52
- ブール型演算子, 167
  
- マッチ オプション, 236, 238, 240-242, 282
  
- 定義済みカテゴリ, 142-143, 152
  - インデント形式, 150
  - コンパクト形式, 148
  - フラット リスト形式, 147
- 積み重ね, 203
  
- 電子メール (固有表現), 260
  
- モニタリング, 79
- 探索的分析モード, 185
- 編集モード, 185
  
- ユニターム, 90
- ユーザー定義の色, 21
  
- ライブラリ, 16, 221, 236
  - 無効化, 227
  - 作成, 223
  - 公開, 55, 232
  - 共有, 79
  - 削除, 227, 229
  - 同期, 230
  - 名前, 226
  - 表示, 225
  - 辞書, 221
  - 追加, 224
  - インポート, 228
  - エクスポート, 229
  - 共有および公開, 230
  - コア ライブラリ, 237
  - 更新する, 233
  - 名前の変更, 226
  - 付属のデフォルト ライブラリ, 221
  - パブリック ライブラリ, 230
  - 予算ライブラリ, 237
  - 意見ライブラリ, 237
  - ライブラリ同期の警告, 230
  - リンク, 224
  - ローカル ライブラリ, 230
  - 予算ライブラリ, 237
  - 意見ライブラリ, 237
  - ライブラリの共有, 230
    - 公開, 55, 232
    - 更新する, 233
  - パブリック ライブラリの追加, 224
  - ライブラリの同期, 230, 232-233
  - ライブラリのフィルタリング, 225
  
- 区切り文字, 19
- 係り受け解析, 271
- リソース
  - 復元, 216
  - アドバンス リソースを編集, 256
  - テンプレート リソースに切り替え, 212
  - 付属のデフォルト ライブラリ, 221
  - バックアップ, 216
- 言語リソース, 27, 221
  - テキスト分析パッケージ, 44, 46
  - リソース テンプレート, 207
- リソース エディタ, 12, 211-212, 256
  - 日本語, 274
  - グローバルな区切りオプション, 19
  - テンプレートの作成, 211
  - テンプレートの更新, 211
  - リソースの切り替え, 212
- リソース テンプレート, 5, 207, 269
- リソースからテンプレートを作成, 211
- リソースのバックアップ, 216
- リソースをテンプレートに置き換え, 212
- 割り当てフラグ, 79

---

最小リンク値, 126  
リンクの例外, 126

ルール

共起規則手法, 132  
作成, 166  
削除, 168  
編集, 168  
シンタックス, 157  
ブール型演算子, 167

レコード, 104  
レポートおよび棒グラフ, 62