

Guide de l'utilisateur IBM SPSS Text
Analytics for Surveys 4.0.1



Remarque : Avant d'utiliser ces informations et le produit concerné, lisez les informations générales de la rubrique Avis sur p. 269.

Cette version concerne IBM® SPSS® Text Analytics for Surveys 4 et toutes les versions et modifications ultérieures sauf mention contraire dans les nouvelles versions.

Les captures d'écran de produits Adobe sont réimprimées avec l'autorisation de Adobe Systems Incorporated.

Les captures d'écran de produits Microsoft sont réimprimées avec l'autorisation de Microsoft Corporation.

Matériel licencié - propriété d'IBM

© Copyright IBM Corporation 2004, 2011.

Droits limités des utilisateurs gouvernementaux américains - l'utilisation, la copie ou la divulgation sont régies par le contrat de marché à prix unitaires entre GSA ADP et IBM Corp.

Préface

Bienvenue dans IBM® SPSS® Text Analytics for Surveys version 4.0.1, une application de codage de texte d'enquête qui permet une analyse pertinente des réponses aux questions ouvertes. Avec ce produit, toute personne exécutant une enquête d'opinion peut rapidement transformer des réponses d'enquête non structurées en données quantitatives. Le “déverrouillage” de ces données de texte peut considérablement améliorer la qualité d'analyse et la possibilité de prise de décision. Cette application vous permet d'importer des données d'enquête, extraire des concepts clés, affiner les résultats, et catégoriser les réponses. Une fois que vous avez catégorisé vos données, vous pouvez exporter vos catégories pour les importer dans des outils analytiques quantitatifs, tels que le système IBM® SPSS® Statistics, pour une analyse et une représentation graphique plus approfondies.

SPSS Text Analytics for Surveys combine des technologies linguistiques avancées conçues pour extraire et classifier de manière fiable les concepts clés dans les réponses d'enquête ouvertes avec des techniques manuelles. A l'aide d'algorithmes de création de catégories et de la simple fonctionnalité de glisser-déposer, vous pouvez créer des catégories, ou des “codes”, dans lesquels seront catégorisées vos réponses d'enquête. Les catégories produites peuvent également être réutilisées pour fournir des résultats homogènes sur des études identiques ou similaires. Les données de réponses ouvertes pouvant varier considérablement d'une enquête à une autre, deux projets ne seront jamais exactement identiques. Vous pouvez toutefois décider de suivre la même procédure de base pour accomplir votre analyse. [Pour plus d'informations, reportez-vous à la section Processus habituel dans le chapitre 2 sur p. 11.](#)

À propos d'IBM Business Analytics

Le logiciel IBM Business Analytics propose des informations complètes, cohérentes et précises que les preneurs de décisions peuvent utiliser en toute confiance pour améliorer les performances de leur entreprise. Un portefeuille complet de [veille économique](#), d'[analyses prédictives](#), de [gestion des performances et de stratégie financière](#) et d'[applications analytiques](#) qui propose des informations claires, immédiates et décisionnelles sur les performances actuelles permettant de prévoir les résultats à venir. Ce logiciel allie des solutions dédiées à l'industrie, des pratiques ayant fait leur preuve et des services professionnels afin que les organisations de toute taille puissent obtenir la meilleure productivité possible, automatiser leurs décisions en toute confiance et améliorer leurs résultats.

Ce portefeuille contient le logiciel IBM SPSS Predictive Analytics qui aide les organisations à prévoir les événements à venir et à agir en fonction de ces informations pour améliorer leurs résultats. Les clients de l'industrie du commerce et des domaines gouvernementaux et de l'éducation du monde entier font confiance à la technologie IBM SPSS et reconnaissent son avantage concurrentiel pour attirer et fidéliser les clients, pour grossir la base de clientèle tout en réduisant la fraude et en limitant les risques. En intégrant le logiciel IBM SPSS à leurs opérations quotidiennes, les organisations deviennent des entreprises prédictives, capables de diriger et d'automatiser leurs décisions afin de répondre à leurs objectifs et d'obtenir des avantages concurrentiels mesurables. Pour des informations supplémentaires ou pour joindre un représentant, consultez le site <http://www.ibm.com/spss>.

Assistance technique

L'assistance technique est disponible pour les clients de la maintenance. Les clients peuvent contacter l'assistance technique pour obtenir de l'aide concernant l'utilisation des produits IBM Corp. ou l'installation dans l'un des environnements matériels pris en charge. Pour joindre l'assistance technique, consultez le site Web IBM Corp. à l'adresse <http://www.ibm.com/support>. Lorsque vous demandez une assistance, veillez à être en mesure de pouvoir vous identifier ainsi que votre organisation et votre contrat d'assistance.

Contenu

Partie I: Pour commencer

1 A propos de l'analyse de texte 1

Nouveautés	1
Données d'enquête ouvertes	2
A propos de Text Mining	3
Fonctionnement de l'extraction	4
Fonctionnement de la catégorisation	6
Préparation à l'analyse de texte	8
Fiabilité et affinement	9
Affinement des ressources linguistiques	9
Affinement des définitions de catégorie	9

2 Pour commencer 11

Processus habituel	11
La fenêtre d'analyse de texte	12
La vue Question	13
La vue Projet complet	15
La fenêtre Editeur de ressources	16
Définition des options	19
Options : Onglet Système	19
Options : Onglet Affichage	21
Options : Onglet Sons	23
Options : Onglet Traduction	23
Paramètres de Microsoft Internet Explorer de l'aide	25

Partie II: Analyse de texte

3 Création de projets et de packages 27

Création de projets	27
Préparation de vos données	28
Démarrage de nouveaux projets	29
Sélection de sources de données	30

Sélection de variables	35
Traduction vers l'anglais	37
Sélection de catégories et de ressources	39
Utilisation des packages d'analyse de texte	43
Création des packages d'analyse de texte	43
Mise à jour des Packages d'analyse de texte	45

4 Utilisation des projets 49

Ouvrir des projets	49
Modifier les propriétés du projet	50
Afficher les données du projet	51
Trier les variables	52
Modifier les propriétés des variables	52
Enregistrer des projets	53
Export des résultats de catégorisation	55
Export vers IBM SPSS Statistics ou IBM SPSS Data Collection	56
Export vers Microsoft Excel	58
Exporter des graphiques récapitulatifs	60
Modifier les sources de données	63
Sélection de sources de données	64
Sélection de variables	69
Correspondance de variables	71
Traduction vers l'anglais	72
Mise à jour des données	74
Traduction vers l'anglais	74
Partage de projets	76
Marquer des réponses	76
Barre d'état du projet	77

5 Extraction de données 80

Résultats extraits : Concepts, types et patrons	80
Extraction de données	85
Enregistrement des résultats de l'extraction	87
Affinage des résultats de l'extraction	88
Ajout de synonymes	89
Ajout de concepts à des types	91

Exclusion de concepts de l'extraction	94
Extraction de mots imposée	94

6 Catégorisation des données textuelles 96

Le panneau Catégories	97
Le panneau Données	100
Pertinence des catégories	102
Stratégies et méthodes de création de catégories	104
Méthodes de création de catégories	105
Stratégies de création de catégories	105
Conseils pour la création de catégories	106
Choix des meilleurs descripteurs	107
A propos des catégories	110
Propriétés de la catégorie	111
Création de catégories	112
Paramètres linguistiques avancés	116
A propos des Techniques linguistiques	121
Paramètres de fréquence avancés	127
Extension de catégories	128
Création de catégories manuellement	133
Création de catégories ou attribution d'un nouveau nom aux catégories	133
Création de catégories par la méthode Glisser-déposer	134
Import et export de catégories prédéfinies	135
Import de catégories prédéfinies	136
Exporter des catégories	144
Utilisation des règles de catégorie	147
Syntaxe des règles de catégorie	148
Utilisation des patrons TLA dans les règles de catégorie	150
Utilisation de caractères génériques dans les règles de catégorie	152
Exemples de règles de catégorie	154
Création de règles de catégorie	156
Modification et suppression des règles	158
Edition et réglage des catégories	158
Modification des propriétés de catégories	159
Ajout de descripteurs aux catégories	160
Modification des descripteurs de catégorie	161
Déplacement de catégories	162
Fusion ou combinaison de catégories	163
Forcer les réponses à entrer dans des catégories	164
Correspondance de texte dans les catégories	165

Copie de catégories	167
Impression de catégories	167
Suppression de catégories	168

7 Visualisation des graphiques 170

Diagramme Barre Catégorie	171
Graphique Relations de catégorie	172
Tableau des relations de catégorie	173
Utilisation des palettes et des barres d'outils de graphiques	173
Modification des visualisations	174
Règles générales d'édition de visualisations	175
Edition et formatage de texte	176
Changement des couleurs, des motifs, des tirets et de la transparence	177
Changement de la forme et du rapport d'aspect des points et rotation des points	179
Changement de la taille des éléments graphiques	179
Spécification des marges et de l'extension	180
Formatage des nombres	180
Changement des paramètres d'axe et d'échelle	182
Modification de catégories	184
Modification de l'orientation des panneaux	185
Transformation du système de coordonnées	186
Modification des statistiques et des éléments graphiques	187
Changement de la position de la légende	190
Copie d'une visualisation et de données de visualisation	191
Raccourcis clavier	191

Partie III: Editeur de ressources

8 Modèles et ressources 194

Interface de l'éditeur	195
Création et mise à jour de modèles	197
Changement des modèles de ressources	198
Gestion des modèles	199
Import et export des modèles	200
Sauvegarde des ressources	202
Import des fichiers de ressources	203

9 Utilisation des bibliothèques 206

Bibliothèques fournies	206
Création de bibliothèques	208
Ajout de bibliothèques publiques	208
Recherche de termes et de types	209
Affichage des bibliothèques	210
Gestion des bibliothèques locales	211
Attribution d'un nouveau nom à une bibliothèque locale	211
Désactivation des bibliothèques locales	211
Suppression des bibliothèques locales	212
Gestion des bibliothèques publiques	212
Partage de bibliothèques	214
Publication de bibliothèques	216
Mise à jour des bibliothèques	217
Résolution des conflits	218

10 A propos des dictionnaires de bibliothèque 220

Déclarations de types	220
Types intégrés	221
Création de types	222
Ajout de termes	224
Ajout des termes forcés	228
Attribution de nouveaux noms aux types	229
Déplacement de types	230
Désactiver et supprimer des types	231
Dictionnaires des substitutions/synonymes	231
Définition de synonymes	233
Définition des éléments optionnels	235
Désactiver et supprimer des substitutions	236
Dictionnaires d'exclusions	236

11 À propos des ressources avancées 239

Recherche	240
Remplacement	241
Regroupement flou	242

Entités non linguistiques	243
Expressions régulières	244
Normalisation	246
Configuration	247
Traitement des langues	248
patrons d'extraction	248
Définitions forcées	249
Abréviations	250

Annexes

A Exceptions pour le texte en japonais **251**

Extraction et catégorisation du texte japonais	251
Fonctionnement de l'extraction	251
Fonctionnement de l'extraction secondaire	253
Fonctionnement de la catégorisation	256
Modification des ressources pour du texte en japonais	256
Panneau Arborescence des bibliothèques japonaises, panneau des types et panneau des termes	258
Types disponibles pour du texte en japonais	261
Modification des propriétés de types japonais	264
Utilisation du dictionnaire des synonymes pour du texte japonais	265
Validation et compilation des ressources en japonais	267
Autres exceptions pour le japonais	267

B Avis **269**

Index **272**

Partie I:
Pour commencer

A propos de l'analyse de texte

Bienvenue dans IBM® SPSS® Text Analytics for Surveys version 4.0.1, une application de codage de texte d'enquête qui permet une analyse pertinente des réponses aux questions ouvertes. Avec ce produit, toute personne exécutant une enquête d'opinion peut rapidement transformer des réponses d'enquête non structurées en données quantitatives. Le "déverrouillage" de ces données de texte peut considérablement améliorer la qualité d'analyse et la possibilité de prise de décision. Cette application vous permet d'importer des données d'enquête, extraire des concepts clés, affiner les résultats, et catégoriser les réponses. Une fois que vous avez catégorisé vos données, vous pouvez exporter vos catégories pour les importer dans des outils analytiques quantitatifs, tels que le système IBM® SPSS® Statistics, pour une analyse et une représentation graphique plus approfondies.

SPSS Text Analytics for Surveys combine des technologies linguistiques avancées conçues pour extraire et classifier de manière fiable les concepts clés dans les réponses d'enquête ouvertes avec des techniques manuelles. A l'aide d'algorithmes de création de catégories et de la simple fonctionnalité de glisser-déposer, vous pouvez créer des catégories, ou des "codes", dans lesquels seront catégorisées vos réponses d'enquête. Les catégories produites peuvent également être réutilisées pour fournir des résultats homogènes sur des études identiques ou similaires. Les données de réponses ouvertes pouvant varier considérablement d'une enquête à une autre, deux projets ne seront jamais exactement identiques. Vous pouvez toutefois décider de suivre la même procédure de base pour accomplir votre analyse. [Pour plus d'informations, reportez-vous à la section Processus habituel dans le chapitre 2 sur p. 11.](#)

Nouveautés

Les nouvelles fonctions de IBM® SPSS® Text Analytics 4.0.1 sont les suivantes :

Catégories hiérarchiques. Les catégories peuvent également avoir une structure hiérarchique, ce qui signifie qu'elles peuvent contenir des sous-catégories et que ces sous-catégories peuvent elles-même contenir des sous-catégories, et ainsi de suite. Vous pouvez importer des structures de catégories prédéfinies, nommées auparavant plans de codage, avec des catégories hiérarchiques ou bien créer ces catégories hiérarchiques à l'intérieur du produit.

En effet, les catégories hiérarchiques vous permettent de créer une structure en arborescence contenant une ou plusieurs sous-catégories afin d'y regrouper plus clairement les éléments par groupe de concepts ou de rubriques. Prenons un exemple simple avec des activités de loisir ; pour répondre à une question telle que *Quelle activité aimeriez-vous faire si vous aviez davantage de temps ?* vous pourriez avoir des catégories principales comme *sport, art et artisanat, pêche, etc.* ; à un niveau inférieur, sous *sport*, vous pourriez avoir des sous-catégories pour voir s'il s'agit de *jeux de ballon, sports nautiques, etc.*

Accès à Language Weaver. Le mode d'accès à l'interface de traduction Language Weaver a été simplifié de façon à utiliser une seule URL et les détails de sécurité associés.

Données d'enquête ouvertes

Les questionnaires d'enquête contiennent généralement différentes sortes de questions, notamment des questions ouvertes et fermées. Les premières, les **questions fermées**, présentent un ensemble limité de réponses qui permet de nombreux types d'analyse quantitative. Les deuxièmes, les **questions ouvertes**, permettent à la personne interrogée de fournir une réponse non structurée d'une longueur et d'un niveau de détails variables.

Les mots utilisés par les personnes pour répondre à une question vous en disent beaucoup sur ce qu'ils pensent et ressentent. C'est pourquoi des questions ouvertes sont souvent incluses dans les enquêtes : elles fournissent des informations plus variées et texturées que les questions fermées et fournissent souvent des informations qui n'avaient pas été prévues par le concepteur de l'enquête.

Toutefois, l'utilisation d'un plus grand nombre de questions ouvertes a un coût généralement considéré comme prohibitif, en raison des frais d'analyse induits par l'interprétation des ressources. De plus, ces longues réponses doivent être codées de manière standardisée, à l'aide d'un ensemble détaillé d'instructions de codage. Cette tâche peut être difficile car les codeurs sont souvent en désaccord sur la manière de catégoriser des réponses spécifiques. Lorsque les codeurs sont en désaccord, la fiabilité des résultats est réduite. Pour toutes ces raisons, le codage de réponses ouvertes a longtemps été considéré comme long et coûteux, ces contraintes l'emportant souvent sur les bénéfices dérivés des données collectées.

IBM® SPSS® Text Analytics for Surveys offre une alternative à cette procédure coûteuse, car il peut accomplir le codage de réponses ouvertes en une fraction du temps nécessaire à l'exécution de cette tâche manuellement. Grâce à une théorie et des technologies linguistiques avancées, SPSS Text Analytics for Surveys analyse le texte des réponses ouvertes sous la forme d'un ensemble d'expressions et de phrases dont la structure grammaticale fournit des indications sur le sens d'une réponse. Une fois ce texte analysé, les concepts clés et patrons de mots sont extraits et classés en catégories.

Vous pouvez utiliser des techniques de création de catégories intégrées pour créer automatiquement des catégories, et des techniques manuelles pour affiner les résultats. La fiabilité des résultats augmente considérablement, car l'extraction et la catégorisation sont toujours exécutées de manière homogène et répétitive. La même réponse est catégorisée dans les mêmes catégories à chaque fois, sauf si vous choisissiez d'affiner vos définitions de catégorie ou bibliothèques linguistiques.

Le succès d'une analyse d'enquête ne repose pas sur une méthode unique. La nature subjective de l'interprétation des réponses ouvertes nécessite l'utilisation de plusieurs techniques. Outre ses techniques intégrées d'extraction et de création de catégories, SPSS Text Analytics for Surveys repose également sur la compréhension par l'utilisateur des objectifs d'analyse de texte spécifiques pour chaque enquête. L'analyse de texte est particulièrement puissante lorsqu'elle s'effectue de manière itérative (extraire, examiner, affiner, réextraire), et son utilité dépend souvent du temps et des efforts consacrés à l'examen et l'affinement manuels des résultats d'extraction et des définitions de catégorie. [Pour plus d'informations, reportez-vous à la section **Fiabilité et affinement** sur p. 9.](#)

Si vous travaillez avec des questions identiques ou similaires dans des enquêtes répétitives, vous pouvez réutiliser les catégories dans d'autres questions ou projets. Le fait de réutiliser des catégories permet une plus grande cohérence dans le codage, en permettant également d'épargner du temps et des efforts.

Vous pouvez également effectuer d'autres analyses. Les catégories que vous produisez peuvent être utilisées dans différents types d'analyses statistiques avec d'autres questions du questionnaire ou d'autres informations personnelles, pour obtenir davantage d'informations sur les personnes interrogées et sur leurs opinions et comportements. Après avoir utilisé SPSS Text Analytics for Surveys pour découvrir les catégories qui sous-tendent un ensemble de réponses, vous pouvez également exporter les catégories pour approfondir l'analyse dans un autre programme tel que IBM® SPSS® Statistics Base.

A propos de Text Mining

L'analyse de texte, sorte d'analyse qualitative, est l'extraction d'informations utiles d'un texte, (par exemple les réponses ouvertes) de manière à regrouper les principaux concepts ou idées qui figurent dans ce texte dans un nombre approprié de catégories. Vous pouvez effectuer une analyse de texte sur tout type et toute longueur de texte, bien que l'approche analytique variera quelque peu.

Etant donné que les enregistrements courts sont moins complexes et contiennent généralement moins de mots et de réponses ambigus, leur catégorisation est plus simple. Par exemple, si nous posons des questions ouvertes et courtes au cours d'une enquête sur les trois activités préférées des personnes interrogées lorsqu'elles sont en vacances, leurs réponses seront pour la plupart courtes : *aller à la plage, visiter des parcs nationaux* ou *ne rien faire*. Des réponses ouvertes plus longues risquent, par contre, d'être plutôt complexes et démesurées, en particulier si les personnes interrogées sont instruites, motivées et qu'elles disposent de suffisamment de temps pour remplir un questionnaire. Si nous interrogeons des personnes sur leurs opinions politiques dans le cadre d'une enquête ou si nous mettons au point un blog concernant la politique, nous nous attendons à recevoir de très longs commentaires sur une grande variété de problèmes et de prises de position.

Les enquêteurs n'analysent généralement pas les très longues réponses. Les réponses de la plupart des enquêtes sont généralement courtes ou d'une longueur moyenne (d'une phrase à un petit paragraphe). IBM® SPSS® Text Analytics for Surveys a été conçu pour gérer cette taille de texte mais peut analyser des réponses beaucoup plus longues.

Il existe différentes méthodes d'analyse de texte. D'abord, il existe l'approche manuelle : des personnes lisent les réponses aux enquêtes, en note leur contenu, déterminent les concepts clés qu'elles contiennent et leur attribuent des codes. Les personnes étant douées pour comprendre les textes, cette approche est plutôt précise. Mais elle prend du temps, demande beaucoup de travail et, avec l'immense volume de texte disponible, est de plus en plus impraticable. Cette approche se base également sur l'interprétation de chaque personne qui analyse les données.

Une autre approche est d'utiliser des solutions automatisées. Vous avez le choix entre de nombreuses solutions automatisées différentes, notamment des solutions statistiques et linguistiques. SPSS Text Analytics for Surveys offre une combinaison de techniques linguistiques et statistiques automatisées afin de générer des résultats aussi fiables que possible à chaque étape du processus. Ce produit utilise des techniques basées sur la linguistique pour extraire automatiquement des concepts clés à partir des réponses, et des techniques linguistiques et statistiques peuvent être utilisées pour créer des définitions de catégorie (codes) qui sont attribuées aux réponses.

Fonctionnement de l'extraction

Lors de l'extraction des principaux concepts et idées de vos réponses, IBM® SPSS® Text Analytics for Surveys s'appuie sur une analyse de texte linguistique. Cette approche a la même efficacité en temps et en argent que les systèmes statistiques. Mais elle offre un plus grand degré de précision tout en ne nécessitant que peu d'intervention humaine. L'analyse de texte linguistique se base sur un domaine d'étude appelé processus de langage naturel, également connu sous le nom de linguistique computationnelle.

Pour illustrer la différence entre la méthode statistique et la méthode linguistique pendant le processus d'extraction, examinons le mode d'action de chacune de ces méthodes dans le cadre d'une requête concernant l'expression `reproduction de documents`. La solution statistique et la solution linguistique doivent toutes les deux étendre le mot `reproduction` à ses synonymes (`copie` et `duplication`, par exemple). Sinon, des informations pertinentes risquent d'être ignorées. Si toutefois une solution statistique tente d'appliquer ce type de synonymie (recherche d'autres termes possédant la même signification), elle inclura vraisemblablement le terme `naissance`, générant ainsi de nombreux résultats inappropriés. Comme la compréhension de la langue permet de lever toute ambiguïté dans le texte, l'exploration de texte linguistique reste par définition la méthode la plus fiable.

Si vous comprenez le fonctionnement du processus d'extraction, vous êtes plus à même de prendre les décisions-clés lorsque vous affinez vos ressources linguistiques (bibliothèques, types, synonymes, etc.). Les principales étapes du processus d'extraction sont les suivantes :

- Conversion des données source en un format standard
- Identification des termes susceptibles d'être extraits
- Identification des classes d'équivalence et intégration des synonymes
- Affectation d'un type
- Indexation
- Mise en correspondance de l'extraction des patrons et des événements

Etape 1. Conversion des données source en un format standard

Au cours de cette première étape, les données que vous importez sont converties dans un format uniforme pouvant être utilisé pour effectuer d'autres analyses. Cette conversion, qui s'effectue en interne, ne modifie pas les données d'origine.

Etape 2. Identification des termes susceptibles d'être extraits

Il est important de comprendre le rôle des ressources linguistiques dans l'identification des termes susceptibles d'être extraits lors de l'extraction linguistique. Les ressources linguistiques sont utilisées lors de chaque exécution d'une extraction. Elles se présentent sous la forme de ressources compilées, de bibliothèques et de modèles. Les bibliothèques comportent des listes de mots, des relations et des informations complémentaires qui permettent de spécifier ou d'affiner l'extraction. Vous ne pouvez pas afficher ni éditer les ressources compilées. Toutefois, les autres ressources (modèles) peuvent être modifiées dans l'Editeur de ressources.

Les ressources compilées sont des composants internes essentiels du moteur du programme d'extraction de IBM® SPSS® Text Analytics for Surveys. Ces ressources comportent un dictionnaire général qui liste les formes de base avec un code concernant la catégorie grammaticale (nom, verbe, adjectif, adverbe, participe, élément de coordination, déterminant ou préposition).

Les ressources comprennent également des types intégrés et réservés qui permettent d'affecter de nombreux termes extraits aux types suivants : <Location>, <Organization>, ou <Person>. [Pour plus d'informations, reportez-vous à la section Types intégrés dans le chapitre 10 sur p. 221.](#)

Outre ces ressources compilées, plusieurs bibliothèques sont fournies avec le produit et peuvent être utilisées pour compléter les types et les définitions de concept figurant dans les ressources compilées, ainsi que pour proposer d'autres types et synonymes. Ces bibliothèques —et toutes les bibliothèques personnalisées que vous créez— comprennent plusieurs dictionnaires : déclarations de types, dictionnaires de substitutions (synonymes et éléments optionnels) et dictionnaires d'exclusions. [Pour plus d'informations, reportez-vous à la section Utilisation des bibliothèques dans le chapitre 9 sur p. 206.](#)

Une fois les données importées et converties, le moteur du programme d'extraction commence à identifier les termes susceptibles d'être extraits. Ces termes sont des mots ou des groupes de mots qui permettent d'identifier des concepts du texte. Lors du traitement du texte, les mots uniques (**unitermes**) qui ne figurent pas dans les ressources compilées sont considérés comme étant des termes susceptibles d'être extraits. Les mots composés (**multitermes**) susceptibles d'être extraits sont identifiés à l'aide d'extracteurs de patrons de catégorie grammaticale. Par exemple, l'expression multiterme `belle voiture`, qui suit le patron de catégorie grammaticale `adjectif nom`, possède deux composants. L'expression multiterme `belle petite voiture`, qui suit le patron de catégorie grammaticale `"adjectif adjectif nom"`, possède trois composants.

Remarque : les termes qui figurent dans le dictionnaire général compilé susmentionné constituent la liste de tous les mots qui risquent de s'avérer inintéressants ou de présenter une ambiguïté linguistique en tant qu'unitermes. Ces mots sont exclus de l'extraction lorsque vous identifiez les unitermes. Ils font toutefois l'objet d'une réévaluation lorsque vous déterminez les catégories grammaticales ou que vous recherchez des mots composés (expressions multitermes) plus longs, susceptibles d'être extraits.

Enfin, un algorithme spécial est appliqué pour traiter les chaînes en majuscules (intitulés de postes, par exemple), de telle sorte que ces patrons puissent être extraits.

Etape 3. Identification des classes d'équivalence et intégration des synonymes

Une fois les expressions multitermes et les unitermes susceptibles d'être extraits identifiés, le logiciel utilise un ensemble d'algorithmes pour les comparer et identifier des classes d'équivalence. Une classe d'équivalence désigne la forme de base d'une phrase ou la forme unique de deux variantes d'une même phrase. L'affectation de phrases à des classes d'équivalence a pour objectif de veiller à ce que, par exemple, `président de l'entreprise` et `président d'entreprise` ne soient pas traités comme des concepts distincts. Pour déterminer le concept à utiliser pour la classe d'équivalence — (à savoir, si `président de l'entreprise` ou `président d'entreprise` est utilisé en tant que terme principal), le moteur du programme d'extraction applique les règles suivantes dans l'ordre indiqué ci-dessous :

- Forme définie par l'utilisateur dans une bibliothèque.
- Forme la plus fréquente dans l'ensemble du corps du texte.
- Forme la plus courte dans l'ensemble du corps du texte (ce qui correspond généralement à la forme de base).

Etape 4. Affectation d'un type

Des types sont ensuite affectés aux concepts extraits. Un type correspond à un regroupement sémantique de concepts. Les ressources compilées et les bibliothèques sont utilisées au cours de cette étape. Les types comprennent des éléments tels que des concepts de niveau supérieur, des mots positifs et négatifs, des prénoms, des lieux, des organisations, etc. Vous pouvez définir d'autres types. [Pour plus d'informations, reportez-vous à la section Déclarations de types dans le chapitre 10 sur p. 220.](#)

Etape 5. Indexation

L'indexation de l'ensemble des enregistrements s'effectue en définissant un pointeur entre une position de texte et le terme représentatif de chaque classe d'équivalence. Cela suppose que toutes les instances de forme fléchies d'un concept susceptible d'être extrait sont indexées en tant que forme de base susceptible d'être extraite. La fréquence globale est calculée pour chaque forme de base.

Etape 6. Mise en correspondance de l'extraction des patrons et des événements.

IBM SPSS Text Analytics for Surveys peut non seulement détecter les types et les concepts, mais également les relations qui existent entre eux. Plusieurs algorithmes et bibliothèques sont fournis avec ce produit ; ils permettent d'extraire les patrons de relations entre les types et les concepts. Ils s'avèrent particulièrement utiles lorsque vous tentez de détecter des opinions spécifiques (relations entre des produits, par exemple) ou les liens relationnels entre des personnes ou des objets (liens entre des groupes politiques ou des génomes, par exemple).

Fonctionnement de la catégorisation

Vous disposez de plusieurs techniques de création de catégories. Etant donné que chaque ensemble de données est unique, le nombre de techniques et leur ordre d'application peuvent varier. Votre interprétation des résultats pouvant être différente de celle d'une autre personne, vous pouvez être amené à essayer plusieurs techniques de manière à déterminer celle qui donne les meilleurs résultats pour vos données textuelles.

Dans ce manuel, la **création de catégories** fait référence à la génération de définitions et de classification de catégories à l'aide d'une ou de plusieurs techniques intégrées et la **catégorisation** fait référence au scoring, ou à l'étiquetage, processus par lequel des identificateurs uniques (nom/ID/valeur) sont affectés aux définitions de catégorie de chaque enregistrement.

Pendant la création de catégories, les concepts et les types qui ont été extraits sont utilisés en tant que blocs de construction de vos catégories. Lorsque vous créez des catégories, les enregistrements sont automatiquement affectés aux catégories s'ils contiennent du texte qui correspond à un élément d'une définition de catégorie.

IBM® SPSS® Text Analytics for Surveys vous propose plusieurs techniques automatisées de classification supervisée qui vous permettent de catégoriser rapidement vos enregistrements.

Techniques de regroupement

Chaque technique disponible convient à certains types de données et de situation. Cependant, il est souvent judicieux de combiner des techniques dans la même analyse afin de capturer le maximum d'enregistrements. Aussi est-il possible de voir un concept figurer dans plusieurs catégories ou de rencontrer des catégories redondantes.

Dérivation des racines de concept. Cette technique crée des catégories en sélectionnant un concept et en recherchant les autres concepts y étant associés. Pour cela, elle analyse si certains composants du concept sont liés d'un point de vue morphologique ou s'ils partagent des racines. Cette technique est très utile pour l'identification de concepts de mots composés synonymes, car les concepts de chaque catégorie générée sont synonymes ou leur signification est proche. Elle utilise des données de différentes longueurs et génère un nombre inférieur de catégories compactes. Par exemple, le concept `opportunités d'avancer` serait regroupé avec les concepts `opportunité d'avancement` et `opportunité d'un avancement`. [Pour plus d'informations, reportez-vous à la section Dérivation des racines de concept dans le chapitre 6 sur p. 122.](#)

Réseau sémantique. Cette technique commence en identifiant les sens possibles de chaque concept à partir de son index complet de relations existant entre les mots puis crée des catégories en regroupant les concepts associés. Cette technique est plus performante lorsque les concepts sont connus dans le réseau sémantique et qu'ils ne sont pas trop ambigus. Son efficacité est cependant amoindrie lorsque le texte contient des termes spécialisés dont le réseau n'a pas connaissance. Par exemple, le concept `pomme granny smith` pourrait être regroupé avec `pomme gala` et `pomme golden` car il s'agit de soeurs de la `granny smith`. Pour donner un autre exemple, le concept `animal` pourrait être regroupé avec `chat` et `kangourou` car il s'agit d'hyponymes d'`animal`. Cette technique est uniquement disponible pour les textes en anglais dans cette version. [Pour plus d'informations, reportez-vous à la section Réseaux sémantiques dans le chapitre 6 sur p. 124.](#)

Inclusion de concept. Cette technique crée des catégories en regroupant les concepts multitermes (mots composés) selon qu'ils contiennent ou non des mots qui sont des sous-ensembles ou des super-ensembles d'un mot dans l'autre. Par exemple, le concept `siège` serait regroupé avec `siège de sécurité`, `siège couchette` et `commande de siège éjectable`. [Pour plus d'informations, reportez-vous à la section Inclusion de concepts dans le chapitre 6 sur p. 123.](#)

Cooccurrence. Cette technique crée des catégories à partir des cooccurrences trouvées dans le texte. Ainsi, lorsque des concepts ou des patrons de concept apparaissent souvent ensemble dans des documents et des enregistrements, la cooccurrence reflète une relation sous-jacente qui a vraisemblablement de l'intérêt dans vos définitions de catégorie. Lorsque des mots font l'objet d'une cooccurrence de manière significative, une règle de cooccurrence est créée et peut être utilisée comme descripteur de catégorie pour une nouvelle sous-catégorie. Par exemple, si de nombreux enregistrements contiennent les mots `prix` et `disponibilité`, (mais que peu d'enregistrements contiennent l'un sans l'autre) alors ces concepts peuvent être regroupés dans une règle de cooccurrence, (`prix & disponible`) et ils peuvent être affectés à une sous-catégorie de la catégorie `prix` par exemple. [Pour plus d'informations, reportez-vous à la section Règles de cooccurrence dans le chapitre 6 sur p. 126.](#)

- **Nombre minimum d'enregistrements.** Pour aider à déterminer l'intérêt des cooccurrences, déterminez le nombre minimum d'enregistrements devant contenir une cooccurrence donnée pour être utilisés en tant que descripteurs dans une catégorie.

Préparation à l'analyse de texte

L'analyse de texte implique plus que l'extraction et la catégorisation des données. Pour réussir l'analyse de texte, considérez les points suivants :

- Comme dans la conception des enquêtes, la qualité des réponses que vous importez dans IBM® SPSS® Text Analytics for Surveys affecte directement la qualité des catégories qui en résultent. En général, les questions vagues ou peu claires entraînent des réponses qui peuvent dériver et s'écarter du sujet, et être assez difficiles à analyser.
- De même que l'analyse statistique, l'analyse de texte doit être effectuée avec des objectifs clairs à l'esprit. Avant de commencer une analyse, vous devez réfléchir sur votre étude et déterminer ce que vous essayez d'apprendre.

Par exemple, supposons qu'une enquête a été menée dans une école de quartier pour mesurer l'attitude des parents concernant la qualité de l'éducation reçue par leurs enfants. Au cours de l'analyse, nous pouvons nous concentrer sur des sujets tels que le nom des professeurs, les programmes scolaires, etc., ou alors nous concentrer sur l'identification et le regroupement des commentaires positifs et négatifs. De la même manière, le niveau de granularité requis pour l'analyse doit être défini. Nous pouvons choisir par exemple de regrouper toutes les remarques sur le financement, ou alors de diviser cette catégorie en financement par programme. Les codes ou les catégories que nous créons doivent refléter la priorité et les objectifs de notre analyse.

- Plus encore que l'analyse statistique, l'analyse de texte ne constitue pas une science exacte, car il n'existe pas de résultat "correct" unique. L'analyse de texte est exécutée avec des objectifs à l'esprit, mais elle est également subjective du fait qu'elle est influencée par l'interprétation par l'analyste du message transmis par la personne interrogée. L'analyste décide par exemple comment identifier et classer des attitudes pleines de sarcasme. Selon leurs objectifs et leur centre d'intérêt, deux personnes compétentes peuvent analyser les mêmes données et parvenir à des conclusions différentes.
- L'analyse de texte est en grande partie un *processus itératif*. Lorsque vous travaillerez avec des réponses d'enquête, vous serez probablement amené à réextraire et recatégoriser vos réponses à l'aide de définitions de catégorie différentes (c'est-à-dire des systèmes de codage), de définitions de concept ou de synonyme différentes, et de regroupements de réponses différentes.

Une fois que vous avez extrait les concepts du texte et créé vos catégories, vous devez examiner vos résultats attentivement. Si vous trouvez des éléments que vous souhaitez modifier, réglez simplement votre analyse pour affiner vos définitions de catégorie et définitions de bibliothèque linguistique. Les réponses seront ensuite automatiquement recatégorisées lorsque vous procéderez de nouveau à l'extraction. Vous devrez peut-être effectuer cette procédure une ou plusieurs fois avant d'être satisfait des résultats de votre analyse.

Remarque : Pour plus d'informations sur les considérations à prendre en compte avant d'importer les données, consultez la rubrique [Préparation de vos données sur p. 28](#)

Fiabilité et affinement

Lorsque vous codez des données, vous souhaitez que les catégories qui en résultent soient fiables. Dans le contexte d'un codage de réponses ouvertes, cela signifie que deux codeurs indépendants, utilisant les mêmes règles (système de codage), doivent coder les mêmes réponses de manière identique. Lorsque l'analyse de texte est effectuée manuellement, ce problème est critique. Un ensemble de catégories de qualité peut être créé, mais s'il ne peut pas être appliqué de manière fiable aux réponses, sa valeur diminue considérablement. Lorsque IBM® SPSS® Text Analytics for Surveys est appliqué aux mêmes données, avec les mêmes ressources linguistiques, il reproduit toujours parfaitement une analyse antérieure. Il est fiable à 100 %.

Cela ne signifie pas pour autant que l'analyse ne contiendra aucune erreur. Mais le codage peut maintenant se concentrer sur autre chose, l'—*affinement*. Dans le codage humain, les codeurs lisent la réponse et peuvent saisir toutes les nuances d'une déclaration (même s'ils ont des difficultés à appliquer les catégories de codage). SPSS Text Analytics for Surveys peut appliquer les catégories de codage, mais les catégories doivent être définies de manière à saisir les nuances et les distinctions. Il existe deux manières de procéder à l'affinement :

- Affinement des ressources linguistiques
- Affinement des définitions de catégorie

Affinement des ressources linguistiques

IBM® SPSS® Text Analytics for Surveys peut facilement créer des catégories sans intervention de votre part, mais celles-ci ne saisissent invariablement pas toutes les informations figurant dans les réponses. Vous devez améliorer la base linguistique sur laquelle le programme s'appuie, de telle sorte que la création de catégories soit encore plus adaptée aux idiosyncrasies du texte. Pour améliorer cette base, vous pouvez personnaliser et affiner les ressources linguistiques utilisées lors de l'extraction effectuée à partir du texte.

L'affinement, dans ce cas, implique l'ajout de mots et de phrases à différentes bibliothèques linguistiques et dictionnaires, en spécifiant des mots à exclure de l'analyse, en définissant des synonymes, ou en créant des bibliothèques personnalisées avec un objectif spécifique à l'esprit. Cet objectif est de saisir avec précision les idées des personnes interrogées, contenues dans le texte, et de supprimer l'ambiguïté des résultats.

Affinement des définitions de catégorie

Outre le réglage des ressources linguistiques, vous devez procéder à l'examen de vos catégories en recherchant des moyens de combiner ou de nettoyer leurs définitions. Vérifiez également certaines réponses catégorisées. Vous pouvez utiliser les techniques de création de catégories automatisées pour créer vos catégories. En revanche, vous serez probablement amené à apporter quelques modifications à ces définitions. Lorsque vous utilisez une technique, un certain nombre de nouvelles catégories apparaissent dans la fenêtre. Vous pouvez développer les catégories pour voir les concepts qui définissent chacune d'elle. Vous pouvez alors examiner les réponses d'une catégorie et les ajuster jusqu'à ce que vos définitions de catégorie vous conviennent.

Aucune des techniques automatiques ne pourra parfaitement catégoriser vos données; c'est pourquoi nous recommandons de trouver et d'appliquer une ou plusieurs techniques automatiques qui correspondent aux besoins de vos données. Vous pouvez utiliser des techniques manuelles pour effectuer des changements mineurs, supprimer toute mauvaise réaffectation, ou ajouter les enregistrements ou concepts ayant été oubliés.

Pour commencer

Cette documentation présente les tâches que vous pouvez effectuer avec IBM® SPSS® Text Analytics for Surveys et les techniques que vous pouvez employer pour catégoriser les réponses. Les informations présentées ici vous guideront dans la réalisation de votre analyse initiale. Elles abordent tous les processus permettant d'analyser complètement vos données, mais chaque ensemble de données étant différent, vous devrez en fin de compte décider à quel moment votre analyse est terminée.

Ce chapitre présente le processus habituel que suivent les utilisateurs lorsqu'ils effectuent une analyse de texte. L'interface est également présentée d'une manière globale, ainsi que les principaux éléments et tâches utilisés.

Processus habituel

Le processus de workflow habituel que vous allez suivre tout en utilisant IBM® SPSS® Text Analytics for Surveys est récapitulé ci-dessous.

- **Créez un projet** en important les données d'enquête (réponses ouvertes, variable d'ID et autres variables de référence) dans SPSS Text Analytics for Surveys. Vous pouvez lire les données à partir des fichiers de données IBM® SPSS® Statistics, dans Microsoft Excel, dans tout programme de base de données conforme à ODBC ou à partir d'une source de données IBM® SPSS® Data Collection. Vous pouvez choisir un package d'analyse de texte pour bénéficier de certaines catégories prédéfinies et de ressources spécialisées pour être opérationnel rapidement.
- **Extrayez** les concepts et patrons de chacune des questions ouvertes importées. Le moteur du programme d'extraction interne identifie et collecte automatiquement les termes clés exprimés dans le texte. Ces termes sont regroupés sous un concept principal. Les concepts sont ensuite regroupés en types, qui sont des ensembles de termes similaires (organisations, produits ou opinions positives, par exemple). Des patrons sont également extraits. Ils correspondent à des combinaisons de termes et de types qui représentent des opinions et des relations (commentaires positifs concernant une organisation, par exemple).
- **Affinez** les concepts extraits et réglez les extractions en manipulant directement une ou plusieurs bibliothèques qui contiennent des types de mot, des termes, des synonymes, des exclusions et d'autres constructions linguistiques. Comme indiqué précédemment, l'analyse de texte est un processus itératif au cours duquel le réglage des bibliothèques et des dictionnaires génère directement des résultats adaptés à vos données.

- **Catégorisez** vos réponses en créant et modifiant manuellement des catégories à l'aide de règles de catégories, de plans de codage et/ou de techniques de création de catégories supervisées. Les catégories représentent des concepts de niveau supérieur qui saisissent les principales idées et attitudes exprimées par les personnes interrogées.
- **Exportez** les catégories et la variable d'ID dans des formats de fichier courants en vue d'effectuer d'autres analyses et de les représenter graphiquement dans d'autres applications. Il peut en résulter un ensemble de variables à réponses multiples, sous la forme d'un fichier SPSS Statistics ou d'un fichier Microsoft Excel.

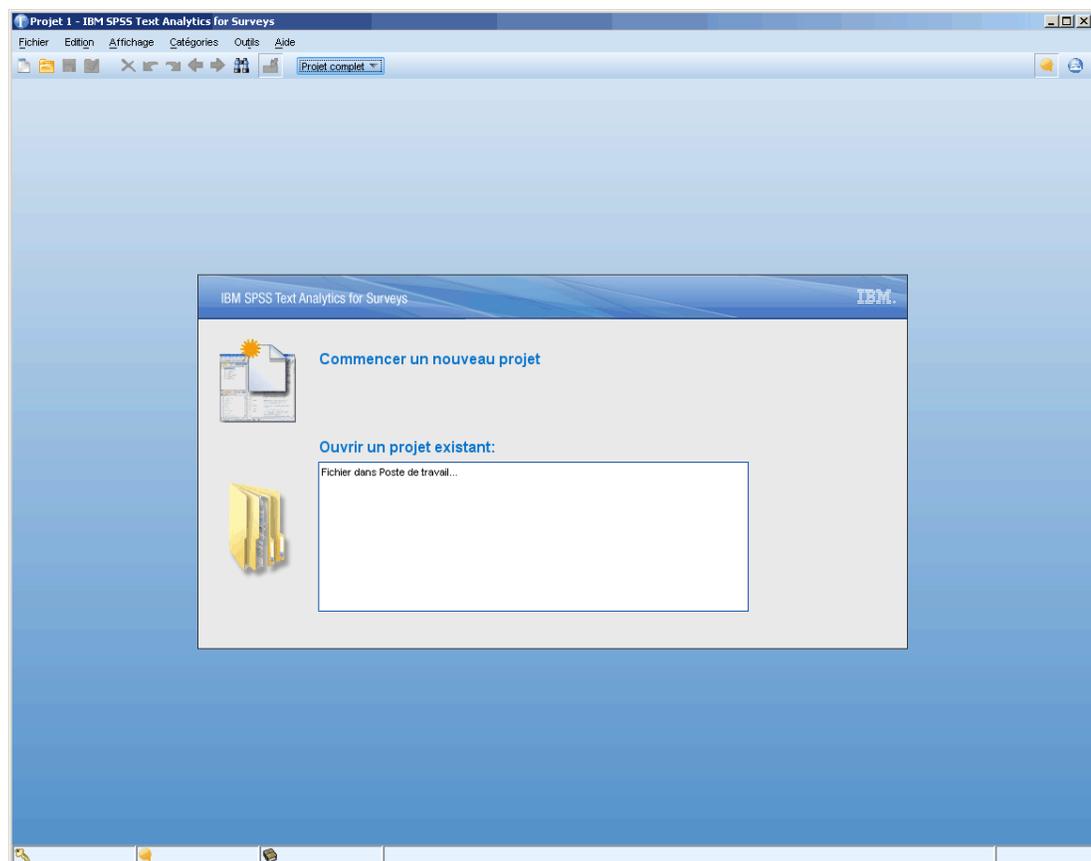
La fenêtre d'analyse de texte

L'interface de l'application comporte deux fenêtres. La première est la fenêtre d'analyse de texte, dans laquelle vous effectuez la plus grande partie du travail. Dans cette fenêtre, vous pouvez analyser chaque question dans vos données. Pour chaque question, vous pouvez extraire des concepts, types, patrons, puis catégoriser vos réponses.

Lorsque vous démarrez l'application, un écran s'affiche dans lequel vous pouvez ouvrir un projet existant ou en créer un nouveau. Si vous choisissez de créer un nouveau projet, un assistant s'ouvre pour vous guider dans le processus de création du projet. [Pour plus d'informations, reportez-vous à la section Création de projets dans le chapitre 3 sur p. 27.](#)

Figure 2-1

Fenêtre d'analyse de texte au lancement du produit



Une fois que vous avez importé les données, vous pouvez afficher : les vues Question ou la vue Projet complet. Vous pouvez changer de vue en en sélectionnant une dans la liste déroulante sur la barre d'outils de la fenêtre d'analyse de texte ou en sélectionnant la vue dans le menu Affichage. Le texte qui apparaît dans la liste provient de l'étiquette de variable pour chaque question.

Figure 2-2

Barre d'outils de l'application avec liste déroulante de sélection de vues



La vue Question

La vue Question vous offre un espace dans lequel vous pouvez analyser et catégoriser les réponses à une question particulière. Après la création d'un nouveau projet, la vue Question apparaît avec les réponses pour la première variable de texte que vous avez importée. Vous pouvez sélectionner cette vue dans la liste déroulante ou choisir un nom de question dans le menu Affichage > Question >.

Figure 2-3

Vue Question

The screenshot displays the 'Vue Question' interface. It features a menu bar (Fichier, Edition, Affichage, Catégories, Outils, Aide) and a toolbar. The main window is divided into several panes:

- Left Pane:** A tree view showing categories and their counts. The 'memory' category is highlighted.
- Bottom-Left Pane:** A list of concepts extracted from the data, such as 'product', 'lightweight', 'long-lasting', etc.
- Top-Right Pane:** A table showing the distribution of categories. The 'Barre Catégorie' tab is active.
- Bottom-Right Pane:** A table of responses with columns for 'Id', 'Réponse', and 'Catégories'. Response 133 is highlighted.

Barre	Sélection %	Personnes interrogées	Total %
size	46,4	13	0,0
load+<Positive>	3,6	1	0,5
sound	3,6	1	0,0
interface+<Positive>	3,6	1	1,0
music	7,1	2	0,0
cds	14,3	4	0,0
look+<Positive>	3,6	1	0,7
design+<Positive>	3,6	1	1,7
capacity	39,3	11	0,0
listening	3,6	1	0,0

Id	Réponse	Catégories	
1	4	Having all my CDs in the palm of my hand!	cds/cds+<Positive>
2	8	portability, capacity, sound quality, durability	capacity/capacity+<Positive>
3	52	i have a Product A. I like the small size and good sound.	size/size+<Positive> sound/sound+<Positive>
4	63	it keeps playing. i don't have to stop and change a CD.	cds/cds+<Contextual>
5	68	Large storage capacity	..storage capacity+<Contextual>
6	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.	load+<Positive> music/music+<Positive> size/size+<Contextual>
7	98	Big storage capacity, also does video	..storage capacity+<Contextual>
8	114	Large storage capacity and a good LCD screen for viewing digital photos	..storage capacity+<Contextual>
9	118	I can listen to where ever I am, I don't have to worry about which Cd to take etc have all my music on it so listen to what I like when I like. Also I like the fact that I can plug it into speakers.	cds/cds+<Positive> listening/listenings+<Positive> music/music+<Positive>
10	133	I like the small size of the device. It allows me to keep it in a pocket while I use it so that I don't have to hold it like I would a larger device	size/size+<Positive>

Les opérations que vous effectuez dans la vue Question concernent trois éléments : résultats d'extraction, catégories et données de réponse. Afin de vous aider à travailler et analyser chacun des éléments de manière indépendante, cette fenêtre est divisée en quatre panneaux :

Panneau Catégories

Ce panneau, situé dans l'angle supérieur gauche, représente une zone dans laquelle vous pouvez créer et gérer toutes vos catégories. Une fois les concepts, types et patrons extraits de vos données textuelles, vous pouvez commencer à créer des catégories en utilisant des techniques automatiques (réseaux sémantiques et inclusion de concept, par exemple) ou en procédant de façon manuelle. Vous pouvez cliquer sur une catégorie pour la développer et voir tous les descripteurs qui composent la définition de cette catégorie (concepts, types, règles). Lorsque vous sélectionnez une catégorie ou un descripteur, vous pouvez ensuite afficher les informations concernant les enregistrements correspondants dans les panneaux Données et Visualisation. [Pour plus d'informations, reportez-vous à la section Le panneau Catégories dans le chapitre 6 sur p. 97.](#)

Figure 2-4
Panneau Catégories : Définition de catégorie développée

Catégorie	Nombre ...	Nombre ...
function	1	3
headphones	2	4
home	3	3
listening	3	27
memory device	13	29
tape		2
memory	8	22
add-on memory		1
amount of memory		1
amount of storage		1
memory card		3
memory space		1
storage capacity	2	9
memory storage		6
recording	4	6
record		1

Panneau Résultats extraits

Cette zone, située dans l'angle inférieur gauche, présente les résultats de l'extraction. Lorsque vous exécutez une extraction, le moteur du programme d'extraction parcourt les données textuelles, identifie les concepts pertinents et affecte un type à chaque concept. Les **concepts** correspondent à des mots ou groupes de mots extraits à partir des données textuelles. Les **types** représentent des regroupements sémantiques de concepts stockés sous la forme de déclarations de types. Une fois l'extraction terminée, les concepts, les types et les patrons apparaissent dans le panneau Résultats d'extraction. Les concepts et les types font l'objet d'un codage de couleur afin que vous puissiez identifier le type auquel ils appartiennent. Vous pouvez utiliser les concepts, types et patrons que vous collectez ici pour créer vos catégories. [Pour plus d'informations, reportez-vous à la section Résultats extraits : Concepts, types et patrons dans le chapitre 5 sur p. 80.](#)

Le Text Mining est un processus itératif au cours duquel les résultats de l'extraction sont passés en revue en fonction du contexte des données textuelles. Ils sont ensuite affinés afin de générer de nouveaux résultats avant d'être réévalués. Vous pouvez affiner les résultats de l'extraction en modifiant les ressources linguistiques. Vous pouvez procéder en partie à ce réglage à partir du panneau Résultats extraits ou Données directement, ou bien directement dans la vue Editeur de ressources. [Pour plus d'informations, reportez-vous à la section La fenêtre Editeur de ressources sur p. 16.](#)

Panneau Données

Situé dans l'angle inférieur droit de cette vue, il présente, sous forme de tableau, les données des réponses à la question ouverte sélectionnée. Par défaut, le panneau Données affiche trois colonnes (ID d'enregistrement, réponses textuelles, et catégories affectées). Les réponses qui apparaissent dans ce panneau sont filtrées en fonction de ce que vous avez sélectionné dans l'autre panneau. Bien que vous puissiez afficher les données que vous avez importées dans ce panneau, vous ne pouvez pas modifier, supprimer ou annexer des données aux enregistrements. [Pour plus d'informations, reportez-vous à la section Le panneau Données dans le chapitre 6 sur p. 100.](#)

Panneau Visualisation

Situé dans l'angle supérieur droit de cette vue, il est masqué par défaut. Vous pouvez afficher ce panneau (choisissez Affichage > Visualisation). Ce panneau offre trois vues uniques de la manière dont les réponses correspondent aux catégories ou de la manière dont elles partagent des réponses (graphique Web, diagramme à barres, et tableau) en fonction des sélections que vous faites dans les autres panneaux. [Pour plus d'informations, reportez-vous à la section Visualisation des graphiques dans le chapitre 7 sur p. 170.](#)

Selon que vous choisissez ou non l'option d'extraction dans l'assistant de Nouveau projet, des résultats d'extraction s'affichent dans le panneau inférieur gauche. Cliquez sur Extraire dans le panneau Résultats d'extraction pour commencer l'extraction. Après l'extraction, vous pouvez examiner les résultats pour voir si un affinage est nécessaire, tel que regrouper des synonymes sous un nom de concept ou exclure de la liste des concepts courants, sans intérêt.

Une fois que vous êtes satisfait des résultats de l'extraction, vous pouvez commencer à catégoriser vos réponses manuellement en faisant glisser et en déposant les concepts en tant que catégories ou en utilisant des techniques de création de catégories, telles que l'inclusion de concepts et le réseau sémantique.

La vue Projet complet

La vue Projet complet offre un aperçu de l'ensemble des variables que vous avez importées dans le projet. Vous pouvez sélectionner cette vue dans la liste déroulante ou à partir du menu Affichage (Affichage > Projet complet). Dans cette vue, vous pouvez examiner les données que vous avez importées, modifier le rôle d'une variable (passer par exemple de question à variable de référence), et affecter des étiquettes aux variables. [Pour plus d'informations, reportez-vous à la section Afficher les données du projet dans le chapitre 4 sur p. 51.](#) Bien que vous puissiez afficher les données que vous avez importées dans cette vue, vous ne pouvez pas modifier, corriger, supprimer ou annexer des données aux enregistrements.

Remarque : Pour afficher tout le contenu d'une cellule dans cette vue, vous pouvez positionner la souris sur la cellule. Une info-bulle affiche le contenu de la cellule.

Figure 2-5
Vue Projet complet

ID	Points torts de votre dernier séjour à l'hôtel	Points faibles de votre dernier séjour à l'hôtel	Tranche d'âge	Sexe	Raison de votre séjour
1	Excellentes prestations générales . J'ai particulière...	Le petit déjeuner n'était pas à la hauteur du tarif pr...	(30 à 40 ans)	(Homme)	(Raison professionnelle)
2	qualité du service, accueil très chaleureux, discrétio...	chambres exigues. pas de service de blanchisserie	(30 à 40 ans)	(Femme)	(Raison personnelle)
3	propreté des lieux, accueil chaleureux, buffet extr...	trop de bruit. pas de chambre fumeur	(30 à 40 ans)	(Femme)	(Raison personnelle)
4	PROXIMITE	ACCUEIL	(30 à 40 ans)	(Homme)	(Raison professionnelle)
5	l'accueil	les chambres	(30 à 40 ans)	(Femme)	(Raison personnelle)
6	vue sur mer	bruyant	(30 à 40 ans)	(Homme)	(Raison personnelle)
7	climatisation agreable, belle vue	bruit	(30 à 40 ans)	(Femme)	(Raison professionnelle)
8	Chambre agreable, restaurant raffiné et service très...	Prix excessif	(40 à 50 ans)	(Homme)	(Raison professionnelle)
9	les chambres bien équipées	l'accueil	(30 à 40 ans)	(Femme)	(Raison professionnelle)
10	la piscine	l'accueil	(20 à 30 ans)	(Femme)	(Raison personnelle)
11	femmes de chambre ravissantes. service ok sans +	piscine trop petite. buffets assez moyens.	(30 à 40 ans)	(Homme)	(Raison professionnelle)
12	Une grande qualité d'accueil	Absence de mini bar	(30 à 40 ans)	(Homme)	(Raison professionnelle)
13	très bon restaurant, service de chambre de qualité	l'eau de la piscine était froide, pas de garde d'enfant	(30 à 40 ans)	(Homme)	(Raison personnelle)
14	belle chambre, service agreable	clim ne fonctionnait pas	(20 à 30 ans)	(Femme)	(Raison professionnelle)
15	Super piscine, accueil top et chambre très bien ent...	la cuisine très très médiocre	(30 à 40 ans)	(Homme)	(Raison professionnelle)
16	La nourriture, la piscine.	La chambre donné sur une cour crasseuse	(40 à 50 ans)	(Homme)	(Raison personnelle)
17	La baie vitrée sur les gratte-ciel de Mahattan	Rien de spécial	(40 à 50 ans)	(Homme)	(Raison professionnelle)
18	C'était tranquille. En pleine cambrousse.	La femme de ménage est entrée sans frapper à 10...	(40 à 50 ans)	(Homme)	(Raison personnelle)
19	Le petit-déjeuner. On était seuls dans la pièce (san...	Pas très clean quand même. Il restait des outils sur...	(40 à 50 ans)	(Homme)	(Raison personnelle)
20	service, environnement, piscine. j'ai surtout appréci...	pas de service de blanchisserie, horaires de petit-...	(40 à 50 ans)	(Homme)	(Raison professionnelle)
21	la piscine, la grande chambre, les hoteliers	pas de restaurant dans l'hotel	(20 à 30 ans)	(Femme)	(Raison personnelle)
22	Les chambres de l'hôtel sont spacieuses. J'ai bien ...	Les petits déjeuners ne sont pas à la hauteur du pr...	(30 à 40 ans)	(Homme)	(Raison personnelle)
23	proche du lieu de travail	nuisances sonores externes (bruit de voitures)	(30 à 40 ans)	(Homme)	(Raison professionnelle)
24	possibilité d'accéder à internet en wifi. La chambre...	Les toilettes étaient sales. Le prix: était exhubérant.	(30 à 40 ans)	(Homme)	(Raison professionnelle)
25	L'accueil et la tranquillité du lieu	aucun	(30 à 40 ans)	(Femme)	(Raison personnelle)
26	la nourriture était très bonne et on pouvait se servi...	Le personnel parlait mal le français	(50 à 60 ans)	(Homme)	(Raison personnelle)
27	Un lit bien confortable et un petit dej avec pleins de...	Rien du tout !	(30 à 40 ans)	(Femme)	(Raison personnelle)
28	Pas grand chose. Le prix peut-etre mais c'est pas ...	Ca pu la clope dans la chambre et le matelas a bien...	(30 à 40 ans)	(Femme)	(Raison professionnelle)
29	Le prix et la proximité à l'airport.	On a perdu ma reservation. J'ai du attendre une de...	(40 à 50 ans)	(Homme)	(Raison personnelle)
30	calme et bonne nourriture. confort et beau décor a...	loin de la ville et peu de transport en commun	(30 à 40 ans)	(Femme)	(Raison personnelle)
31	La qualité du restaurant. Le bon accueil	Nuisances sonores	(30 à 40 ans)	(Femme)	(Raison professionnelle)

La fenêtre Editeur de ressources

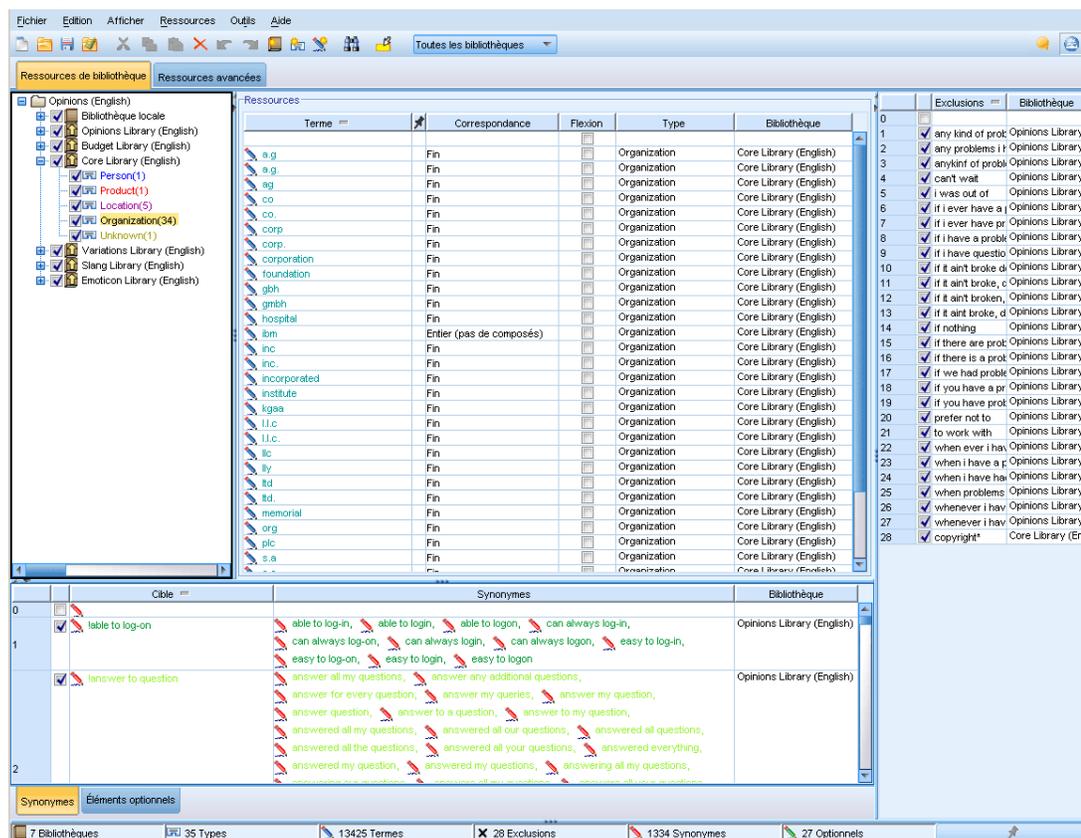
IBM® SPSS® Text Analytics for Surveys utilise un moteur d'extraction fiable pour saisir avec rapidité et précision les principaux concepts des données textuelles. Ce moteur s'appuie essentiellement sur les ressources linguistiques pour déterminer la quantité de données textuelles non structurées à analyser et à interpréter.

Dans la vue de l'Editeur de ressources, vous pouvez afficher et affiner les ressources linguistiques utilisées pour extraire les concepts, les regrouper en types, découvrir des patrons dans les données textuelles, et bien plus encore. IBM® SPSS® Text Analytics for Surveys propose plusieurs modèles de ressources préconfigurés. Pour certaines langues, vous pouvez également utiliser les ressources dans un package d'analyse de texte. [Pour plus d'informations, reportez-vous à la section Utilisation des packages d'analyse de texte dans le chapitre 3 sur p. 43.](#)

Comme ces ressources risquent de ne pas toujours être parfaitement adaptées au contexte de vos données, vous pouvez créer, modifier et gérer vos propres ressources pour un contexte ou un domaine particulier dans l'Editeur de ressources. [Pour plus d'informations, reportez-vous à la section Utilisation des bibliothèques dans le chapitre 9 sur p. 206.](#) Certains utilisateurs n'utilisent que rarement cette fenêtre car les ressources proposées avec le produit sont suffisantes. De plus, la plupart des tâches simples de bibliothèque peuvent être effectuées directement à partir du panneau Résultats d'extraction dans la fenêtre d'analyse de texte.

Pour simplifier le processus d'adaptation de vos ressources linguistiques, vous pouvez effectuer les tâches courantes relatives aux dictionnaires directement à partir de la vue Text Analytics à l'aide des menus contextuels des panneaux Résultats d'extraction et Données. [Pour plus d'informations, reportez-vous à la section Affinage des résultats de l'extraction dans le chapitre 5 sur p. 88.](#)

Figure 2-6
Vue Editeur de ressources



Les opérations que vous effectuez dans la vue Editeur de ressources concernent la gestion et l'adaptation des ressources linguistiques. Ces ressources sont stockées sous la forme de modèles et de bibliothèques. La vue de l'Editeur de ressources comporte quatre parties : le panneau Arborescence des bibliothèques, le panneau Déclaration de types, le panneau Dictionnaire de substitutions et le panneau Dictionnaire d'exclusions.

L'interface comporte quatre parties :

1. Panneau Arborescence des bibliothèques. Situé dans l'angle supérieur gauche, ce plan présente l'arborescence des bibliothèques. Vous pouvez activer et désactiver les bibliothèques figurant dans cette arborescence. Vous pouvez également filtrer les vues des autres panneaux en sélectionnant une bibliothèque dans l'arborescence. Vous pouvez effectuer plusieurs opérations dans cette arborescence à l'aide des menus contextuels. Lorsque vous développez une bibliothèque figurant dans l'arborescence, vous pouvez consulter l'ensemble des types qu'elle comporte. Vous pouvez également filtrer cette liste avec le menu Affichage si vous souhaitez vous concentrer sur une bibliothèque spécifique.

2. Listes des termes du panneau Déclarations de types. Ce panneau, situé à droite de l'arborescence des bibliothèques, affiche les listes de termes des déclarations de types correspondant aux bibliothèques sélectionnées dans l'arborescence des bibliothèques. Une **déclaration de types** est un ensemble de termes devant être regroupés sous une même étiquette ou sous un même nom de type. Lorsque le moteur du programme d'extraction lit vos données textuelles, il compare les mots trouvés dans le texte aux termes figurant dans les déclarations de types. Si un concept extrait figure sous la forme d'un terme dans une déclaration de types, le nom du type en question lui est alors affecté. Vous pouvez considérer la déclaration de types comme étant un dictionnaire distinct qui regroupe les termes présentant des points communs. Par exemple, le type <Location> qui figure dans Core Library comporte des concepts tels que `new orleans`, `great britain` et `new york`. Ces termes désignent tous des lieux géographiques. Une bibliothèque peut comporter une ou plusieurs déclarations de types. [Pour plus d'informations, reportez-vous à la section Déclarations de types dans le chapitre 10 sur p. 220.](#)

3. Panneau Dictionnaire d'exclusions. Ce panneau, situé sur le côté droit, affiche la collection de termes qui seront exclus des résultats finaux de l'extraction. Les termes apparaissant dans ce dictionnaire d'exclusions n'apparaissent pas dans le panneau Résultats d'extraction. Les termes exclus peuvent être stockés dans la bibliothèque de votre choix. Mais, le panneau Dictionnaire d'exclusions affiche tous les termes exclus de toutes les bibliothèques visibles dans l'arborescence des bibliothèques. [Pour plus d'informations, reportez-vous à la section Dictionnaires d'exclusions dans le chapitre 10 sur p. 236.](#)

4. Panneau Dictionnaire de substitutions. Ce panneau, situé en bas à gauche, affiche les synonymes et les éléments optionnels, chacun dans leur propre onglet. Les synonymes et les éléments optionnels permettent de grouper les termes similaires sous un concept principal ou cible dans les résultats d'extraction finaux. Ce dictionnaire peut comporter des synonymes connus, des synonymes et des éléments définis par l'utilisateur, ainsi que les fautes d'orthographe courantes associées à leur correction. Les définitions des synonymes et les éléments optionnels peuvent être stockés dans la bibliothèque de votre choix. Mais, le panneau Dictionnaire de substitutions affiche tous les contenus de toutes les bibliothèques visibles dans l'arborescence des bibliothèques. Pendant que ce panneau affiche tous les synonymes et éléments facultatifs de toutes les bibliothèques, les substitutions de toutes les bibliothèques de l'arborescence apparaissent ensemble dans ce panneau. Une bibliothèque ne peut comporter qu'un seul dictionnaire de substitutions. [Pour plus d'informations, reportez-vous à la section Dictionnaires des substitutions/synonymes dans le chapitre 10 sur p. 231.](#)

Remarque :

- pour procéder à un filtrage visant à n'afficher que les informations propres à une seule bibliothèque, vous pouvez modifier l'affichage de la bibliothèque à l'aide de la liste déroulante figurant dans la barre d'outils. Elle comporte une entrée de niveau supérieur appelée Toutes les bibliothèques, ainsi qu'une autre entrée supplémentaire pour chacune des bibliothèques. [Pour plus d'informations, reportez-vous à la section Affichage des bibliothèques dans le chapitre 9 sur p. 210.](#)

Définition des options

Vous pouvez définir les options générales de IBM® SPSS® Text Analytics for Surveys dans la boîte de dialogue Options. Cette boîte de dialogue comporte les onglets suivants :

L'onglet **Système** contient des options relatives aux listes de bibliothèque par défaut, à l'enregistrement automatique, à l'enregistrement des résultats d'extraction, aux délimiteurs et à la langue de l'interface.

L'onglet **Afficher** contient les options relatives aux couleurs utilisées dans l'interface.

L'onglet **Sons** contient les options relatives aux signaux sonores.

L'onglet **Traduction** contient des options pour les connexions de traduction.

Pour éditer les options

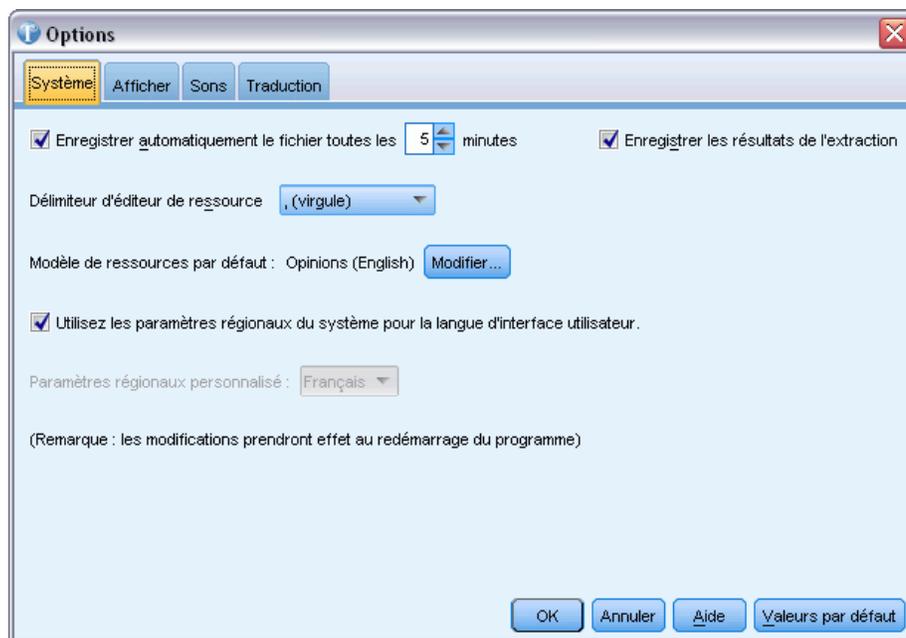
- ▶ A partir des menus, sélectionnez Outils> Options. La boîte de dialogue Options apparaît.
- ▶ Sélectionnez l'onglet contenant les informations à modifier.
- ▶ Modifiez les options.
- ▶ Cliquez sur OK pour enregistrer les modifications.

Options : Onglet Système

Cet onglet permet de définir plusieurs paramètres du projet, notamment :

- L'ajout ou la suppression de bibliothèques qui doivent apparaître dans tous les nouveaux projets par défaut
- L'activation ou la désactivation de la fonction de restauration par enregistrement automatique
- L'activation ou la désactivation de l'enregistrement des résultats de l'extraction
- La définition du délimiteur global qui sera utilisé dans l'Editeur de ressources pour séparer les éléments.

Figure 2-7
Boîte de dialogue Options : Onglet Système



Enregistrer automatiquement le fichier toutes les n minutes. Sélectionnez cette option pour que le produit crée automatiquement une version enregistrée temporaire du fichier de projet ouvert en cas de défaillance de l'ordinateur. Définissez également le nombre de minutes entre chaque sauvegarde. Si vous activez cette fonction et que le produit se ferme de manière inattendue ou que vous rencontrez un problème avec votre ordinateur, la prochaine fois que vous lancerez le produit, vous aurez la possibilité d'ouvrir et de travailler avec une version de restauration de votre fichier.

Enregistrer les résultats de l'extraction. Sélectionnez cette option pour enregistrer les résultats de vos extractions dans votre projet. Cela permet de gagner du temps lorsque vous travaillez sur vos catégories. Toutefois, cela peut allonger la durée de téléchargement, et augmenter la taille de votre projet. A des fins de sécurité, ces résultats d'extraction sont codés durant le processus d'enregistrement et placés dans la base de données. Cette procédure rend difficile pour les utilisateurs, même les utilisateurs avancés, la détection de données dans la base. De plus, les résultats de l'extraction ne sont jamais présentés dans IBM® SPSS® Text Analytics for Surveys tant que la source de données n'a pas été localisée par l'application. Par conséquent, si les données sont protégées par un mot de passe, l'utilisateur doit entrer le nom d'utilisateur et le mot de passe pour cette source de données afin que les résultats de l'extraction apparaissent à l'écran.

- Le principal avantage de l'*enregistrement* est le gain de temps. Etant donné que le processus d'extraction peut prendre un certain temps lorsque vous travaillez avec des ensembles de données volumineux, l'enregistrement fournit un accès immédiat aux résultats à chaque réouverture du projet. Cependant, vous pouvez remarquer un temps d'attente légèrement plus long à l'ouverture d'un projet.
- Le *non enregistrement* est utilisé quand vous ne souhaitez pas que le texte de réponse réside ailleurs que dans le fichier de données d'origine, même si des mesures de sécurité sont appliquées.

Editeur de ressources Séparateur. Sélectionnez le caractère à utiliser en tant que séparateur lors de la saisie d'éléments, tels que des termes, des synonymes et des éléments optionnels dans l'Editeur de ressources.

Modèle de ressources. Si vous n'avez pas sélectionné de package d'analyse de texte, un ensemble de ressources par défaut sera utilisé. Ces ressources sont stockées dans un modèle. Cliquez sur Modifier pour sélectionner un modèle de ressources par défaut différent. Puis, dans la boîte de dialogue Modifier les modèles, sélectionnez la ligne du tableau correspondant au modèle que vous voulez utiliser et cliquez sur OK.

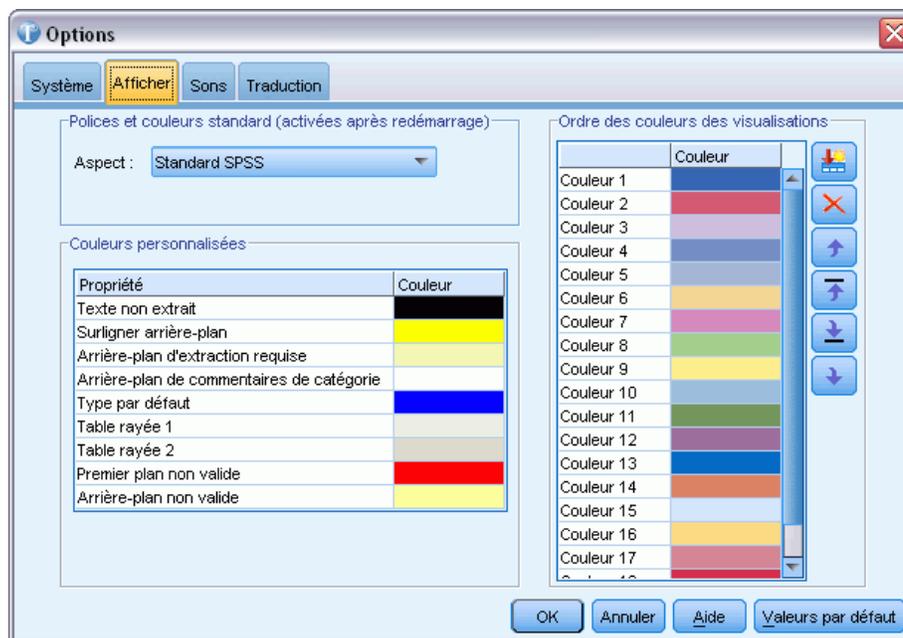
Utiliser la langue du système comme langue de l'interface utilisateur. Sélectionnez cette option pour que SPSS Text Analytics for Surveys utilise les détails de la langue de votre système afin de fournir la langue utilisée sur l'interface. Autrement, vous pouvez désélectionner cette option et choisir une langue d'interface différente. Par exemple, vous pouvez faire cela si vous analysez des informations enregistrées dans une langue différente de celle de votre système et si vous souhaitez exécuter SPSS Text Analytics for Surveys dans cette langue.

Notez que les modifications apportées à cette option ne prendront effet qu'après un redémarrage SPSS Text Analytics for Surveys.

Options : Onglet Affichage

Dans ce deuxième onglet, vous pouvez éditer les options qui affectent la présentation globale de l'application et les couleurs utilisées pour différencier les éléments.

Figure 2-8
Boîte de dialogue Options : Onglet Affichage



Polices et couleurs standard (activées après redémarrage). Les options de cette boîte de commande servent à spécifier le modèle de couleurs et la présentation affichés. Les options sélectionnées ici ne prennent effet qu'après fermeture et redémarrage de l'application. Vous avez le choix entre :

- **SPSS Standard (par défaut)**, un plan courant parmi les produits de marque SPSS (qui fait partie de IBM Corp.).
- **SPSS Classique**, un plan bien connu des utilisateurs de versions antérieures de cette application.
- **Microsoft Windows**, un plan Microsoft Windows pouvant être utile pour un contraste accru dans le canevas de flux et les palettes.

Couleurs personnalisées. Editez les couleurs des éléments apparaissant à l'écran. Vous pouvez modifier la couleur de chacun des éléments du tableau. Pour indiquer une couleur personnalisée, cliquez sur la zone Couleur à droite de l'élément à modifier et choisissez une couleur dans la liste déroulante des couleurs.

- **Texte non extrait.** Texte des réponses, qui n'ont pas encore été extraites, visible dans le panneau Données.
- **Surligner l'arrière-plan.** Couleur d'arrière-plan de sélection de texte utilisée lors de la sélection d'éléments dans les panneaux ou de texte dans le panneau Données.
- **Arrière-plan d'extraction nécessaire.** Couleur d'arrière-plan du panneau Résultats d'extraction indiquant que des modifications ont été apportées aux bibliothèques et qu'une extraction est nécessaire.
- **Arrière-plan des commentaires de catégorie.** Couleur d'arrière-plan des catégories qui apparaît après une opération, telle que faire glisser et déposer des réponses et forcer des réponses du panneau Données à entrer dans le panneau Catégories.
- **Type par défaut.** Couleur par défaut des types et des termes apparaissant dans les panneaux Données et Résultats d'extraction. Cette couleur apparaît dans l'interface à chaque fois que le type `Unknown` ou l'un des concepts associés s'affiche. Cette couleur sera également appliquée à tous les types personnalisés que vous créerez dans l'Editeur de ressources. Vous pouvez remplacer cette couleur par défaut pour les déclarations de types personnalisés en éditant leurs propriétés dans l'Editeur de ressources. [Pour plus d'informations, reportez-vous à la section Création de types dans le chapitre 10 sur p. 222.](#)
- **Table rayée 1.** Première des deux couleurs utilisées en alternance dans le tableau de la boîte de dialogue Editer les termes forcés afin de différencier chaque ligne.
- **Table rayée 2.** Seconde des deux couleurs utilisées en alternance dans le tableau de la boîte de dialogue Editer les termes forcés afin de différencier chaque ligne.
- **Premier plan non valide.** Couleur du texte des entrées en double dans le Plan de codage, indiquant une erreur.
- **Arrière-plan non valide.** Couleur de la surbrillance d'arrière-plan des entrées en double dans le Plan de codage, indiquant une erreur.

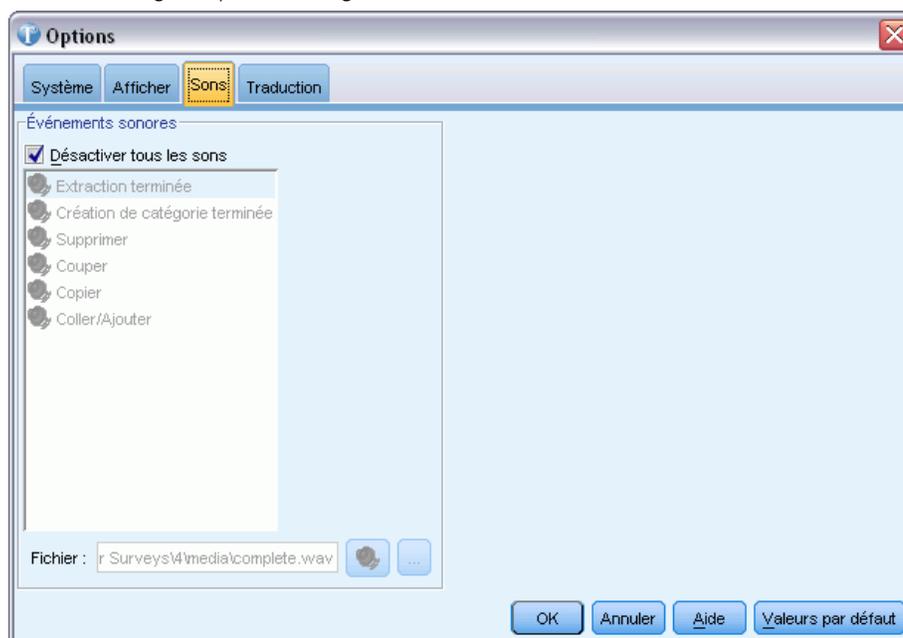
Ordre des couleurs des visualisations. Si vous utilisez le diagramme en bâtons des catégories dans le panneau de visualisation et que vous sélectionnez également une variable de référence, vous pouvez voir chacune des valeurs possibles de la variable de référence dans une légende au bas du panneau. Chaque valeur est également affichée selon un code de couleur pour vous aider à la distinguer visuellement dans le diagramme en bâtons. Vous pouvez modifier ici ces couleurs

par défaut. [Pour plus d'informations, reportez-vous à la section Visualisation des graphiques dans le chapitre 7 sur p. 170.](#)

Options : Onglet Sons

Dans cet onglet, vous pouvez éditer les sons utilisés dans le produit. Dans Sons, vous pouvez indiquer un son à utiliser pour vous avertir lorsqu'un événement se produit. Vous pouvez contrôler le type d'avertissement utilisé pour certaines opérations dans le logiciel en activant ou en désactivant les sons ou en affectant des sons spécifiques. Vous pouvez par exemple activer des sons pour des événements tels que la fin du processus d'extraction, la fin d'une technique de catégorisation automatique, ou d'autres tâches courantes, telles que couper, coller, copier, et supprimer.

Figure 2-9
Boîte de dialogue Options : Onglet Sons



Un grand nombre de sons sont disponibles. Utilisez le bouton ... pour sélectionner un son. Les fichiers .wav utilisés pour créer des sons pour IBM® SPSS® Text Analytics for Surveys sont stockés dans le sous-répertoire /media du répertoire d'installation. Si vous ne souhaitez pas que les sons soient lus, sélectionnez Désactiver tous les sons. Les sons sont désactivés par défaut.

Options : Onglet Traduction

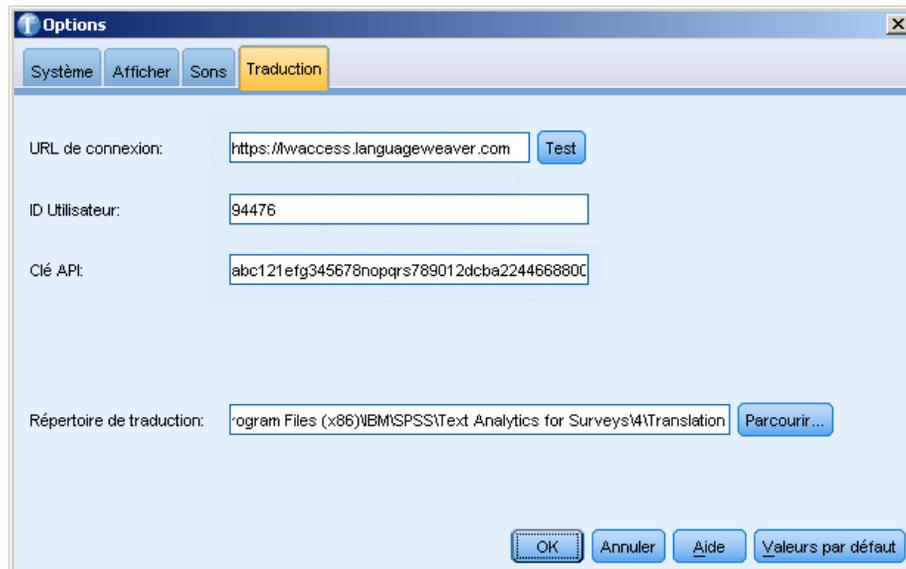
Important ! La traduction n'est disponible qu'en anglais.

Dans cette boîte de dialogue, vous pouvez définir et gérer la connexion du serveur de traduction Language Weaver que vous pouvez réutiliser à chaque traduction. Lorsqu'une connexion a été définie ici, vous pouvez rapidement choisir une connexion de paire de langues au moment de la traduction sans avoir à saisir de nouveau tous les paramètres de connexion.

Une connexion de paire de langues identifie les langue source et langue de traduction ainsi que les détails de la connexion URL vers le serveur. Par exemple, *Chinois - Anglais* signifie que le texte source est en chinois et que la traduction effectuée est en anglais. Vous devez définir manuellement la connexion au serveur Language Weaver auquel vous accédez via les services en ligne Language Weaver.

Les résultats de la traduction sont stockés dans l'emplacement de répertoire défini dans cette boîte de dialogue. Vous pouvez gérer vos fichiers de traduction directement dans ce répertoire et/ou spécifier ici un répertoire différent.

Figure 2-10
Boîte de dialogue Options : Onglet Traduction



Connexion URL. Saisissez l'URL du serveur ou l'adresse Web du serveur en ligne Language Weaver.

ID utilisateur. Entrez l'ID unique qui vous a été fourni par Language Weaver.

Clé API. Entrez la clé unique qui vous a été fournie par Language Weaver.

Test. Cliquez sur Tester pour vérifier que la connexion est configurée correctement et pour afficher les paires de langues trouvées sur cette connexion.

Répertoire de traduction. Cliquez sur Parcourir pour accéder à un répertoire différent ou saisissez le chemin du dossier directement dans le champ.

Figure 2-11
Message de connexion réussie



Paramètres de Microsoft Internet Explorer de l'aide

Paramètres Microsoft Internet Explorer

La plupart des fonctions d'Aide de cette application utilisent des techniques basées sur Microsoft Internet Explorer. Certaines versions d'Internet Explorer (y compris celle fournie avec Microsoft Windows XP, Service Pack 2) bloquent par défaut ce qu'elles considèrent comme du « contenu actif » dans les fenêtres Internet Explorer de votre ordinateur local. Ce paramètre par défaut risque de bloquer du contenu dans les fonctions d'Aide. Pour visualiser tout le contenu de l'Aide, modifiez le comportement par défaut d'Internet Explorer.

- ▶ Dans les menus Internet Explorer, choisissez :
Outils > Options Internet...
- ▶ Cliquez sur l'onglet Avancé.
- ▶ Accédez à la section Sécurité.
- ▶ Sélectionnez (activez) Autoriser le contenu actif à s'exécuter dans les fichiers de la zone Ordinateur local.

Partie II:
Analyse de texte

Création de projets et de packages

Création de projets

Dans IBM® SPSS® Text Analytics for Surveys, vous travaillez avec des données d'enquête et vous les catégorisez. Pour ce faire, vous créez des projets dans lesquels vous importerez des données à partir de votre source de données, sélectionnerez certaines variables, et choisirez des catégories et ressources. Une fois que vous avez créé un projet, vous pouvez ajuster vos ressources et catégories jusqu'à obtenir votre ensemble final de catégories. Un projet peut contenir les éléments suivants : données d'enquête, ressources linguistiques, résultats d'extraction et catégories.

Données d'enquête

La source de données d'enquête importée est référencée dans le projet, mais les données d'enquête ne sont pas enregistrées dans le projet. Au lieu de cela, lorsqu'un projet est ouvert, les données d'enquête sont relues à partir de la source de données d'origine. Bien que les variables les plus importantes dans ce contexte soient les variables de texte, seule la variable d'ID est conservée, ainsi que les variables de référence (telles que les variables démographiques) spécifiées au moment de l'import des données. Les valeurs de toutes ces variables peuvent être affichées dans le panneau Données de la vue Question, dans la fenêtre d'analyse de texte, ou dans la vue Projet complet de la fenêtre d'analyse de texte.

Ressources linguistiques

Les bibliothèques propriétaires et personnalisées par l'utilisateur contenant des listes de termes, des synonymes, des listes de mots exclus et des déclarations de type sont enregistrées dans un projet et peuvent être modifiées. En outre, certaines ressources compilées, utilisées pour traiter du texte, sont également enregistrées dans le projet et ne peuvent pas être modifiées. Les bibliothèques peuvent être publiées, ce qui les met à la disposition de tous dans la base de données pour être utilisées dans d'autres projets. Une bibliothèque publiée peut être ajoutée à d'autres projets.

Résultats extraits

Les résultats d'extraction sont des mots et expressions clés (concepts), leurs regroupements sémantiques (types) et leurs relations (patrons) qui sont identifiés et extraits des réponses textuelles. Ces résultats d'extraction font partie du projet et constituent la base de la création de catégorie. Par défaut, les résultats d'extraction sont enregistrés dans le projet, mais si vous pensez qu'ils rendent les fichiers du projet trop volumineux, vous pouvez désactiver cette fonction et procéder de nouveau à l'extraction la prochaine fois que vous ouvrirez le projet. [Pour plus d'informations, reportez-vous à la section Enregistrement des résultats de l'extraction dans le chapitre 5 sur p. 87.](#)

Categories

Les réponses textuelles sont placées dans des catégories qui peuvent être créées automatiquement à l'aide de techniques de création de catégories, manuellement au moyen d'opérations glisser-déposer, en important des fichiers de définition de catégorie, ou encore en utilisant le Plan de codage. Si vous choisissez de ne pas enregistrer les résultats d'extraction, à chaque fois qu'un projet sera rouvert, les définitions de catégorie seront conservées mais les nombres de réponses correspondant aux différentes parties des définitions et issus des résultats d'extraction seront affichés accompagnés d'un point d'interrogation (?) jusqu'à ce que vous procédiez de nouveau à l'extraction.

Préparation de vos données

Avant d'importer vos données dans IBM® SPSS® Text Analytics for Surveys, veuillez examiner les points suivants :

- **Données de saisie :** Pour que vous puissiez importer votre source de données dans SPSS Text Analytics for Surveys, elle doit contenir certains éléments de base, tels qu'une variable d'ID et au moins une question ouverte. La variable d'ID ne doit contenir que des valeurs uniques. Les doublons peuvent faire échouer l'import. Vous pouvez importer plusieurs questions ouvertes ainsi que des variables de référence. [Pour plus d'informations, reportez-vous à la section Sélection de variables sur p. 35.](#)
- **Fautes d'orthographe.** Bien que le programme prenne en compte certaines fautes d'orthographe, nous vous conseillons de corriger ces fautes avant d'importer vos données dans le programme. Les fautes d'orthographe peuvent provoquer des problèmes dans l'analyse de texte, pour les humains comme pour les logiciels. Plus vous pourrez corriger de fautes d'orthographe à l'avance, plus les catégories obtenues seront fiables. Vous pouvez également créer dans le programme des synonymes avec la correction d'un mot et les variations mal orthographiées les plus fréquentes. Plusieurs fautes d'orthographe courantes sont en fait prédéfinies dans Core Library. Si vous n'êtes pas sûr de la quantité d'effort à consacrer à la correction orthographique, vous pouvez exécuter quelques expériences avec un petit échantillon de réponses pour voir à quel point l'analyse est affectée par les fautes d'orthographe.
- **Réponses vierges.** Il n'est pas rare de trouver des réponses vierges dans les données d'enquête ouvertes. Bien que les réponses vierges ne fournissent aucune information, elles peuvent tout de même être utiles. Vous pouvez par exemple juger intéressant de savoir combien de personnes n'ont pas répondu à une question, ou quel type de personne n'a pas répondu. Toutefois, du fait que SPSS Text Analytics for Surveys utilise du texte pour extraire les termes et catégoriser les réponses, ces réponses vierges ne peuvent pas facilement être catégorisées. Une méthode consiste à remplacer toutes les réponses vierges par le mot *vierge* ou tout autre terme approprié dans vos données avant l'import. Ensuite, une fois les données importées, vous pouvez créer un nouveau type qui représente une réponse vierge, le mot *vierge* (ou le mot que vous aurez inséré) étant le terme qui représente ce type. Une autre option consiste à forcer l'intégration de réponses vierges dans une catégorie. Une fois que vous aurez catégorisé les réponses, les réponses vierges se retrouveront d'abord sans catégorie. Vous pouvez créer une nouvelle catégorie Vierge manuellement en cliquant avec le

bouton droit dans l'écran Catégories. Puis, après avoir sélectionné l'ensemble des réponses vierges, vous pouvez les intégrer de force dans la nouvelle catégorie Vierge.

- **Questions à réponses multiples.** Bien que les questions ouvertes soient généralement indépendantes, ce n'est pas toujours le cas. Les questions ouvertes servent parfois d'ensemble à réponses multiples. Par exemple, si vous demandez à une personne interrogée de "Citer trois choses que nous pouvons améliorer concernant cet hôtel" et fournissez trois espaces séparés (variables) dans lesquels apporter une réponse, cela représente une question à réponses multiples.

Du fait que SPSS Text Analytics for Surveys analyse chaque variable de question séparément, vous pouvez réutiliser les catégories et ressources linguistiques créées pour analyser la première réponse, afin de catégoriser les deuxième et troisième réponses. Toutefois, cette méthode n'est pas systématiquement la plus efficace. Vous pouvez envisager de combiner l'ensemble des trois variables de réponse dans une seule variable avant d'importer les données dans le programme. Si vous les combinez, veuillez vérifier que vous laissez au moins un espace entre le dernier mot d'une réponse et le premier mot de la suivante, ou de préférence un point. Etant donné que la tâche peut être longue avec des ensembles de données volumineux, il est préférable de combiner les réponses lors de la création du fichier de données plutôt qu'après.

- **Echantillons de réponses.** Plus le nombre de réponses est important, et plus ces réponses sont longues en moyenne, plus la durée de l'extraction ou de la catégorisation est longue. Pour travailler plus efficacement, lorsque la taille de votre fichier est importante (peut-être 1500 observations ou plus), vous pouvez prendre d'abord un échantillon au hasard et utiliser ce petit sous-ensemble de réponses pour effectuer une première analyse.

Un échantillon de taille moins importante est souvent parfaitement indiqué pour décider de la façon dont vous modifiez les ressources linguistiques. Et une fois que vous avez effectué une catégorisation sur le fichier de données moins volumineux, vous pouvez lire le fichier entier et procéder de nouveau à l'extraction, ce qui permettra de catégoriser de nombreuses réponses. Vous pouvez ensuite rechercher des réponses qui ne font pas partie des catégories créées et faire les ajustements nécessaires. La taille de l'échantillon aléatoire peut varier, mais un volume de 300 observations environ est généralement suffisant.

Important ! Il existe d'autres considérations à prendre en compte concernant l'ensemble du processus d'analyse de texte. [Pour plus d'informations, reportez-vous à la section Préparation à l'analyse de texte dans le chapitre 1 sur p. 8.](#)

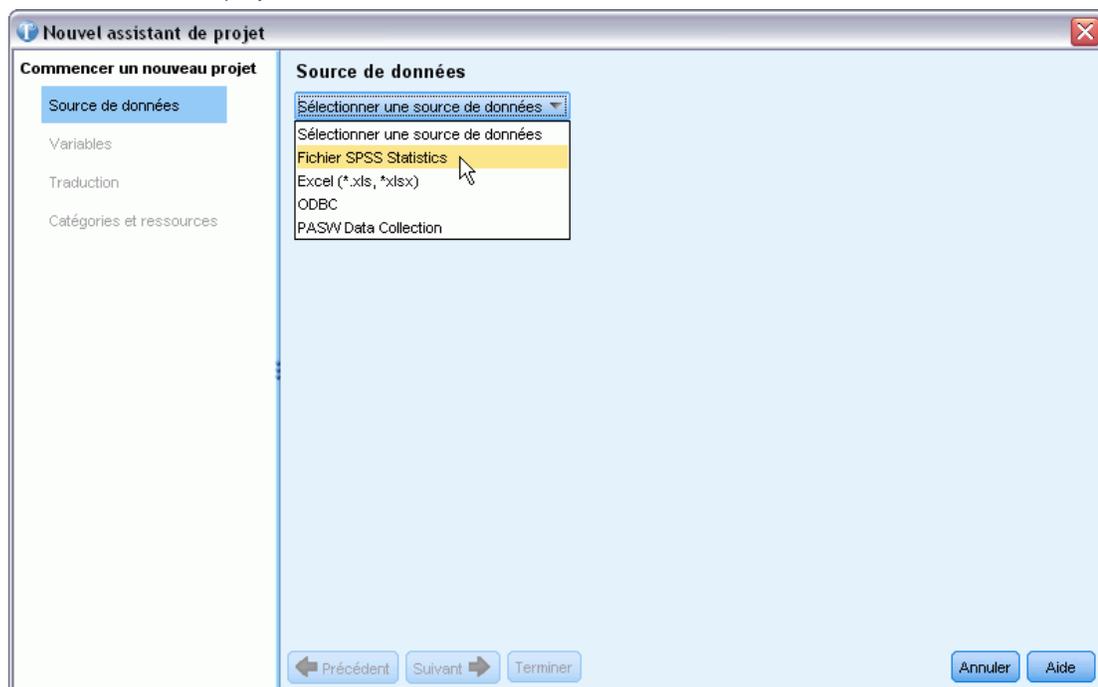
Démarrage de nouveaux projets

Afin de commencer à catégoriser vos données de réponse, vous devez d'abord créer un projet. Un assistant vous guide dans la sélection de source de données et de variables, dans les spécifications de catégories et de ressources, et plus encore. Avant de commencer la création de votre projet, vous pouvez préparer vos données.

Pour démarrer un nouveau projet

- Dans les menus, choisissez Fichier > Nouveau projet. Vous pouvez aussi cliquer sur Démarrer un nouveau projet dans l'écran de démarrage si aucun projet n'est ouvert. L'Assistant de nouveau projet apparaît.

Figure 3-1
Nouvel assistant de projet



- Commencez par sélectionner le type de source de données dans la liste déroulante Sélectionnez une source de données. [Pour plus d'informations, reportez-vous à la section Sélection de sources de données sur p. 30.](#)

Sélection de sources de données

Lorsque l'assistant s'ouvre, vous commencez par sélectionner une source de données. IBM® SPSS® Text Analytics for Surveys a été optimisé pour traiter des ensembles de données allant jusqu'à 10 000 enregistrements, bien que la performance dépende du volume de texte contenu dans ces enregistrements. Reportez-vous aux instructions d'installation pour obtenir des statistiques et des recommandations sur les performances.

Important ! Une variable ID avec une valeur unique pour chaque enregistrement doit être présente afin d'importer les données.

Vous pouvez choisir l'une des sources de données suivantes :

- **SPSS Statistics fichiers** (*.sav).
- **fichiers Microsoft Excel** (*.xls / *.xlsx).

- Base de données **ODBC** (protocole Open Database Connectivity de Microsoft)
- **Data Collection** modèle de données. Cette option n'est disponible que si le modèle de données est installé.

Utilisation de fichiers IBM SPSS Statistics

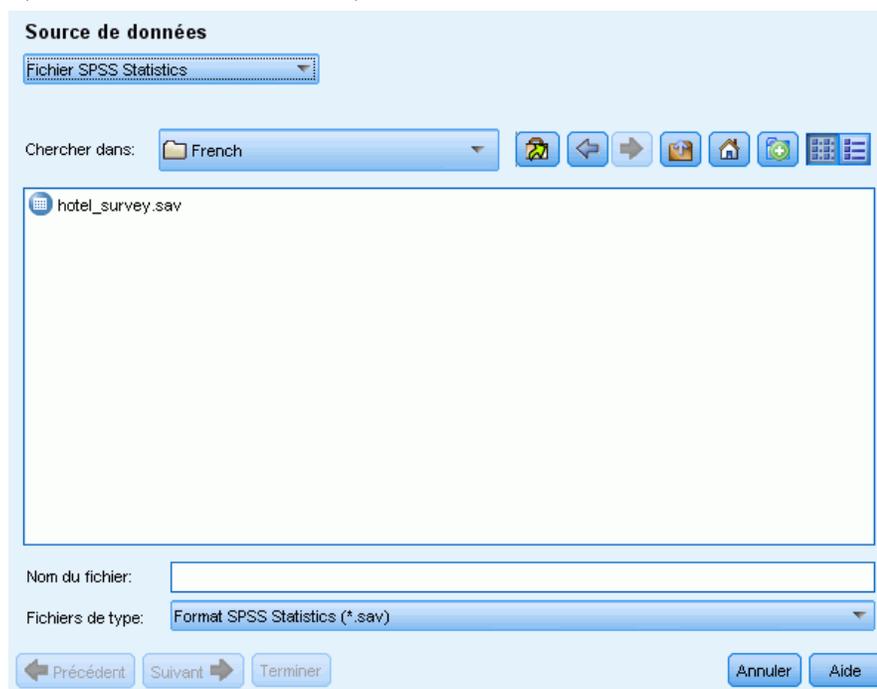
Vous pouvez importer un fichier IBM® SPSS® Statistics (.sav) dans IBM® SPSS® Text Analytics for Surveys. Une variable ID avec une valeur unique pour chaque enregistrement doit être présente afin d'importer les données.

Important ! Vous ne pouvez pas importer un fichier SPSS Statistics (.sav) avec des enregistrements d'une longueur supérieure à 4000 caractères.

Remarque : SPSS Text Analytics for Surveys a été optimisé pour traiter des ensembles de données allant jusqu'à 10 000 enregistrements, bien que la performance dépende du volume de texte contenu dans ces enregistrements. Reportez-vous aux instructions d'installation pour obtenir des statistiques et des recommandations sur les performances.

Figure 3-2

Options de sources de données pour les fichiers IBM SPSS Statistics



Obtenir les données de IBM SPSS Statistics

- ▶ Dans le premier écran de l'assistant, sélectionnez un fichier SPSS Statistics dans la liste déroulante. L'assistant affiche les options disponibles pour les fichiers SPSS Statistics.
- ▶ Dans la liste déroulante Rechercher dans, sélectionnez le pilote et le dossier dans lesquels se trouve le fichier.

- ▶ Sélectionnez le fichier dans la liste. Il apparaîtra dans la zone de texte Nom du fichier.
- ▶ Cliquez sur Suivant pour sélectionner des variables. [Pour plus d'informations, reportez-vous à la section Sélection de variables sur p. 35.](#)

Utilisation de fichiers Microsoft Excel

Vous pouvez importer un fichier Microsoft Excel (.xls / .xlsx) dans IBM® SPSS® Text Analytics for Surveys. Une variable ID avec une valeur unique pour chaque enregistrement doit être présente afin d'importer les données.

Important ! Au cours de l'import du fichier Microsoft Excel, vous pouvez sélectionner une option pour afficher les Noms de colonne dans la première ligne. Pour utiliser cette option, la toute première ligne de la feuille de calcul doit contenir des noms de colonne, et pas la ligne juste dessous où commencent les données. Par exemple, si vos données et noms de colonne commencent à la ligne 7, vous devez supprimer les lignes 1 à 6 avant de commencer l'import, afin d'importer le fichier correctement.

Remarque : SPSS Text Analytics for Surveys a été optimisé pour traiter des ensembles de données allant jusqu'à 10 000 enregistrements, bien que la performance dépende du volume de texte contenu dans ces enregistrements. Reportez-vous aux instructions d'installation pour obtenir des statistiques et des recommandations sur les performances.

Figure 3-3

Options de sources de données pour les fichiers Microsoft Excel

Source de données

Excel (*.xls, *.xlsx)

Chercher dans: French

hotel_survey.xls

Nom du fichier: testProject.xls

Fichiers de type: Fichiers Excel (*.xls, *.xlsx)

Feuille de calcul:

Noms de colonnes dans la première ligne

Précédent Suivant Terminer Annuler Aide

Obtenir les données de Microsoft Excel

- ▶ Dans le premier écran de l'assistant, sélectionnez Excel dans la liste déroulante. L'assistant affiche les options disponibles pour les fichiers Microsoft Excel.
- ▶ Dans la liste déroulante Rechercher dans, sélectionnez le pilote et le dossier dans lesquels se trouve le fichier.
- ▶ Sélectionnez le fichier dans la liste. Il apparaîtra dans la zone de texte Nom du fichier.
- ▶ Sélectionnez la feuille de calcul dans la liste déroulante. Vous ne pouvez importer des données que d'une seule feuille de calcul. Pour travailler avec les données de plusieurs feuilles de calcul, vous devez créer plusieurs projets.
- ▶ Si la première ligne de cette feuille de calcul contient les titres de colonne, sélectionnez Nom de colonne dans la première ligne. Pour utiliser cette option, la toute première ligne de la feuille de calcul doit contenir des noms de colonne, et pas la ligne juste dessous où commencent les données. Par exemple, si vos données et noms de colonne commencent à la ligne 7, vous devez supprimer les lignes 1 à 6 avant de commencer l'import, afin d'importer le fichier correctement. L'application peut les utiliser comme noms de variable (ou une version convertie si les titres de colonne ne sont pas conformes aux conventions de dénomination des variables IBM® SPSS® Statistics). Sinon, l'application utilisera les lettres de colonne de la feuille de calcul comme identificateurs.
- ▶ Cliquez sur Suivant pour sélectionner des variables. [Pour plus d'informations, reportez-vous à la section Sélection de variables sur p. 35.](#)

Utilisation de données via ODBC

Les données issues de sources de base de données, généralement des bases de données, peuvent facilement être importées dans IBM® SPSS® Text Analytics for Surveys. Toute base de données utilisant des pilotes Open Database Connectivity (ODBC) peut être lue directement par le produit, une fois ces pilotes installés sur l'ordinateur sur lequel SPSS Text Analytics for Surveys est installé. Une variable ID avec une valeur unique pour chaque enregistrement doit être présente afin d'importer les données.

Remarque : SPSS Text Analytics for Surveys a été optimisé pour traiter des ensembles de données allant jusqu'à 10 000 enregistrements, bien que la performance dépende du volume de texte contenu dans ces enregistrements. Reportez-vous aux instructions d'installation pour obtenir des statistiques et des recommandations sur les performances.

Figure 3-4
Options de source de données pour ODBC

Source de données

ODBC

Source(DSN) :

Nom	Description
MS Access Database	Microsoft Access Driver (*.mdb)
Excel Files	Microsoft Excel Driver (*.xls)
dBASE Files	Microsoft dBase Driver (*.dbf)
Visio Database Samples	Microsoft Access Driver (*.mdb)

ODBC

Utilisateur: _____

Mot de passe: _____

Table _____

SQL _____

← Précédent Suivant → Terminer Annuler Aide

Utiliser via ODBC

- ▶ Dans le premier écran de l'assistant, sélectionnez ODBC dans la liste déroulante. L'assistant affiche les options disponibles pour ODBC.
- ▶ Spécifiez la source de données en la sélectionnant dans la liste de sources ODBC enregistrées ou en saisissant le nom dans la zone de texte Source (DSN). Si vous avez besoin d'enregistrer de nouvelles sources de données qui n'apparaissent pas dans la liste, cliquez sur ODBC. Cela permet d'ouvrir l'administrateur de source de données ODBC, présent sur la plupart des ordinateurs Microsoft Windows. S'il est introuvable, vous ne pouvez pas utiliser l'import ODBC. Consultez le système d'Aide de Microsoft Windows pour plus d'informations.
- ▶ Si la source de données est protégée par un mot de passe, vous devez saisir un nom d'utilisateur et un mot de passe. Vous devrez faire de même à chaque fois que vous ouvrirez un projet car pour des raisons de sécurité, le nom d'utilisateur et le mot de passe ne sont pas stockés dans le projet.
- ▶ Sélectionnez vos données de l'une des deux manières suivantes : Table ou SQL. Vous pouvez sélectionner une table directement ou utiliser les commandes SQL pour sélectionner des données.
- ▶ Cliquez sur Suivant pour sélectionner des variables. [Pour plus d'informations, reportez-vous à la section Sélection de variables sur p. 35.](#)

Utilisation des données IBM SPSS Data Collection

Importer via IBM SPSS Data Collection

- ▶ Dans le premier écran de l'assistant, sélectionnez Data Collection dans la liste déroulante. L'option de modèle de données IBM® SPSS® Data Collection n'est disponible que si le modèle de données est installé avec un autre produit.

Sélection de variables

Après la sélection de la source de données, l'étape suivante consiste à spécifier les variables à importer. Trois types de variables peuvent être importés dans un projet.

Variable ID unique (Requise)

La variable ID est une clé numérique ou chaîne qui identifie chaque personne interrogée. Le fichier de données n'a pas besoin d'être ordonné par la variable ID unique pour la lire correctement. Une fois lus dans le programme, les enregistrements peuvent être triés par différents critères. [Pour plus d'informations, reportez-vous à la section Trier les variables dans le chapitre 4 sur p. 52.](#) Cette variable ID est requise pour l'import des données. Chaque enregistrement importé (ou observation) doit avoir une valeur ID unique.

Deux situations peuvent faire échouer l'import :

- Valeurs ID en double détectées
- Enregistrements contenant des valeurs ID non renseignées

Remarque : Si un ID en double est détecté et que IBM® SPSS® Statistics est installé sur votre ordinateur, vous pouvez utiliser la procédure Identifier les observations dupliquées de ce produit pour identifier les doublons, puis utiliser les options pour indiquer quels enregistrements doivent être conservés (observations principales).

Variable(s) de texte (Requises)

Les variables de texte représentent les réponses textuelles aux questions de l'enquête. Au moins une de ces variables est requise pour importer des données. Ces variables peuvent être des variables chaîne ou des variables chaîne longues dans SPSS Statistics, des colonnes contenant des cellules de contenu général ou de texte d'Microsoft Excel, ou des champs de texte ou de notes de bases de données. Chaque variable de texte sera analysée séparément. La taille (largeur) de chaque variable texte importée d'un fichier .SAV est limitée à 4 000 caractères.

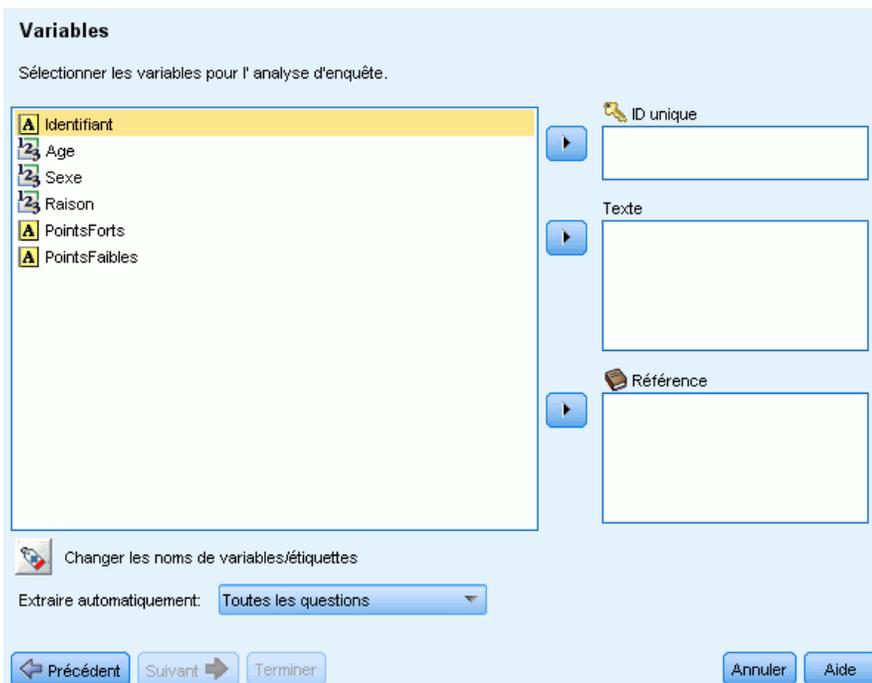
Variable(s) de référence (Optionnelles)

Les variables de référence sont des variables supplémentaires, optionnelles, généralement catégorielles, pouvant être importées à des fins de référence. Les variables de référence ne sont pas utilisées dans l'analyse de texte mais fournissent des informations supplémentaires décrivant la personne interrogée, qui peuvent aider à la compréhension et à l'interprétation. Les variables démographiques sont souvent incluses comme variables de référence, car elle peuvent aider à

comprendre quels termes ou catégories sont utilisés par quels groupes d'individus. Il s'agit par exemple du sexe, du département, de la profession et du niveau d'études (pour les évaluations des étudiants et d'apprentissage). Vous pouvez afficher l'ensemble des variables de référence après leur import, dans la vue Projet complet. Vous pouvez également afficher les variables de référence dans le panneau Données de la vue Question. De plus, vous pouvez sélectionner des variables de référence dans le diagramme à barres du panneau de visualisation pour pouvoir passer à un sous-ensemble de personnes interrogées.

Remarque : Les variables de référence lues à partir d'un fichier de données SPSS Statistics auront des étiquettes de variable (le cas échéant) qui apparaîtront en tant que titres de colonne, tandis que leurs étiquettes de valeur (le cas échéant) s'afficheront dans les cellules du panneau Données.

Figure 3-5
Sélection de variables



Pour sélectionner des variables et des options d'extraction

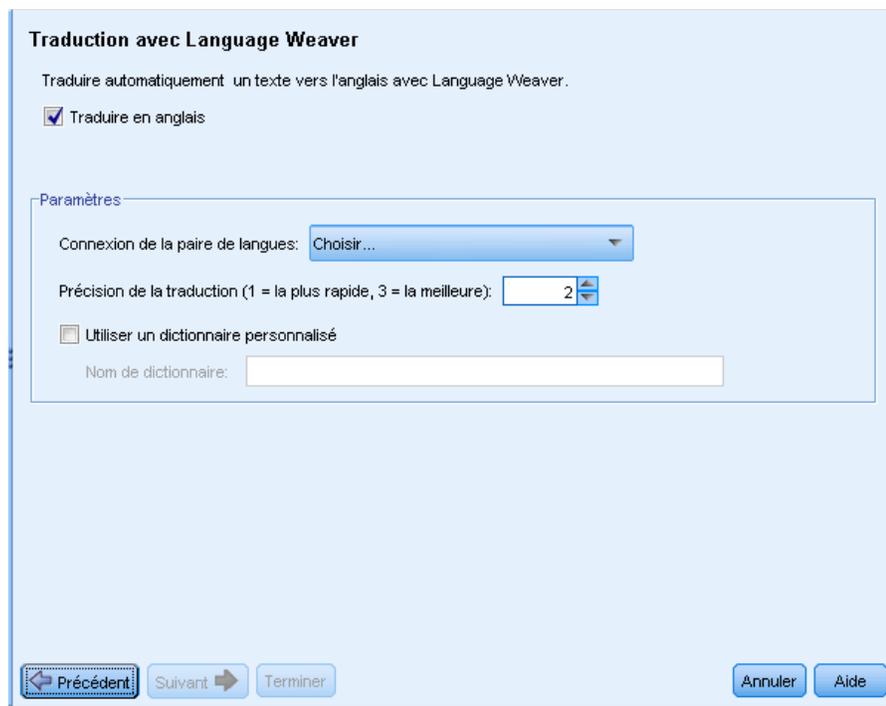
- ▶ Dans la liste des variables disponibles, sélectionnez la variable qui correspond à la variable ID de votre ensemble de données et cliquez sur la flèche pour la déplacer vers la zone ID unique. L'ID doit être une chaîne numérique ou alphanumérique unique qui distingue un enregistrement d'un autre. Si votre ensemble de données contient des ID en double, un message d'erreur apparaît. Dans ce cas, vous devez d'abord nettoyer vos données avant de réessayer.
- ▶ Dans la liste des variables disponibles, sélectionnez une ou plusieurs variables qui correspondent aux variables de réponse ouverte et cliquez sur la flèche pour déplacer les variables vers la liste Texte. Chacune des variables sera importée en tant que question distincte dont vous analyserez et catégoriserez les réponses.

- ▶ Dans la liste des variables disponibles, sélectionnez une ou plusieurs variables qui correspondent aux variables de référence et cliquez sur la flèche pour déplacer les variables vers la liste Référence. Les variables de référence ne sont pas utilisées par les techniques de création de catégorie automatisées. Vous pouvez toutefois afficher leur contenu et les utiliser pour vous aider à prendre des décisions judicieuses lors de la catégorisation de vos réponses.
- ▶ Pour afficher les étiquettes de variable au lieu des noms de variable, cliquez sur le bouton sous la liste des variables à gauche.
- ▶ Pour modifier le paramètre d'extraction, faites une sélection dans la liste déroulante. Par défaut, l'option Première question uniquement est sélectionnée, ce qui signifie que si plus d'une variable de texte est sélectionnée, le processus d'extraction démarre automatiquement pour la première question dès que l'assistant prend fin. L'extraction peut prendre du temps avec les ensembles de données plus importants. Vous pouvez par conséquent choisir d'extraire Aucune ou Toutes les questions en fonction du temps disponible.
- ▶ Cliquez sur Suivant> lorsque vous avez sélectionné toutes les variables.

Traduction vers l'anglais

Si vous travaillez avec un texte source non anglais, vous pouvez vous connecter à Language Weaver pour le traduire en anglais. La traduction n'est disponible qu'en anglais. Vous devez avoir configuré Language Weaver de manière appropriée et avec les connexions définies pour réaliser la traduction. [Pour plus d'informations, reportez-vous à la section Options : Onglet Traduction dans le chapitre 2 sur p. 23.](#)

Figure 3-6
Options de traduction



Pour traduire en anglais

- ▶ Pour traduire en anglais les données textuelles à partir d'une langue bénéficiant d'une licence, cochez la case Traduire en anglais.
- ▶ Dans la liste connexion des paires de langues, sélectionnez la connexion pour la paire de langues Language Weaver que vous souhaitez utiliser. Si Language Weaver est configuré sur votre ordinateur local, ces paires de langues apparaîtront automatiquement dans cette liste. Vous pouvez ajouter, modifier ou tester la connexion des services en ligne dans l'onglet Traduction de la boîte de dialogue Options. [Pour plus d'informations, reportez-vous à la section Options : Onglet Traduction dans le chapitre 2 sur p. 23.](#)
- ▶ Spécifiez la Précision de la traduction de votre choix. Choisissez une valeur de 1 à 3 indiquant le rapport entre le niveau de vitesse et de précision souhaité. Une valeur faible permet d'obtenir des résultats de traduction plus rapides mais moins précis. Une valeur élevée permet, quant à elle, d'obtenir des résultats plus précis mais après une durée de traitement plus longue. Pour optimiser le temps de traitement, nous vous recommandons de commencer par un niveau peu élevé et de l'augmenter seulement si, après avoir examiné les résultats, vous pensez avoir besoin d'une plus grande précision.
- ▶ Si vous aviez précédemment créé des dictionnaires personnalisés, détenus par Language Weaver, vous pouvez les utiliser dans le cadre de la traduction. Pour choisir un dictionnaire personnalisé, cochez la case Utiliser un dictionnaire personnalisé et entrez le Nom du dictionnaire. Pour utiliser plusieurs dictionnaires, séparez les noms avec une virgule.

- ▶ Dans l'assistant Nouveau projet, cliquez sur Suivant > pour commencer à sélectionner des catégories et des ressources. [Pour plus d'informations, reportez-vous à la section Sélection de catégories et de ressources sur p. 39.](#)
- ▶ Dans l'assistant Modifier l'ensemble de données, cliquez sur Terminer pour terminer la modification de l'ensemble de données et pour démarrer le processus de traduction.

Pour ignorer la traduction :

- ▶ Désélectionnez l'option Traduire en anglais.
- ▶ Dans l'assistant Nouveau projet, cliquez sur Suivant > pour commencer à sélectionner des catégories et des ressources. [Pour plus d'informations, reportez-vous à la section Sélection de catégories et de ressources sur p. 39.](#)
- ▶ Dans l'assistant Modifier l'ensemble de données, cliquez sur Terminer pour terminer la modification de l'ensemble de données.

Sélection de catégories et de ressources

Dans cette dernière étape, vous pouvez sélectionner les ressources linguistiques qui seront utilisées pour extraire les concepts et patrons prépondérants de votre texte. Autrement, vous pouvez charger un package d'analyse de texte (TAP), qui inclut non seulement les ressources linguistiques mais également un ou plusieurs ensembles de catégories prédéfinies qui représentent des plans de codage améliorés. [Pour plus d'informations, reportez-vous à la section Utilisation des packages d'analyse de texte sur p. 43.](#) IBM® SPSS® Text Analytics for Surveys offre plusieurs fichiers TAP prédéfinis pour le texte en langue anglaise. Chaque fichier TAP fourni avec ce produit est ajusté pour un type d'enquête spécifique, tel que des enquêtes de type employé, produit ou satisfaction de la clientèle. Vous pouvez également créer vos propres TAP pour n'importe quelle langue prise en charge par le produit.

Par défaut, un modèle de ressources est préchargé. Vous pouvez changer le modèle de ressources par défaut proposé dans le premier onglet de la boîte de dialogue Options. [Pour plus d'informations, reportez-vous à la section Définition des options dans le chapitre 2 sur p. 19.](#) Vous pouvez charger un modèle de ressources différent ou choisir un TAP à la place.

Figure 3-7
Sélection des ressources

Catégories et ressources

Charger une copie des ressources que vous souhaitez utiliser pour l'extraction de texte. Les résultats de l'extraction sont utilisés pour construire vos catégories. Ces ressources dictent la manière dont le texte est manipulé et extrait.

Copier les ressources depuis :

Modèle de ressource.
Je créerai des catégories ultérieurement.

Pack d'analyse de texte (TAP) comprenant des ensembles de catégories prédéfinis et des ressources associées. Je peux modifier les catégories ultérieurement.

Selected: Opinions (French)

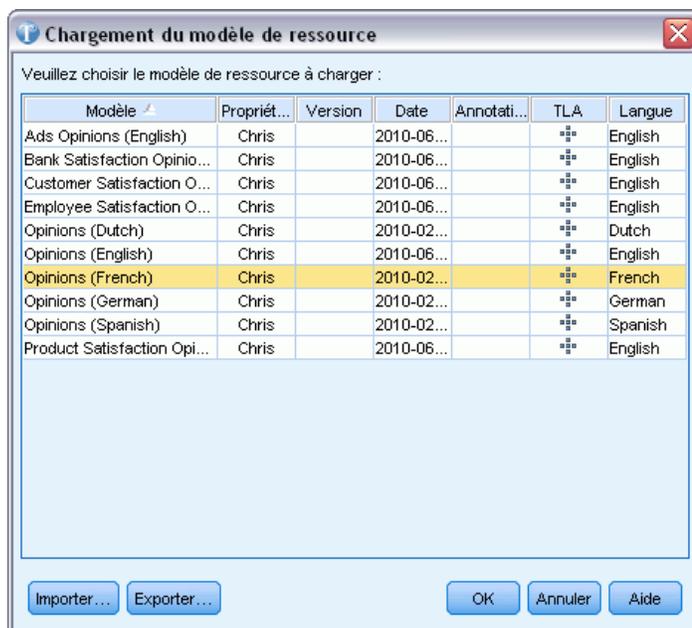
Chargement...

← Précédent Suivant → Terminer Annuler Aide

Pour sélectionner un modèle de ressource différent :

- Pour charger un modèle de ressources différent, assurez-vous que l'option **Modèle de ressources** est sélectionnée et cliquez sur **Charger**. La boîte de dialogue **Charger le modèle de ressources** apparaît.

Figure 3-8
Charger le modèle de ressources



- Dans la boîte de dialogue Charger le modèle de ressources, sélectionnez le modèle à utiliser et cliquez sur OK. La boîte de dialogue se ferme et l'assistant affiche le nouveau modèle choisi.

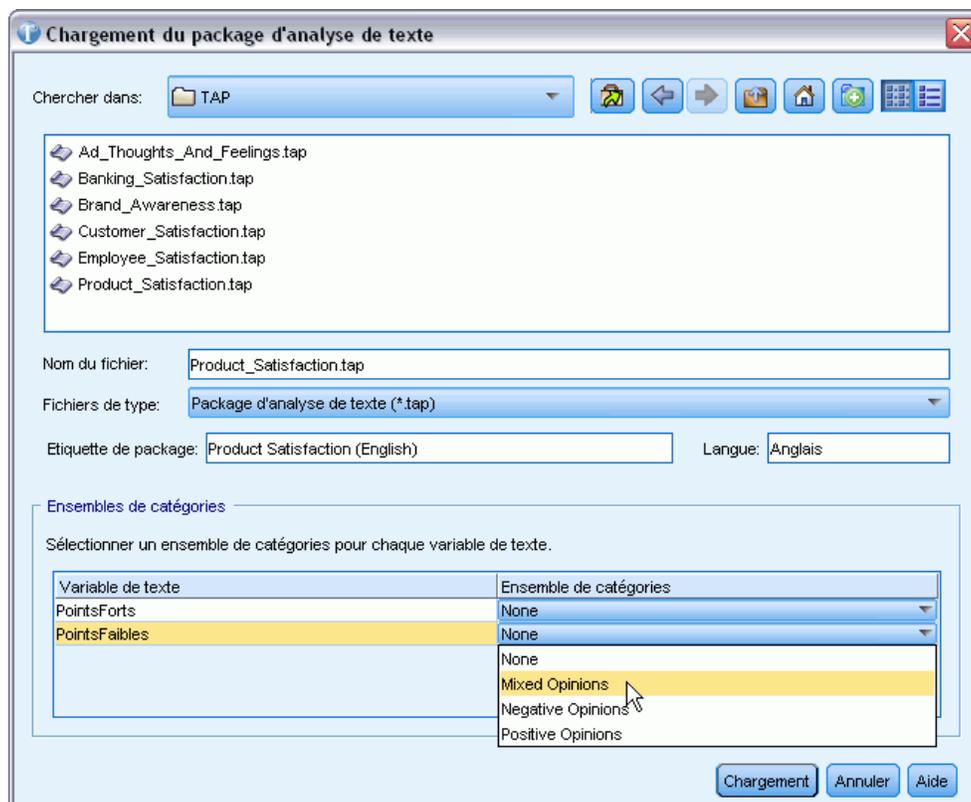
Notez que si vous avez des modèles dans des langues pour lesquelles vous ne disposez pas de licence, une case à cocher au bas de la boîte de dialogue vous permet de masquer les modèles de langues non autorisées.

- Cliquez sur Terminer pour lancer l'importation des données. Si vous choisissez cette option, le projet créé possédera les ressources linguistiques par défaut, et une fois les extractions exécutées, vous pourrez créer vos catégories manuellement ou au moyen d'une technique automatisée. [Pour plus d'informations, reportez-vous à la section Catégorisation des données textuelles dans le chapitre 6 sur p. 96.](#)

Pour sélectionner un package d'analyse de texte :

- Pour charger un package d'analyse de texte, assurez-vous que l'option Package d'analyse de texte est sélectionnée et cliquez sur Charger. La boîte de dialogue Charger un package d'analyse de texte apparaît.

Figure 3-9
Charger un package d'analyse de texte



- ▶ Dans la boîte de dialogue, sélectionnez le TAP à utiliser. Seuls les packages stockés dans le répertoire <installation_directory>\TAP par défaut apparaissent directement dans la liste. Les champs de dessous sont mis à jour et affichent les détails spécifiques du TAP sélectionné.
- ▶ Dans la table Ensembles de catégories, vous pouvez affecter un ensemble de catégories à chacune des variables de texte. Dans la colonne Ensemble de catégories, cliquez sur la liste déroulante de chaque cellule pour choisir un ensemble de catégories disponible. Si vous sélectionnez Aucun, vous n'aurez aucune catégorie pour cette variable de texte, tant que vous n'en créez pas ultérieurement.
- ▶ Cliquez sur OK. La boîte de dialogue se ferme et l'assistant affiche le nouveau TAP choisi. Après la sélection de TAP ou d'ensembles de catégories, l'assistant a terminé et en quelques instants vous pouvez voir vos enregistrements codés dans les catégories prédéfinies. A partir de là, vous pouvez exporter les résultats ou utiliser les catégories comme point de départ pour votre analyse.
- ▶ Cliquez sur Terminer pour fermer la boîte de dialogue et créer le projet. Une fois terminé, l'application ouvre automatiquement la vue Question pour la première question de texte de votre projet. Si vous choisissez d'exécuter une extraction, une boîte de dialogue de progression de l'extraction s'affiche et le processus d'extraction peut prendre un certain temps. Vous pouvez maintenant commencer à analyser vos questions. Pour passer à une autre question, choisissez Affichage > Question dans les menus.

Utilisation des packages d'analyse de texte

Un package d'analyse de texte, également appelé TAP (Text Analysis Package), sert en tant que modèle de catégorisations de réponses texte. L'utilisation d'un TAP est un moyen facile de catégoriser vos données textuelles en intervenant le moins possible car il contient les catégories et les ressources linguistiques nécessaires au codage d'un grand nombre d'enregistrements rapidement et automatiquement. Les ressources linguistiques permettent d'analyser et d'exploiter les données textuelles pour en extraire des concepts clés. En fonction des concepts clés et des patrons trouvés dans le texte, les enregistrements peuvent être catégorisés dans l'ensemble de catégories sélectionné dans le TAP. Vous pouvez créer votre propre TAP ou en mettre un à jour.

Un TAP est composé des éléments suivants :

- **Ensemble(s) de catégories.** Un ensemble de catégories est principalement composé de catégories prédéfinies, de codes de catégorie, de descripteurs pour chaque catégorie et enfin d'un nom pour l'ensemble en son entier. Les descripteurs sont des éléments linguistiques (concepts, types, patrons et règles) comme le terme *cher* ou le patron *bon prix*. Les descripteurs sont utilisés pour définir une catégorie. Ainsi, lorsque le texte correspond à un descripteur de catégorie, l'enregistrement est placé dans cette catégorie.
- **Ressources linguistiques.** Les ressources linguistiques sont un ensemble de bibliothèques et de ressources avancées qui permettent d'extraire des concepts et patrons clés. Ces concepts et patrons d'extraction sont utilisés comme descripteurs qui permettent aux enregistrements d'être placés dans une catégorie de l'ensemble de catégories.

Vous pouvez créer et mettre à jour des packages d'analyse de texte.

Après avoir sélectionné un TAP et un ensemble de catégories à chaque variable de texte dans l'assistant Nouveau projet, IBM® SPSS® Text Analytics for Surveys peut effectuer une extraction et catégoriser vos documents. Ensuite, vous pouvez soit exporter les résultats ou continuer à affiner la catégorisation pour obtenir les résultats souhaités.

Remarque : Vous pouvez créer et interchanger les TAP entre SPSS Text Analytics for Surveys et IBM® SPSS® Text Analytics.

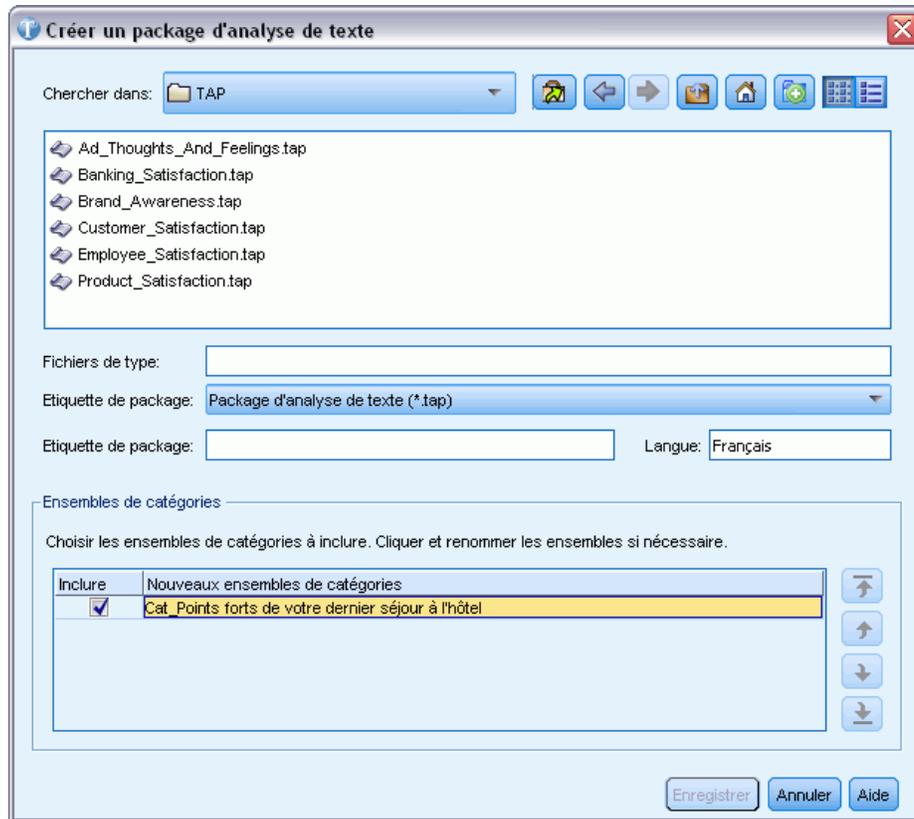
Création des packages d'analyse de texte

Lorsque vous avez un projet contenant au moins une catégorie et des ressources, vous pouvez créer un package d'analyse de texte (TAP) à partir du contenu de projet ouvert. L'ensemble de catégories et de descripteurs (concepts, types, règles ou résultats de patrons TLA) dans chaque question peut être transformé en TAP utilisant toutes les ressources linguistiques ouvertes dans l'éditeur de ressources.

Vous pouvez voir la langue pour laquelle les ressources ont été créées. La langue est définie dans l'onglet Ressources avancées de l'éditeur de ressources.

Important ! Si vos catégories contiennent des correspondances textuelles, des enregistrements forcés ou des marqueurs, ils ne seront pas enregistrés dans des ensembles de catégories car ils sont propres à certaines données et ne peuvent généralement pas être utilisés pour d'autres données. Cependant, les codes et étiquettes de catégorie sont enregistrés.

Figure 3-10
Boîte de dialogue Créer un package d'analyse de texte

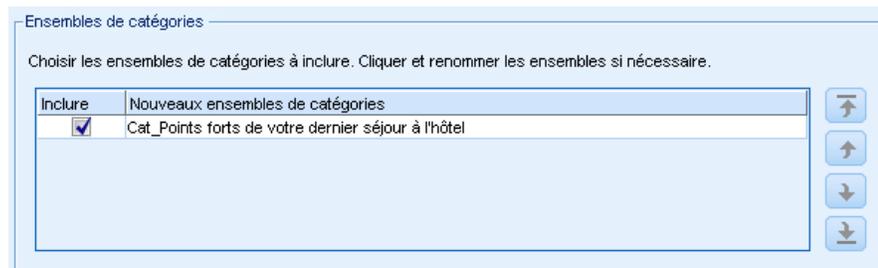


Pour créer un package d'analyse de texte

- ▶ Dans les menus, choisissez Fichier > Packages d'analyse de texte > Créer un package. La boîte de dialogue Créer un package apparaît.
- ▶ Accédez au répertoire dans lequel vous avez enregistré le TAP. Par défaut, les TAP sont enregistrés dans le sous-répertoire \TAP du répertoire d'installation du produit.
- ▶ Entrez un nom pour le TAP dans le champ Nom du fichier.
- ▶ Saisissez une étiquette dans le champ Etiquette du package. Lorsque vous saisissez un nom de fichier, ce nom apparaît automatiquement comme étiquette mais vous pouvez la modifier. Si vous enregistrez un TAP dans ce répertoire par défaut, l'étiquette du package apparaîtra comme option dans la liste déroulante de l'assistant Nouveau projet.
- ▶ Pour exclure un ensemble de catégories du TAP, désélectionnez la case Inclure. Ainsi, il ne sera pas ajouté à votre package. Par défaut, un ensemble de catégories par question est inclus dans le TAP. Le TAP doit contenir au moins un ensemble de catégories.
- ▶ Renommer des ensembles de catégories. La colonne Nouvel ensemble de catégories contient des noms génériques par défaut qui sont générés en ajoutant le préfixe Cat_ au nom de la variable de texte. Un simple clic dans la cellule permet de modifier ce nom. Appuyer sur entrée ou cliquer à

un autre droit permet d'appliquer la modification du nom. Si vous renommez un ensemble de catégories, le nom est uniquement modifié dans le TAP et ne modifie pas le nom de la variable dans le projet ouvert.

Figure 3-11
Changement de nom des ensembles de catégories

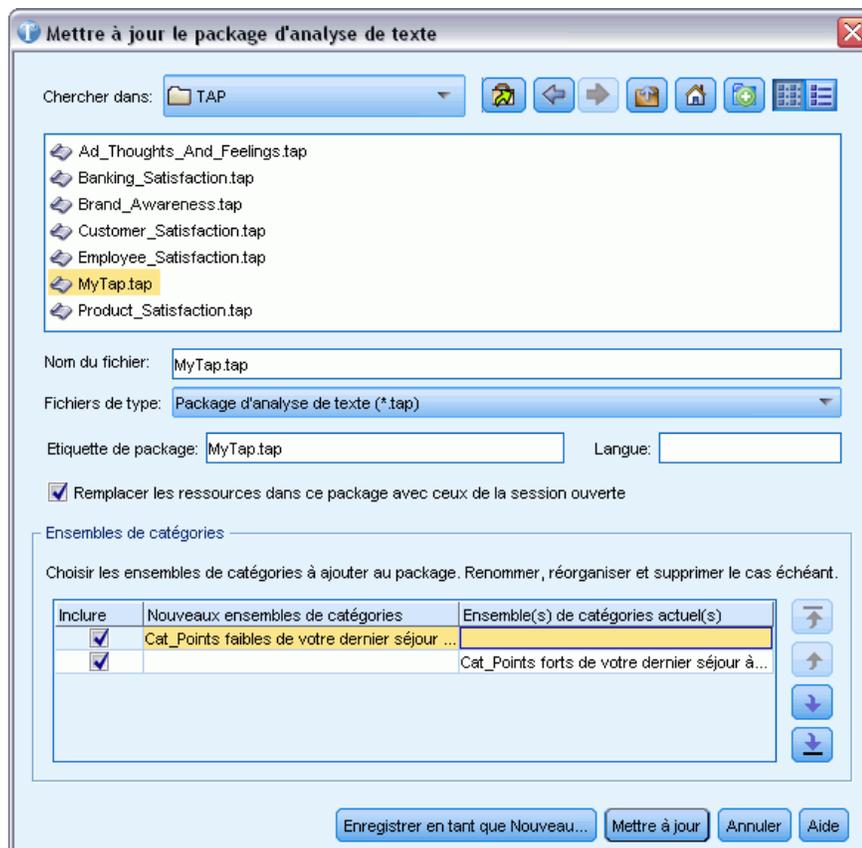


- ▶ Vous pouvez changer l'ordre des ensembles de catégories avec les flèches situées à droite du tableau des ensembles de catégories.
- ▶ Cliquez sur Enregistrer pour créer le package d'analyse de texte. La boîte de dialogue se ferme.

Mise à jour des Packages d'analyse de texte

Si vous apportez des améliorations à un ensemble de catégories, à des ressources linguistiques ou que vous créez un tout nouvel ensemble de catégories, vous pouvez mettre à jour un package d'analyse de texte (TAP) pour que ces améliorations soient plus faciles à réutiliser ultérieurement. Pour ce faire, allez dans la session ouverte contenant les informations que vous souhaitez intégrer au TAP. Lorsque vous effectuez une mise à jour, vous pouvez choisir d'ajouter des ensembles de catégories, de remplacer des ressources, de changer l'étiquette du package ou de renommer/réorganiser les ensembles.

Figure 3-12
Boîte de dialogue Mettre à jour un package d'analyse de texte



Pour mettre à jour un package d'analyse de texte

- ▶ Dans les menus, choisissez Fichier > Packages d'analyse de texte > Mise à jour d'un package. La boîte de dialogue Mettre à jour un package d'analyse de texte apparaît.
- ▶ Accédez au répertoire contenant le package d'analyse de texte à mettre à jour.
- ▶ Entrez un nom pour le TAP dans le champ Nom du fichier.
- ▶ Pour remplacer les ressources linguistiques à l'intérieur du TAP par celles du projet en cours, sélectionnez l'option Remplacer les ressources de ce package par celles de la session ouverte. Généralement, il est utile de mettre à jour les ressources linguistiques car elles ont servi à extraire les concepts et patrons clés utilisés pour créer les définitions de catégories. Utiliser les ressources linguistiques les plus récentes permet d'obtenir de meilleurs résultats lors de la catégorisation de vos enregistrements. Si vous ne sélectionnez pas cette option, les ressources linguistiques déjà contenues dans le package sont conservées en l'état.
- ▶ Pour modifier uniquement les ressources linguistiques, vérifiez que l'option Remplacer les ressources de ce package par celles de la session ouverte est bien sélectionnée puis sélectionnez uniquement les ensembles de catégories actuels qui se trouvaient dans le TAP.

- ▶ Pour inclure les nouveaux ensembles de catégories du projet en cours dans le TAP, cochez la case de chaque ensemble de catégories à ajouter. Vous pouvez ajouter un, plusieurs ou aucun des ensembles de catégories.
- ▶ Pour supprimer des ensembles de catégorie du TAP, désélectionnez la case Inclure correspondante. Vous pouvez choisir de supprimer un ensemble de catégories qui se trouvait déjà dans le TAP car vous en ajoutez un ayant été amélioré. Pour ce faire, désélectionnez la case Inclure de l'ensemble de catégories correspondant dans la colonne Ensemble de catégories actuel. Le TAP doit contenir au moins un ensemble de catégories.
- ▶ Si nécessaire, modifiez les noms des ensembles de catégories. Un simple clic dans la cellule permet de modifier ce nom. Appuyer sur entrée ou cliquer à un autre droit permet d'appliquer la modification du nom. Si vous renommez un ensemble de catégories, le nom est uniquement modifié dans le TAP et ne modifie pas le nom de la variable dans le projet ouvert. Si deux ensembles de catégories ont le même nom, ces noms apparaissent en rouge tant que le doublon n'a pas été corrigé.

Figure 3-13
Noms en double

Inclure	Nouveaux ensembles de catégories	Ensemble(s) de catégories actuel(s)
<input checked="" type="checkbox"/>	Cat_Points forts de votre dernier séjour à ...	
<input checked="" type="checkbox"/>	Cat_Points faibles de votre dernier séjour ...	
<input checked="" type="checkbox"/>		Cat_Points forts de votre dernier séjour à...
<input checked="" type="checkbox"/>		Cat_Points faibles de votre dernier séjour...

- ▶ Pour créer un nouveau package en intégrant le contenu de la session au contenu du TAP sélectionné, cliquez sur Enregistrer en tant que Nouveau. La boîte de dialogue Enregistrer comme package d'analyse de texte apparaît. Suivez les instructions suivantes.
- ▶ Cliquez sur Mettre à jour pour enregistrer les modifications effectuées dans le TAP sélectionné.

Figure 3-14
Boîte de dialogue Enregistrer comme package d'analyse de texte



Pour enregistrer un package d'analyse de texte

- ▶ Accédez au répertoire dans lequel vous avez enregistré le fichier du TAP. Par défaut, les fichiers TAP sont enregistrés dans le sous-répertoire TAP du répertoire d'installation.
- ▶ Entrez un nom pour le fichier TAP dans le champ Nom du fichier.
- ▶ Saisissez une étiquette dans le champ Etiquette du package. Le nom de fichier choisi est automatiquement utilisé comme étiquette. Mais vous pouvez renommer cette étiquette. Une étiquette est nécessaire. Si vous enregistrez un TAP dans ce répertoire par défaut, l'étiquette du package apparaîtra comme option dans la liste déroulante de l'assistant Nouveau projet.
- ▶ Cliquez sur Enregistrer pour créer le nouveau package.

Utilisation des projets

Dans IBM® SPSS® Text Analytics for Surveys, vous travaillez avec des données d'enquête et vous les catégorisez. Pour cela, vous créez des projets dans lesquels vous construisez et stockez des définitions de catégories et les réponses auxquelles elles correspondent. Un projet peut contenir les éléments suivants :

- **Données d'enquête.** Variable de réponse sous forme de texte pour les questions ouvertes, une variable d'ID unique et toute variable de référence optionnelle. Les données d'enquête ne sont pas stockées dans le projet, elles sont lues depuis la source de données d'origine lorsque le projet est ouvert.
- **Ressources linguistiques.** Modèles et bibliothèques propriétaires et personnalisés (dictionnaires de synonymes, d'exclusions et déclarations de types) utilisés lors de l'extraction de concepts et de modèles du texte.
- **Résultats extraits.** Après une extraction, ce sont les mots clés et les phrases identifiés et extraits de vos données de réponse. Vous utiliserez ces concepts pour créer vos catégories.
- **Catégories.** Proviennent des ensembles de catégories TAP, de la création manuelle et/ou de la technique de création de catégories automatisée. Les réponses aux enquêtes sont classées selon ces catégories.

Ouvrir des projets

Vous pouvez retourner à un projet existant en l'ouvrant. Vous ne pouvez ouvrir qu'un seul projet à la fois. Si vous essayez d'ouvrir un projet quand un autre est déjà ouvert, le système vous invite à enregistrer d'abord l'autre projet si nécessaire.

Lorsqu'un projet est ouvert, IBM® SPSS® Text Analytics for Surveys contrôle vos ressources linguistiques pour déterminer si une bibliothèque publique plus récente que celles dans le projet existe. Si c'est le cas, le système vous demande si vous souhaitez mettre à jour les bibliothèques. Vous pouvez choisir de garder votre version et de ne pas effectuer la mise à jour, ou de fusionner les mises à jour dans votre projet. [Pour plus d'informations, reportez-vous à la section Mise à jour des bibliothèques dans le chapitre 9 sur p. 217.](#)

Attention ! SPSS Text Analytics for Surveys ne stocke pas physiquement la donnée source dans ses projets. En revanche, une référence à cette donnée sur votre ordinateur est stockée dans le projet. Si quelqu'un modifie l'une des variables d'origine importées dans la source de données, un avertissement s'affiche pour vous prévenir que les données sont introuvables. Si cela se produit, vous devez réimporter les données et associer les "nouvelles" questions aux noms de variables d'origine pour continuer à travailler sur ce projet. En général, il n'est pas recommandé de renommer les variables ou les en-têtes de colonnes dans vos données sources.

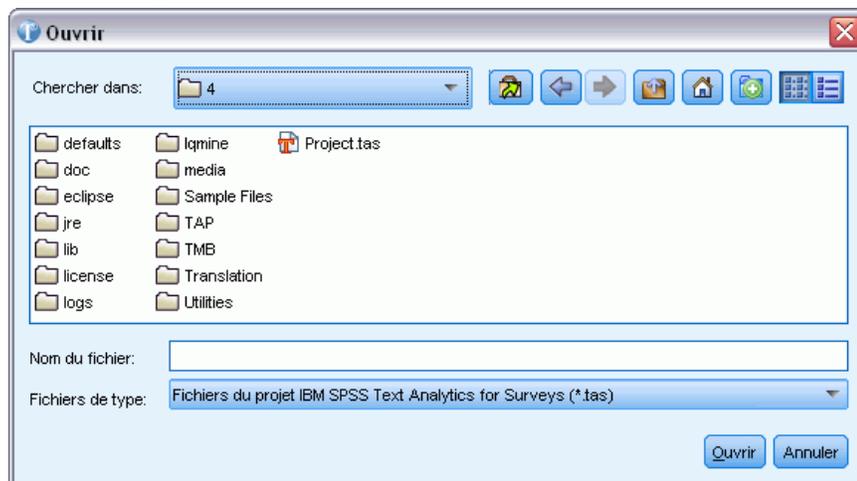
Remarque : Les résultats de l'extraction sont enregistrés dans vos projets, sauf si vous choisissez de ne pas le faire (Outils > Options). Lorsque vous fermez un projet, vos définitions de catégorie sont enregistrées mais le panneau Résultats d'extraction est effacé. Lorsque vous ouvrez ce projet, vous devez exécuter une extraction si vous voulez continuer à catégoriser vos réponses. Les

définitions de catégorie existantes affichent un point d'interrogation (?) à la place d'un nombre d'un réponses. Le nombre de réponses réapparaît après la réextraction.

Pour ouvrir un projet

- Dans les menus, choisissez Fichier > Ouvrir projet. La boîte de dialogue Ouvrir apparaît.

Figure 4-1
Boîte de dialogue Ouvrir



- Dans la liste, sélectionnez le répertoire et le nom du projet que vous voulez ouvrir. Vous ne pouvez pas trier les détails comme la taille de fichier et la date, dans cette boîte de dialogue.
- Cliquez sur OK pour ouvrir le projet dans la fenêtre principale.
- Si votre projet contient des données d'une base de données protégée par mot de passe, le système vous demande le mot de passe à chaque fois que vous ouvrez ce projet.
- Si le contenu de votre projet vient d'une version précédente du produit, le système vous demande de convertir vos ressources dans le nouveau format. Ceci implique qu'après avoir enregistré votre projet, vous ne pourrez plus l'ouvrir dans une version précédente de ce produit.
- Si des bibliothèques publiques dans votre projet ont été modifiées depuis votre dernière ouverture du projet, une alerte vous avertit de ces changements.

Important ! Dès que vous ouvrez un projet, l'ensemble de données correspondant est ouvert. Si ces données sont introuvables, un message d'erreur apparaît. Vous devez réimporter les données afin de continuer à les utiliser. [Pour plus d'informations, reportez-vous à la section Modifier les sources de données sur p. 63.](#)

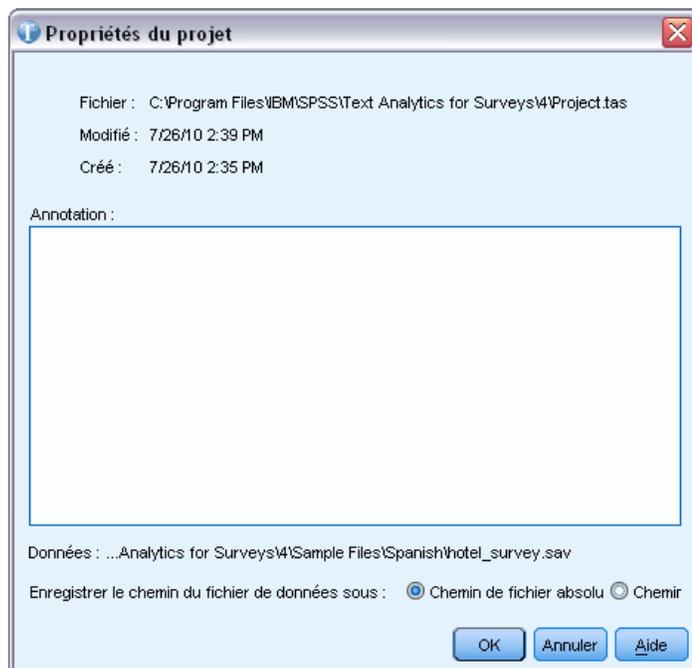
Modifier les propriétés du projet

Par défaut, chaque nouveau projet est nommé *Projet 1*. Vous pouvez examiner les propriétés de base de votre projet et ajouter ou modifier une annotation pour celui-ci.

Pour modifier les propriétés du projet

- Dans les menus, sélectionnez Fichier > Propriétés du projet. La boîte de dialogue Propriétés du projet apparaît.

Figure 4-2
Boîte de dialogue Propriétés du projet



- Si vous le souhaitez, insérez un commentaire ou la description du projet dans le champ Annotation.
Remarque : Le nom du fichier de données s'affiche dans cette boîte de dialogue. Comme vous créez un projet et n'avez pas encore importé de données, le nom du fichier est inconnu. Après l'import, seuls les 60 derniers caractères du nom du fichier de données apparaissent ici. Si le nom du fichier est plus long, vous pouvez passer votre curseur sur le nom pour l'afficher en entier.
- Cliquez sur OK pour accepter les nouvelles propriétés. La boîte de dialogue se ferme et les propriétés du projet sont appliquées.

Afficher les données du projet

Lorsque vous avez importé le fichier de données, la vue Question s'affiche pour la première question texte dans votre projet. Néanmoins, vous souhaitez peut-être examiner toutes les données que vous avez importées. Vous pouvez le faire dans la vue Projet complet, qui offre un affichage global de vos données. Pour accéder à cette vue, sélectionnez Affichage > Projet complet dans le menu. Dans cette vue, vous pouvez effectuer les opérations suivantes :

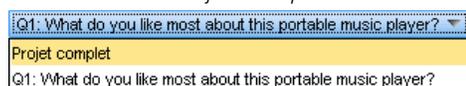
- Examiner le contenu de toutes les variables importées.
- Attribuer des valeurs et des étiquettes aux variables.
- Changer les types des variables.

- Trier les variables.
- Copier des données de cellules contiguës et les coller dans d'autres applications.
- Redimensionner les colonnes de variables.

Important ! Les données que vous avez importées dans le projet sont en lecture seule ; vous ne pouvez pas modifier ces données dans IBM® SPSS® Text Analytics for Surveys.

Vous pouvez ensuite commencer à extraire des concepts de ces réponses, avec lesquels vous allez créer vos catégories. [Pour plus d'informations, reportez-vous à la section Catégorisation des données textuelles dans le chapitre 6 sur p. 96.](#)

Figure 4-3
Accès à la vue *Projet Complet*



Trier les variables

Vous pouvez trier vos données dans la vue *Projet complet* par ordre alphabétique ou par longueur des données.

Pour trier les données dans la vue *Projet complet*

- ▶ Sélectionnez la colonne que vous voulez trier et cliquez avec le bouton droit de la souris sur le titre de la colonne pour ouvrir un menu contextuel.
- ▶ Sélectionnez l'option de tri désirée parmi les choix suivants :
 - **Tri naturel.** Tri des résultats tels qu'ils ont été lus au cours de l'import.
 - **Tri croissant : A–Z.** Tri alphabétique des résultats en commençant par les cellules vides, les nombres puis de *A* à *Z*.
 - **Tri décroissant : A–Z.** Tri alphabétique des résultats en commençant de *Z* à *A*, les nombres puis les cellules vides.
 - **Tri croissant : longueur.** Tri des résultats par longueur, les réponses les plus courtes apparaissant en premier.
 - **Tri décroissant : longueur.** Tri des résultats par longueur, les réponses les plus longues apparaissant en premier.

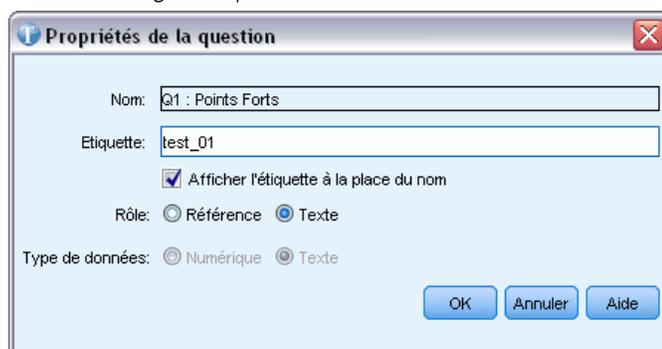
Modifier les propriétés des variables

Lors de la définition de vos données au cours de l'import, le système vous demande d'identifier la variable représentant les ID uniques, celles représentant les questions que vous souhaitez analyser et, le cas échéant, les variables de référence que vous voulez inclure. Une fois les données importées, vous voudrez peut-être ajouter des informations aux propriétés pour ces variables ou modifier leur rôle dans le projet. Par exemple, vous pouvez souhaiter analyser une variable

que vous avez importée comme variable de référence. Vous pouvez modifier les propriétés de variable suivantes :

- Ajouter ou modifier le nom ou l'étiquette d'une variable.
- Transformer une variable de référence en variable texte.
- Transformer une variable texte en variable de référence.
- Modifier le type de données d'une variable d'ID ou d'une variable de référence.

Figure 4-4
Boîte de dialogue Propriétés de référence



Pour modifier les propriétés de variable

- ▶ Dans la vue Projet Complet, sélectionnez la colonne de la variable dont vous voulez modifier les propriétés et cliquez avec le bouton droit de la souris sur le titre de la colonne pour ouvrir un menu contextuel.
- ▶ Sélectionnez Propriétés dans le menu. La boîte de dialogue Propriétés apparaît.
- ▶ Si vous le souhaitez, ajoutez un nom ou une étiquette de variable ou modifiez les.
- ▶ Pour utiliser les étiquettes de variable à la place du nom de la variable dans le produit, sélectionnez l'option Afficher l'étiquette à la place du nom.
- ▶ Si vous le souhaitez, transformez le rôle de la variable dans l'analyse en variable de référence ou variable texte. Vous ne pouvez pas modifier le rôle de la variable ID. Si vous avez commencé à travailler sur une variable (ou question) texte et transformer son rôle en variable de référence, le travail de catégorisation que vous avez effectué sera perdu.
- ▶ Transformez le type de données de la variable en Texte ou Numérique.

Enregistrer des projets

Quand vous fermez un projet ou mettez fin à la session, le système vous demande d'enregistrer les modifications si nécessaire. Les projets sont enregistrés dans des fichiers avec l'extension *.tas.

Pour enregistrer des projets

- ▶ Dans les menus, choisissez Fichier > Enregistrer le projet. Le projet est enregistré.

Pour enregistrer un projet en quittant

- ▶ Quand vous fermez un projet, une boîte de dialogue s'ouvre et vous demande si vous voulez enregistrer les modifications apportées au projet et si vous voulez (re)publier les bibliothèques.

Figure 4-5

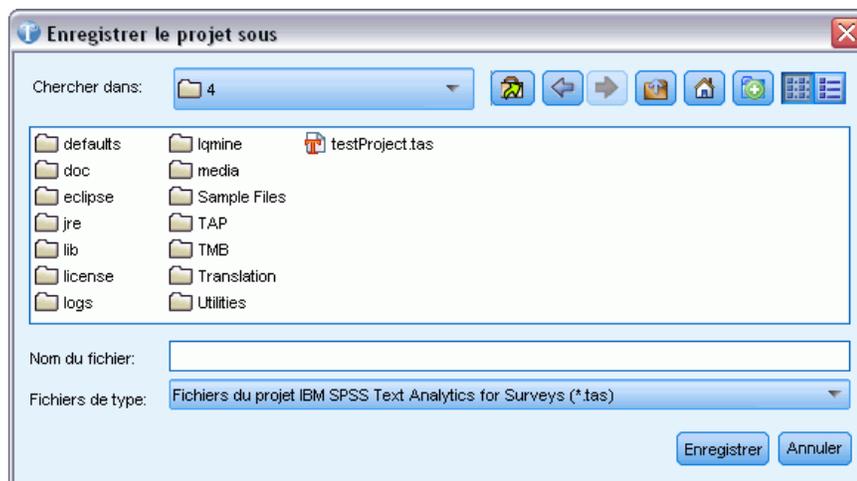
Boîte de dialogue Enregistrer le projet en cours

- ▶ Sélectionner Enregistrer les modifications du projet.
- ▶ Si vous voulez publier des bibliothèques pour une utilisation ultérieure ou pour mettre à jour des versions publiques, sélectionnez aussi Publier les bibliothèques. Si aucune bibliothèque n'a besoin d'être publiée, cette option est désactivée. [Pour plus d'informations, reportez-vous à la section Publication de bibliothèques dans le chapitre 9 sur p. 216.](#)
- ▶ Cliquez sur Oui pour enregistrer. Si vous avez choisi de publier les bibliothèques, une autre boîte de dialogue s'ouvre. [Pour plus d'informations, reportez-vous à la section Partage de bibliothèques dans le chapitre 9 sur p. 214.](#)

Pour enregistrer sous un autre nom de projet

Si vous avez reçu une alerte de nom en double ou si vous choisissez d'enregistrer le projet sous un nom différent, la boîte de dialogue Enregistrer le projet sous s'ouvre.

Figure 4-6

Boîte de dialogue Enregistrer le projet sous

- ▶ Saisissez le nouveau nom unique du projet dans la zone de texte Fichier.
- ▶ Cliquez sur Enregistrer pour enregistrer le nouveau nom.

Export des résultats de catégorisation

Dans certains cas, la création de catégories constitue le but de l'analyse. Le simple fait de prendre connaissance des grands thèmes exposés par les personnes interrogées et de savoir combien d'entre elles ont cité chaque thème peut correspondre aux objectifs de l'analyse de texte. Toutefois, vous serez souvent amené à générer des rapports ou à poursuivre l'analyse des catégories, en créant notamment des tableaux et des graphiques pour présenter les résultats. Vous pouvez même utiliser d'autres variables du questionnaire pour continuer à définir les personnes interrogées dans chaque catégorie, et utiliser les catégories pour l'étude d'autres réponses.

Si vous souhaitez continuer à travailler avec vos nouveaux résultats de catégorisation, vous pouvez exporter les catégories au format texte en vue de les importer dans une application d'analyse quantitative, telle que le système IBM® SPSS® Statistics Base. Le fichier obtenu contient les ID des réponses, ainsi que les noms et étiquettes des catégories. En revanche, il ne contient pas les valeurs des variables de référence, ni les réponses ouvertes.

Remarque : Vous pouvez aussi générer des graphiques récapitulatifs, comme un diagramme en bâton des 5 premières catégories. Ces graphiques, qui sont exportés en HTML, peuvent ensuite être utilisés dans des présentations. [Pour plus d'informations, reportez-vous à la section Exporter des graphiques récapitulatifs sur p. 60.](#)

Types de fichiers exportés

A l'export, vous pouvez créer l'un des types de fichier suivants :

- SPSS Statistics fichiers (*.sav). [Pour plus d'informations, reportez-vous à la section Export vers IBM SPSS Statistics ou IBM SPSS Data Collection sur p. 56.](#)
- fichiers Microsoft Excel (*.xls / *.xlsx). [Pour plus d'informations, reportez-vous à la section Export vers Microsoft Excel sur p. 58.](#)
- IBM® SPSS® Data Collection. [Pour plus d'informations, reportez-vous à la section Export vers IBM SPSS Statistics ou IBM SPSS Data Collection sur p. 56.](#) Reportez-vous également à Data Collection Development Library dans *Data Collection Modèle de données*.

Dichotomies et sortie de catégories

Les données textuelles codées avec IBM® SPSS® Text Analytics for Surveys constituent un ensemble de réponses multiples, dans la mesure où chaque personne interrogée peut donner plusieurs réponses et peut être affectée, pour une seule question, à plusieurs catégories. Ceci implique que les données doivent être codées sous un format spécial lors de l'export. Vous disposez de deux formats de sortie pour l'export : les **dichotomies** et les **catégories**.

Dichotomies. Les résultats se concentrent sur les booléens d'appartenance à une catégorie de chaque ID de réponse. Pour chaque catégorie de données, chaque personne interrogée (par ID) appartient ou n'appartient pas à la catégorie. Un booléen binaire signale cet état en renvoyant le code vrai ou faux. Les données sont structurées sous la forme d'un tableau, dont la colonne de gauche contient les ID et chaque colonne supplémentaire une catégorie. Ce type de données accepte un nombre illimité de catégories par réponse. S'il existe 10 catégories, 10 variables sont créées.

Catégories. Les résultats se concentrent sur l'ensemble de catégories auquel une réponse appartient. La catégorie à laquelle chaque réponse est affectée apparaît sous la forme d'une valeur (pour SPSS Statistics) ou de la catégorie elle-même (pour Microsoft Excel). Les données d'export de catégories sont structurées sous la forme d'un tableau, qui affiche les ID dans la colonne de gauche, puis une colonne par catégorie à laquelle au moins une réponse appartient. Ces colonnes ne représentent pas une catégorie en particulier, mais plutôt une propriété dans laquelle un code de catégorie affecté est enregistré. Le code de la catégorie à laquelle appartient chaque réponse est stocké dans une propriété distincte. La réponse à laquelle le nombre maximal de catégories a été attribué détermine le nombre de variables à créer. S'il y a 10 catégories, mais qu'aucune personne interrogée n'est associée aux codes de plus de 4 catégories, 4 variables s'avèrent alors nécessaires pour représenter les catégories.

- **Pour SPSS Statistics/Data Collection.** Chaque catégorie à laquelle un ID de réponse est affecté apparaît sous la forme d'une valeur distincte entre 1 et N , où N représente la valeur de code de catégorie la plus élevée. Si vous n'avez pas affecté de codes dans le Plan de codage, ils ont été affectés automatiquement lors de la création de la catégorie. Si une personne interrogée est affectée à un nombre de catégories inférieur au nombre maximal de catégories, les variables inutilisées restantes sont représentées par le code de valeur manquante du système SPSS Statistics (un point).
- **Pour Microsoft Excel.** Chaque catégorie à laquelle chaque ID de réponse des données est affecté apparaît sous la forme du nom ou de l'étiquette de la catégorie, en fonction de ce que vous utilisez dans l'interface du produit. Dans Microsoft Excel, si une personne interrogée est affectée à un nombre de catégories inférieur au nombre maximal de catégories, les variables inutilisées restantes sont codées par un espace.

Export vers IBM SPSS Statistics ou IBM SPSS Data Collection

Une fois les réponses classées en catégories, l'étape suivante consiste à analyser vos résultats au moyen de procédures statistiques. Dans IBM® SPSS® Text Analytics for Surveys, vous pouvez créer un fichier de données formaté pour l'utilisation de différents produits—les présentes instructions concernent les exports destinés à IBM® SPSS® Statistics (programme d'analyse statistique) et à divers produits IBM® SPSS® Data Collection. SPSS Text Analytics for Surveys crée automatiquement la variable à réponses multiples dans le fichier exporté. Le format exact du fichier dépend du type de données que vous sélectionnez, à savoir dichotomies ou catégories.

Remarque : Le fichier obtenu contient les ID des réponses, ainsi que les noms et étiquettes des catégories. En revanche, il ne contient pas les valeurs des variables de référence, ni les réponses ouvertes.

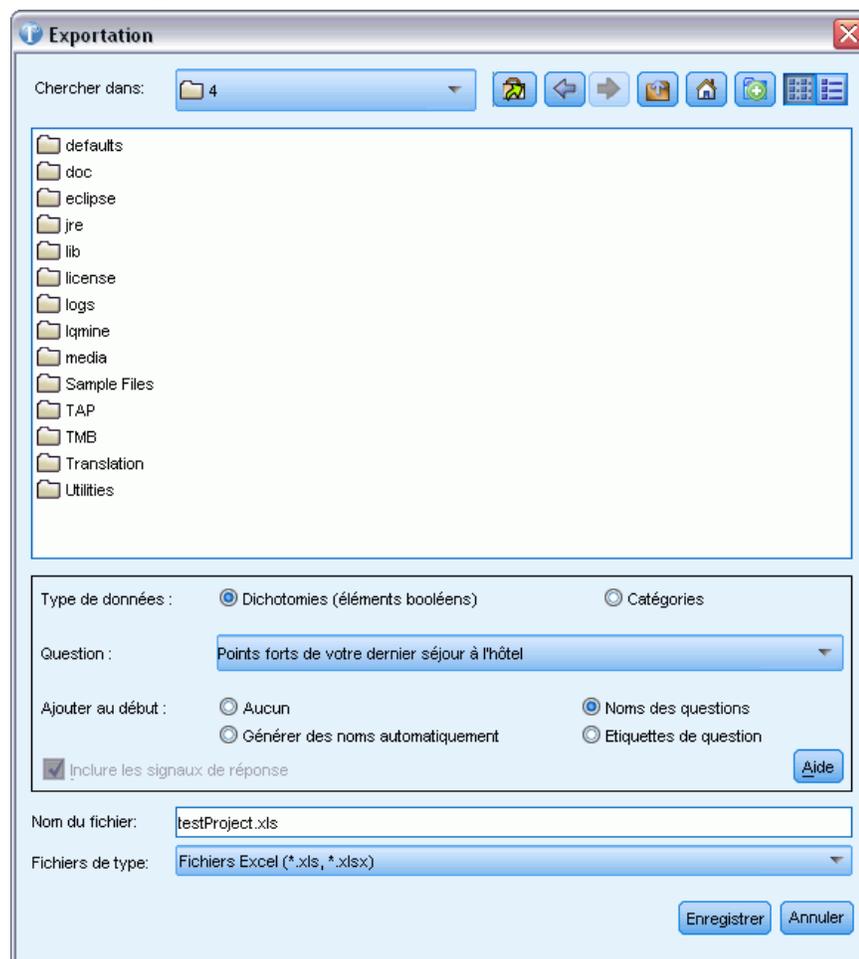
SPSS Statistics uniquement. Pour la sortie, si votre ensemble de données contient des données manquantes ou des observations dans lesquelles une personne interrogée n'a pas répondu à une question particulière, l'application attribue la valeur manquante système à ces observations. Les fichiers SPSS Statistics exportés par SPSS Text Analytics for Surveys ne sont pas pris en charge par les versions de SPSS Statistics antérieures à 7.5.

Pour exporter des données

- ▶ Dans le menu Fichier > Exporter les résultats, choisissez l'une des options suivantes pour ouvrir la boîte de dialogue Exporter :
 - SPSS Statistics Fichier
 - Data Collection Fichier

Figure 4-7

Boîte de dialogue Exporter pour les formats de fichier IBM SPSS Statistics



- ▶ Dans la liste déroulante Enregistrer dans, sélectionnez le pilote et le dossier dans lesquels vous voulez enregistrer le fichier.
- ▶ Sélectionnez une option sous Type de données. [Pour plus d'informations, reportez-vous à la section Export des résultats de catégorisation sur p. 55.](#)
 - Dichotomies.
 - Catégories. Cette option n'est pas disponible pour le fichier de données Data Collection. L'option Dichotomies est sélectionnée par défaut.

- ▶ Dans la liste déroulante Question, sélectionnez la question que vous souhaitez exporter. Vous pouvez choisir d'exporter les résultats de catégorisation d'une seule question ou de l'ensemble du projet. Si vous souhaitez exporter les questions séparément, sélectionnez-les et exportez-les une par une. Vous pouvez également sélectionner Projet complet pour exporter les résultats de toutes les questions ouvertes.
- ▶ Sélectionnez une option d'ajout au début pour choisir un préfixe lorsque vous exportez les noms des catégories du *projet complet*. Cette option s'avère des plus utiles lors de l'export des données de questions multiples. L'option d'ajout au début place un préfixe devant le nom de la catégorie ou de l'étiquette de variable d'origine. Vous êtes ainsi sûr de ne pas avoir de doublons lors de la combinaison des résultats de questions multiples à l'export du projet complet. Choisissez l'une des options suivantes :
 - Aucun. Comme le nom de l'option l'indique, aucun préfixe n'est ajouté.
 - Noms des questions. Préfixe ajouté aux noms de variables de catégorie renvoyés (le nom ou l'étiquette de la catégorie, en fonction de ce qui est utilisé dans le projet) le nom de variable du texte Le nom de variable de la question provient de la source de données d'origine. Si le nom de variable de catégorie renvoyé ne respecte pas les conventions de dénomination des variables ou dépasse 40 caractères, un nom par défaut est créé par l'option Générer automatiquement.
 - Nommer automatiquement. Ajoute automatiquement les préfixes *Q1*, *Q2*, *Q3*, et ainsi de suite, aux noms de catégorie. *Q1* désigne la première question que vous exportez, et ainsi de suite.
 - Etiquettes de question. Préfixe ajouté aux noms de variables de catégorie renvoyés (le nom ou l'étiquette de la catégorie, en fonction de ce qui est utilisé dans le projet) l'étiquette de variable du texte (question). Le nom de variable de la question provient de la source de données d'origine. Si le nom de variable de catégorie renvoyé ne respecte pas les conventions de dénomination des variables ou dépasse 40 caractères, un nom par défaut est créé par l'option Générer automatiquement.
- ▶ Si vos données contiennent des booléens de réponse, vous pouvez également les exporter. Pour exporter les booléens de réponse, sélectionnez cette option. [Pour plus d'informations, reportez-vous à la section Marquer des réponses sur p. 76.](#)
- ▶ Dans la zone de texte Fichier, sélectionnez le nom de projet qui apparaît par défaut ou entrez un autre nom pour ce fichier.
- ▶ Cliquez sur Enregistrer pour exporter les résultats.

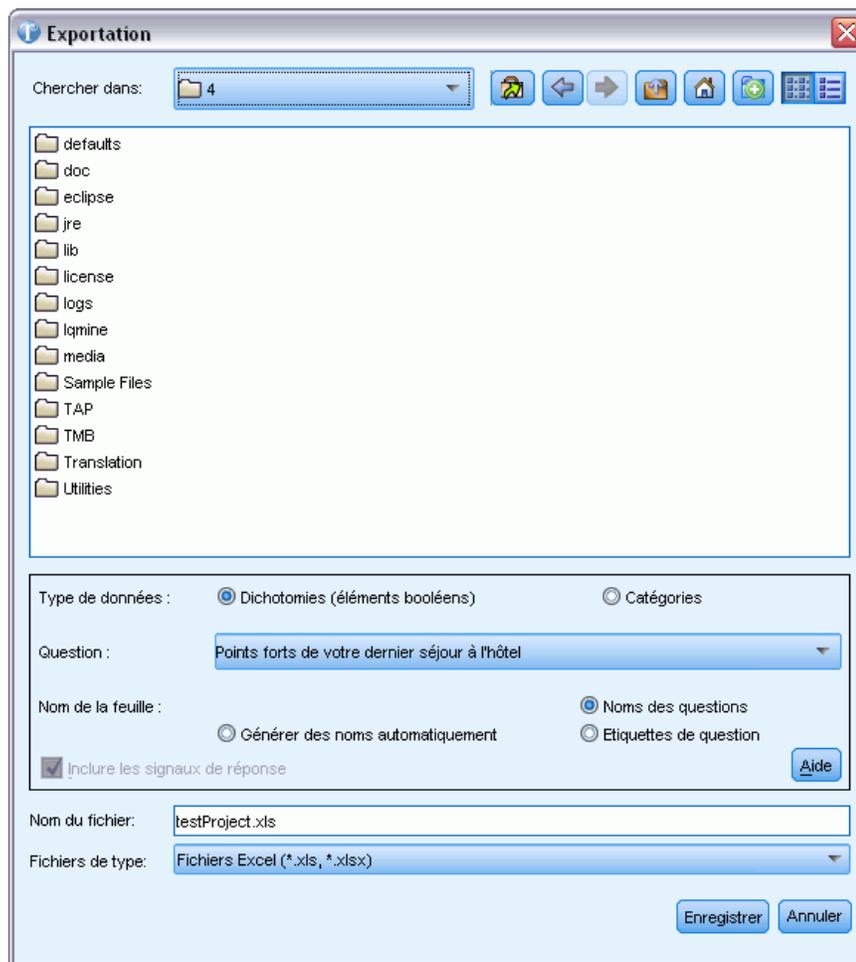
Export vers Microsoft Excel

Une fois les réponses classées en catégories, l'étape suivante consiste à analyser les résultats au moyen de procédures statistiques. Dans IBM® SPSS® Text Analytics for Surveys, vous pouvez créer un fichier de données dans un format compatible avec différents produits. Les instructions suivantes concernent l'export au format Microsoft Excel. SPSS Text Analytics for Surveys crée automatiquement la variable à réponses multiples dans votre fichier exporté. Le format exact du fichier dépend du type de données que vous sélectionnez, à savoir dichotomies ou catégories. Le fichier obtenu contient les ID des réponses, ainsi que les noms et étiquettes des catégories. En revanche, il ne contient pas les valeurs des variables de référence, ni les réponses ouvertes.

Pour exporter des données

- ▶ Dans les menus, choisissez Fichier > Résultats d'exportation > Fichier Microsoft Excel. La boîte de dialogue Exporter apparaît.

Figure 4-8
Boîte de dialogue Exporter pour les fichiers Microsoft Excel



- ▶ Dans la liste déroulante Enregistrer dans, sélectionnez le pilote et le dossier dans lesquels vous voulez enregistrer le fichier.
- ▶ Sélectionnez une option sous Type de données. [Pour plus d'informations, reportez-vous à la section Export des résultats de catégorisation sur p. 55.](#)
 - Dichotomies.
 - Catégories. Cette option n'est pas disponible pour le fichier de données IBM® SPSS® Data Collection. L'option Dichotomies est sélectionnée par défaut.
- ▶ Dans la liste déroulante Question, sélectionnez la question que vous souhaitez exporter. Vous pouvez choisir d'exporter les résultats de catégorisation d'une seule question ou de l'ensemble du projet. Si vous souhaitez exporter les questions séparément, sélectionnez-les et exportez-les

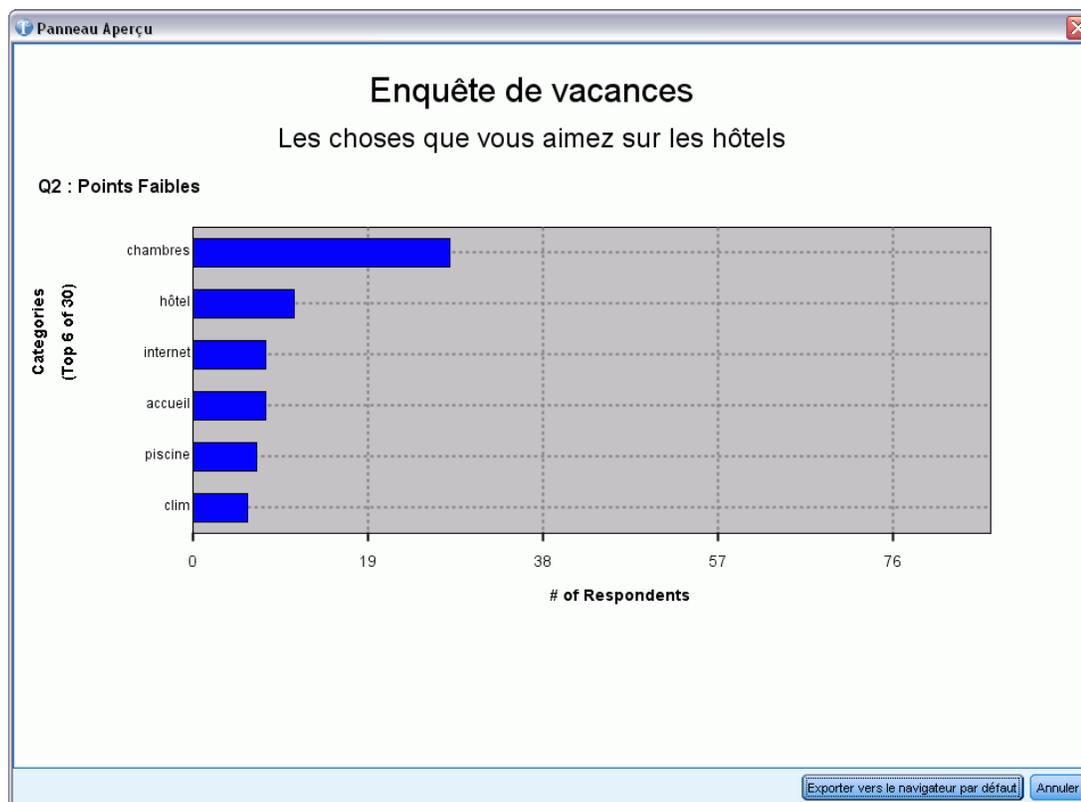
une par une. Vous pouvez également sélectionner **Projet complet** pour exporter les résultats de toutes les questions ouvertes.

- ▶ Sélectionnez une option de dénomination de la feuille de calcul pour déterminer la manière dont chaque feuille générée dans le fichier Microsoft Excel exporté doit être nommée. Choisissez l'une des options suivantes :
 - **Noms des questions.** Utilisez le nom de la variable de texte (question) pour la feuille. Le nom de la question est tiré de la source de données d'origine. Si le nom de variable de catégorie renvoyé ne respecte pas les conventions de dénomination des variables ou dépasse 40 caractères, un nom par défaut est créé par l'option **Générer automatiquement**.
 - **Nommer automatiquement.** Nomme automatiquement les feuilles de calcul *Q1*, *Q2*, *Q3*, et ainsi de suite. *Q1* désigne la première question que vous exportez, et ainsi de suite.
 - **Étiquettes de question.** Utilisez l'étiquette de la variable de texte (question), s'il en existe une, pour nommer la feuille. Si le nom de variable de catégorie renvoyé ne respecte pas les conventions de dénomination des variables ou dépasse 40 caractères, un nom par défaut est créé par l'option **Générer automatiquement**.
- ▶ Si vos données contiennent des booléens de réponse, vous pouvez également les exporter. Pour exporter les booléens de réponse, sélectionnez cette option. [Pour plus d'informations, reportez-vous à la section Marquer des réponses sur p. 76.](#)
- ▶ Dans la zone de texte **Fichier**, sélectionnez le nom de projet qui apparaît par défaut ou entrez un autre nom pour ce fichier.

Exporter des graphiques récapitulatifs

Lorsque vous avez terminé de travailler avec vos catégories et données, vous pouvez exporter des rapports de résumés graphiques afin de partager vos résultats d'analyse avec d'autres. La sortie produit un diagramme en bâtons par question. Vous pouvez choisir le nombre de premières catégories à utiliser dans chaque graphique afin de pouvoir présenter visuellement les 5 ou 10 premières catégories pour une question donnée. Le graphique peut être exporté dans votre navigateur par défaut à partir duquel vous pouvez enregistrer l'image à utiliser dans d'autres produits ou présentations.

Figure 4-9
Graphiques récapitulatifs d'échantillons dans une fenêtre de navigateur



Pour exporter les graphiques récapitulatifs

- ▶ Dans les menus, choisissez Catégories > Exporter des graphiques récapitulatifs. La boîte de dialogue Exporter des graphiques récapitulatifs apparaît.
- ▶ Configurez votre graphique récapitulatif à l'aide des options décrites dans cette rubrique.
- ▶ Cliquez sur Créer pour générer le graphique et l'afficher dans un panneau Aperçu.
- ▶ Cliquez sur Exporter vers le navigateur par défaut pour afficher le graphique dans une fenêtre de navigateur.

Figure 4-10
Boîte de dialogue Exporter des graphiques récapitulatifs

Options de rapport

Exporter le(s) graphique(s) de. Choisissez si vous voulez générer un graphique récapitulatif pour toutes les questions dans votre projet ou seulement pour la question actuellement sélectionnée.

Nombre de premières catégories à inclure. Sélectionnez le nombre maximum de catégories à afficher dans le graphique. Les catégories ayant le plus grand nombre d'enregistrements sont utilisées en premier.

Recommencez l'extraction si des résultats d'extraction obsolètes sont trouvés. Sélectionnez cette option pour forcer une nouvelle extraction avant de générer le graphique si les résultats de l'extraction ne sont pas à jour.

Titre principal. Entrez un titre principal pour vos graphiques. Par exemple, il peut s'agir du nom de votre enquête.

Sous-titre. Entrez un sous-titre pour vos graphiques. Par exemple, il peut s'agir du nom de l'entreprise ou de l'année de l'enquête.

Titre de question. Pour vous aider à identifier chaque graphique, le titre est dérivé de la question. Choisissez d'utiliser le nom de variable de la question, l'étiquette ou aucun nom.

Étiquette de l'axe X. Définissez une étiquette pour l'axe X des graphiques. Une étiquette est proposée par défaut.

Étiquette de l'axe Y. Définissez une étiquette pour l'axe Y des graphiques. Une étiquette est proposée par défaut.

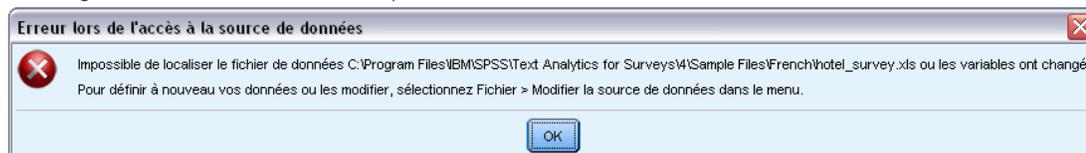
Couleur de barre. Choisissez une couleur pour les barres dans le graphique récapitulatif. Cette couleur s'applique à toutes les questions.

Police. Choisissez une police pour les titres et les étiquettes dans le graphique.

Modifier les sources de données

Lorsque vous ouvrez un projet, l'ensemble de données correspondant est ouvert. Si ces données sont introuvables, un message d'erreur apparaît. Parfois les données sont introuvables car elles ont été déplacées à un autre endroit, accidentellement supprimées ou renommées. Vous pouvez aussi vouloir changer de sources de données.

Figure 4-11
Message d'erreur de données manquantes

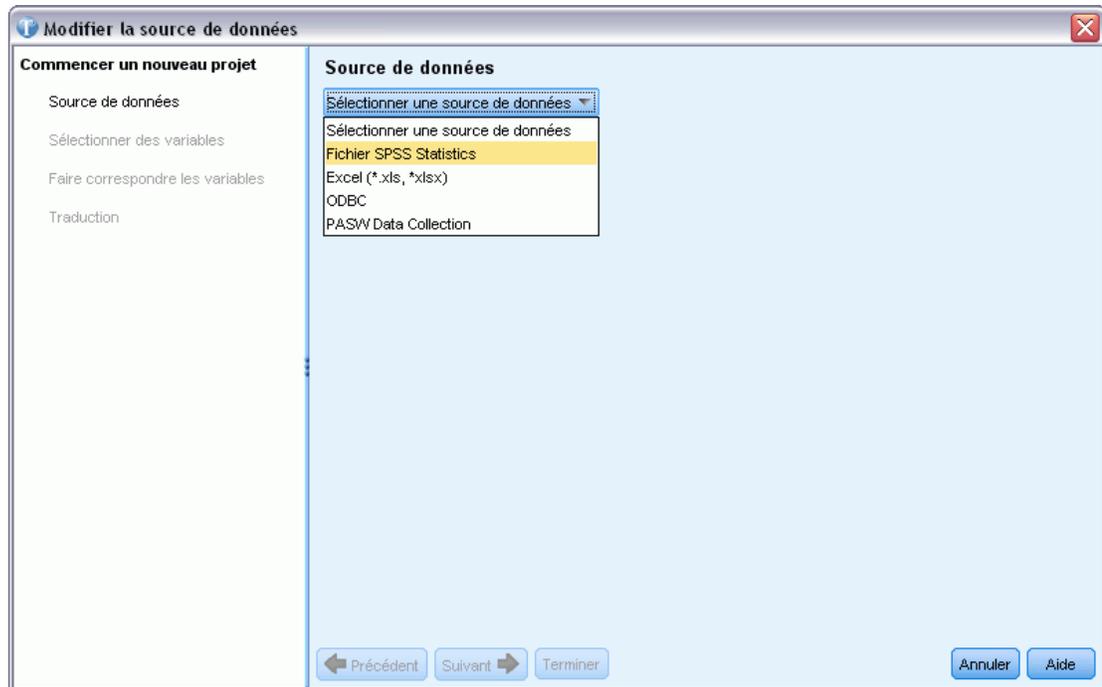


Pour continuer à travailler avec vos données, vous devez modifier l'emplacement vers la source de données réelle. Si des modifications concernant les variables sont trouvées dans les données, comme de nouvelles variables, des variables renommées ou manquantes, vous devrez faire correspondre les variables précédemment importées aux nouvelles.

Pour modifier votre source de données :

- ▶ Quand vous recevez ce message d'erreur, cliquez sur OK.
- ▶ Dans les menus, sélectionnez Fichier > Modifier la source de données. La boîte de dialogue de l'assistant Modifier la source de données apparaît.

Figure 4-12
Assistant Modifier la source de données



Sélection de sources de données

Lorsque l'assistant s'ouvre, vous commencez par sélectionner une source de données. IBM® SPSS® Text Analytics for Surveys a été optimisé pour traiter des ensembles de données allant jusqu'à 10 000 enregistrements, bien que la performance dépende du volume de texte contenu dans ces enregistrements. Reportez-vous aux instructions d'installation pour obtenir des statistiques et des recommandations sur les performances.

Important ! Une variable ID avec une valeur unique pour chaque enregistrement doit être présente afin d'importer les données.

Vous pouvez choisir l'une des sources de données suivantes :

- **SPSS Statistics fichiers** (*.sav).
- **fichiers Microsoft Excel** (*.xls / *.xlsx).
- Base de données **ODBC** (protocole Open Database Connectivity de Microsoft)
- **Data Collection** modèle de données. Cette option n'est disponible que si le modèle de données est installé.

Utilisation de fichiers IBM SPSS Statistics

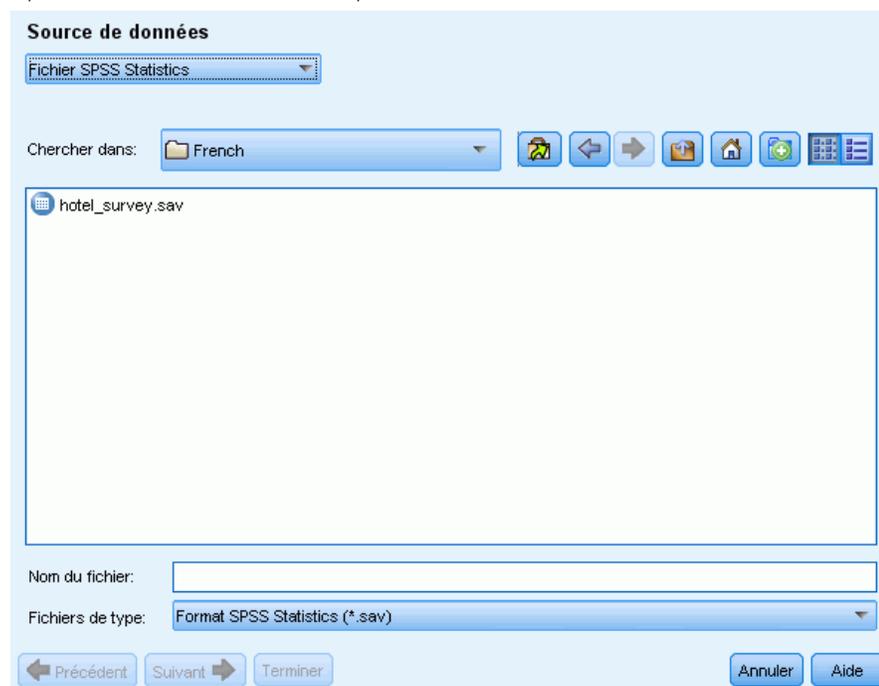
Vous pouvez importer un fichier IBM® SPSS® Statistics (.sav) dans IBM® SPSS® Text Analytics for Surveys. Une variable ID avec une valeur unique pour chaque enregistrement doit être présente afin d'importer les données.

Important ! Vous ne pouvez pas importer un fichier SPSS Statistics (.sav) avec des enregistrements d'une longueur supérieure à 4000 caractères.

Remarque : SPSS Text Analytics for Surveys a été optimisé pour traiter des ensembles de données allant jusqu'à 10 000 enregistrements, bien que la performance dépende du volume de texte contenu dans ces enregistrements. Reportez-vous aux instructions d'installation pour obtenir des statistiques et des recommandations sur les performances.

Figure 4-13

Options de sources de données pour les fichiers IBM SPSS Statistics



Obtenir les données de IBM SPSS Statistics

- ▶ Dans le premier écran de l'assistant, sélectionnez un fichier SPSS Statistics dans la liste déroulante. L'assistant affiche les options disponibles pour les fichiers SPSS Statistics.
- ▶ Dans la liste déroulante Rechercher dans, sélectionnez le pilote et le dossier dans lesquels se trouve le fichier.
- ▶ Sélectionnez le fichier dans la liste. Il apparaîtra dans la zone de texte Nom du fichier.
- ▶ Cliquez sur Suivant pour sélectionner des variables. [Pour plus d'informations, reportez-vous à la section Sélection de variables sur p. 35.](#)

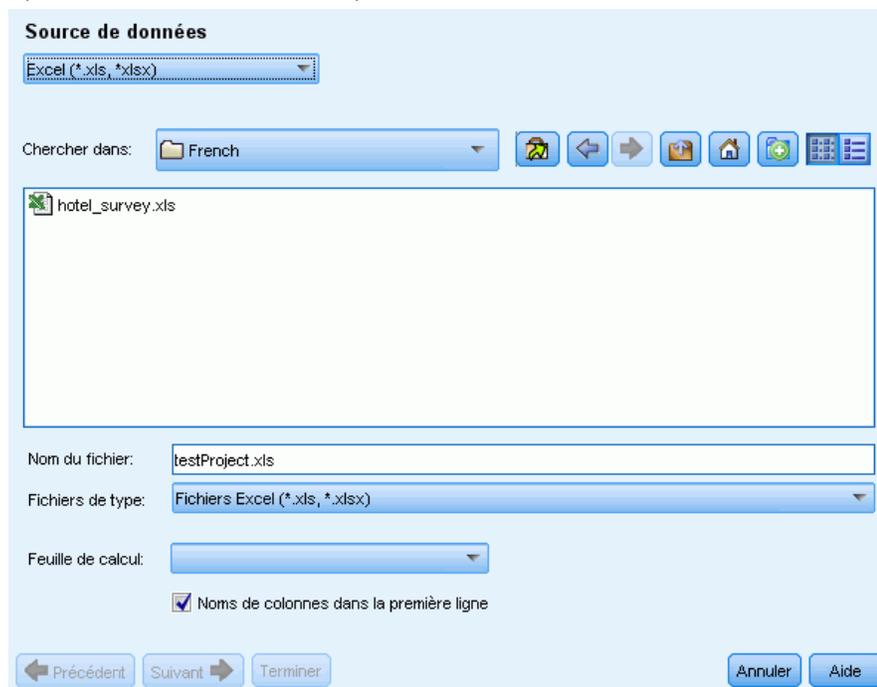
Utilisation de fichiers Microsoft Excel

Vous pouvez importer un fichier Microsoft Excel (.xls / .xlsx) dans IBM® SPSS® Text Analytics for Surveys. Une variable ID avec une valeur unique pour chaque enregistrement doit être présente afin d'importer les données.

Important ! Au cours de l'import du fichier Microsoft Excel, vous pouvez sélectionner une option pour afficher les Noms de colonne dans la première ligne. Pour utiliser cette option, la toute première ligne de la feuille de calcul doit contenir des noms de colonne, et pas la ligne juste dessous où commencent les données. Par exemple, si vos données et noms de colonne commencent à la ligne 7, vous devez supprimer les lignes 1 à 6 avant de commencer l'import, afin d'importer le fichier correctement.

Remarque : SPSS Text Analytics for Surveys a été optimisé pour traiter des ensembles de données allant jusqu'à 10 000 enregistrements, bien que la performance dépende du volume de texte contenu dans ces enregistrements. Reportez-vous aux instructions d'installation pour obtenir des statistiques et des recommandations sur les performances.

Figure 4-14
Options de sources de données pour les fichiers Microsoft Excel



Obtenir les données de Microsoft Excel

- ▶ Dans le premier écran de l'assistant, sélectionnez Excel dans la liste déroulante. L'assistant affiche les options disponibles pour les fichiers Microsoft Excel.
- ▶ Dans la liste déroulante Rechercher dans, sélectionnez le pilote et le dossier dans lesquels se trouve le fichier.
- ▶ Sélectionnez le fichier dans la liste. Il apparaîtra dans la zone de texte Nom du fichier.

- ▶ Sélectionnez la feuille de calcul dans la liste déroulante. Vous ne pouvez importer des données que d'une seule feuille de calcul. Pour travailler avec les données de plusieurs feuilles de calcul, vous devez créer plusieurs projets.
- ▶ Si la première ligne de cette feuille de calcul contient les titres de colonne, sélectionnez Nom de colonne dans la première ligne. Pour utiliser cette option, la toute première ligne de la feuille de calcul doit contenir des noms de colonne, et pas la ligne juste dessous où commencent les données. Par exemple, si vos données et noms de colonne commencent à la ligne 7, vous devez supprimer les lignes 1 à 6 avant de commencer l'import, afin d'importer le fichier correctement. L'application peut les utiliser comme noms de variable (ou une version convertie si les titres de colonne ne sont pas conformes aux conventions de dénomination des variables IBM® SPSS® Statistics). Sinon, l'application utilisera les lettres de colonne de la feuille de calcul comme identificateurs.
- ▶ Cliquez sur Suivant pour sélectionner des variables. [Pour plus d'informations, reportez-vous à la section Sélection de variables sur p. 35.](#)

Utilisation de données via ODBC

Les données issues de sources de base de données, généralement des bases de données, peuvent facilement être importées dans IBM® SPSS® Text Analytics for Surveys. Toute base de données utilisant des pilotes Open Database Connectivity (ODBC) peut être lue directement par le produit, une fois ces pilotes installés sur l'ordinateur sur lequel SPSS Text Analytics for Surveys est installé. Une variable ID avec une valeur unique pour chaque enregistrement doit être présente afin d'importer les données.

Remarque : SPSS Text Analytics for Surveys a été optimisé pour traiter des ensembles de données allant jusqu'à 10 000 enregistrements, bien que la performance dépende du volume de texte contenu dans ces enregistrements. Reportez-vous aux instructions d'installation pour obtenir des statistiques et des recommandations sur les performances.

Figure 4-15
Options de source de données pour ODBC

Source de données

ODBC

Source(DSN)

Nom	Description
MS Access Database	Microsoft Access Driver (*.mdb)
Excel Files	Microsoft Excel Driver (*.xls)
dBASE Files	Microsoft dBase Driver (*.dbf)
Visio Database Samples	Microsoft Access Driver (*.mdb)

ODBC

Utilisateur:

Mot de passe:

Table

SQL

← Précédent Suivant → Terminer Annuler Aide

Utiliser via ODBC

- ▶ Dans le premier écran de l'assistant, sélectionnez ODBC dans la liste déroulante. L'assistant affiche les options disponibles pour ODBC.
- ▶ Spécifiez la source de données en la sélectionnant dans la liste de sources ODBC enregistrées ou en saisissant le nom dans la zone de texte Source (DSN). Si vous avez besoin d'enregistrer de nouvelles sources de données qui n'apparaissent pas dans la liste, cliquez sur ODBC. Cela permet d'ouvrir l'administrateur de source de données ODBC, présent sur la plupart des ordinateurs Microsoft Windows. S'il est introuvable, vous ne pouvez pas utiliser l'import ODBC. Consultez le système d'Aide de Microsoft Windows pour plus d'informations.
- ▶ Si la source de données est protégée par un mot de passe, vous devez saisir un nom d'utilisateur et un mot de passe. Vous devrez faire de même à chaque fois que vous ouvrirez un projet car pour des raisons de sécurité, le nom d'utilisateur et le mot de passe ne sont pas stockés dans le projet.
- ▶ Sélectionnez vos données de l'une des deux manières suivantes : Table ou SQL. Vous pouvez sélectionner une table directement ou utiliser les commandes SQL pour sélectionner des données.
- ▶ Cliquez sur Suivant pour sélectionner des variables. [Pour plus d'informations, reportez-vous à la section Sélection de variables sur p. 35.](#)

Utilisation des données IBM SPSS Data Collection

Importer via IBM SPSS Data Collection

- ▶ Dans le premier écran de l'assistant, sélectionnez Data Collection dans la liste déroulante. L'option de modèle de données IBM® SPSS® Data Collection n'est disponible que si le modèle de données est installé avec un autre produit.

Sélection de variables

Après la sélection de la source de données, l'étape suivante consiste à spécifier les variables à importer. Trois types de variables peuvent être importés dans un projet.

Variable ID unique (Requise)

La variable ID est une clé numérique ou chaîne qui identifie chaque personne interrogée. Le fichier de données n'a pas besoin d'être ordonné par la variable ID unique pour la lire correctement. Une fois lus dans le programme, les enregistrements peuvent être triés par différents critères. [Pour plus d'informations, reportez-vous à la section Trier les variables sur p. 52.](#) Cette variable ID est requise pour l'import des données. Chaque enregistrement importé (ou observation) doit avoir une valeur ID unique.

Deux situations peuvent faire échouer l'import :

- Valeurs ID en double détectées
- Enregistrements contenant des valeurs ID non renseignées

Remarque : Si un ID en double est détecté et que IBM® SPSS® Statistics est installé sur votre ordinateur, vous pouvez utiliser la procédure Identifier les observations dupliquées de ce produit pour identifier les doublons, puis utiliser les options pour indiquer quels enregistrements doivent être conservés (observations principales).

Variable(s) de texte (Requises)

Les variables de texte représentent les réponses textuelles aux questions de l'enquête. Au moins une de ces variables est requise pour importer des données. Ces variables peuvent être des variables chaîne ou des variables chaîne longues dans SPSS Statistics, des colonnes contenant des cellules de contenu général ou de texte d'Microsoft Excel, ou des champs de texte ou de notes de bases de données. Chaque variable de texte sera analysée séparément. La taille (largeur) de chaque variable texte importée d'un fichier .SAV est limitée à 4 000 caractères.

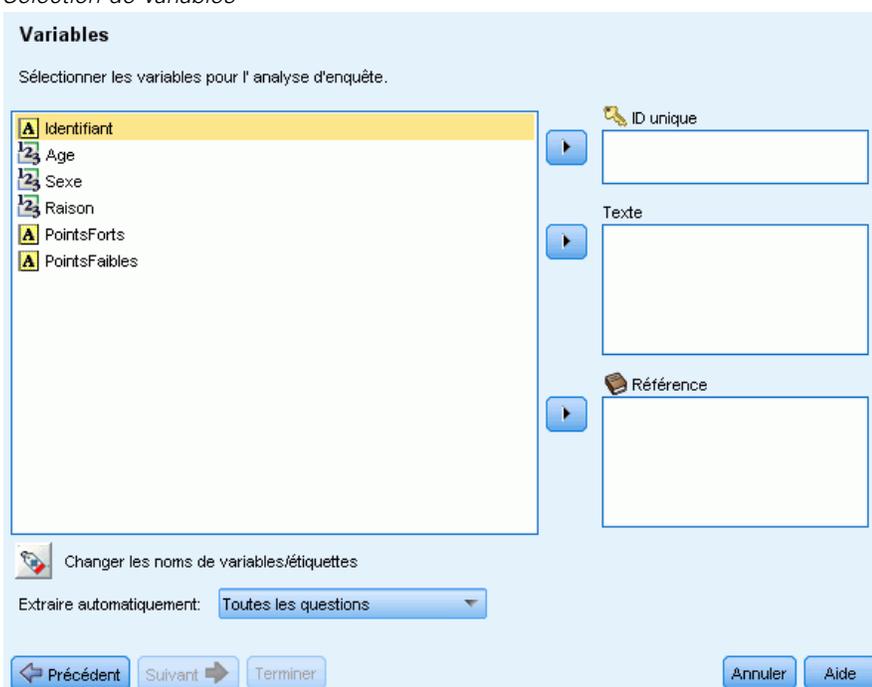
Variable(s) de référence (Optionnelles)

Les variables de référence sont des variables supplémentaires, optionnelles, généralement catégorielles, pouvant être importées à des fins de référence. Les variables de référence ne sont pas utilisées dans l'analyse de texte mais fournissent des informations supplémentaires décrivant la personne interrogée, qui peuvent aider à la compréhension et à l'interprétation. Les variables démographiques sont souvent incluses comme variables de référence, car elle peuvent aider à

comprendre quels termes ou catégories sont utilisés par quels groupes d'individus. Il s'agit par exemple du sexe, du département, de la profession et du niveau d'études (pour les évaluations des étudiants et d'apprentissage). Vous pouvez afficher l'ensemble des variables de référence après leur import, dans la vue Projet complet. Vous pouvez également afficher les variables de référence dans le panneau Données de la vue Question. De plus, vous pouvez sélectionner des variables de référence dans le diagramme à barres du panneau de visualisation pour pouvoir passer à un sous-ensemble de personnes interrogées.

Remarque : Les variables de référence lues à partir d'un fichier de données SPSS Statistics auront des étiquettes de variable (le cas échéant) qui apparaîtront en tant que titres de colonne, tandis que leurs étiquettes de valeur (le cas échéant) s'afficheront dans les cellules du panneau Données.

Figure 4-16
Sélection de variables



Pour sélectionner des variables et des options d'extraction

- ▶ Dans la liste des variables disponibles, sélectionnez la variable qui correspond à la variable ID de votre ensemble de données et cliquez sur la flèche pour la déplacer vers la zone ID unique. L'ID doit être une chaîne numérique ou alphanumérique unique qui distingue un enregistrement d'un autre. Si votre ensemble de données contient des ID en double, un message d'erreur apparaît. Dans ce cas, vous devez d'abord nettoyer vos données avant de réessayer.
- ▶ Dans la liste des variables disponibles, sélectionnez une ou plusieurs variables qui correspondent aux variables de réponse ouverte et cliquez sur la flèche pour déplacer les variables vers la liste Texte. Chacune des variables sera importée en tant que question distincte dont vous analyserez et catégoriserez les réponses.

- ▶ Dans la liste des variables disponibles, sélectionnez une ou plusieurs variables qui correspondent aux variables de référence et cliquez sur la flèche pour déplacer les variables vers la liste Référence. Les variables de référence ne sont pas utilisées par les techniques de création de catégorie automatisées. Vous pouvez toutefois afficher leur contenu et les utiliser pour vous aider à prendre des décisions judicieuses lors de la catégorisation de vos réponses.
- ▶ Pour afficher les étiquettes de variable au lieu des noms de variable, cliquez sur le bouton sous la liste des variables à gauche.
- ▶ Pour modifier le paramètre d'extraction, faites une sélection dans la liste déroulante. Par défaut, l'option Première question uniquement est sélectionnée, ce qui signifie que si plus d'une variable de texte est sélectionnée, le processus d'extraction démarre automatiquement pour la première question dès que l'assistant prend fin. L'extraction peut prendre du temps avec les ensembles de données plus importants. Vous pouvez par conséquent choisir d'extraire Aucune ou Toutes les questions en fonction du temps disponible.
- ▶ Cliquez sur Suivant> lorsque vous avez sélectionné toutes les variables.

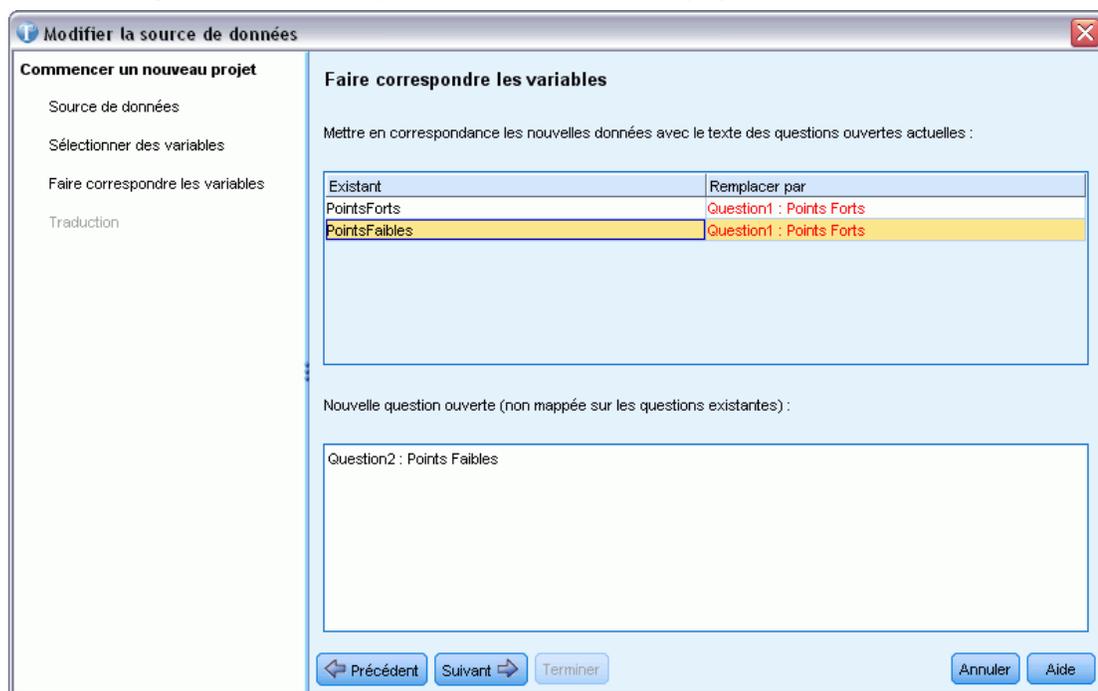
Correspondance de variables

Après avoir sélectionné des variables à l'étape précédente, IBM® SPSS® Text Analytics for Surveys essaie de mettre en correspondance les variables précédemment importées avec celles que vous venez de sélectionner. Des correspondances sont automatiquement proposées mais vous pouvez associer les variables différemment en cliquant dans la colonne Remplacer par et en choisissant une autre variable.

Si le nouveau fichier de données ne contient pas de variable texte qui existait déjà dans le projet, vous pouvez sélectionner AUCUN dans la liste, et toute donnée associée à l'ancienne question sera éliminée du projet.

Toutes les variables du nouveau fichier de données qui ne correspondent pas avec le projet existant s'affichent dans la liste Nouvelle question ouverte au bas de la boîte de dialogue. Une fois les ensembles de données modifiés, ces variables restantes apparaissent comme nouvelles questions dans votre projet.

Figure 4-17
Boîte de dialogue Mettre en correspondance le contenu avec le projet existant



Pour mettre en correspondance des variables à des variables existantes :

- ▶ Bien que le produit essaie d'associer vos nouvelles variables à celles précédemment trouvées dans le fichier de données, vous pouvez modifier la manière dont les variables sont associées en cliquant dans la colonne Remplacer par et en sélectionnant la variable correspondante. Si vous avez sélectionné plus de variables à l'étape précédente que le nombre spécifié dans la colonne Remplacer par, celles-ci apparaissent dans la liste Nouvelles questions ouvertes (qui ne sont pas mise en correspondance avec des questions existantes). Une fois les ensembles de données modifiés, ces variables restantes apparaissent comme nouvelles questions dans votre projet.
- ▶ Si la variable existante n'a pas de correspondance dans le nouvel ensemble de données, sélectionnez AUCUN et les données de cette question seront éliminées.
- ▶ Si votre projet contient des réponses qui ont été forcées à entrer ou à sortir de catégories ou de booléens, le système vous demande de garder ou de jeter ces résultats d'ID de réponse spécifiques. Généralement, si vous importez différentes données (nouvelles questions ou différentes personnes interrogées par exemple), vous voudrez ignorer ces informations de manière à ne pas produire de faux résultats. Si vous importez le même fichier de données, vous souhaitez généralement garder ces informations car les ID de réponse correspondent aux anciennes données.

Traduction vers l'anglais

Si vous travaillez avec un texte source non anglais, vous pouvez vous connecter à Language Weaver pour le traduire en anglais. La traduction n'est disponible qu'en anglais. Vous devez avoir configuré Language Weaver de manière appropriée et avec les connexions définies pour réaliser

la traduction. [Pour plus d'informations, reportez-vous à la section Options : Onglet Traduction dans le chapitre 2 sur p. 23.](#)

Figure 4-18
Options de traduction



Pour traduire en anglais

- Pour traduire en anglais les données textuelles à partir d'une langue bénéficiant d'une licence, cochez la case Traduire en anglais.
- Dans la liste connexion des paires de langues, sélectionnez la connexion pour la paire de langues Language Weaver que vous souhaitez utiliser. Si Language Weaver est configuré sur votre ordinateur local, ces paires de langues apparaîtront automatiquement dans cette liste. Vous pouvez ajouter, modifier ou tester la connexion des services en ligne dans l'onglet Traduction de la boîte de dialogue Options. [Pour plus d'informations, reportez-vous à la section Options : Onglet Traduction dans le chapitre 2 sur p. 23.](#)
- Spécifiez la Précision de la traduction de votre choix. Choisissez une valeur de 1 à 3 indiquant le rapport entre le niveau de vitesse et de précision souhaité. Une valeur faible permet d'obtenir des résultats de traduction plus rapides mais moins précis. Une valeur élevée permet, quant à elle, d'obtenir des résultats plus précis mais après une durée de traitement plus longue. Pour optimiser le temps de traitement, nous vous recommandons de commencer par un niveau peu élevé et de l'augmenter seulement si, après avoir examiné les résultats, vous pensez avoir besoin d'une plus grande précision.
- Si vous aviez précédemment créé des dictionnaires personnalisés, détenus par Language Weaver, vous pouvez les utiliser dans le cadre de la traduction. Pour choisir un dictionnaire personnalisé,

cochez la case Utiliser un dictionnaire personnalisé et entrez le Nom du dictionnaire. Pour utiliser plusieurs dictionnaires, séparez les noms avec une virgule.

- ▶ Dans l'assistant Nouveau projet, cliquez sur Suivant > pour commencer à sélectionner des catégories et des ressources. [Pour plus d'informations, reportez-vous à la section Sélection de catégories et de ressources dans le chapitre 3 sur p. 39.](#)
- ▶ Dans l'assistant Modifier l'ensemble de données, cliquez sur Terminer pour terminer la modification de l'ensemble de données et pour démarrer le processus de traduction.

Pour ignorer la traduction :

- ▶ Désélectionnez l'option Traduire en anglais.
- ▶ Dans l'assistant Nouveau projet, cliquez sur Suivant > pour commencer à sélectionner des catégories et des ressources. [Pour plus d'informations, reportez-vous à la section Sélection de catégories et de ressources dans le chapitre 3 sur p. 39.](#)
- ▶ Dans l'assistant Modifier l'ensemble de données, cliquez sur Terminer pour terminer la modification de l'ensemble de données.

Mise à jour des données

Lorsque vous travaillez avec vos données de projet, vous pouvez modifier la source de données d'origine. Par exemple, vous pouvez ajouter ou supprimer des enregistrements. Vous pouvez mettre à jour et actualiser les données à l'aide de la fonction Mettre à jour les données. Néanmoins, si vous avez modifié les noms de variable ou le nom du fichier, vous devrez réimporter complètement vos données. [Pour plus d'informations, reportez-vous à la section Modifier les sources de données sur p. 63.](#)

Pour mettre à jour et actualiser vos données :

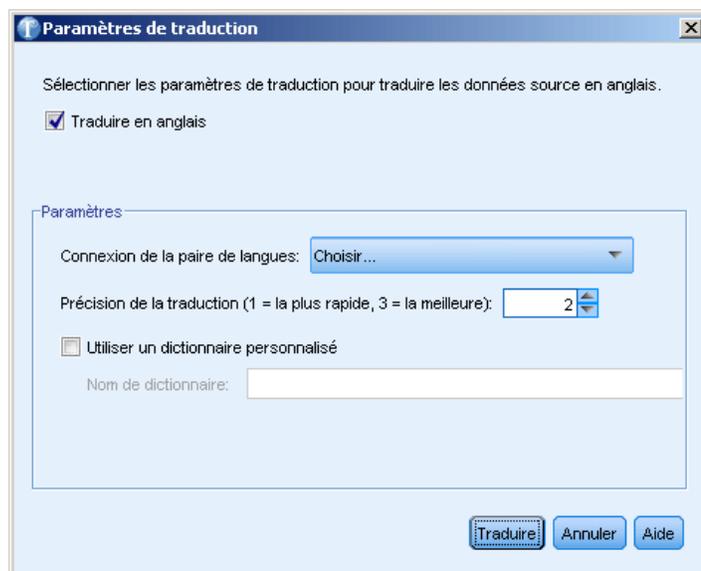
- ▶ Dans les menus, sélectionnez Fichier > Mettre à jour les données. Les données sont relues pour prendre en compte vos nouvelles modifications.
- ▶ Si une traduction en anglais a été réalisée auparavant, la boîte de dialogue Paramètres de traduction s'affiche, de sorte que vous pouvez choisir la paire de langues et réaliser une nouvelle traduction. [Pour plus d'informations, reportez-vous à la section Traduction vers l'anglais sur p. 74.](#)

Traduction vers l'anglais

Vous pouvez actualiser une traduction quand vous le souhaitez. Après la traduction, vous devez procéder à une nouvelle extraction car les résultats de votre traduction ne seront plus synchronisés avec votre nouvelle traduction.

Remarque : Si vous souhaitez traduire de nouvelles données, vous pouvez le faire directement dans l'assistant de Nouveau projet lorsque vous créez un nouveau projet. [Pour plus d'informations, reportez-vous à la section Traduction vers l'anglais sur p. 37.](#)

Figure 4-19
Paramètres de traduction



Pour traduire en anglais

- ▶ A partir des menus, sélectionnez Outils > Paramètres de traduction. La boîte de dialogue Paramètres de traduction s'affiche.
- ▶ Pour traduire en anglais les données textuelles à partir d'une langue bénéficiant d'une licence, cochez la case Traduire en anglais.
- ▶ Dans la liste connexion des paires de langues, sélectionnez la connexion pour la paire de langues Language Weaver que vous souhaitez utiliser. Si Language Weaver est configuré sur votre ordinateur local, ces paires de langues apparaîtront automatiquement dans cette liste. Vous pouvez ajouter ou tester les connexions réseau (WAN) ou les connexions des services en ligne (HTTP) dans l'onglet Traduction de la boîte de dialogue Options. [Pour plus d'informations, reportez-vous à la section Options : Onglet Traduction dans le chapitre 2 sur p. 23.](#)
- ▶ Spécifiez la Précision de la traduction de votre choix. Choisissez une valeur de 1 à 3 indiquant le rapport entre le niveau de vitesse et de précision souhaité. Une valeur faible permet d'obtenir des résultats de traduction plus rapides mais moins précis. Une valeur élevée permet, quant à elle, d'obtenir des résultats plus précis mais après une durée de traitement plus longue. Pour optimiser le temps de traitement, nous vous recommandons de commencer par un niveau peu élevé et de l'augmenter seulement si, après avoir examiné les résultats, vous pensez avoir besoin d'une plus grande précision.
- ▶ Si vous aviez précédemment créé des dictionnaires personnalisés, détenus par Language Weaver, vous pouvez les utiliser dans le cadre de la traduction. Pour choisir un dictionnaire personnalisé, cochez la case Utiliser un dictionnaire personnalisé et entrez le Nom du dictionnaire. Pour utiliser plusieurs dictionnaires, séparez les noms avec une virgule.

- ▶ Cliquez sur Traduire pour lancer le processus de traduction. La boîte de dialogue de progression de la traduction s'affiche.

Partage de projets

Vous pouvez partager vos projets avec d'autres utilisateurs, ou pour travailler sur un projet sur un autre ordinateur.

Pour partager un projet :

- ▶ Dans les menus, choisissez Fichier > Enregistrer le projet. Le projet est enregistré.
- ▶ Envoyez le fichier du projet vers un autre ordinateur ou à une personne. Ce fichier de projet contient une référence au fichier de données que vous avez importé à l'origine. Si vous voulez que l'autre utilisateur puisse utiliser les mêmes données sources pour ce projet, vous devez aussi lui fournir le fichier de données d'origine et l'informer de l'emplacement vers lequel il doit copier ce fichier de données afin que IBM® SPSS® Text Analytics for Surveys puisse trouver le fichier de données quand le fichier de projet est ouvert.
- ▶ Quand l'autre utilisateur ouvre le fichier du projet dans SPSS Text Analytics for Surveys, il ou elle peut utiliser les bibliothèques locales contenues dans le fichier du projet ou utiliser les versions publiques de ces bibliothèques. Généralement, pour garantir les mêmes résultats, les versions locales doivent être utilisées.
- ▶ Si SPSS Text Analytics for Surveys ne peut pas localiser le fichier de données, un message apparaît et avertit l'utilisateur que les données doivent être réimportées. [Pour plus d'informations, reportez-vous à la section Modifier les sources de données sur p. 63.](#)

Marquer des réponses

Pour vous aider à surveiller votre progression lorsque vous analysez votre enquête, vous pouvez marquer les réponses à l'aide de booléens dans le panneau Données. Il existe de nombreuses raisons de vouloir marquer une réponse, par exemple :

- Pour distinguer les réponses que vous avez examinées manuellement afin de savoir où continuer plus tard
- Pour distinguer une réponse que vous ne savez pas comment traiter
- Pour marquer et exporter les booléens dans un autre programme

Quand vous marquez une réponse avec un booléen, vous pouvez continuer à travailler avec cette réponse. Ils servent uniquement à votre propre usage. Vous avez le choix entre les booléens suivants :

Table 4-1

Descriptions des booléens

Booléen	Description
	Booléen Terminé pour marquer les réponses que vous considérez comme finies.
	Booléen Important pour marquer les réponses que vous considérez comme importantes.

Figure 4-20
Booléens de réponse dans le panneau Données

	Id	Réponse	Catégories
1	1	Excellentes prestations générales. J'ai particulièrement apprécié les salons où l'on pouvait travailler ou se reposer en toute tranquillité	prestations
2	3	propreté des lieux, accueil chaleureux, buffet extraordinaire	accueil buffet lieu
3	15	Super piscine, accueil top et chambre très bien entretenue	accueil chambre piscine
4	50	ABSOLUMENT TOUT. PARFAIT. DU DEBUT JUSQU'À LA FIN.	
5	67	excellent rapport qualité-prix	
6	97	ambiance, restauration propreté, accueil. Tout parfait	restauration accueil ambiance
7	127	room-service impeccable	service
8	130	service de chambre formidable et super rapide	chambre

Pour marquer une réponse avec un booléen :

- ▶ Dans le panneau Données, sélectionnez la réponse que vous voulez marquer.
- ▶ Dans les menus, sélectionnez Edition > Marquer les réponses avec puis sélectionnez le type de booléen à utiliser (Booléen important ou Booléen terminé). Le booléen sélectionné est attribué. Si la colonne Booléen du panneau Données n'est pas visible, cela s'affiche. La barre d'état est mise à jour avec le nombre de réponses marquées.

Pour supprimer les booléens :

- ▶ Dans le panneau Données, cliquez avec le bouton droit de la souris sur les réponses pour lesquelles vous voulez supprimer un booléen.
- ▶ Dans le menu contextuel, sélectionnez Marquer les réponses avec > Supprimer les booléens. Les booléens sélectionnés sont supprimés.

Barre d'état du projet

Différentes barres d'état existent en fonction de la fenêtre ou la vue dans laquelle vous travaillez. Par défaut, une barre d'état s'affiche dès qu'un projet est ouvert. Cette barre d'état fournit des informations récapitulatives à propos du projet et des éléments qu'il contient. Vous pouvez activer et désactiver la barre d'état lorsque vous le souhaitez.

Pour activer ou désactiver la barre d'état dans une fenêtre :

- ▶ Dans les menus, choisissez Affichage > Barre d'état.

Fenêtre d'analyse de texte

Cette barre d'état fournit des informations récapitulatives à propos des questions et réponses du projet. Selon l'endroit où vous vous trouvez dans la fenêtre d'analyse de texte, les informations dans la barre d'état changent. Vous pouvez aussi voir le nombre de réponses qui ont été marquées comme importantes ou terminées.

Figure 4-21

Barre d'état dans la fenêtre d'analyse de texte : Vue Question



Quand vous êtes dans la vue Question, vous voyez le nombre de catégories pour cette question et la catégorisation Pourcentage de réponses. Lorsque vous êtes dans la vue Projet complet, vous voyez les informations du projet complet.

Figure 4-22

Barre d'état dans la fenêtre d'analyse de texte : Vue Projet complet



Le tableau suivant décrit chaque élément dans la barre d'état.

Table 4-2

Fenêtre d'analyse de texte : Description de la barre d'état

Élément	Description
Records	Le nombre d'enregistrements dans vos données.
Questions	Le nombre de questions dans votre projet.
Référence	Le nombre de variables de référence importées. Les variables de référence sont des variables supplémentaires qui sont importées pour référence mais ne sont pas analysées.
Catégories	Le nombre de catégories pour une question donnée. Si des catégories vides existent, leur nombre s'affiche entre parenthèses.
Catégorisé	Le nombre de réponses à la question suivi du pourcentage de catégorisation entre parenthèses.

Editeur de ressources Fenêtre

Cette barre d'état fournit des informations à propos des ressources linguistiques du projet. La zone des termes forcés dans la barre est activable, c'est-à-dire que vous pouvez cliquer dessus pour effectuer une action. Lorsque vous travaillez avec les bibliothèques, vous pouvez désactiver des éléments dans les bibliothèques pour les exclure du traitement. [Pour plus d'informations, reportez-vous à la section Désactivation des bibliothèques locales dans le chapitre 9 sur p. 211.](#) Si le projet contient des éléments désactivés, deux comptes numériques apparaissent dans la barre d'état. Le premier est le nombre d'éléments présents et le deuxième est le nombre d'éléments activés. Par exemple, si votre barre d'état affiche 5(2) bibliothèques, cela signifie que cinq bibliothèques se trouvent dans votre projet mais seulement deux sont activées.

Figure 4-23

Barre d'état dans la fenêtre Editeur de dictionnaire



Le tableau suivant décrit chaque élément dans la barre d'état.

Table 4-3
Editeur de ressources fenêtre : Description de la barre d'état

Elément	Description
Bibliothèque	Le nombre de bibliothèques dans le projet.
Type	Le nombre de types dans le projet complet.
Terme	Le nombre de termes dans toutes les bibliothèques. Si un terme est dans les Exclusions, il fait toujours partie du compte. Remarque : si un type est désactivé, tous les termes de ce type sont aussi désactivés.
Exclusions	Le nombre d'éléments exclus dans toutes les bibliothèques du projet.
Synonyme	Le nombre de cibles synonymes dans toutes les bibliothèques du projet.
Optionnels	Le nombre d'éléments optionnels définis dans toutes les bibliothèques du projet. Remarque : le compte comprend toute entrée délimitée individuellement dans une cellule.
Termes forcés	Un bouton qui est activé quand des termes forcés se trouvent dans les bibliothèques de votre projet. Cliquez sur ce bouton pour afficher la boîte de dialogue Modifier les termes forcés. Pour plus d'informations, reportez-vous à la section Ajout des termes forcés dans le chapitre 10 sur p. 228.

Extraction de données

Lorsque vous créez un projet avec l'assistant de Nouveau projet, le choix par défaut est de réaliser une extraction automatiquement pour la première question. Si vous souhaitez actualiser une extraction ou extraire une nouvelle question, vous pouvez le faire manuellement (Outils > Extraire) ou choisir d'extraire au moment où vous commencez la création de catégories. Le résultat final de cette extraction correspond à un ensemble de concepts, de types et de patrons. Vous pouvez afficher et utiliser ces résultats dans le panneau Résultats d'extraction.

Si aucune extraction n'a été réalisée lorsque vous avez créé votre projet ou si vous avez choisi de ne pas enregistrer les résultats d'extraction, vous pouvez naviguer vers la question avec laquelle vous souhaitez commencer à travailler (Affichage > Question > "Question") puis procéder à l'extraction.

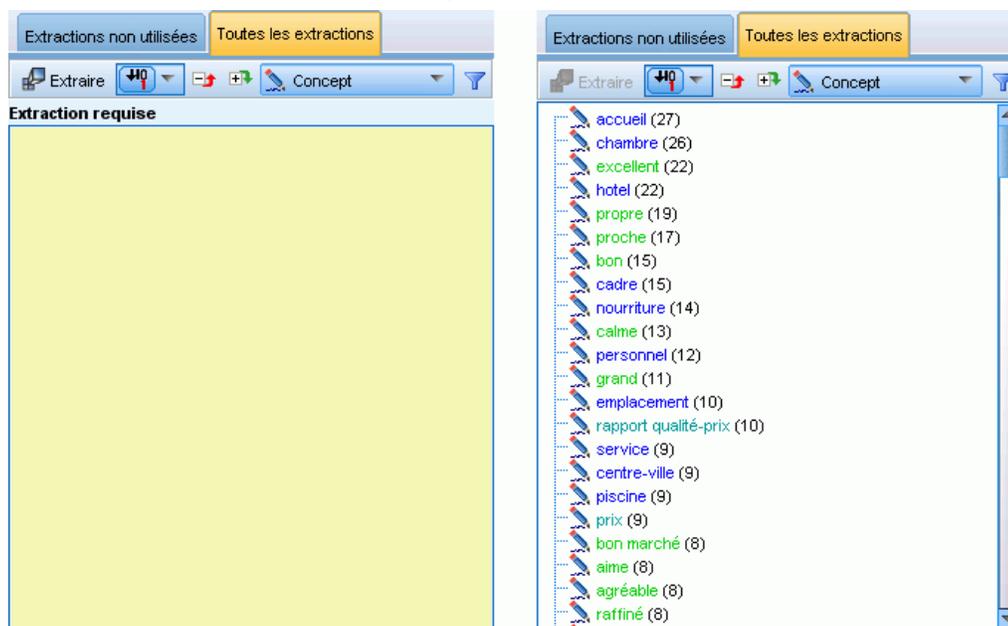
Après l'extraction, passez en revue les résultats et apportez les modifications que vous estimez nécessaires. [Pour plus d'informations, reportez-vous à la section Affinage des résultats de l'extraction sur p. 88.](#) Vous pouvez ensuite procéder à une nouvelle extraction pour voir les nouveaux résultats. Lors du codage manuel des réponses, il se peut que deux individus regroupent des réponses légèrement différentes. Toutefois, l'exactitude et la continuité sont extrêmement importantes pour regrouper en catégories des réponses à des enquêtes.

La puissance de IBM® SPSS® Text Analytics for Surveys réside dans sa capacité à fournir la réapplication régulière de vos définitions de catégorie. En affinant vos résultats d'extraction dès le départ, vous êtes assuré d'obtenir à chaque nouvelle extraction des résultats identiques et bien adaptés au contexte des données dans vos définitions de catégorie. De cette manière, l'attribution des réponses à vos définitions de catégorie est plus précise et plus à même d'être répétée.

Résultats extraits : Concepts, types et patrons

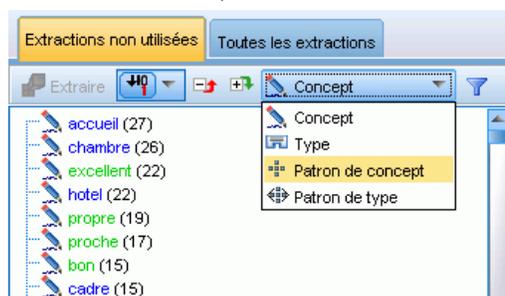
Une fois que vous avez créé un projet, la fenêtre affiche automatiquement la première question ouverte importée. Le panneau Résultats d'extraction se situe dans l'angle inférieur gauche de la vue Question. Vous pouvez accéder à cette vue à partir du menu Affichage (Affichage > Question > "Question"). Si aucun résultat d'extraction n'existe, vous devez procéder à l'extraction pour commencer à travailler. [Pour plus d'informations, reportez-vous à la section Extraction de données sur p. 85.](#)

Figure 5-1
Panneau Résultats d'extraction avant et après l'extraction



Si le panneau Résultats d'extraction est vide ou ne s'applique plus, il est coloré en jaune. Cliquez sur le bouton Extraire pour lancer le processus d'extraction. Après l'extraction, vous pouvez examiner les résultats en sélectionnant les éléments que vous souhaitez voir dans la liste déroulante.

Figure 5-2
Liste déroulante du panneau Résultats d'extraction



Les concepts, les types et les patrons TLA extraits sont désignés collectivement par le terme **résultats d'extraction** et servent de descripteurs ou blocs de construction pour vos catégories. Vous pouvez également utiliser des concepts, des types et des patrons dans vos règles de catégorie. De plus, les techniques automatiques s'appuient sur des concepts et des types pour créer des catégories.

L'analyse est un processus itératif au cours duquel les résultats de l'extraction sont passés en revue en fonction du contexte des données textuelles, puis affinés afin de générer de nouveaux résultats avant d'être réévalués. Après l'extraction, passez en revue les résultats et apportez les modifications que vous estimez nécessaires en modifiant les ressources linguistiques. Vous pouvez directement affiner une partie des ressources dans le panneau Résultats d'extraction, ou

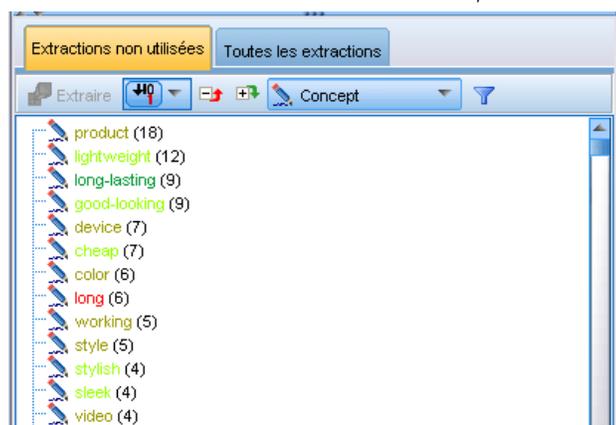
le panneau Données à l'aide des menus contextuels. [Pour plus d'informations, reportez-vous à la section Affinage des résultats de l'extraction sur p. 88.](#) Vous pouvez également travailler directement dans la vue de l'Editeur de ressources. [Pour plus d'informations, reportez-vous à la section La fenêtre Editeur de ressources dans le chapitre 2 sur p. 16.](#)

Après cette optimisation, vous pouvez procéder à une nouvelle extraction pour voir les nouveaux résultats. En affinant vos résultats d'extraction dès le départ, vous êtes assuré d'obtenir à chaque nouvelle extraction des résultats identiques et parfaitement adaptés au contexte des données dans vos définitions de catégorie. De cette manière, l'attribution des enregistrements à vos définitions de catégorie est plus précise et plus à même d'être répétée.

Concepts

Lors du processus d'extraction, les données textuelles sont analysées pour découvrir des mots isolés intéressants ou pertinents (par exemple, *élection* ou *paix*) et des groupes de mots (par exemple, *élection présidentielle*, *élection du président* ou *traités de paix*). Ces mots et groupes de mots sont appelés des *termes*. En utilisant les ressources linguistiques, les termes pertinents sont extraits et les termes similaires sont regroupés sous un terme principal appelé **concept**.

Figure 5-3
Panneau Résultats d'extraction : Vue Concept



Par défaut, les concepts sont affichés en minuscules et triés dans l'ordre décroissant en fonction du nombre d'enregistrements dans lesquels le concept apparaît. Quand les concepts sont extraits, un type leur est affecté pour regrouper les concepts similaires. Ils apparaissent sous différents codes de couleurs en fonction de ce type. Le choix des couleurs s'effectue sous les propriétés du type, dans l'Editeur de ressources. [Pour plus d'informations, reportez-vous à la section Déclarations de types dans le chapitre 10 sur p. 220.](#)

Lorsqu'un concept, un type ou un patron est utilisé dans une définition de catégorie, il apparaît en italique dans le tableau.

Types

Les **types** correspondent à des regroupements sémantiques de concepts. Quand les concepts sont extraits, un type leur est affecté pour regrouper les concepts similaires. Plusieurs types intégrés sont fournis avec IBM® SPSS® Text Analytics for Surveys, dont <Location>, <Organization>, <Person>, <Positive>, <Negative> etc. Par exemple, le type <Location> regroupe des mots clés géographiques et des lieux. Ce type est affecté à des concepts tels que *chicago*, *paris* et *tokyo*. Les concepts qui ne se trouvent dans aucune déclaration de type mais sont extraits du texte sont automatiquement dotés du type <Unknown>. [Pour plus d'informations, reportez-vous à la section Types intégrés dans le chapitre 10 sur p. 221.](#)

Quand vous sélectionnez la vue Type, les types extraits apparaissent par défaut dans l'ordre décroissant, par fréquence. Lorsque que l'arborescence est développée, les concepts extraits pour ce type apparaissent. Vous pouvez également remarquer que les types font l'objet d'un codage couleur pour être plus faciles à distinguer. Les couleurs font partie des propriétés du type. [Pour plus d'informations, reportez-vous à la section Création de types dans le chapitre 10 sur p. 222.](#) Vous pouvez aussi créer vos propres types.

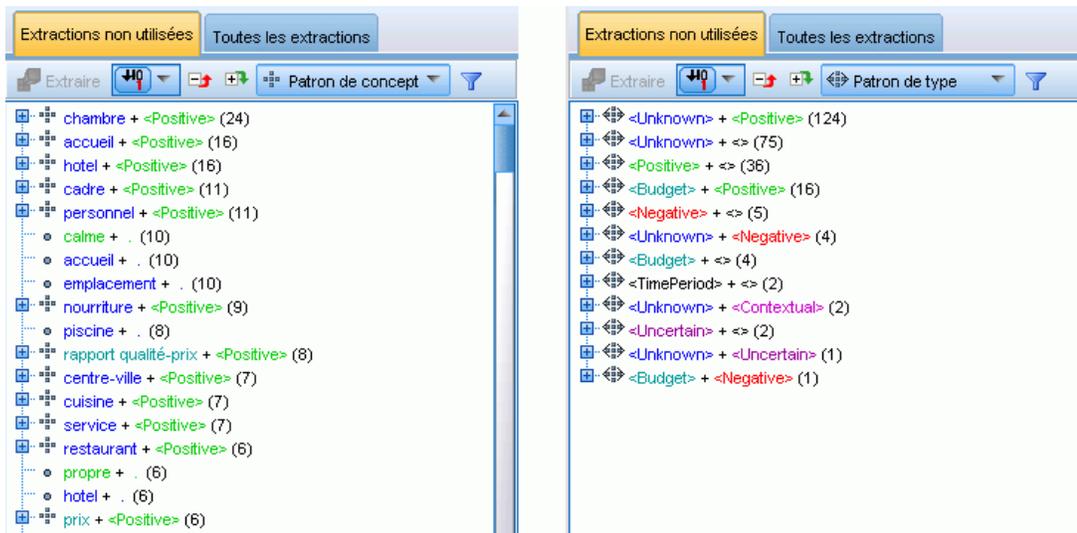
Figure 5-4
Panneau Résultats d'extraction : Vue Type



Patrons

Les patrons sont composés de deux parties : une combinaison de concepts et de types intégrés représentant les qualificatifs et les adjectifs. Les patrons s'avèrent particulièrement utiles lorsque vous tentez de découvrir des opinions sur un sujet donné. Pour vous, l'intérêt de l'extraction du nom du produit de votre concurrent peut être limité. Dans ce cas, vous pouvez examiner les patrons extraits pour voir s'il est possible de trouver des exemples dans lesquels les personnes interrogées ont trouvé le produit bon, mauvais ou cher. Il existe deux vues de patrons différentes : Les patrons de concept et les patrons de type.

Figure 5-5
 Panneau Résultats d'extraction : Vue Patron de concept et vue Patron de type



Patron de concept. Dans cette vue, le niveau supérieur de l'arborescence dans le panneau Résultats d'extraction affiche les patrons avec la structure suivante : `concept1 + <Type1>` pour les patrons de concept, comme `analyse de texte + <Positive>` ou `coût + <Negative>`. Lorsque l'arborescence est développée, les patrons exacts apparaissent, comme `analyse de texte + puissante` ou `coût + cher`.

Les patrons peuvent également être utiles lorsqu'il n'ont pas de deuxième partie. Par exemple, vous souhaitez trouver des occurrences dans lesquelles la personne interrogée n'a exprimé ni une opinion positive ni une opinion négative sur le sujet. Dans ce cas, cela est représenté par `concept1 + .`, où `.` désigne un qualificatif vide. Par exemple, si une personne interrogée a répondu « *Coût, emplacement de stockage* » en réponse à la question, « *Quels facteurs ont influencé votre choix de lecteur de musique ?* », l'extraction pourra renvoyer `coût + .` et `emplacement_stockage + .` comme patrons vides. Lorsque les patrons sont affichés, des couleurs sont attribuées à chaque élément du patron en fonction de son type.

Patron de type. Dans cette vue, le niveau supérieur de l'arborescence dans le panneau Résultats d'extraction affiche les patrons avec la structure suivante : `<Type> + <Type>`, comme `<Budget> + <Positive>`. Si vous continuez à développer l'arborescence, vous pourrez afficher les relations décrites et présentées dans la vue Patron de concept. Lorsque les patrons sont affichés, des couleurs sont attribuées à chaque élément du patron en fonction de son type.

Onglets Extractions non utilisées et Toutes les extractions

Le panneau Résultats d'extraction présente les résultats du processus d'extraction. Lorsque vous commencerez à créer des catégories, certains des résultats d'extraction (concepts, types et patrons) seront intégrés aux descripteurs de catégorie. Pour cette raison, SPSS Text Analytics for Surveys présente ces informations de deux manières au moyen d'onglets. Vous pouvez basculer entre l'affichage des éléments qui sont déjà utilisés dans les définitions de catégorie ou l'affichage de l'ensemble des concepts extraits. Pour ce faire, cliquez sur les onglets Extractions non utilisées et Toutes les extractions. L'onglet Extractions non utilisées affiche tous les éléments qui ne font

pas partie d'un descripteur de catégorie. L'onglet Toutes les extractions affiche tous les éléments extraits, les éléments utilisés apparaissant en italiques.

Extraction de données

Chaque fois qu'une extraction est nécessaire, le panneau Résultats d'extraction devient jaune et le message Appuyer sur le bouton Extraire pour extraire les concepts apparaît sous la barre d'outils de ce panneau.

Vous aurez besoin de procéder à une extraction si vous n'avez pas encore de résultat d'extraction, si vous avez modifié les ressources linguistiques et devez mettre à jour les résultats d'extraction, ou si vous avez réouvert un projet dans lequel vous n'aviez pas enregistré les résultats de l'extraction (Outils > Options).

Remarque : Si vous modifiez le noeud source de votre flux après la mise en cache des résultats d'extraction à l'aide de l'option Utiliser le travail d'une session..., vous devrez exécuter une nouvelle extraction une fois la session interactive démarrée, pour obtenir les résultats d'extraction mis à jour.

Lorsque vous effectuez une extraction, un indicateur de progression apparaît pour vous fournir des informations sur l'état de l'extraction. Pendant ce temps, le moteur du programme d'extraction analyse toutes les données textuelles et identifie les termes et patrons pertinents puis les extrait et les attribue à un type. Ensuite, le moteur tente de regrouper les synonymes sous un terme principal, appelé un concept. Une fois le processus terminé, les concepts, les types et les patrons obtenus apparaissent dans le panneau Résultats d'extraction. Vous pouvez commencer à travailler avec les résultats et à les examiner.

Remarque : Il existe un rapport entre la taille de votre ensemble de données et la durée nécessaire à l'exécution du processus d'extraction. Reportez-vous aux instructions d'installation pour obtenir des statistiques et des recommandations sur les performances.

Pour extraire des données

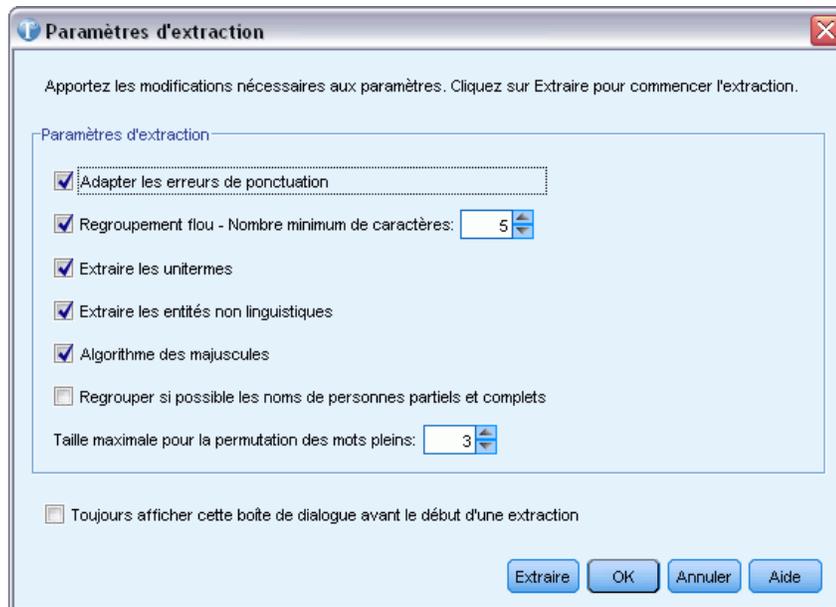
- ▶ A partir des menus, sélectionnez Outils> Extraire. Vous pouvez également cliquer sur le bouton Extraire de la barre d'outils.
- ▶ Si vous choisissez de toujours afficher la boîte de dialogue Paramètres d'extraction, elle apparaît pour vous permettre d'apporter des modifications. Vous trouverez plus loin dans cette rubrique des descriptions de chaque paramètre.
- ▶ Cliquez sur Extraire pour lancer le processus d'extraction. Dès le début de l'extraction, une boîte de dialogue indique la progression du processus. Après l'extraction, les résultats apparaissent dans le panneau Résultats d'extraction. Par défaut, les concepts sont affichés en minuscules et triés dans l'ordre décroissant en fonction du nombre d'enregistrements dans lesquels le concept apparaît.

Vous pouvez examiner les résultats à l'aide des options de la barre d'outils pour trier les résultats différemment, les filtrer, ou encore afficher une autre vue (concepts, patrons, ou types). Vous pouvez aussi redéfinir vos résultats d'extraction en travaillant avec les ressources linguistiques.

[Pour plus d'informations, reportez-vous à la section Affinage des résultats de l'extraction sur p. 88.](#)

La boîte de dialogue Paramètres d'extraction contient des options d'extraction élémentaires.

Figure 5-6
Boîte de dialogue Paramètres d'extraction



Adapter les erreurs de ponctuation. Cette option normalise temporairement le texte contenant des erreurs de ponctuation (par exemple, usage incorrect) au cours de l'extraction pour améliorer l'extraction des concepts. Cette option s'avère extrêmement utile lorsque le texte est court et de qualité médiocre (réponses ouvertes, messages électroniques, données CRM, etc.) ou qu'il contient de nombreuses abréviations.

Traitement des fautes de frappe. Nombre de caractères minimum requis ([n]). Cette option applique une technique de regroupement flou qui permet de regrouper les mots mal orthographiés ou d'orthographe similaire sous un seul concept. L'algorithme de regroupement flou retire temporairement toutes les voyelles (à l'exception de la première) et les consonnes doubles/triples des mots extraits pour comparer ces derniers et voir s'ils sont identiques, afin par exemple que *modélisation* et *modélisation* soient regroupés. Néanmoins, si chaque type est affecté à un type différent et que vous excluez le type <Unknown>, le regroupement flou n'est pas appliqué.

Vous pouvez également définir le nombre minimum de caractères *racines* requis avant d'utiliser le regroupement flou. Le nombre de caractères racine d'un terme est calculé en ajoutant l'ensemble des caractères et en soustrayant les caractères formant des suffixes flexionnels et, dans le cas des termes apparaissant sous la forme de mots composés, les déterminants et les prépositions. Par exemple, le terme *exercices* serait déterminé comme comportant 8 caractères racine dans la forme « exercice », étant donné que la lettre *s* située à la fin du mot représente une flexion (forme plurielle). De même, *jus énergétique* comporte 14 caractères racine « jus énergétique » et *conception de dessins*, 16 caractères racine « conception dessin ». Cette méthode de calcul est uniquement utilisée pour vérifier si le regroupement flou doit être appliqué, mais n'influe pas sur la mise en correspondance des mots.

Remarque : Si, d'après vous, certains mots sont ensuite groupés de manière incorrecte, vous pouvez exclure des paires de mots de cette technique en les déclarant de manière explicite dans la section Regroupement flou : Exceptions dans l'onglet Ressources avancées. [Pour plus d'informations, reportez-vous à la section Regroupement flou dans le chapitre 11 sur p. 242.](#)

Extraire les unitermes. Cette option extrait les mots uniques (unitermes) dans la mesure où le mot ne fait pas déjà partie d'un mot composé et si c'est un nom ou une catégorie grammaticale non reconnue.

Extraction des entités non linguistiques. Cette option extrait les entités non linguistiques, telles que les numéros de téléphone, numéros de sécurité sociale, heures, dates, devises, chiffres, pourcentages, adresses électroniques, adresses HTTP, etc. Vous pouvez inclure ou exclure certains types d'entités non linguistiques dans la section Entités non linguistiques : Configuration dans l'onglet Ressources avancées. Désactivez les entités dont vous n'avez pas besoin pour éviter au moteur d'extraction un temps de traitement inutile. [Pour plus d'informations, reportez-vous à la section Configuration dans le chapitre 11 sur p. 247.](#)

Algorithme des noms propres. Cette option extrait les termes simples et composés ne figurant pas dans les dictionnaires intégrés, si la première lettre est une majuscule. Cette option offre une bonne manière d'extraire la plupart des noms propres.

Regroupement éventuel des noms de personnes partiels et complets. Cette option regroupe les noms qui apparaissent différemment dans le texte. Cette fonction est utile car les noms sont souvent cités dans leur forme complète au début du texte puis uniquement en version abrégée. Cette option essaye de faire correspondre tout uniterme ayant le type <Unknown> avec le dernier mot de tout terme composé entré comme <Person>. Par exemple, si *martin* est trouvé et initialement saisi comme <Unknown>, le moteur d'extraction vérifie si un terme composé de type <Person> contient *martin* comme dernier mot, tel que *pierre martin*. Cette option ne s'applique pas aux prénoms, car la plupart ne sont jamais extraits comme unitermes.

Nombre de mots pleins soumis à une permutation pour le regroupement. Cette option indique le nombre maximal de mots utiles pouvant être présents lorsque s'applique la technique de permutation. Cette technique de permutation regroupe les expressions similaires qui ne diffèrent les unes des autres que par la présence de mots utiles (par exemple, de et la), quelle que soit leur flexion. Par exemple, si vous avez paramétré cette valeur sur au moins deux mots et si les deux expressions *représentants d'entreprise* et *représentants de l'entreprise* ont été extraites. Dans ce cas, les deux termes extraits sont regroupés dans la liste de concepts finale car les deux termes sont considérés comme étant les mêmes lorsque *de l'* est ignoré.

Toujours afficher cette boîte de dialogue avant le début d'une extraction. Vous permet de spécifier si vous souhaitez que la boîte de dialogue Paramètres d'extraction apparaisse à chaque fois que vous effectuez une extraction, si vous ne voulez pas qu'elle s'affiche sauf si vous l'ouvrez via le menu Outils ou si vous voulez être interrogé à chaque fois que vous effectuez une extraction si vous souhaitez modifier les paramètres d'extraction.

Enregistrement des résultats de l'extraction

A chaque extraction, les résultats apparaissent dans le panneau Résultats d'extraction et peuvent être utilisés pour catégoriser vos réponses. Au cours d'une session IBM® SPSS® Text Analytics for Surveys, ces résultats d'extraction sont conservés en mémoire de façon à ce que vous puissiez

les exploiter. Par défaut, les résultats de l'extraction sont enregistrés dans les projets. Leur enregistrement à la fermeture du projet constitue un paramètre global qu'il est possible de modifier à tout moment dans la boîte de dialogue Options (Outils > Options). [Pour plus d'informations, reportez-vous à la section Options : Onglet Système dans le chapitre 2 sur p. 19.](#)

A des fins de sécurité, ces résultats d'extraction sont codés durant le processus d'enregistrement et placés dans la base de données. Cette procédure rend difficile pour les utilisateurs la détection de données dans la base. De plus, les résultats de l'extraction ne sont jamais présentés dans SPSS Text Analytics for Surveys tant que la source de données n'a pas été localisée par l'application. Par conséquent, si les données sont protégées par un mot de passe, l'utilisateur doit entrer le nom d'utilisateur et le mot de passe pour cette source de données afin que les résultats de l'extraction apparaissent à l'écran.

Affinage des résultats de l'extraction

L'extraction est un processus itératif selon lequel vous pouvez extraire des résultats, les examiner, les modifier, puis procéder à une nouvelle extraction pour les mettre à jour. Etant donné que la précision et la continuité sont essentielles à la réussite du Text Mining et de la catégorisation, le fait d'affiner les résultats de l'extraction dès le départ garantit qu'à chaque nouvelle extraction, vous obtenez exactement les mêmes résultats dans vos définitions de catégorie. De cette manière, l'attribution des enregistrements à vos catégories est plus précise et plus répétable.

Les résultats de l'extraction font office de blocs de construction pour les catégories. Lorsque vous créez des catégories à l'aide de ces résultats d'extraction, les enregistrements sont automatiquement attribués aux catégories s'ils contiennent du texte qui correspond à un ou plusieurs descripteurs de catégorie. Bien que vous puissiez commencer la catégorisation avant d'affiner les ressources linguistiques, il est utile d'examiner les résultats de l'extraction au moins une fois au préalable.

Durant l'examen des résultats, vous pouvez trouver des éléments pour lesquels vous souhaitez que le moteur du programme d'extraction se comporte de manière différente. Prenons les exemples suivants :

- **Synonymes non reconnus.** Supposons que vous trouviez plusieurs concepts qui, selon vous, sont synonymes, par exemple `intelligent`, `astucieux`, `brillant` et `ingénieux`, et qu'ils apparaissent tous en tant que concepts individuels dans les résultats d'extraction. Vous pouvez créer une définition de synonyme dans laquelle les concepts `astucieux`, `brillant` et `ingénieux` sont regroupés sous le concept cible `intelligent`. Cette action regroupe ainsi tous les concepts avec `intelligent` et la fréquence globale est alors supérieure. [Pour plus d'informations, reportez-vous à la section Ajout de synonymes sur p. 89.](#)
- **Modification du type des concepts.** Supposons que les concepts de vos résultats d'extraction apparaissent sous un type et vous souhaitez qu'ils soient affectés à un autre type. Dans un autre exemple, imaginez que vous trouviez 15 concepts relatifs aux légumes dans vos résultats d'extraction et que vous souhaitiez tous les ajouter à un nouveau type appelé `<Légume>`. Les concepts qui ne se trouvent dans aucune déclaration de type mais sont extraits du texte sont automatiquement dotés du type `<Unknown>`. Vous pouvez ajouter des concepts aux types. [Pour plus d'informations, reportez-vous à la section Ajout de concepts à des types sur p. 91.](#)

- **Concepts non pertinents.** Supposons qu'un concept extrait présente une fréquence très élevée, ce qui signifie qu'il apparaît dans de nombreux enregistrements. Toutefois, vous considérez que ce concept n'est pas pertinent pour l'analyse. Vous pouvez l'exclure de l'extraction. [Pour plus d'informations, reportez-vous à la section Exclusion de concepts de l'extraction sur p. 94.](#)
- **Correspondances incorrectes.** Supposons que lors de l'examen des enregistrements qui contiennent un certain concept, vous découvriez que deux mots ont été regroupés de façon incorrecte, par exemple *faculté* et *facilité*. Cette correspondance peut être due à un algorithme interne, connu sous le nom de regroupement flou, qui ignore temporairement les consonnes et voyelles doubles ou triples, de façon à regrouper les fautes d'orthographe courantes. Vous pouvez ajouter ces mots à une liste de paires de mots qui ne doivent pas être regroupés. [Pour plus d'informations, reportez-vous à la section Regroupement flou dans le chapitre 11 sur p. 242.](#)
- **Concepts non extraits.** Supposons que vous remarquiez, au moment de l'examen du texte de l'enregistrement, que quelques mots ou expressions n'ont pas été extraits alors que vous vous attendiez à ce qu'ils le soient. Souvent, ces mots sont des verbes ou des adjectifs qui ne présentent pas d'intérêt pour vous. Cependant, il arrive que vous souhaitiez utiliser un mot ou une expression, qui n'a pas été extrait, comme partie intégrante d'une définition de catégorie. Pour extraire le concept, vous pouvez imposer un terme dans une déclaration de types. [Pour plus d'informations, reportez-vous à la section Extraction de mots imposée sur p. 94.](#)

Il est possible d'exécuter un grand nombre de ces modifications directement à partir du panneau Résultats d'extraction ou du panneau Données en sélectionnant un ou plusieurs éléments et en cliquant avec le bouton droit de la souris pour accéder aux menus contextuels.

Une fois les modifications apportées, la couleur d'arrière-plan du panneau change pour indiquer que vous devez procéder à une nouvelle extraction pour visualiser ces modifications. [Pour plus d'informations, reportez-vous à la section Extraction de données sur p. 85.](#) Si vous travaillez avec des ensembles de données volumineux, il peut s'avérer plus efficace de procéder à une nouvelle extraction après plusieurs modifications plutôt qu'après chaque modification.

Remarque : vous pouvez visualiser l'ensemble complet des ressources linguistiques modifiables utilisées pour produire les résultats de l'extraction dans la vue de l'Editeur de ressources (Affichage > Editeur de ressources). Dans cette vue, les ressources apparaissent sous la forme de bibliothèques et de dictionnaires. Vous pouvez personnaliser les concepts et les types directement au sein des bibliothèques et des dictionnaires. [Pour plus d'informations, reportez-vous à la section Utilisation des bibliothèques dans le chapitre 9 sur p. 206.](#)

Ajout de synonymes

Les **synonymes** sont des mots ayant le même sens. Les synonymes sont souvent utilisés pour regrouper des termes et leurs abréviations, ou pour réunir les mots fréquemment mal orthographiés sous la version correcte du mot. Grâce à l'utilisation de synonymes, la fréquence du concept cible est plus élevée, ce qui facilite largement la recherche d'informations similaires qui se présentent sous différentes formes dans vos données textuelles.

Les modèles et bibliothèques de ressources linguistiques fournis avec le produit contiennent de nombreux synonymes prédéfinis. Néanmoins, si vous découvrez des synonymes non reconnus, vous pouvez les définir de façon à ce qu'ils soient reconnus lors de la prochaine extraction.

La première étape consiste à décider du concept cible, ou concept principal. Le **concept cible** est l'expression ou le mot sous lequel vous souhaitez regrouper tous les termes synonymes dans les résultats finaux. Au cours de l'extraction, les synonymes sont regroupés sous ce concept cible. La deuxième étape consiste à identifier tous les synonymes de ce concept. Le concept cible vient remplacer tous les synonymes dans l'extraction finale. Pour être un synonyme, un terme doit être extrait. En revanche, il n'est pas nécessaire que le concept cible soit extrait pour que la substitution se produise. Par exemple, si vous souhaitez que le terme *astucieux* soit remplacé par *intelligent*, alors *astucieux* est le synonyme et *intelligent* est le concept cible.

Si vous créez une définition de synonyme, un nouveau concept cible est ajouté au dictionnaire. Vous devez ensuite ajouter des synonymes à ce concept cible. Lorsque vous créez ou éditez des synonymes, ces modifications sont enregistrées dans les dictionnaires de synonymes de l'Editeur de ressources. Pour visualiser la totalité du contenu de ces dictionnaires de synonymes ou pour apporter un nombre important de modifications, travaillez plutôt directement dans l'Editeur de ressources. [Pour plus d'informations, reportez-vous à la section Dictionnaires des substitutions/synonymes dans le chapitre 10 sur p. 231.](#)

Les nouveaux synonymes sont automatiquement stockés dans la première bibliothèque répertoriée dans l'arborescence de bibliothèques de la vue de l'Editeur de ressources—par défaut, il s'agit de la *bibliothèque locale*.

Remarque : si vous recherchez une définition de synonyme et que vous ne la trouvez ni via les menus contextuels ni directement dans l'Editeur de ressources, une correspondance a peut-être été obtenue à partir d'une technique interne de regroupement flou. [Pour plus d'informations, reportez-vous à la section Regroupement flou dans le chapitre 11 sur p. 242.](#)

Pour créer un synonyme

- ▶ Dans le panneau Résultats d'extraction ou dans le panneau Données, sélectionnez les concepts pour lesquels vous souhaitez créer un nouveau synonyme.
- ▶ Dans les menus, choisissez Edition > Ajouter au synonyme > Nouveau. La boîte de dialogue Créer un synonyme apparaît.

Figure 5-7

Boîte de dialogue Créer un synonyme



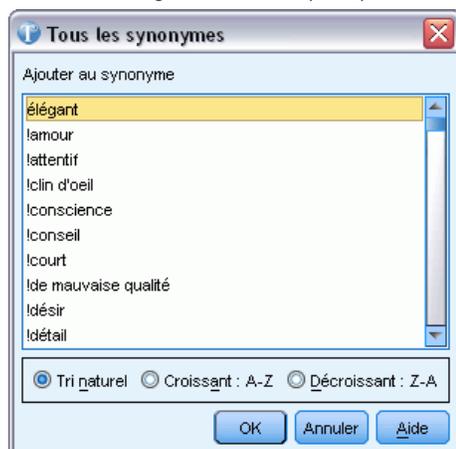
- ▶ Entrez un concept cible dans la zone de texte Cible. Il s'agit du concept sous lequel tous les synonymes seront regroupés.

- ▶ Si vous souhaitez ajouter davantage de synonymes, entrez-les dans la zone de liste Synonymes. Utilisez le séparateur global pour séparer chaque terme synonyme. [Pour plus d'informations, reportez-vous à la section Options : Onglet Système dans le chapitre 2 sur p. 19.](#)
- ▶ Cliquez sur OK pour appliquer les modifications. La boîte de dialogue se ferme et la couleur d'arrière-plan du panneau Résultats d'extraction change, indiquant que vous devez procéder à une nouvelle extraction pour visualiser les modifications. Si vous prévoyez plusieurs modifications, apportez-les avant de procéder à une nouvelle extraction.

Pour ajouter des éléments à un synonyme

- ▶ Dans le panneau Résultats d'extraction ou dans le panneau Données, sélectionnez les concepts à ajouter à une définition de synonyme existante.
- ▶ Dans les menus, choisissez Edition > Ajouter au synonyme > . Le menu affiche un ensemble de synonymes, plaçant en début de liste les synonymes créés le plus récemment. Sélectionnez le nom du synonyme auquel vous souhaitez ajouter les concepts sélectionnés. Si vous trouvez le synonyme recherché, sélectionnez-le ; les concepts sélectionnés sont ajoutés à cette définition de synonyme. Si vous ne le trouvez pas, sélectionnez Plus pour afficher la boîte de dialogue Tous les synonymes.

Figure 5-8
Boîte de dialogue Tous les synonymes



- ▶ Dans la boîte de dialogue Tous les synonymes, vous pouvez trier la liste selon l'ordre de tri naturel (ordre de création), ou selon l'ordre croissant ou décroissant. Sélectionnez le nom du synonyme auquel vous souhaitez ajouter les concepts sélectionnés et cliquez sur OK. La boîte de dialogue se ferme et les concepts sont ajoutés à la définition de synonyme.

Ajout de concepts à des types

Durant chaque processus d'extraction, les concepts extraits sont affectés à des types de façon à regrouper les termes qui présentent un élément commun. IBM® SPSS® Text Analytics for Surveys est livré avec de nombreux types intégrés. [Pour plus d'informations, reportez-vous à la](#)

[section Types intégrés dans le chapitre 10 sur p. 221](#). Les concepts qui ne se trouvent dans aucune déclaration de type mais sont extraits du texte sont automatiquement dotés du type <Unknown>.

Lors de l'analyse de vos résultats, vous pouvez découvrir que certains concepts apparaissent dans un type que vous voulez attribuer à un autre ou qu'un groupe de mots appartient en fait à un nouveau type propre. Dans ces cas de figure, vous pouvez réattribuer les concepts à un autre type ou encore créer un type.

Par exemple, supposons que vous travaillez avec des données d'enquête liées aux automobiles et que vous souhaitez procéder à une catégorisation en vous centrant sur différentes parties des véhicules. Vous pouvez créer un type appelé <Tableau de bord> pour regrouper tous les concepts liés aux indicateurs et boutons situés sur le tableau de bord des véhicules. Ensuite, vous pouvez attribuer des concepts tels que jauge de carburant, chauffage, radio et compteur kilométrique à ce nouveau type.

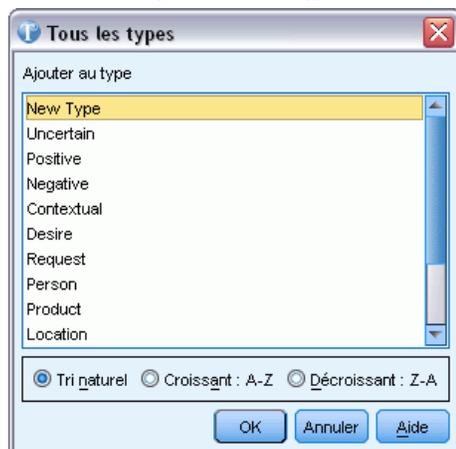
Autre exemple : supposons que vous travaillez avec des données d'enquête liées aux universités et que l'extraction a défini Jean Moulin (l'université) en tant que type <Person> plutôt qu'en tant que type <Organization>. Dans ce cas, vous pouvez ajouter ce concept au type <Organization>.

Lorsque vous créez un type ou ajoutez des concepts à une liste de termes d'un type, ces modifications sont enregistrées dans des déclarations de types dans les bibliothèques de ressources linguistiques de l'Editeur de ressources. Pour visualiser le contenu de ces bibliothèques ou pour apporter un nombre important de modifications, travaillez plutôt directement dans l'Editeur de ressources. [Pour plus d'informations, reportez-vous à la section Ajout de termes dans le chapitre 10 sur p. 224](#).

Pour ajouter un concept à un type

- ▶ Dans le panneau Résultats d'extraction ou dans le panneau Données, sélectionnez les concepts à ajouter à un type existant.
- ▶ Cliquez avec le bouton droit de la souris pour ouvrir le menu contextuel.
- ▶ Dans les menus, choisissez Edition > Ajouter au type >. Le menu affiche un ensemble de types, plaçant en début de liste les types créés le plus récemment. Sélectionnez le nom du type auquel vous souhaitez ajouter les concepts sélectionnés. Si vous trouvez le nom du type recherché, sélectionnez-le ; les concepts sélectionnés sont ajoutés à ce type. Si vous ne le trouvez pas, sélectionnez Plus pour afficher la boîte de dialogue Tous les types.

Figure 5-9
Boîte de dialogue Tous les types

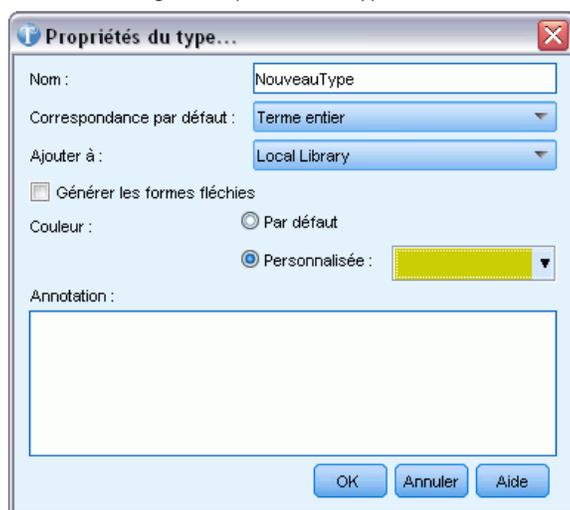


- Dans la boîte de dialogue Tous les types, vous pouvez trier la liste selon l'ordre de tri naturel (ordre de création) ou selon l'ordre croissant ou décroissant. Sélectionnez le nom du type auquel vous souhaitez ajouter les concepts sélectionnés et cliquez sur OK. La boîte de dialogue se ferme et les concepts sont ajoutés au type en tant que termes.

Pour créer un type

- Dans le panneau Résultats d'extraction ou dans le panneau Données, sélectionnez les concepts pour lesquels vous souhaitez créer un nouveau type.
- Dans les menus, choisissez Edition > Ajouter au type > Nouveau. La boîte de dialogue Propriétés de type apparaît.

Figure 5-10
Boîte de dialogue Propriétés de type



- ▶ Dans la zone de texte Nom, entrez le nom de ce nouveau type et apportez éventuellement des modifications aux autres champs. [Pour plus d'informations, reportez-vous à la section Création de types dans le chapitre 10 sur p. 222.](#)
- ▶ Cliquez sur OK pour appliquer les modifications. La boîte de dialogue se ferme et la couleur d'arrière-plan du panneau Résultats d'extraction change, indiquant que vous devez procéder à une nouvelle extraction pour visualiser les modifications. Si vous prévoyez plusieurs modifications, apportez-les avant de procéder à une nouvelle extraction.

Exclusion de concepts de l'extraction

Lors de l'examen des résultats, vous pouvez éventuellement découvrir des concepts dont vous ne souhaitez pas l'extraction ou l'utilisation par une technique de création de catégorie automatisée. Dans certains cas, ces concepts présentent une fréquence très élevée et ne sont pas du tout pertinents pour votre analyse. Vous pouvez dès lors marquer un concept à exclusion de l'extraction finale. En règle générale, les concepts que vous entrez dans cette liste sont des mots ou des expressions de liaison utilisés pour la continuité du texte, mais qui ne lui apportent rien d'important et risquent d'encombrer les résultats de l'extraction. En ajoutant ces concepts au dictionnaire d'exclusions, vous êtes assuré qu'ils ne seront jamais extraits.

Le processus d'exclusion implique que toutes les variantes du concept exclu disparaissent des résultats lors de la prochaine extraction. Si ce concept apparaît déjà en tant que descripteur dans une catégorie, il restera dans la catégorie avec un effectif nul après la nouvelle extraction.

Lorsque vous excluez, ces modifications sont enregistrées dans un dictionnaire d'exclusions dans l'Editeur de ressources. Pour visualiser toutes les définitions d'exclusion et les éditer directement, travaillez plutôt directement dans l'Editeur de ressources. [Pour plus d'informations, reportez-vous à la section Dictionnaires d'exclusions dans le chapitre 10 sur p. 236.](#)

Pour exclure des concepts

- ▶ Dans le panneau Résultats d'extraction ou dans le panneau Données, sélectionnez les concepts à exclusion de l'extraction.
- ▶ Cliquez avec le bouton droit de la souris pour ouvrir le menu contextuel.
- ▶ Sélectionnez Exclure de l'extraction. Le concept est ajouté au dictionnaire d'exclusions dans l'Editeur de ressources et la couleur d'arrière-plan du panneau Résultats d'extraction change, indiquant que vous devez procéder à une nouvelle extraction pour visualiser les modifications. Si vous prévoyez plusieurs modifications, apportez-les avant de procéder à une nouvelle extraction.

Remarque. Les mots exclus sont automatiquement stockés dans la première bibliothèque répertoriée dans l'arborescence de bibliothèques de l'Editeur de ressources—par défaut, il s'agit de la *bibliothèque locale*.

Extraction de mots imposée

Lors de l'examen des données textuelles dans le panneau Données après l'extraction, vous pouvez découvrir que certains mots ou expressions n'ont pas été extraits. Souvent, ces mots sont des verbes ou des adjectifs qui ne présentent pas d'intérêt pour vous. Cependant, il arrive que vous

souhaitez utiliser un mot ou une expression, qui n'a pas été extrait, comme partie intégrante d'une définition de catégorie.

Si vous souhaitez extraire ces mots et phrases, vous disposez de deux options :

- Imposer un terme dans une bibliothèque de types. [Pour plus d'informations, reportez-vous à la section Ajout des termes forcés dans le chapitre 10 sur p. 228.](#)
- Ajout direct de mots à une définition de catégorie existante. Cette option est généralement utilisée si la première n'a pas donné les résultats escomptés. [Pour plus d'informations, reportez-vous à la section Correspondance de texte dans les catégories dans le chapitre 6 sur p. 165.](#)

Important ! Le marquage d'un terme dans un dictionnaire comme étant imposé n'est pas infaillible. En effet, même si vous avez explicitement ajouté un terme à un dictionnaire, il arrive qu'il n'apparaisse pas dans le panneau Résultats d'extraction après la nouvelle extraction ou qu'il apparaisse bien, mais pas exactement tel que vous l'avez déclaré. Bien que cet événement soit rare, il peut avoir lieu lorsqu'un mot ou une expression a déjà été extrait dans le cadre d'une expression plus longue. Pour éviter cela, appliquez l'option de mise en correspondance Entier (pas de composés) à ce terme dans la déclaration de types. [Pour plus d'informations, reportez-vous à la section Ajout de termes dans le chapitre 10 sur p. 224.](#)

Catégorisation des données textuelles

Dans IBM® SPSS® Text Analytics for Surveys, vous pouvez créer des **catégories** qui représentent, essentiellement, des rubriques ou des concepts de niveau supérieur qui captureront les principales idées, connaissances et attitudes exprimées dans le texte.

Les catégories peuvent également avoir une structure hiérarchique, ce qui signifie qu'elles peuvent contenir des sous-catégories et que ces sous-catégories peuvent elles-même contenir des sous-catégories, et ainsi de suite. Vous pouvez importer des structures de catégories prédéfinies, nommées auparavant plans de codage, avec des catégories hiérarchiques ou bien créer ces catégories hiérarchiques à l'intérieur du produit.

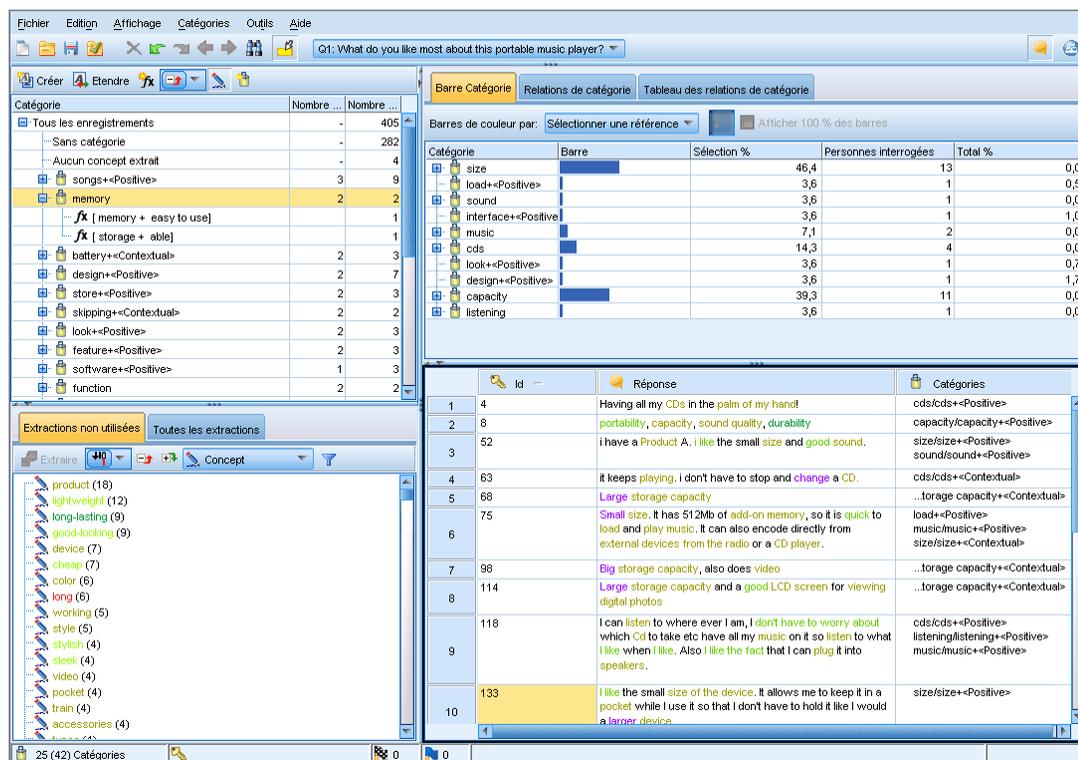
En effet, les catégories hiérarchiques vous permettent de créer une structure en arborescence contenant une ou plusieurs sous-catégories afin d'y regrouper plus clairement les éléments par groupe de concepts ou de rubriques. Prenons un exemple simple avec des activités de loisir ; pour répondre à une question telle que *Quelle activité aimeriez-vous faire si vous aviez davantage de temps ?* vous pourriez avoir des catégories principales comme *sport, art et artisanat, pêche, etc.* ; à un niveau inférieur, sous *sport*, vous pourriez avoir des sous-catégories pour voir s'il s'agit de *jeux de ballon, sports nautiques, etc.*

Les catégories sont constituées d'un ensemble de descripteurs, tels que des *concepts*, des *types*, des *patrons* et des *règles de catégorie*. Ensemble, ces descripteurs permettent d'identifier si un enregistrement appartient ou non à une catégorie. Le texte d'un enregistrement peut être analysé afin de déterminer s'il correspond à un descripteur. Si une correspondance est détectée, l'enregistrement est attribué à cette catégorie. Ce processus est appelé **catégorisation**.

Vous pouvez créer des catégories, les utiliser et les explorer à l'aide des données affichées dans les quatre panneaux, que vous pouvez masquer ou afficher en sélectionnant leur nom dans le menu Affichage.

- **Panneau Catégories.** Dans ce panneau, créez et gérez vos catégories. [Pour plus d'informations, reportez-vous à la section Le panneau Catégories sur p. 97.](#)
- **Panneau Résultats d'extraction.** Dans ce panneau, explorez et utilisez les concepts et les types extraits. [Pour plus d'informations, reportez-vous à la section Résultats extraits : Concepts, types et patrons dans le chapitre 5 sur p. 80.](#)
- **Panneau Visualisation.** Dans ce panneau, explorez visuellement vos catégories et analysez la manière dont elles interagissent. [Pour plus d'informations, reportez-vous à la section Visualisation des graphiques dans le chapitre 7 sur p. 170.](#)
- **Panneau Données.** Dans ce panneau, explorez et passez en revue le texte contenu dans les enregistrements qui correspondent aux sélections effectuées. [Pour plus d'informations, reportez-vous à la section Le panneau Données sur p. 100.](#)

Figure 6-1
Vue Question



Vous pouvez commencer avec un ensemble de catégories provenant d'un package d'analyse de texte (TAP), ou importer depuis un fichier de catégories prédéfinies, vous pouvez également avoir besoin de créer votre propre ensemble. Les catégories peuvent être créées automatiquement à l'aide des techniques fiables et automatisées qui utilisent les résultats d'extraction (concepts, types et patrons) pour générer des catégories et leurs descripteurs. Les catégories peuvent aussi être créées manuellement en utilisant des informations supplémentaires que vous pouvez avoir au sujet des données. Vous pouvez également créer des définitions de catégorie manuellement en faisant glisser les résultats de l'extraction dans les catégories. Vous pouvez enrichir ces catégories ou une catégorie vide par l'ajout de règles de catégorie à une catégorie, par l'utilisation de vos propres catégories prédéfinies, par l'ajout d'un mot ou de groupes de mots n'ayant jamais été extraits (**mis en correspondance de texte**), par l'intégration forcée de réponses directement dans une catégorie ou par une combinaison.

Chaque technique convient à un type de données et à certaines situations. Cependant, il est souvent judicieux de combiner plusieurs méthodes dans la même analyse afin de capturer le maximum d'enregistrements. En cours de catégorisation, vous pouvez également envisager d'apporter d'autres modifications aux ressources linguistiques.

Le panneau Catégories

Le panneau Catégories est la zone dans laquelle vous pouvez créer et gérer vos catégories. Ce panneau est situé en haut à gauche de la vue Question et est accessible à partir du menu AffichageAffichage > Question > "Votre_Question"). Après avoir extrait les concepts et les types de

vos données textuelles, vous pouvez créer des catégories automatiquement à l'aide de technique telles que l'inclusion de concept, la co-occurrence, etc. ou vous pouvez les créer manuellement. [Pour plus d'informations, reportez-vous à la section Création de catégories sur p. 112.](#)

Figure 6-2
panneau Catégories avec et sans catégories

Catégorie	Nombre d...	Nombre d...
Tous les enregistrements	-	200
Sans catégorie	-	200
Aucun concept extrait	-	200

Catégorie	Nombre ...	Nombre ...
Tous les enregistrements	-	405
Sans catégorie	-	282
Aucun concept extrait	-	4
songs+«Positive»	3	9
memory	2	2
battery+«Contextual»	2	3
design+«Positive»	2	7
store+«Positive»	2	3
skipping+«Contextual»	2	2
look+«Positive»	2	3
feature+«Positive»	2	3
software+«Positive»	1	3
function	2	2
playlists+«Positive»	1	2
battery life+«Positive»	1	5

Chaque fois qu'une catégorie est créée ou mise à jour, les enregistrements sont analysés automatiquement pour déterminer si le texte correspond à un descripteur dans une catégorie donnée. Si une correspondance est détectée, le l'enregistrement est attribué à cette catégorie. Le résultat final est que la plupart, sinon l'intégralité, des enregistrements est affectée à des catégories en fonction des descripteurs des catégories.

Table d'arborescence des catégories

La table d'arborescence de ce panneau présente l'ensemble de catégories, des sous-catégories et des descripteurs. L'arborescence possède également plusieurs colonnes présentant des informations pour chaque élément de l'arborescence. Les colonnes pouvant être affichées sont les suivantes :

- **Code.** Répertorie la valeur de code de chaque catégorie. Cette colonne est sélectionnée par défaut. Vous pouvez afficher cette colonne en cliquant avec le bouton droit de la souris dans la table d'arborescence et en sélectionnant **Affichage > Code de catégorie**.
- **Catégorie.** Contient l'arborescence des catégories avec le nom de la catégorie et des sous-catégories. De plus si l'on clique sur l'icône de la barre d'outils des descripteurs, l'ensemble des descripteurs sera aussi affiché.
- **Descripteurs.** Donne le nombre de descripteurs qui la définissent. Ce nombre ne comprend pas le nombre de descripteurs dans les sous-catégories. Aucun nombre n'est donné lorsqu'un nom de descripteur apparaît dans la colonne Catégories. Vous pouvez afficher cette colonne en cliquant avec le bouton droit de la souris dans la table d'arborescence et en sélectionnant **Affichage > Tous les descripteurs**.
- **Docs.** Après le scoring, cette colonne affiche le nombre d'enregistrements qui sont catégorisés dans une catégorie et toutes ses sous-catégories. Ainsi, si 5 enregistrements correspondent à votre catégorie de niveau supérieur en fonction de ses descripteurs et 7 autres enregistrements correspondent à une sous-catégorie en fonction de ses descripteurs, le nombre de documents total pour la catégorie de niveau supérieur est la somme des deux : dans ce cas, ce nombre est

de 12. Mais, si le même enregistrement correspondait à la catégorie de niveau supérieure et à sa sous-catégorie, alors ce nombre serait 11.

S'il n'existe aucune catégorie, le tableau contient tout de même deux lignes. La première ligne, Tous les enregistrements, correspond au nombre total d'enregistrements. Une seconde ligne, Sans catégorie, représente le nombre de documents/d'enregistrements devant être catégorisés.

Pour chaque catégorie présente dans le panneau, une icône représentant un petit seau jaune précède le nom de la catégorie. Si vous double-cliquez sur une catégorie, ou cliquez avec le bouton droit de la souris dans la table d'arborescence et sélectionnez Définitions des catégories, la boîte de dialogue Définitions des catégories s'ouvre et présente tous les éléments, appelés **descripteurs**, qui la définissent, comme les concepts, les types, les patrons et les règles de catégorie. [Pour plus d'informations, reportez-vous à la section A propos des catégories sur p. 110.](#) Par défaut, la table d'arborescence des catégories n'affiche pas les descripteurs dans les catégories. Si vous souhaitez voir les descripteurs directement dans l'arborescence plutôt que dans la boîte de dialogue Définitions de catégories, cliquez sur le bouton bascule à l'aide de l'icône du stylo dans la barre d'outils. Lorsque le bouton bascule est sélectionné, vous pouvez développer votre arborescence pour afficher également les descripteurs.

Scoring des catégories

La colonne Documents, dans la table d'arborescence des catégories affiche le nombre d'enregistrements qui sont catégorisés dans cette catégorie spécifique. Si les nombres sont périmés ou ne sont pas calculés, une icône apparaît dans cette colonne. Gardez à l'esprit que le processus de scoring peut prendre un certain temps lorsque vous utilisez des ensembles de données volumineux.

Sélection des catégories dans l'arborescence

Lorsque vous effectuez des sélections dans l'arborescence, vous ne pouvez sélectionner que des catégories Frère, c'est-à-dire, si vous sélectionnez des catégories de niveau supérieur, vous ne pouvez pas également sélectionner une sous-catégorie. Ou si vous sélectionnez 2 sous-catégories d'une catégorie donnée, vous ne pouvez pas sélectionner en même temps une sous-catégorie d'une autre catégorie. La sélection d'une catégorie discontinue provoquera la perte de votre sélection précédente.

Affichage dans les panneaux Données et Visualisation

Lorsque vous sélectionnez une ligne dans la table, les panneaux Visualisation et Données sont automatiquement actualisés et affichent les informations correspondant à votre sélection.

Réglage de vos catégories

La catégorisation peut ne pas générer des résultats parfaits pour vos données lors de votre première tentative, et il se peut que vous souhaitiez supprimer des catégories ou les combiner avec d'autres catégories. Vous pouvez également vous apercevoir, en consultant les résultats de l'extraction, que certaines catégories n'ayant pas été créées vous seraient utiles. Dans ce cas, vous pouvez apporter des modifications manuelles aux résultats afin de les adapter à votre contexte. [Pour plus d'informations, reportez-vous à la section Edition et réglage des catégories sur p. 158.](#)

- Modifier ou ajouter aux définitions de catégories et déplacer, fusionner ou supprimer des catégories. Pour plus d'informations, reportez-vous à [Edition et réglage des catégories](#) ci-dessous.
- Forcer des ID de réponse spécifiques à entrer ou sortir des catégories. [Pour plus d'informations, reportez-vous à la section Forcer les réponses à entrer dans des catégories](#) sur p. 164.
- Ajoutez des correspondances de texte aux catégories pour capturer les réponses contenant le même texte dans la catégorie. [Pour plus d'informations, reportez-vous à la section Correspondance de texte dans les catégories](#) sur p. 165.
- Ajouter des règles de catégorie à une catégorie pour classifier automatiquement les réponses dans une catégorie en fonction d'une expression logique. [Pour plus d'informations, reportez-vous à la section Utilisation des règles de catégorie](#) sur p. 147.
- Visualiser l'interaction de vos catégories. [Pour plus d'informations, reportez-vous à la section Visualisation des graphiques dans le chapitre 7](#) sur p. 170.
- Exporter les résultats de votre catégorisation. [Pour plus d'informations, reportez-vous à la section Export des résultats de catégorisation dans le chapitre 4](#) sur p. 55.

Le panneau Données

Lorsque vous créez des catégories, vous pouvez parfois souhaiter examiner certaines des données textuelles utilisées. Par exemple, si vous créez une catégorie dans laquelle 640 enregistrements sont catégorisés, vous pouvez souhaiter consulter une partie ou l'intégralité de ces enregistrements afin de découvrir le texte qui était effectivement rédigé. Vous pouvez consulter les enregistrements dans le panneau Données, situé dans l'angle inférieur droit. S'il n'apparaît pas par défaut, sélectionnez Affichage > Panneaux > Données dans les menus.

Ce panneau présente, sous forme de tableau, les enregistrements des réponses de vos données ouvertes. En fonction de ce qui est sélectionné dans les autres panneaux de cette vue, seuls les enregistrements correspondants s'affichent dans le panneau. Par exemple, si vous sélectionnez un concept dans le panneau Résultats d'extraction, seuls les enregistrements contenant ce concept (et les termes associés) s'affichent dans le panneau Données.

Lorsque vous sélectionnez un concept ou une catégorie dans un autre panneau et que vous affichez les données, les concepts (et les termes associés) présents dans ces enregistrements sont mis en surbrillance en couleur afin que vous puissiez les repérer plus facilement dans le texte. Le code couleur correspond aux types auxquels les concepts appartiennent. Vous pouvez également pointer sur un élément faisant l'objet d'un codage couleur pour afficher le concept sous lequel il a été extrait et le type auquel il a été affecté. Tout texte n'ayant pas été extrait apparaît en noir. En règle générale, ces mots non extraits sont souvent des connecteurs (*et* ou *avec*), des pronoms (*me* ou *ils*), et des verbes (*être*, *avoir* ou *prendre*).

Figure 6-3
panneau Données

	Id	Réponse	Catégories
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 playlists
10	121	Storage capacity (40GB)	...vice/memory/storage capacity
11	146	It has a lot of storage capacity. I can fit a lot of songs on it. Also, it's very lightweight.	...vice/memory/storage capacity songs
	151	This has 256MB of memory; it holds about 50 songs. I've got another chip in my bag with another 50 songs on it. The cool thing about this is it cost \$200 plus \$60 for a portable	music computer network listening

Remarque : Pour afficher tous les enregistrements pour une question donnée dans le panneau Données, cliquez sur le noeud Tous les enregistrements en haut du panneau Catégories.

Par défaut, le panneau Données affiche trois colonnes (ID, Réponse et Catégories). Vous pouvez néanmoins ajouter des colonnes supplémentaires à ce panneau. Les colonnes possibles sont les suivantes :

- **ID.** Répertorie l'identificateur de l'enregistrement ou du document (ID), si un enregistrement ou un document a été importé.
- **Réponse.** Répertorie les données textuelles à partir desquelles les concepts et le type ont été extraits.
- **Catégories.** Répertorie chacune des catégories à laquelle appartient un enregistrement. Lorsque cette colonne est affichée, l'actualisation du panneau Données peut prendre plus de temps afin d'afficher les informations les plus récentes. Les catégories sont répertoriées dans cette colonne en fonction de leur correspondance avec l'enregistrement. [Pour plus d'informations, reportez-vous à la section Pertinence des catégories sur p. 102.](#)
- **Forcer à entrer.** Répertorie les catégories dans lesquelles vous avez forcé une réponse à entrer. Vous pouvez forcer les réponses à entrer dans la catégorie à travers la sélection de menusEdition > Forcer à entrer. [Pour plus d'informations, reportez-vous à la section Forcer les réponses à entrer dans des catégories sur p. 164.](#)
- **Forcer à sortir.** Répertorie les catégories à partir desquelles vous avez forcé une réponse à sortir. Vous pouvez forcer les réponses à sortir d'une catégorie à travers la sélection de menusEdition > Forcer à sortir. Ceci est généralement utilisé lorsque le sarcasme d'une personne interrogée

entraîne la mauvaise catégorisation d'une réponse. [Pour plus d'informations, reportez-vous à la section Forcer les réponses à entrer dans des catégories sur p. 164.](#)

- **Correspondance de texte.** Répertorie toute Correspondance de texte trouvée pour chaque réponse. Des chaînes peuvent être définies pour forcer du texte spécifique à faire partie d'une définition de catégorie, que cette chaîne ait été extraite ou non. [Pour plus d'informations, reportez-vous à la section Correspondance de texte dans les catégories sur p. 165.](#)
- **Effectifs des catégories.** Donne le nombre total de catégories auxquelles appartient la réponse à cette question.
- **Rang de pertinence.** Donne un rang pour chaque enregistrement dans une seule catégorie. Ce rang montre dans quelle mesure l'enregistrement correspond à la catégorie par rapport aux autres enregistrements dans cette catégorie. Sélectionnez une catégorie dans le panneau Catégories (panneau supérieur gauche) pour voir le rang dans cette colonne. [Pour plus d'informations, reportez-vous à la section Pertinence des catégories sur p. 102.](#)
- **Booléens de réponse.** Ajoute une colonne qui montre tous les booléens de réponse que vous pouvez utiliser. En cliquant dans cette colonne, vous pouvez changer le type de booléen que vous attribuez à chaque réponse.
- **<tous les noms de variables de référence>.** Ajoute une colonne au panneau Données pour la variable de référence sélectionnée. Si vous n'avez importé aucune variable de référence, aucune colonne ne sera affichée dans ce cas. Une colonne distincte est disponible pour chaque variable de référence. [Pour plus d'informations, reportez-vous à la section Sélection de variables dans le chapitre 3 sur p. 35.](#)

Pour afficher les autres colonnes du panneau Données

- ▶ Dans le panneau Données, cliquez avec le bouton droit de la souris sur l'en-tête de colonne pour ouvrir un menu contextuel.
- ▶ Dans le menu, cliquez sur Afficher les colonnes, puis sélectionnez la colonne que vous souhaitez afficher dans le panneau Données. La nouvelle colonne apparaît dans le panneau.

Remarque : Le fait de forcer les réponses à entrer et sortir des catégories vous permet de remplacer les définitions de catégorie créées par les techniques de création de catégorie automatisées sans modifier la définition de catégorie actuelle. [Pour plus d'informations, reportez-vous à la section Forcer les réponses à entrer dans des catégories sur p. 164.](#)

Pertinence des catégories

Pour vous aider à créer de meilleures catégories, vous pouvez examiner la pertinence des enregistrements dans chaque catégorie ainsi que la pertinence de toutes les catégories auxquelles appartient un enregistrement.

Pertinence d'une catégorie pour un enregistrement

Quand un enregistrement s'affiche dans le panneau Données, toutes les catégories auxquelles il appartient sont répertoriées dans la colonne Catégories. Quand un enregistrement appartient à plusieurs catégories, les catégories dans cette colonne s'affichent dans l'ordre de la correspondance la plus pertinente à la moins pertinente. La première catégorie est considérée

comme correspondant le mieux à cet enregistrement. [Pour plus d'informations, reportez-vous à la section Le panneau Données sur p. 100.](#)

Pertinence d'un enregistrement pour une catégorie

Quand vous sélectionnez une catégorie, vous pouvez examiner la pertinence de chacun de ses enregistrements dans la colonne Rang de pertinence dans le panneau Données. Ce rang de pertinence indique à quel point l'enregistrement correspond à la catégorie sélectionnée par rapport aux autres enregistrements de cette catégorie. Pour voir le rang des enregistrements pour une seule catégorie, sélectionnez celle-ci dans le panneau Catégories (panneau supérieur gauche) et le rang de l'enregistrement s'affiche dans la colonne. Cette colonne n'est pas visible par défaut mais vous pouvez choisir de l'afficher. [Pour plus d'informations, reportez-vous à la section Le panneau Données sur p. 100.](#)

Plus le rang de l'enregistrement est bas, plus il correspond à la catégorie sélectionnée de façon à ce que 1 soit la meilleure correspondance. Si plusieurs enregistrements ont la même pertinence, chacun est affiché avec le même rang suivi d'un signe égal (=) pour montrer qu'ils ont la même pertinence. Par exemple, vous pouvez avoir les rangs suivants 1=, 1=, 3, 4, etc., ce qui signifie qu'il existe deux enregistrements considérés de manière égale comme étant les meilleures correspondances pour cette catégorie.

Astuce : Vous pouvez ajouter le texte de l'enregistrement le plus pertinent à l'annotation de catégorie pour donner une meilleure description de la catégorie. Ajoutez le texte directement à partir du panneau Données en sélectionnant le texte et en choisissant Catégories > Ajouter à l'annotation dans le menu.

Figure 6-4
 Panneau Données affichant les catégories et le rang de pertinence

	Id	Réponse	Catégories	Rang d'importance
1	2	chambres exigues. pas de service de blanchisserie	chambres	10=
2	3	trop de bruit. pas de chambre fumeur	chambres	22=
3	5	les chambres	chambres	18=
4	16	La chambre donné sur une cour crasseuse	chambres	22=
5	18	La femme de ménage est entrée sans frapper à 10 du matin croyant qu'on avait libéré la chambre (erreur de numéro)	chambres	15=
6	22	Les petits déjeuners ne sont pas à la hauteur du prix payé: les viennoiseries ne sont pas toujours très fraîches et les buffets ne sont pas régulièrement réapprovisionnés, ce qui fait qu'on est constamment obligé de réclamer. Le ménage des chambres laisse lui aussi à désirer.	buffets petit-déjeuner chambres	28
7	28	Ca pu la clope dans la chambre et le matelas a bien 36 ans.	chambres	15=
8	34	ma chambre donnait sur les immeubles. J'aurais préféré avoir une vue sur le parc.	chambres	27
9	47	il n'y a pas d'ascenseur et ma chambre était au 11ème étage	chambres	13=
10	60	isolation nul, on entend tout dans les chambres	isolation chambres	22=
11	64	isolation phonique des chambres a revoir	chambres	18=
12	75	Les chambres sont un peu trop petites. La propreté de la salle de bains laisse a désirer.	chambres salle de bains	13=
13	76	Les chambres ne sont pas entretenues: le papier peint n'est pas d'actualité, la television et le radio-veille ne fonctionnent pas correctement. L'ascenseur est trop petit, et trop lent.	chambres télévision	17
14	77	restauration. propreté des chambres	chambres	10=
15	91	chambres trop petites	chambres	1=
16	95	chambre trop étroite	chambres	1=
17	121	climatisation défectueuse dans les chambres	climatisation chambres	22=
18	122	chambre trop étroite, du moins la mienne;-)	chambres	1=
19	123	service en chambre trop long. j'ai attendu mon repas plus d'une heure	repas chambres	22=

Stratégies et méthodes de création de catégories

Si vous n'avez pas encore effectué d'extraction ou que vos résultats d'extraction ne sont pas à jour, l'utilisation de l'une des techniques de création ou d'extension de catégorie vous invitera automatiquement à effectuer une extraction. Après avoir appliqué une technique, les concepts et les types qui ont été regroupés dans une catégorie restent disponibles et pourront être classés par le biais d'autres techniques. Cela signifie que vous pouvez voir un concept dans plusieurs catégories sauf si vous choisissez de ne pas les réutiliser.

Afin de vous aider à créer les catégories les plus pertinentes, veuillez examiner les points suivants :

- Méthodes de création de catégories
- Stratégies de création de catégories
- Conseils pour la création de catégories

Méthodes de création de catégories

Puisque chaque ensemble de données est unique, le nombre de méthodes de création de catégories et l'ordre dans lequel vous les appliquez peut varier. De plus, dans la mesure où vos objectifs d'exploration de texte peuvent être différents d'un ensemble de données à un autre, vous pouvez tester les différentes méthodes afin de déterminer celle qui génère les meilleurs résultats pour les données textuelles en question. Aucune des techniques automatiques ne pourra parfaitement catégoriser vos données; c'est pourquoi nous recommandons de trouver et d'appliquer une ou plusieurs techniques automatiques qui correspondent aux besoins de vos données.

Outre l'utilisation de packages d'analyse de texte (TAP, *.tap) avec des ensembles de catégories prédéfinis, vous pouvez également catégoriser vos réponses à l'aide de toute combinaison des méthodes suivantes :

- **Techniques de création automatique.** Plusieurs options de catégorie basées sur la linguistique et la fréquence sont disponibles pour créer automatiquement des catégories. [Pour plus d'informations, reportez-vous à la section Création de catégories sur p. 112.](#)
- **Techniques d'extension automatiques.** Plusieurs techniques linguistiques sont disponibles pour étendre les catégories existantes en ajoutant et en améliorant les descripteurs afin qu'ils capturent davantage d'enregistrements. [Pour plus d'informations, reportez-vous à la section Extension de catégories sur p. 128.](#)
- **Techniques manuelles.** Il existe plusieurs méthodes manuelles, telles que le glisser-déposer. [Pour plus d'informations, reportez-vous à la section Création de catégories manuellement sur p. 133.](#)
- **Catégories.** Importez vos propres catégories ou copiez/collez dans le plan de codage. [Pour plus d'informations, reportez-vous à la section Import de catégories prédéfinies sur p. 136.](#)

Stratégies de création de catégories

La liste de stratégies suivante n'est en aucun cas exhaustive mais elle peut vous donner une idée de l'approche de la création de vos catégories.

- Lorsque vous démarrez un projet, sélectionnez un ensemble de catégories dans un package d'analyse de texte (TAP) afin de débiter votre analyse avec des catégories prédéfinies. Il est possible que ces catégories suffisent à l'analyse de votre texte depuis le début. Cependant, si vous souhaitez ajouter d'autres catégories, vous pouvez modifier les paramètres de création de catégories (Catégories > Configurer les paramètres). Ouvrez la boîte de dialogue Paramètres avancés : Linguistique et choisissez l'option d'entrée Catégorie Résultats d'extraction inutilisés et créez les catégories supplémentaires.
- Lorsque vous démarrez un projet, sélectionnez un ensemble de catégories dans un TAP. Ensuite, faites glisser et déposer les concepts ou patrons inutilisés dans les catégories qui vous semblent appropriées. Ensuite, étendez les catégories existantes que vous venez de modifier (Catégories > Étendre les catégories) pour obtenir davantage de descripteurs associés aux descripteurs de catégorie existants.
- Créez des catégories automatiquement à l'aide des paramètres linguistiques avancés (Catégories > Créer des catégories). Ensuite, affinez manuellement les catégories en supprimant des descripteurs, des catégories ou en fusionnant des catégories similaires jusqu'à ce que vous soyez satisfait des catégories résultantes. De plus, si vous créez des catégories **sans** utiliser l'option Généraliser avec des caractères génériques lorsque cela est possible,

vous pouvez également essayer de simplifier automatiquement les catégories à l'aide de la fonctionnalité Etendre les catégories, en activant l'option Généraliser.

- Importez un fichier de catégories prédéfinies avec des noms de catégories très descriptifs et/ou des annotations. De plus, si à l'origine vous importez **sans** choisir l'option d'importer ou de générer des descripteurs à partir des noms de catégorie, vous pouvez ensuite utiliser la boîte de dialogue Etendre les catégories et choisir l'option Etendre les catégories vides avec des descripteurs générés à partir du nom de la catégorie. Puis, étendez ces catégories une deuxième fois, mais utilisez alors les techniques de regroupement.
- Créez manuellement un premier ensemble de catégories en triant les concepts ou patrons de concept par fréquence puis en faisant glisser et déposer les plus intéressants dans le panneau Catégories. Une fois que vous avez cet ensemble initial de catégories, utilisez la fonction Etendre (Catégories > Etendre les catégories) pour développer et affiner toutes les catégories sélectionnées afin qu'elles comprennent les autres descripteurs associés et correspondent ainsi à plus d'enregistrements.

Après avoir appliqué ces techniques, nous vous recommandons de passer en revue les catégories résultantes et d'utiliser des techniques manuelles pour effectuer des changements mineurs, supprimer toute mauvaise réaffectation ou ajouter les enregistrements ou mots ayant été oubliés. En outre, puisque l'utilisation de techniques différentes peut être à l'origine de catégories redondantes, vous avez la possibilité de fusionner ou de supprimer des catégories si nécessaire. [Pour plus d'informations, reportez-vous à la section Edition et réglage des catégories sur p. 158.](#)

Conseils pour la création de catégories

Afin de vous aider à créer de meilleures catégories, vous pouvez examiner certains conseils qui peuvent vous aider à prendre des décisions à propos de votre approche.

Conseils concernant le rapport Catégorie-Réponse

Lorsque les codes sont créés pour une question fermée, telle que "*Quand avez-vous rendu visite à notre magasin pour la dernière fois ?*", les catégories auxquelles les réponses appartiennent doivent être mutuellement exclusives et exhaustives. Ce principe ne s'applique pas nécessairement aux analyses de texte qualitatives, pour au moins deux raisons :

- En premier lieu, il est généralement admis que plus un enregistrement texte est long, plus les idées et opinions qu'il exprime sont diverses. Ainsi, un enregistrement a plus de chances de se voir attribuer plusieurs catégories.
- En second lieu, il existe souvent plusieurs manières de regrouper ou d'interpréter des enregistrements texte qui ne sont pas logiquement distincts. Dans le cas d'une enquête comportant une question ouverte sur les opinions politiques de la personne interrogée, nous pourrions créer des catégories, telles que *libéral* et *conservateur*, ou *Républicain* et *Démocrate*, ainsi que des catégories plus spécifiques, telles que *libéral pour l'aspect social*, *conservateur pour l'aspect fiscal* et ainsi de suite. Il n'est pas nécessaire que ces catégories s'excluent mutuellement, ni qu'elles soient exhaustives.

Conseils sur le nombre de catégories à créer

A l'exception des cas où une question ouverte extrêmement simple est posée, il n'est jamais évident de savoir combien de catégories créer. Le nombre de catégories n'est pas d'une réelle importance. Il est plus important que la création de catégories découle directement des données—lorsque vous notez un point intéressant concernant les objectifs de votre enquête, vous pouvez créer une catégorie qui représente ces attitudes et ces idées.

- **Fréquence des catégories.** Pour qu'une catégorie soit utile, elle doit contenir une quantité minimale de d'enregistrements. Un ou deux enregistrements peuvent contenir des éléments intéressants. Cependant, si cela ne concerne qu'un ou deux enregistrements sur 1 000, les informations qu'ils contiennent, peuvent ne pas être pas assez fréquentes par rapport à la quantité de documents pour présenter une utilité pratique.
- **Complexité.** Plus vous créez de catégories, plus vous avez d'informations à examiner et à résumer après analyse. Toutefois, un trop grand nombre de catégories risque de créer une trop grande complexité sans apporter de détails utiles.

Malheureusement, aucune règle ne détermine à partir de quel nombre les catégories sont jugées trop nombreuses, ni ne définit le nombre minimal d'enregistrements par catégorie. C'est à vous d'effectuer ces déterminations, en fonction des exigences de votre situation.

Nous pouvons toutefois vous fournir nos conseils et vous indiquer où commencer. Bien que le nombre de catégories ne doive pas être trop élevé, il est préférable d'en avoir trop que pas assez lors des premières étapes de l'analyse. Il est plus simple de regrouper des catégories relativement similaires que de diviser des observations en catégories nouvelles. La stratégie consistant à travailler d'un plus grand nombre à un moins grand nombre de catégories est donc généralement la meilleure. Etant donné la nature itérative du Text Mining et la facilité avec laquelle il peut être exécuté avec ce logiciel, il est acceptable de créer plus de catégories au départ.

Choix des meilleurs descripteurs

Les informations suivantes vous fourniront quelques conseils afin de choisir ou de créer les meilleurs descripteurs (concepts, types, patrons TLA et règles de catégorie) pour vos catégories. Les descripteurs sont les blocs de construction des catégories. Lorsqu'une partie ou l'ensemble du texte d'un enregistrement correspond à un descripteur, l'enregistrement est mis en correspondance avec la catégorie.

Un descripteur n'est mis en correspondance avec des enregistrements que s'il contient ou correspond à un concept ou à un patron extrait. Par conséquent, utilisez les concepts, les types, les patrons et les règles de catégorie de la manière indiquée dans les paragraphes suivants.

Les concepts représentant un ensemble de termes sous-jacents (en plus de se représenter eux-mêmes) qui peuvent comprendre un ensemble vaste de termes allant des formes singulier/pluriel, aux synonymes, ou aux variations orthographiques, seul le concept lui-même doit être utilisé en tant que descripteur ou comme partie d'un descripteur. Pour en savoir plus sur les termes sous-jacents de chaque concept donné, cliquez sur le nom du concept dans le panneau Résultats d'extraction. Lorsque vous placez la souris sur le nom d'un concept, une info-bulle apparaît et affiche tous les termes sous-jacents trouvés dans votre texte lors de la dernière extraction. Les concepts ne possèdent pas tous des termes sous-jacents. Par exemple, si `voiture` et `véhicule` sont des synonymes mais que `voiture` est extrait comme concept et `véhicule` comme terme sous-jacent, alors vous devez utiliser `voiture` comme descripteur,

puisqu'il permettra de mettre en correspondance les enregistrements qui contiennent également le terme `véhicule`.

Les concepts et les types utilisés comme descripteurs

Vous devez utiliser un concept comme descripteur lorsque vous souhaitez trouver tous les enregistrements contenant ce concept (ou ses termes sous-jacents). Dans ce cas, l'utilisation d'une règle de catégorie plus complexe n'est pas nécessaire puisque le nom du concept exact est suffisant. Souvenez-vous que lorsque vous utilisez des ressources qui extraient des opinions, il se peut que les concepts changent lors de l'extraction des patrons TLA, afin de capturer le sens exact de la phrase (reportez-vous à l'exemple de la section qui suit concernant les patrons TLA).

Par exemple, une réponse à une enquête indiquant les fruits préférés de chaque personne telles que "*Les pommes et l'ananas sont les meilleurs fruits*" peut engendrer l'extraction de `pommes` et `ananas`. En ajoutant le concept `pommes` en tant que descripteur à votre catégorie, toutes les réponses contenant le concept `pommes` (ou ses termes sous-jacents) sont mises en correspondance avec cette catégorie.

Toutefois, si vous voulez juste connaître les réponses qui mentionnent le terme *pommes* de quelque manière que ce soit, vous pouvez écrire une règle de catégorie, comme par exemple `* pommes *`, pour capturer toutes les réponses contenant des concepts tels que `pommes`, `jus de pommes`, ou `tarte aux pommes`.

Vous pouvez également capturer tous les enregistrements qui contiennent des concepts de même type en utilisant un type directement en tant que descripteur comme par exemple `<Fruit>`. Remarque : vous ne pouvez pas utiliser la fonction `*` avec les types.

Pour plus d'informations, reportez-vous à la section [Résultats extraits : Concepts, types et patrons dans le chapitre 5 sur p. 80](#).

Patrons d'analyse des liens du texte (TLA) utilisés comme descripteurs

Utilisez un résultat de patron TLA comme descripteur si vous souhaitez capturer des idées plus nuancées ou plus fines. Lorsque le texte est analysé lors de l'extraction TLA, le texte (enregistrement) est traité phrase par phrase ou proposition par proposition, plutôt que dans son ensemble. En prenant en considération l'ensemble des parties d'une phrase, l'analyse des liens du texte peut identifier les opinions, les relations entre deux éléments, ou une négation, par exemple et comprendre ainsi le sens exact de la phrase. Vous pouvez utiliser des patrons de concepts ou des patrons de type en tant que descripteurs.

Par exemple, si l'on considère le texte "*la pièce n'était pas propre*", les concepts suivants peuvent en être extraits : `pièce` et `propre`. Toutefois, si l'extraction TLA avait été activée dans les paramètres de l'extraction, elle aurait pu détecter que `propre` est utilisé de manière négative et correspond en fait au concept `pas propre`, qui est aussi synonyme du concept `sale`. Vous pouvez voir dans cet exemple que l'utilisation du concept `propre` seul, en tant que descripteur sera mis en correspondance avec ce texte, mais capturera aussi d'autres enregistrements qui mentionnent la propreté. Par conséquent, il peut être plus judicieux d'utiliser le patron de concept TLA `sale` comme concept de sortie, car il sera mis en correspondance avec ce texte et sera plus approprié en tant que descripteur.

Règles métier de catégories utilisées comme descripteurs

Les règles de catégorie sont des instructions qui classifient automatiquement les enregistrements en une catégorie basée sur une expression logique à l'aide de concepts, de types et de patrons extraits ou d'opérateurs booléens. Par exemple, vous pouvez créer une expression qui signifie *inclure tous les enregistrements contenant le concept extrait `ambassade` mais ne pas inclure `argentine` dans cette catégorie.*

Vous pouvez créer et utiliser des règles de catégorie en tant que descripteurs dans vos catégories pour exprimer différentes idées à l'aide de `&`, `|`, et `!` () Booléens. Pour obtenir des informations sur la syntaxe de ces règles et la manière de les rédiger et de les modifier, consultez [Utilisation des règles de catégorie sur p. 147](#)

- Utilisez une règle de catégorie avec l'opérateur booléen `&` (AND) pour trouver des enregistrements dans lesquels se trouvent 2 concepts ou plus. Les concepts connectés par des opérateurs `&` ne doivent pas obligatoirement se trouver dans la même phrase, ils peuvent apparaître n'importe où dans le même enregistrement pour être mis en correspondance avec la catégorie. Par exemple, si vous créez la règle de catégorie suivante : `nourriture & bon marché` comme descripteur, elle correspondra à un enregistrement comportant le texte *“la nourriture était très chère, mais les chambres étaient bon marché”* bien que le terme `nourriture` ne soit pas le nom auquel s'applique l'adjectif `bon marché`, et ce car le texte contient à la fois les termes `nourriture` et `bon marché`.
- Utilisez une règle de catégorie comportant l'opérateur booléen `!` () (NOT) comme descripteur pour trouver des enregistrements dans lesquels se trouvent certains termes mais pas d'autres Ceci peut permettre de ne pas regrouper des informations qui semblent être reliées en fonction des mots mais pas en fonction du contexte. Par exemple, si vous créez la règle de catégorie suivante : `<Société> & !(ibm)` comme descripteur, elle correspondra au texte suivant : *SPSS Inc. est une société fondée en 1967* et ne correspondra pas au texte suivant : *la société de logiciels a été achetée par IBM.*
- Utilisez une règle de catégorie avec l'opérateur booléen (OR) pour trouver des enregistrements contenant un concept ou type parmi plusieurs. Par exemple, si vous créez la règle de catégorie suivante : `(personnel|équipe|employés|collaborateurs) & mauvais` comme descripteur, elle correspondra à tout enregistrement dans lequel l'un de ces noms de trouvent en même temps que le concept de `mauvais`.
- Utilisez les types dans les règles de catégorie afin de les rendre plus génériques et plus déployables. Par exemple, si vous travaillez sur des données hôtelières, il se peut que vous soyez intéressé par l'opinion des clients sur le personnel de l'hôtel. Des termes associés peuvent inclure des mots comme `réceptionniste`, `serveur`, `serveuse`, `réception`, etc... Dans ce cas, vous pouvez créer un nouveau type nommé `<PersonnelHôtel>` et lui ajouter tous les termes précédents. Alors qu'il est possible de créer une règle de catégorie pour chaque type d'employés tels que `[* serveuse* & aimable]`, `[* réception* & amical]`, `[* réceptionniste * & accueillant]`, vous pouvez également créer une règle de catégorie unique, plus générale à l'aide du type `<PersonnelHôtel>` pour capturer toutes les réponses contenant des opinions positives sur le personnel de l'hôtel sous la forme de `[<PersonnelHôtel> & <Positif>]`.

Remarque : Vous pouvez utiliser à la fois `+` et `&` dans les règles de catégorie lorsque des patrons TLA sont inclus dans ces règles. [Pour plus d'informations, reportez-vous à la section Utilisation des patrons TLA dans les règles de catégorie sur p. 150.](#)

Exemple des différences de mise en correspondance lors de l'utilisation des concepts, des patrons TLA ou des règles de catégories comme descripteurs.

L'exemple suivant montre la manière dont l'utilisation des concepts, des règles de catégorie ou des patrons TLA en tant que descripteurs affecte la classification des enregistrements en catégories. Supposons que vous ayez les cinq enregistrements suivants :

- A : *“personnel du restaurant fantastique, nourriture excellente et chambres propres et confortables.”*
- B : *“le personnel du restaurant était horrible, mais les chambres étaient propres.”*
- C : *“Chambres confortables et propres.”*
- D : *“Ma chambre n'était pas très propre.”*
- E : *“Propre.”*

Dans la mesure où les enregistrements contiennent le mot *propre* et que vous souhaitez capturer cette information, vous pouvez créer un des descripteurs indiqué dans le tableau ci-dessous. En fonction de la nature des informations que vous essayez de capturer, vous pouvez utiliser un type de descripteur et voir les résultats produits.

Table 6-1
Mise en correspondance des exemples d'enregistrements avec les descripteurs

Descripteur	A	B	C	D	E	Explication
propre	<i>mise en correspondance</i>	Le descripteur est un concept extrait. Chaque enregistrement contient le concept <code>propre</code> , même l'enregistrement D puisque sans TLA, le concept <i>“pas propre”</i> ne signifie pas automatiquement <code>sale</code> dans les règles TLA.				
propre + .	-	-	-	-	<i>mise en correspondance</i>	Le descripteur est un patron TLA qui représente <code>propre</code> lui-même. Ne correspond qu'à l'enregistrement dans lequel <code>propre</code> a été extrait sans concept associé lors de l'extraction TLA.
[propre]	<i>mise en correspondance</i>	<i>mise en correspondance</i>	<i>mise en correspondance</i>	-	<i>mise en correspondance</i>	Le descripteur est une règle de catégorie qui cherche une règle TLA contenant <code>propre</code> seul ou associé à autre chose. Correspond à tous les enregistrements dans lesquels une sortie TLA contenant <code>propre</code> a été trouvée, que <code>propre</code> soit lié à un autre concept, comme par exemple <code>chambre</code> , ou non et indépendamment de sa position dans la phrase.

A propos des catégories

Les **catégories** font référence à un groupe de concepts, d'opinions ou d'attitudes étroitement associés. Pour être utile, une catégorie doit également être décrite par une étiquette ou une phrase courte évoquant l'essentiel de sa signification.

Par exemple, si vous analysez des réponses de consommateurs à une enquête sur une nouvelle lessive, vous pouvez créer une catégorie étiquetée *odeur* qui contient toutes les réponses décrivant l'odeur de ce produit. Mais cette catégorie ne fera pas la différence entre ceux qui ont aimé l'odeur et ceux qui ne l'ont pas aimée. Dans la mesure où IBM® SPSS® Text Analytics for Surveys permet d'extraire des opinions si vous utilisez les ressources appropriées, vous pouvez alors créer deux autres catégories qui identifieront les personnes interrogées qui *ont aimé l'odeur* et celles qui *n'ont pas aimé l'odeur*.

Vous pouvez créer et travailler avec vos catégories dans le panneau Catégories situé dans le panneau supérieur gauche de la fenêtre de l'analyse de texte. Chaque catégorie est définie par un ou plusieurs descripteurs. Les **descripteurs** sont des concepts, des types, des patrons et des règles de catégorie, utilisés pour définir une catégorie.

Si vous souhaitez voir les descripteurs qui forment une catégorie donnée, vous pouvez développer la catégorie dans l'arborescence. Des icônes s'affichent pour vous permettre d'identifier facilement chaque élément. Seul le premier niveau définit la catégorie. Si vous continuez à développer les définitions, vous trouverez des exemples de ce qui a été trouvé dans les données.

Lorsque vous créez automatiquement des catégories à l'aide des techniques de création, telles que l'inclusion de concept, les différentes techniques utiliseront les concepts et les types comme descripteurs pour créer vos catégories. vous pouvez ajouter ces patrons ou une partie d'entre eux en tant que descripteurs de catégorie. Enfin, vous pouvez créer manuellement des règles de catégorie à utiliser en tant que descripteurs dans vos catégories. [Pour plus d'informations, reportez-vous à la section Utilisation des règles de catégorie sur p. 147.](#)

Par exemple, si vous ajoutez un type à une définition de catégorie, les concepts affectés à ce type seront automatiquement inclus, même si les données n'en contiennent actuellement qu'un faible nombre. Cela permet de réutiliser les définitions de catégorie avec de nouvelles données. [Pour plus d'informations, reportez-vous à la section Copie de catégories sur p. 167.](#) Vous pouvez également créer manuellement des règles de catégorie à utiliser en tant que descripteurs dans vos catégories. [Pour plus d'informations, reportez-vous à la section Utilisation des règles de catégorie sur p. 147.](#)

Table 6-2

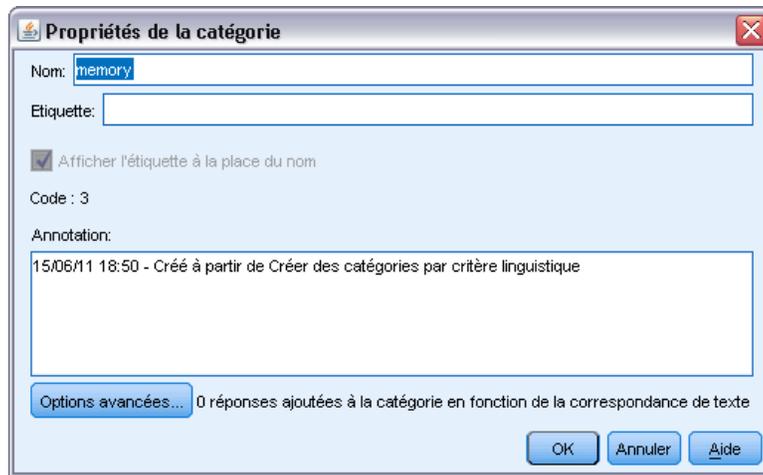
Icônes permettant d'identifier les éléments dans les définitions

Icône	Description
	Concept.
	Type, qui peut être développé pour voir les concepts qu'il contient.
	Patron de concept, qui peut être développé pour voir les concepts spécifiques des patrons.
	Patron de type, qui peut être développé jusqu'au niveau de patrons de concepts.
	Règles de catégorie dans la catégorie. Faites un clic droit sur le nom de la règle pour l'éditer.

Propriétés de la catégorie

En plus des descripteurs, les catégories ont également des propriétés que vous pouvez éditer afin de renommer les catégories, d'ajouter une étiquette ou une annotation, ou d'accéder au dialogue correspondant.

Figure 6-5
Boîte de dialogue Propriétés de catégorie



Les propriétés suivantes sont disponibles :

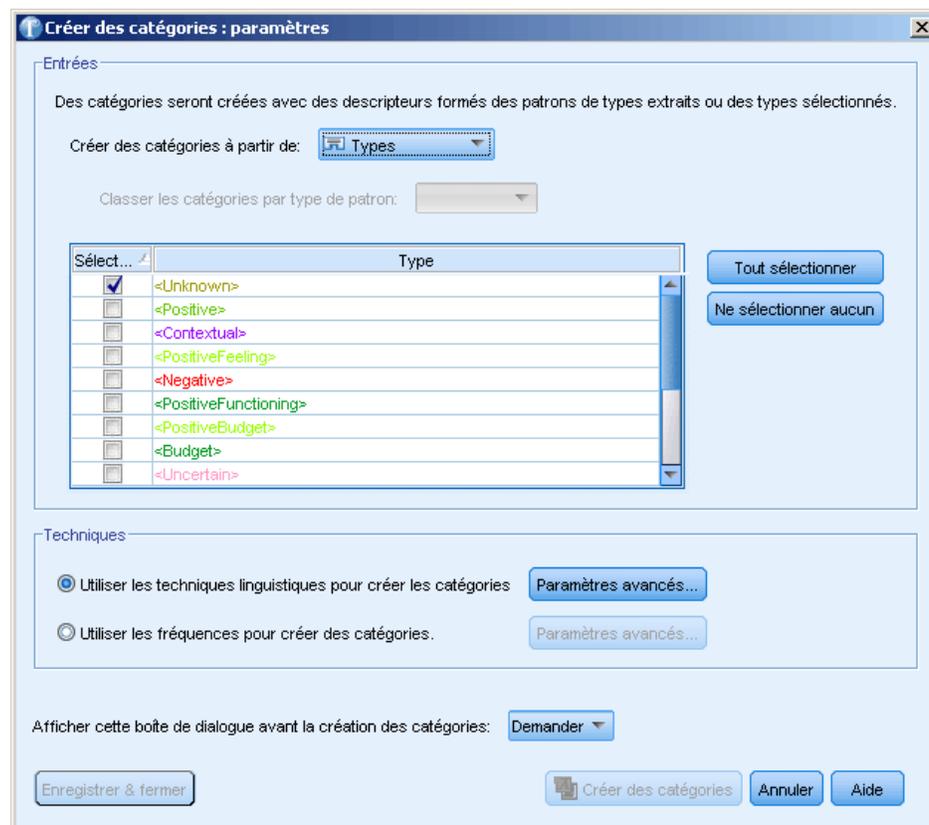
- **Nom.** Ce nom apparaît dans l'arborescence par défaut. Lorsqu'une catégorie est créée à l'aide d'une technique automatique, un nom lui est attribué automatiquement.
- **Etiquette.** L'utilisation d'étiquettes permet de créer des descriptions de catégorie plus significatives en vue de les utiliser dans d'autres produits, ou dans d'autres tableaux ou graphiques. Si vous sélectionnez l'option permettant d'afficher l'étiquette, celle-ci est alors utilisée dans l'interface pour désigner la catégorie. Dans IBM® SPSS® Statistics, les étiquettes sont exportées en tant qu'étiquettes de variable. Dans Microsoft Excel, elles sont exportées en tant que ligne distincte.
- **Code.** Le numéro de code correspond à la valeur de code de cette catégorie. Vous modifiez ce code dans le Plan de codage. Le Plan de codage vous permet d'éditer le nom, l'étiquette et le code de chaque catégorie et de copier et coller des catégories entières. .
- **Annotations.** Vous pouvez ajouter une description courte pour chaque catégorie dans ce champ. Lorsqu'une catégorie est générée par la boîte de dialogue Créer des catégories, une note est ajoutée automatiquement à cette annotation. Vous pouvez aussi ajouter un échantillon de texte à une annotation directement dans le panneau Données en sélectionnant le texte et en sélectionnant Catégories > Ajouter à l'annotation dans le menu.
- **Correspondance de texte avancée.** Le bouton Avancé permet d'ajouter des mots ou des phrases à une définition de catégorie. Cela permet généralement d'ignorer une extraction manquée. [Pour plus d'informations, reportez-vous à la section Correspondance de texte dans les catégories sur p. 165.](#)

Création de catégories

Bien que vous puissiez avoir des catégories d'un package d'analyse de texte, vous pouvez aussi créer automatiquement des catégories à l'aide de diverses techniques linguistiques et de fréquence. Dans la boîte de dialogue Créer des paramètres de catégorie, vous pouvez appliquer

les techniques linguistiques et de fréquence automatiques pour créer des catégories à partir de concepts ou de patrons de concept.

Figure 6-6
Boîte de dialogue Créer des catégories



En général, les catégories peuvent être constituées de différents types de descripteurs (types, concepts, patrons TLA, règles de catégorie). Quand vous créez des catégories à l'aide des techniques de création de catégorie automatiques, les catégories créées sont nommées selon un concept ou un patron de concept (en fonction de l'entrée que vous sélectionnez) et chacune d'elle contient un ensemble de descripteurs. Ces descripteurs peuvent avoir la forme de règles de catégorie ou de concepts et comprennent tous les concepts associés découverts par les techniques.

Après avoir créé les catégories, vous pouvez en apprendre beaucoup à leur sujet en les examinant dans le panneau Catégories ou en les explorant à travers les graphiques et les diagrammes. Vous pouvez utiliser des techniques manuelles pour effectuer des changements mineurs, supprimer toute mauvaise réaffectation, ou ajouter les enregistrements ou mots ayant été oubliés. Après avoir appliqué une technique, les concepts, les types et les patrons qui ont été regroupés dans une catégorie peuvent encore être classés par le biais d'autres techniques. En outre, puisque l'utilisation de techniques différentes peut également être à l'origine de catégories redondantes ou inappropriées, vous avez la possibilité de fusionner ou de supprimer des catégories. [Pour plus d'informations, reportez-vous à la section Edition et réglage des catégories sur p. 158.](#)

Important ! Dans les versions précédentes, les règles de cooccurrence et de synonymes étaient entre crochets. Dans la version actuelle, les crochets indiquent désormais un résultat de patrons. Les règles de cooccurrence et de synonymes sont maintenant entourées de parenthèses, comme dans (enceintes acoustiques|enceintes).

Pour créer des catégories

- ▶ Dans les menus, sélectionnez Catégories > Créer des catégories. Un message apparaît, sauf si vous avez choisi de ne pas recevoir d'invite.

Figure 6-7

Invite avant la création



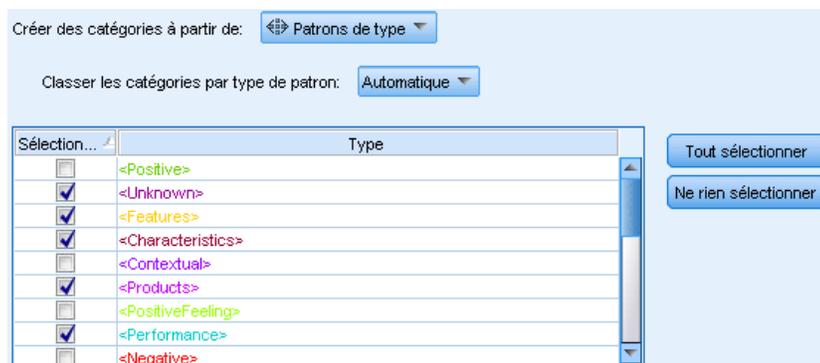
- ▶ Choisissez si vous voulez créer maintenant ou éditer d'abord les paramètres.
 - Cliquez sur Créer Maintenant pour commencer à créer des catégories à l'aide des paramètres actuels. Les paramètres sélectionnés par défaut sont souvent suffisants pour commencer le processus de catégorisation. Le processus de création de catégories commence et une boîte de dialogue de progression apparaît.
 - Cliquez sur Editer pour examiner et modifier les paramètres de création.

Entrées

Les catégories sont créées à partir de descripteurs dérivés de patrons de type ou de types. Par défaut, les patrons de type sont sélectionnés dans la liste déroulante. Dans le tableau, vous pouvez sélectionner les types ou patrons individuels à inclure dans le processus de création de catégories.

Patrons de type. Si vous sélectionnez les patrons de type, des catégories sont créées à partir de patrons plutôt qu'à partir de types et de concepts uniquement. De cette manière, tous les enregistrements contenant un patron de concept appartenant au patron de type sélectionné sont classés en catégories. Ainsi, si vous sélectionnez le patron de type <Budget> et <Positive> dans le tableau, des catégories telles que coût & <Positive ou taux & excellent peuvent être créées. Le tableau affiche une seule ligne pour chaque combinaison de types telle que <Location> + <Positive> et <Positive> + < Location et leur ordre n'a pas d'importance pour la manière dont les catégories sont générées.

Figure 6-8
Boîte de dialogue Créer des catégories montrant les patrons de type disponibles



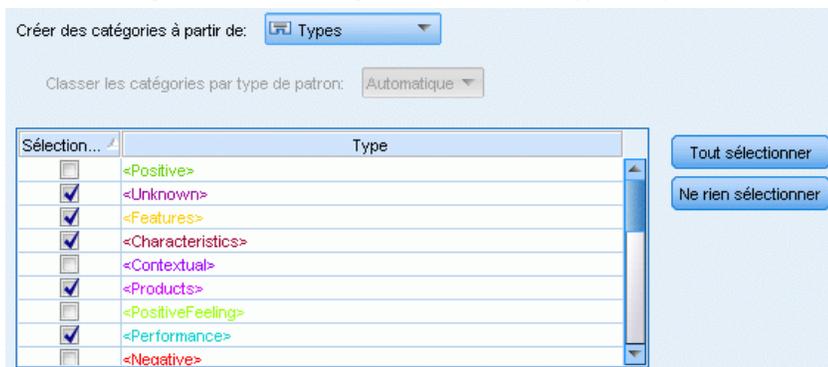
Lorsque des patrons de type sont utilisés comme entrées pour la création de catégories automatiques, les techniques identifient parfois plusieurs façons de former la structure des catégories. Techniquement, il n'existe pas une bonne façon de créer des catégories, mais une structure peut être plus adaptée à votre analyse qu'une autre. Pour aider à personnaliser la sortie dans ce cas, vous pouvez choisir un type préféré. Choisissez ce type dans le champ Structurer les catégories par type de patron : et le tableau sera mis à jour et n'affichera que les patrons applicables contenant le type sélectionné. La plupart du temps, le type <Unknown> sera présélectionné. Ainsi tous les patrons contenant le type <Unknown> seront sélectionnés. Le tableau affiche les types dans l'ordre décroissant en commençant par celui contenant le plus grand nombre d'enregistrements.

Types. Si vous sélectionnez les types, les catégories seront créées à partir des concepts appartenant aux types sélectionnés. Ainsi, si vous sélectionnez le type <Budget> dans le tableau, des catégories telles que coût ou prix peuvent être créées car coût et prix sont des concepts attribués au type <Budget>.

Par défaut, seuls les types qui capturent le plus d'enregistrements sont sélectionnés. Cette présélection vous permet de considérer rapidement les types les plus intéressants et d'éviter de créer des catégories sans intérêt. Le tableau affiche les types dans l'ordre décroissant en commençant par celui contenant le plus grand nombre d'enregistrements. Les types de la bibliothèque Opinions sont désélectionnés par défaut dans le tableau des types.

Les entrées que vous choisissez ont une influence sur les catégories obtenues. Lorsque vous choisissez d'utiliser Types comme entrée, vous pouvez plus facilement voir les concepts associés. Par exemple, si vous créez des catégories en utilisant Types comme entrée, vous pourrez obtenir une catégorie Fruit avec des concepts comme ananas, poire, agrumes, orange etc. Si vous choisissez Patrons de type comme entrée et que vous sélectionnez le patron <Unknown> + <Positive>, par exemple, alors vous pouvez obtenir une catégorie fruit + <Positive> avec une ou deux sortes de fruits comme fruit + savoureux et pomme + bonne. Ce deuxième résultat n'affiche que 2 patrons de concept car les autres occurrences de fruit ne sont pas nécessairement considérés comme positives. Et bien que cela puisse être suffisant pour vos données textuelles actuelles, dans les enquêtes longitudinales où différents ensembles de documents sont utilisés, il est conseillé d'ajouter manuellement d'autres descripteurs comme agrume + positif ou d'utiliser des types. La simple utilisation de types comme entrée vous permettra de trouver tous les fruits possibles.

Figure 6-9
Boîte de dialogue Créer des catégories montrant les types disponibles



Techniques

Puisque chaque ensemble de données est unique, le nombre de méthodes et l'ordre dans lequel vous les appliquez peut varier. Dans la mesure où vos objectifs d'exploration de texte peuvent être différents d'un ensemble de données à un autre, vous pouvez tester les différentes techniques afin de déterminer celle qui génère les meilleurs résultats pour les données textuelles en question.

Vous n'avez pas besoin d'être un expert de ces paramètres pour les utiliser. Par défaut, les paramètres les plus communs et moyens sont déjà sélectionnés. C'est pourquoi vous pouvez contourner la boîte de dialogue Paramètres avancés et créer directement vos catégories. De même, si vous effectuez des modifications ici, vous n'avez pas besoin de revenir à la boîte de dialogue Paramètres à chaque fois car les derniers paramètres sont toujours conservés.

Sélectionnez les techniques linguistiques ou de fréquence et cliquez sur le bouton Paramètres avancés pour afficher les paramètres des techniques sélectionnées. Aucune des techniques automatiques ne pourra parfaitement catégoriser vos données; c'est pourquoi nous recommandons de trouver et d'appliquer une ou plusieurs techniques automatiques qui correspondent aux besoins de vos données. Vous ne pouvez pas créer de catégories en utilisant en même temps les techniques de fréquence et linguistique.

- **Techniques linguistiques avancées.** Pour plus d'informations, reportez-vous à sur p. 116.
- **Techniques de fréquence avancées.** Pour plus d'informations, reportez-vous à sur p. 127.

Paramètres linguistiques avancés

Quand vous créez des catégories, vous pouvez choisir parmi diverses techniques de création de catégorie linguistiques avancées, dont la *dérivation des racines de concept*, l'*inclusion de concepts*, les *réseaux sémantiques* (anglais uniquement) et les *règles de cooccurrence*. Ces techniques peuvent être utilisées individuellement ou conjointement pour créer des catégories.

Notez que puisque chaque ensemble de données est unique, le nombre de méthodes et l'ordre dans lequel vous les appliquez peut varier. Dans la mesure où vos objectifs d'exploration de texte peuvent être différents d'un ensemble de données à un autre, vous pouvez tester les différentes techniques afin de déterminer celle qui génère les meilleurs résultats pour les données textuelles en question. Aucune des techniques automatiques ne pourra parfaitement catégoriser

vos données; c'est pourquoi nous recommandons de trouver et d'appliquer une ou plusieurs techniques automatiques qui correspondent aux besoins de vos données.

Figure 6-10

Paramètres avancés : La boîte de dialogue Linguistique pour créer des catégories

Paramètres avancés: Linguistique

Entrée et sortie

Entrée de la catégorie: Résultats d'extraction non utilisés Tous les résultats d'extraction

Sortie de la catégorie:

Hiérarchique avec des sous-catégories Catégories plates (niveau unique seulement)

Niveaux maximum créés: 5

Techniques de regroupement

Dérivation des racines de concept Inclusion de concept Co-occurrence

Nombre minimal d'enregistrements: 2

Distance de recherche maximale: 1 2 3 4

Empêcher l'appariement de concepts spécifiques

Généraliser avec des caractères génériques lorsque cela est possible (par exemple, pomme*)

Autres options

Nombre maximal de catégories de niveau supérieur créées: 30

Nombre minimum de descripteurs et/ou de sous-catégories par catégorie: 2

Permettre aux descripteurs d'apparaître dans plus d'une catégorie

Résoudre les noms de catégories en double en: Fusionner nouvelles avec existantes

Restaurer les paramètres par défaut Enregistrer & fermer Annuler Aide

Entrée et sortie

Entrée de catégorie. Sélectionnez à partir de quoi les catégories seront créées :

- Résultats d'extraction non utilisés. Cette option permet aux catégories d'être créées à partir des résultats d'extraction qui ne sont pas encore utilisés dans une catégorie existante. Ceci réduit la tendance des enregistrements à correspondre à plusieurs catégories et limite le nombre de catégories produites.
- Tous les résultats d'extraction. Cette option permet aux catégories d'être créées à l'aide de tous les résultats d'extraction. Ceci est particulièrement utile quand aucune ou peu de catégories existent déjà.

Sortie de catégorie. Sélectionnez la structure générale des catégories qui seront créées :

- Hiérarchique avec des sous-catégories. Cette option active la création de sous-catégories et de sous-sous-catégories. Vous pouvez définir la profondeur de vos catégories en choisissant le nombre de niveaux maximum (champ Nombre maximum de niveaux créés) pouvant être créés. Si vous choisissez 3, les catégories peuvent contenir des sous-catégories et ces sous-catégories peuvent également contenir des sous-catégories.
- Catégories plates (un seul niveau). Cette option n'active qu'un seul niveau de catégories à créer, ce qui signifie qu'aucune sous-catégorie ne sera créée.

Techniques de regroupement

Chaque technique disponible convient à certains types de données et de situation. Cependant, il est souvent judicieux de combiner des techniques dans la même analyse afin de capturer le maximum d'enregistrements. Aussi est-il possible de voir un concept figurer dans plusieurs catégories ou de rencontrer des catégories redondantes.

Dérivation des racines de concept. Cette technique crée des catégories en sélectionnant un concept et en recherchant les autres concepts y étant associés. Pour cela, elle analyse si certains composants du concept sont liés d'un point de vue morphologique ou s'ils partagent des racines. Cette technique est très utile pour l'identification de concepts de mots composés synonymes, car les concepts de chaque catégorie générée sont synonymes ou leur signification est proche. Elle utilise des données de différentes longueurs et génère un nombre inférieur de catégories compactes. Par exemple, le concept `opportunités d'avancer` serait regroupé avec les concepts `opportunité d'avancement` et `opportunité d'un avancement`. [Pour plus d'informations, reportez-vous à la section Dérivation des racines de concept sur p. 122.](#)

Réseau sémantique. Cette technique commence en identifiant les sens possibles de chaque concept à partir de son index complet de relations existant entre les mots puis crée des catégories en regroupant les concepts associés. Cette technique est plus performante lorsque les concepts sont connus dans le réseau sémantique et qu'ils ne sont pas trop ambigus. Son efficacité est cependant amoindrie lorsque le texte contient des termes spécialisés dont le réseau n'a pas connaissance. Par exemple, le concept `pomme granny smith` pourrait être regroupé avec `pomme gala` et `pomme golden` car il s'agit de soeurs de la granny smith. Pour donner un autre exemple, le concept `animal` pourrait être regroupé avec `chat` et `kangourou` car il s'agit d'hyponymes d'`animal`. Cette technique est uniquement disponible pour les textes en anglais dans cette version. [Pour plus d'informations, reportez-vous à la section Réseaux sémantiques sur p. 124.](#)

Inclusion de concept. Cette technique crée des catégories en regroupant les concepts multitermes (mots composés) selon qu'ils contiennent ou non des mots qui sont des sous-ensembles ou des super-ensembles d'un mot dans l'autre. Par exemple, le concept `siège` serait regroupé avec `siège de sécurité`, `siège couchette` et `commande de siège éjectable`. [Pour plus d'informations, reportez-vous à la section Inclusion de concepts sur p. 123.](#)

Cooccurrence. Cette technique crée des catégories à partir des cooccurrences trouvées dans le texte. Ainsi, lorsque des concepts ou des patrons de concept apparaissent souvent ensemble dans des documents et des enregistrements, la cooccurrence reflète une relation sous-jacente qui a vraisemblablement de l'intérêt dans vos définitions de catégorie. Lorsque des mots font l'objet d'une cooccurrence de manière significative, une règle de cooccurrence est créée et peut être utilisée comme descripteur de catégorie pour une nouvelle sous-catégorie. Par exemple, si de nombreux enregistrements contiennent les mots `prix` et `disponibilité`, (mais que peu d'enregistrements contiennent l'un sans l'autre) alors ces concepts peuvent être regroupés dans une règle de cooccurrence, (`prix & disponible`) et ils peuvent être affectés à une sous-catégorie de la catégorie `prix` par exemple. [Pour plus d'informations, reportez-vous à la section Règles de cooccurrence sur p. 126.](#)

- **Nombre minimum d'enregistrements.** Pour aider à déterminer l'intérêt des cooccurrences, déterminez le nombre minimum d'enregistrements devant contenir une cooccurrence donnée pour être utilisés en tant que descripteurs dans une catégorie.

Distance de recherche maximale. Sélectionnez jusqu'où vous souhaitez que les techniques effectuent la recherche avant de créer des catégories. Plus la valeur est basse, moins vous obtenez de résultats. Cependant, ces résultats produisent moins de parasites, et sont davantage susceptibles d'être liés ou associés de façon significative. Plus la valeur est élevée, plus vous obtenez de résultats. Toutefois, ces résultats risquent d'être moins fiables ou moins pertinents. Bien que cette option soit généralement appliquée à toutes les techniques, son effet est maximal sur les cooccurrences et les réseaux sémantiques.

Éviter l'appariement de concepts spécifiques. Cochez cette case pour arrêter le processus de regroupement ou d'appariement de deux concepts dans les résultats. Pour créer ou gérer des paires de concepts, cliquez sur `Gérer les paires...` [Pour plus d'informations, reportez-vous à la section Gestion des paires d'exceptions de liens sur p. 120.](#)

Généralisez avec des caractères génériques lorsque cela est possible. Sélectionnez cette option pour permettre au produit de créer des règles génériques dans les catégories à l'aide du caractère générique astérisque. Par exemple, au lieu de produire plusieurs descripteurs tels que `[tarte aux pommes + .]` et `[tarte aux fraises + .]`, l'utilisation de caractères génériques peut donner `[tarte * + .]`. Si vous généralisez avec des caractères génériques, vous obtenez souvent exactement le même nombre d'enregistrements que précédemment. Toutefois, cette option a l'avantage de réduire le nombre de descripteurs de catégorie et de les simplifier. De plus, cette option augmente les possibilités de catégoriser davantage d'enregistrements en utilisant ces catégories sur de nouvelles données textuelles (par exemple dans les enquêtes longitudinales/par vagues).

Autres options de création de catégories

En plus de sélectionner les techniques de regroupement à appliquer, vous pouvez éditer plusieurs autres options de création comme suit :

Le nombre maximum de catégories de créées. Cette option sert à limiter le nombre de catégories pouvant être générées lorsque vous cliquez sur le bouton Créer des catégories. Dans certains cas, vous pouvez obtenir de meilleurs résultats si vous réglez cette valeur élevée puis supprimez les catégories sans intérêt.

Le nombre minimum de descripteurs et/ou de sous-catégories par catégorie. Utilisez cette option pour définir le nombre minimum de descripteurs et de sous-catégories qu'une catégorie doit contenir pour être créée. Cette option permet de limiter la création de catégories qui ne capturent pas un nombre assez important d'enregistrements .

Permettre aux descripteurs d'apparaître dans plusieurs catégories. Lorsqu'elle est sélectionnée, cette option permet aux descripteurs d'être utilisés dans plusieurs des catégories qui seront créées ensuite. Cette option est généralement sélectionnée car les éléments entrent fréquemment ou "naturellement" dans deux catégories ou plus, et le fait de leur donner cette possibilité permet d'obtenir des catégories d'une plus grande qualité. Si vous ne sélectionnez pas cette option, vous réduisez le chevauchement d'enregistrements dans plusieurs catégories et en fonction du type de données que vous avez, ceci peut être souhaitable. Toutefois, avec la plupart des types de données, le fait de limiter les descripteurs à une seule catégorie entraîne une perte de la qualité ou de la diversité des catégories. Par exemple, imaginons que vous avez un concept fabricant sièges-auto. Avec cette option, ce concept peut apparaître dans une catégorie basée sur le texte sièges-auto et dans une autre basée sur fabricant. Mais si cette option n'est pas sélectionnée, bien que vous puissiez quand même obtenir les deux catégories, le concept fabricant sièges-auto n'apparaîtra comme descripteur que dans la catégorie à laquelle il correspond le mieux en fonction de plusieurs facteurs, notamment du nombre d'enregistrements dans lesquels sièges-auto et fabricant apparaissent respectivement.

Résoudre les noms de catégorie en double. Choisissez la manière de manipuler les nouvelles catégories ou sous-catégories dont le nom sera identique dans des catégories existantes. Vous pouvez fusionner les nouvelles catégories ou sous-catégories (et leur descripteurs) avec les catégories existantes qui portent le même nom. Vous pouvez également choisir d'ignorer la création de ces catégories si un nom en double se rencontre dans les catégories existantes.

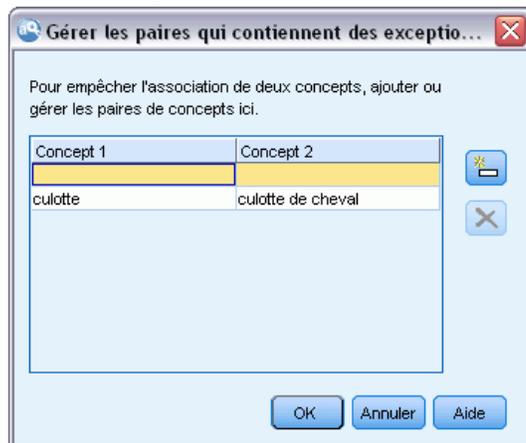
Gestion des paires d'exceptions de liens

Au cours de la création de catégorie et du regroupement de concepts, les algorithmes internes regroupent les mots par associations connues. Pour ne pas associer deux concepts par paires, ou pour ne pas les lier, vous pouvez activer cette fonctionnalité dans la boîte de dialogue Paramètres avancés de création de catégories, et la boîte de dialogue Paramètres de l'index de la carte de concepts puis cliquer sur le bouton Gérer les paires.

Dans la boîte de dialogue Gérer les exceptions de liens, vous pouvez ajouter, modifier ou supprimer les paires de concepts. Tapez une paire par ligne. Si vous entrez les paires ici, cela évitera la génération d'association par paires lors de la création ou de l'extension des catégories, ou du regroupement de concepts. Saisissez les mots exactement comme vous les voulez, par exemple la version accentuée d'un mot n'est pas la même que la version non accentuée.

Par exemple, si vous voulez vous assurer que *foudre* et *coup de foudre* ne soient pas regroupés, vous pouvez ajouter la paire dans une ligne séparée du tableau :

Figure 6-11
Boîte de dialogue Gérer des paires d'exceptions de liens



A propos des Techniques linguistiques

Quand vous créez ou développez des catégories, vous pouvez choisir parmi diverses techniques de création de catégorie linguistiques avancées, dont *ladérivation des racines de concept*, *l'inclusion de concepts*, les *réseaux sémantiques* (anglais uniquement) et les *règles de cooccurrence*. Ces techniques peuvent être utilisées individuellement ou conjointement pour créer des catégories.

Vous n'avez pas besoin d'être un expert de ces paramètres pour les utiliser. Par défaut, les paramètres les plus communs et moyens sont déjà sélectionnés. Vous pouvez contourner la boîte de dialogue Paramètres avancés et créer ou étendre directement vos catégories. De même, si vous effectuez des modifications ici, vous n'avez pas besoin de revenir à la boîte de dialogue Paramètres à chaque fois car les derniers paramètres sont toujours conservés.

Néanmoins, notez que puisque chaque ensemble de données est unique, le nombre de méthodes et l'ordre dans lequel vous les appliquez peut varier. Dans la mesure où vos objectifs d'exploration de texte peuvent être différents d'un ensemble de données à un autre, vous pouvez tester les différentes techniques afin de déterminer celle qui génère les meilleurs résultats pour les données textuelles en question. Aucune des techniques automatiques ne pourra parfaitement catégoriser vos données; c'est pourquoi nous recommandons de trouver et d'appliquer une ou plusieurs techniques automatiques qui correspondent aux besoins de vos données.

Les principales techniques linguistiques automatiques pour la création de catégories sont :

- **Dérivation des racines de concept.** Cette technique crée des catégories en sélectionnant un concept et en recherchant les autres concepts y étant associés. Pour cela, elle analyse si certains composants du concept sont liés d'un point de vue morphologique. [Pour plus d'informations, reportez-vous à la section Dérivation des racines de concept sur p. 122.](#)
- **Inclusion de concept.** Cette technique crée des catégories en sélectionnant un concept et en recherchant les autres concepts le contenant. [Pour plus d'informations, reportez-vous à la section Inclusion de concepts sur p. 123.](#)

- **Réseau sémantique.** Cette technique commence en identifiant les sens possibles de chaque concept à partir de son index complet de relations existant entre les mots puis crée des catégories en regroupant les concepts associés. [Pour plus d'informations, reportez-vous à la section Réseaux sémantiques sur p. 124.](#) Cette option n'est disponible que pour le texte en anglais.
- **Cooccurrence.** Cette technique crée des règles de cooccurrence qui peuvent être utilisées pour créer une nouvelle catégorie, étendre une catégorie ou comme entrée pour une autre technique de catégorie. [Pour plus d'informations, reportez-vous à la section Règles de cooccurrence sur p. 126.](#)

Dérivation des racines de concept

La technique de dérivation des racines de concept crée des catégories en sélectionnant un concept et en recherchant les autres concepts y étant associés. Pour cela, elle analyse si certains composants du concept sont liés d'un point de vue morphologique. Un composant est un mot. La technique tente de regrouper les concepts en observant les terminaisons (suffixes) de chaque composant d'un concept et en recherchant les autres concepts ayant pu être créés à partir d'eux. Ainsi, lorsque des mots dérivent d'autres mots, il est vraisemblable qu'ils partagent la même signification ou qu'ils s'en rapprochent. Des règles internes, propres à chaque langue, identifient les terminaisons. Par exemple, le concept *opportunités d'avancer* serait regroupé avec les concepts *opportunité d'avancement* et *opportunité d'un avancement*.

Vous pouvez utiliser la dérivation des racines de concept sur n'importe quel type de texte. Utilisée seule, elle renvoie un nombre assez peu élevé de catégories et chacune d'entre elles contient généralement peu de concepts. Les concepts de chaque catégorie sont synonymes ou associés dans le cadre de la situation. Cet algorithme peut s'avérer utile, même si vous créez les catégories manuellement ; les synonymes qu'il repère peuvent être synonymes des concepts qui vous intéressent particulièrement.

Remarque : Vous pouvez éviter que les concepts ne se groupent entre eux en les spécifiant de manière explicite. [Pour plus d'informations, reportez-vous à la section Gestion des paires d'exceptions de liens sur p. 120.](#)

Fractionnement des termes en composants et suppression de la flexion

Lorsque vous appliquez les techniques de dérivation des racines de concept ou d'inclusion de concepts, les termes sont tout d'abord fractionnés en composants (mots), puis la flexion des composants est supprimée. Lorsque vous appliquez une technique, les concepts et leurs termes associés sont chargés et fractionnés en composants en fonction de leurs séparateurs, tels que les espaces, les traits d'union et les apostrophes. Par exemple, le terme *administrateur système* est fractionné en composants comme suit : {administrateur, système}.

Cependant, certaines parties du terme d'origine peuvent ne pas être utilisées. C'est ce que l'on appelle des mots vides. Voici certains des composants pouvant être ignorés en français : un, une, et, par, pour, de, du, des, dans, ou, le, à et avec.

Par exemple, le terme *analyse des données* est constitué de l'ensemble de composants {données, analyse} et tant les mots *de* et *le* sont considérés comme pouvant être ignorés. Par ailleurs, l'ordre des mots dans un ensemble de composants n'est pas significatif. Ainsi, les trois termes suivants peuvent être équivalents : *voile pour bateau de compétition*,

compétition de bateau à voile et bateau à voile de compétition, dans la mesure où ils ont tous le même ensemble de composants {voile, bateau, compétition}. Chaque fois que deux termes sont considérés comme étant équivalents, les concepts correspondants sont fusionnés pour constituer un nouveau concept qui référence l'ensemble des termes.

De plus, puisque les composants d'un terme peuvent être déclinés, des règles propres à la langue sont appliquées en interne pour identifier les termes équivalents, quelle que soit la variation occasionnée par leur flexion (formes au pluriel, par exemple). De cette manière, les termes niveau de fiabilité et niveaux de fiabilité peuvent être identifiés comme étant équivalents, puisque la forme au singulier sans flexion est niveau.

Fonctionnement de la dérivation des racines de concept

Après que les termes ont été fractionnés en composants et que leur flexion a été supprimée (voir la section précédente), l'algorithme de dérivation des racines de concept analyse les terminaisons, ou suffixes, des composants pour retrouver leur radical, puis regroupe les concepts dont les radicaux sont identiques ou similaires. Les terminaisons sont identifiées à l'aide d'un ensemble de règles de dérivation linguistique propres à la langue du texte. Par exemple, une règle de dérivation concernant un texte rédigé en langue française détermine qu'un composant de concept présentant le suffixe *ique* peut être dérivé d'un concept ayant le même radical, mais se terminant par le suffixe *ie*. Grâce à cette règle (et à la suppression de la flexion), l'algorithme peut regrouper les concepts *étude épidémiologique* et *études épidémiologiques*.

Etant donné que les termes ont déjà été fractionnés en composants et que les composants pouvant être ignorés (par exemple, *d'*) ont été identifiés, l'algorithme de dérivation des racines de concept est également capable de regrouper le concept *étude d'épidémiologie* avec *étude épidémiologique*.

L'ensemble de règles de dérivation de composant a été choisi de telle sorte que la plupart des concepts regroupés par cet algorithme soient synonymes : les concepts *étude épidémiologique*, *études épidémiologiques*, *étude d'épidémiologie* sont tous des termes équivalents. Pour augmenter le degré d'exhaustivité, certaines règles de dérivation laissent l'algorithme regrouper des concepts qui sont associés dans le cadre de la situation. Par exemple, l'algorithme peut regrouper des concepts tels que *réparer les voitures* et *réparation de voitures*.

Inclusion de concepts

La technique d'inclusion de concepts crée des catégories en prenant un concept et, à l'aide d'algorithmes de série lexicale, identifie les concepts inclus dans d'autres concepts. Ainsi, lorsque les mots d'un concept constituent un sous-ensemble d'un autre concept, ces algorithmes reflètent une relation sémantique sous-jacente. L'inclusion est une technique puissante qui peut être utilisée avec n'importe quel type de texte.

Elle fonctionne bien en conjonction avec des réseaux sémantiques, mais elle peut être utilisée séparément. L'inclusion de concepts peut aussi fournir de meilleurs résultats lorsque les enregistrements contiennent de nombreux termes appartenant à un domaine spécifique. Ceci se confirme particulièrement lorsque vous avez préalablement affiné les dictionnaires de telle sorte que les termes spéciaux soient extraits et regroupés de façon appropriée (avec leurs synonymes).

Fonctionnement de l'inclusion de concept

Avant l'application de l'algorithme d'inclusion de concept, les termes sont fragmentés en composants et leur flexion est supprimée. [Pour plus d'informations, reportez-vous à la section Dérivation des racines de concept sur p. 122.](#) Ensuite, l'algorithme d'inclusion de concepts analyse les ensembles de composants. Pour chacun d'entre eux, l'algorithme recherche un ensemble de composants qui soit un sous-ensemble du premier.

Par exemple, si vous disposez du concept `déjeuner diététique`, qui comporte l'ensemble de composants `{déjeuner, diététique}`, et le concept `déjeuner`, qui comporte l'ensemble de composants `{déjeuner}`, l'algorithme en conclurait que `déjeuner diététique` est un type de `déjeuner` et les regrouperait.

Pour citer un exemple plus étendu, si le concept `siège` figure dans le panneau Résultats d'extraction et que vous appliquez cet algorithme, les concepts tels que `siège de sécurité`, `siège en cuir`, `siège couchette`, `commande de siège éjectable`, `siège-auto coque` et `instructions pour siège-auto` seraient également regroupés dans cette catégorie.

Puisque les termes sont déjà fractionnés en composants et que les composants pouvant être ignorés (par exemple, `en` et `d'`) ont été identifiés, l'algorithme d'inclusion de concepts reconnaît que le concept `cours d'espagnol avancé` inclut le concept `cours en espagnol`.

Remarque : Vous pouvez éviter que les concepts ne se groupent entre eux en les spécifiant de manière explicite. [Pour plus d'informations, reportez-vous à la section Gestion des paires d'exceptions de liens sur p. 120.](#)

Réseaux sémantiques

Dans cette version, la technique de réseau sémantique n'est disponible que pour les textes rédigés en anglais.

Cette technique crée des catégories à l'aide d'un réseau intégré de relations entre les mots. De ce fait, cette technique peut produire d'excellents résultats lorsque les termes sont précis et ne sont pas trop ambigus. En revanche, ne vous attendez pas à identifier de nombreux liens entre des concepts hautement techniques/spécialisés. En travaillant avec des concepts de ce type, il est probable que les techniques d'inclusion de concepts et de dérivation des racines de concept s'avèrent plus utiles.

Fonctionnement du réseau sémantique

La technique de réseau sémantique vise à exploiter les relations connues existant entre les mots pour créer des catégories de synonymes ou d'hyponymes. Un **hyponyme** est un concept considéré comme étant un type de second concept, de façon à ce qu'une relation hiérarchique, également appelée relation ISA, soit établie. Par exemple, si `animal` est un concept, `chat` et `kangourou` sont des hyponymes d'`animal`, puisque ce sont des types d'animaux.

Outre les relations de synonyme et d'hyponyme, la technique du réseau sémantique examine également les liens partie-tout existant entre les concepts du type `<Location>`. Par exemple, cette technique regroupe les concepts `normandie`, `provence` et `france` en une seule catégorie, car la Normandie et la Provence sont des parties de la France.

Les réseaux sémantiques commencent par identifier les sens possibles de chaque concept du réseau. Lorsque des concepts sont identifiés comme étant synonymes ou hyponymes, ils sont regroupés dans une seule catégorie. Par exemple, la technique crée une seule catégorie contenant ces trois concepts : pomme à couteau, pomme à croquer et granny smith, puisque le réseau sémantique contient les informations suivantes : 1) pomme à croquer est synonyme de pomme à couteau et 2) granny smith est un type de pomme à couteau (c'est-à-dire un hyponyme de pomme à couteau).

Considérés individuellement, de nombreux concepts, en particulier les termes univoques, sont ambigus. Par exemple, le concept `buffet` peut désigner un type de repas ou un meuble. Si l'ensemble des concepts inclut `repas`, `meuble` et `buffet`, l'algorithme est forcé de choisir entre regrouper `buffet` avec `repas` ou avec `meuble`. Sachez que, dans certains cas, les choix effectués par l'algorithme peuvent ne pas être appropriés dans le contexte d'un ensemble particulier d'enregistrements.

La technique du réseau sémantique peut offrir de bien meilleures performances que l'inclusion de concepts avec certains types de données. Alors que le réseau sémantique et l'inclusion de concepts reconnaissent que `pinceau rouge` est une sorte de `pinceau`, seul le réseau sémantique reconnaît que `rouleau` est également une sorte de `pinceau`.

Les réseaux sémantiques peuvent être utilisés conjointement avec les autres techniques. Par exemple, imaginons que vous ayez sélectionné les techniques de réseau sémantique et d'inclusion, et que le réseau sémantique ait regroupé le concept `professeur` avec le concept `tuteur` (car un tuteur est un type de professeur). L'algorithme d'inclusion peut regrouper le concept `directeur d'étude` avec `tuteur`. En conséquence, les deux algorithmes collaborent afin de produire une catégorie de sortie contenant les trois concepts : `tuteur`, `directeur d'étude` et `professeur`.

Options du réseau sémantique

Il existe de nombreux paramètres supplémentaires pouvant être intéressants avec cette technique.

- **Changer la distance de recherche maximale.** Sélectionnez jusqu'où vous souhaitez que les techniques effectuent la recherche avant de créer des catégories. Plus la valeur est basse, moins vous obtenez de résultats. Cependant, ces résultats produisent moins de parasites, et sont davantage susceptibles d'être liés ou associés entre eux de façon significative. Plus la valeur est élevée, plus vous obtenez de résultats. Toutefois, ces résultats risquent d'être moins fiables ou moins pertinents.

Par exemple, en fonction de la distance, l'algorithme recherche de `pain au chocolat` jusqu'à `viennoiserie` (son parent), puis `petit pain` (grand-parent) et vers le haut jusqu'à `pain`.

En réduisant la distance de recherche, cette technique produit des catégories plus petites qui peuvent faciliter le travail lorsque les catégories produites sont trop grandes ou regroupent trop d'éléments.

Important ! Il est recommandé, lorsque vous utilisez cette technique, de ne pas appliquer l'option Nombre de caractères minimum requis (qui figure dans la boîte de dialogue Extraire) pour les regroupements flous, car certains mauvais regroupements peuvent avoir un impact très négatif sur les résultats.

Règles de cooccurrence

Les règles de cooccurrence vous permettent d'identifier et de regrouper les concepts étroitement liés au sein de l'ensemble de d'enregistrements. Ainsi, lorsque des concepts apparaissent souvent ensemble dans des enregistrements, la cooccurrence reflète une relation sous-jacente qui a vraisemblablement de l'intérêt dans vos définitions de catégories. Cette technique crée des règles de cooccurrence qui peuvent être utilisées pour créer une nouvelle catégorie, étendre une catégorie ou comme entrée pour une autre technique de catégorie. Deux concepts co-existent fortement s'ils apparaissent fréquemment ensemble dans un ensemble d'enregistrements et rarement séparément dans les autres enregistrements. Cette technique génère de bons résultats avec de plus grands ensembles de données contenant au moins plusieurs centaines d'enregistrements.

Par exemple, si de nombreux enregistrements contiennent les mots `prix` et `disponibilité`, ces concepts peuvent être regroupés dans une règle de cooccurrence, (`prix & disponible`). Dans un autre exemple, si les concepts `beurre`, `confiture`, `tartine` apparaissent plus souvent ensemble que séparément, ils sont regroupés dans une règle de cooccurrence de concepts (`beurre&confiture & tartine`).

Important ! Dans les versions précédentes, les règles de cooccurrence et de synonymes étaient entre crochets. Dans la version actuelle, les crochets indiquent désormais un résultat de patrons. Les règles de cooccurrence et de synonymes sont maintenant entourées de parenthèses, comme dans (`enceintes acoustiques|enceintes`).

Fonctionnement des règles de cooccurrence

Cette technique analyse les enregistrements à la recherche d'au moins deux concepts apparaissant souvent ensemble. Au moins deux concepts présentent une forte cooccurrence s'ils apparaissent souvent l'un avec l'autre dans un ensemble d'enregistrements, et s'ils apparaissent rarement séparément dans d'autres enregistrements.

Une règle de catégorie se crée dès que le système identifie des concepts cooccurrents. Ces règles comportent au moins deux concepts reliés par l'opérateur booléen `&`. Il s'agit d'instructions logiques qui classent automatiquement un enregistrement dans une catégorie, si l'ensemble des concepts qu'elles régissent sont cooccurrents dans ces enregistrements.

Options des règles de cooccurrence

Si vous utilisez la technique des règles de cooccurrence, vous pouvez régler plusieurs paramètres ayant une influence sur les règles obtenues :

- **Changer la distance de recherche maximale.** Sélectionnez jusqu'où vous souhaitez que les techniques effectuent la recherche avant de créer des catégories. Plus la valeur est basse, moins vous obtenez de résultats. Cependant, ces résultats produisent moins de parasites, et sont davantage susceptibles d'être liés ou associés de façon significative. Plus la valeur est élevée, plus vous obtenez de résultats. Toutefois, ces résultats risquent d'être moins fiables ou moins pertinents. Lorsque vous travaillez sur des cooccurrences, la valeur par défaut de la distance de recherche produit de très nombreuses cooccurrences et un grand nombre d'entre elles n'ont qu'un lien distendu et sont par conséquent, peu intéressantes. Lorsque vous

réduisez la distance de recherche, vous filtrez les cooccurrences les moins intéressantes et vous obtenez les résultats les plus probants.

- **Nombre minimum d'enregistrements.** Pour aider à déterminer l'intérêt des cooccurrences, déterminez le nombre minimum d'enregistrements devant contenir une cooccurrence donnée, pour être utilisés en tant que descripteurs dans une catégorie. Pour les plus petits ensembles de données, plus la configuration de cette option est basse, plus il est facile de trouver des cooccurrences.

Remarque : Vous pouvez éviter que les concepts ne se groupent entre eux en les spécifiant de manière explicite. [Pour plus d'informations, reportez-vous à la section Gestion des paires d'exceptions de liens sur p. 120.](#)

Paramètres de fréquence avancés

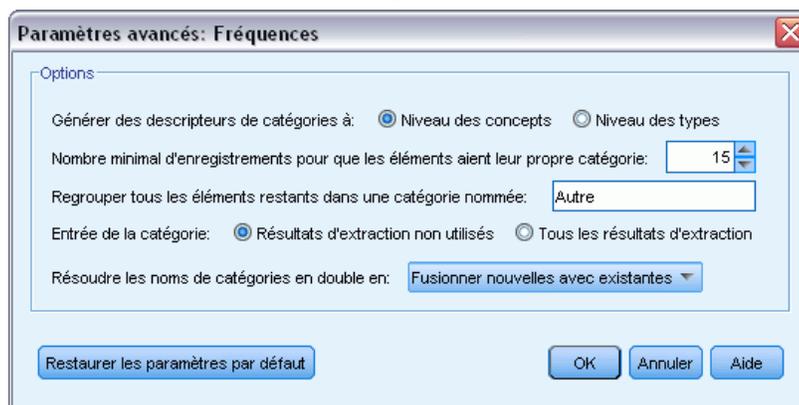
Vous pouvez créer des catégories en fonction d'une technique de fréquence simple et mécanique. Avec cette technique, vous pouvez créer une catégorie pour chaque élément (type, concept ou patron) situé en amont des effectifs des enregistrements. Vous pouvez également créer une catégorie regroupant tous les termes moins fréquents. Les effectifs désignent le nombre d'enregistrements contenant le concept extrait (et tous ses synonymes), le type ou le patron dans la question par rapport au nombre total d'occurrences d'un concept, d'un type ou d'un patron dans l'ensemble du texte.

Le regroupement d'éléments fréquents peut produire des résultats intéressants, car il peut indiquer une réponse courante ou importante. Elle est très efficace si elle a été exécutée sur les résultats d'extraction inutilisés après que d'autres techniques ont été appliquées. Une autre application consiste à exécuter cette technique immédiatement après l'extraction lorsqu'aucune autre catégorie n'existe, éditer les résultats pour supprimer les catégories sans intérêt, puis étendre ces catégories afin de leur faire correspondre toujours plus d'enregistrements. [Pour plus d'informations, reportez-vous à la section Extension de catégories sur p. 128.](#)

A la place de cette technique, vous pouvez également trier les concepts et les patrons de concept par nombre décroissant des enregistrements dans le panneau Résultats d'extraction puis en faisant glisser et en déposant les premiers dans le panneau Catégories pour créer les catégories correspondantes.

Figure 6-12

Paramètres avancés : boîte de dialogue des fréquences



Créer des descripteurs de catégories. Sélectionnez le type d'entrée pour les descripteurs. [Pour plus d'informations, reportez-vous à la section Création de catégories sur p. 112.](#)

- Niveau de concept. Le fait de sélectionner cette option signifie que les fréquences de concepts ou de patrons de concept seront utilisées. Les concepts sont utilisés si les types ont été sélectionnés comme entrée pour la création de catégorie et les patrons de concept sont utilisés si les patrons de type ont été sélectionnés. En général, appliquer cette technique au niveau du concept produira des résultats plus précis car les concepts et patrons de concept représentent un niveau de mesure moins élevé.
- Niveau de type. Le fait de sélectionner cette option signifie que les fréquences de types ou de patrons de type seront utilisées. Les types sont utilisés si les types ont été sélectionnés comme entrée pour la création de catégorie et les patrons de type sont utilisés si les patrons de type ont été sélectionnés. L'application de cette technique au niveau de type vous permet d'obtenir une vue rapide de l'ensemble des réponses données.

Effectif minimum d'enregistrements pour que les éléments aient leur propre catégorie. Cette option vous permet de créer des catégories à partir des éléments fréquents. Cette option limite les résultats aux seules catégories qui contiennent un descripteur que l'on rencontre dans au moins X enregistrements, où X est la valeur à saisir pour cette option.

Regroupez tous les éléments restants dans une catégorie nommée. Cette option vous permet de regrouper tous les concepts ou types peu fréquents en une seule catégorie 'fourre-tout' portant le nom de votre choix. Par défaut, cette catégorie se nomme *Autre*.

Entrée de catégorie. Sélectionnez le groupe auquel appliquer ces techniques :

- Résultats d'extraction non utilisés. Cette option permet aux catégories d'être créées à partir des résultats d'extraction qui ne sont pas encore utilisés dans une catégorie existante. Ceci réduit la tendance des enregistrements à correspondre à plusieurs catégories et limite le nombre de catégories produites.
- Tous les résultats d'extraction. Cette option permet aux catégories d'être créées à l'aide de tous les résultats d'extraction. Ceci est particulièrement utile quand aucune ou peu de catégories existent déjà.

Résoudre les noms de catégorie en double. Choisissez la manière de manipuler les nouvelles catégories dont le nom sera identique dans des catégories existantes. Vous pouvez fusionner les nouvelles catégories ou sous-catégories (et leur descripteurs) avec les catégories existantes qui portent le même nom. Vous pouvez également choisir d'ignorer la création de ces catégories si un nom en double se rencontre dans les catégories existantes.

Extension de catégories

L'extension est un processus au cours duquel des descripteurs sont ajoutés ou améliorés automatiquement pour « agrandir » les catégories existantes. L'objectif est de produire une meilleure catégorie qui capture les enregistrements associés qui n'ont pas été attribués à cette catégorie à l'origine.

Les techniques de regroupement automatiques que vous sélectionnez tenteront d'identifier les concepts, patrons TLA et règles de catégorie associées aux descripteurs de catégorie existants. Ces nouveaux concepts, patrons et règles de catégorie sont ensuite ajoutés comme nouveaux descripteurs ou ajoutés aux descripteurs existants. Les techniques de regroupement pour l'extension incluent la *dérivation des racines de concept*, l'*inclusion de concepts*, les *réseaux sémantiques* (uniquement en anglais) et les *règles de cooccurrence*. La méthode Étendre les catégories vides avec des descripteurs générés à partir du nom de la catégorie génère des descripteurs à l'aide des mots dans les noms de catégories. Ainsi, plus les noms de catégories sont descriptifs, meilleurs sont les résultats.

Remarque : Les techniques de fréquence ne sont pas disponibles lors de l'extension des catégories.

L'extension est une excellente manière d'améliorer vos catégories de manière interactive. Voici quelques exemples de cas où vous pouvez étendre une catégorie :

- Après avoir glissé/déposé les patrons de concept pour créer des catégories dans le panneau Catégories
- Après avoir créé des catégories manuellement et avoir ajouté des règles de catégorie et des descripteurs simples
- Après avoir importé un plan de codage dans lequel les catégories avaient des noms très descriptifs
- Après avoir affiné les catégories provenant du TAP que vous avez choisi au cours de la création du projet

Vous pouvez étendre plusieurs fois une catégorie. Par exemple, si vous avez importé un fichier de catégorie prédéfinie avec des noms très descriptifs, vous pouvez effectuer l'extension avec l'option Étendre les catégories vides avec des descripteurs créés à partir du nom de la catégorie pour obtenir un premier ensemble de descripteurs puis étendre de nouveau ces catégories. Néanmoins, dans d'autres cas, l'extension multiple peut entraîner une catégorie trop générique si les descripteurs sont de plus en plus étendus. Comme les techniques de création et d'extension de regroupement utilisent des algorithmes sous-jacents similaires, l'extension directement après la création de catégories ne produira probablement pas de résultats plus intéressants.

Astuces :

- Si vous tentez une extension et si vous ne voulez pas utiliser les résultats, vous pouvez toujours annuler l'opération (Edition > Annuler) immédiatement après l'extension.
- L'extension peut produire au moins deux règles de catégorie dans une catégorie qui correspondent exactement au même ensemble de documents, les règles étant créées indépendamment pendant le processus. Si besoin est, vous pouvez afficher les catégories et supprimer les redondances en modifiant manuellement la description des catégories. [Pour plus d'informations, reportez-vous à la section Modification des descripteurs de catégorie sur p. 161.](#)

Pour étendre des catégories

- ▶ Dans le panneau Catégories, sélectionnez les catégories à étendre.
- ▶ Dans les menus, sélectionnez Catégories > Étendre les catégories. Un message apparaît, sauf si vous avez choisi de ne pas recevoir d'invite.

- ▶ Choisissez si vous voulez créer maintenant ou éditer d'abord les paramètres.
 - Cliquez sur **Etendre** pour commencer à étendre des catégories à l'aide des paramètres actuels. Le processus commence et une boîte de dialogue de progression apparaît.
 - Cliquez sur **Editer** pour examiner et modifier les paramètres.

Après la tentative d'extension, toutes les catégories pour lesquelles de nouveaux descripteurs ont été trouvés sont signalées par le mot **Etendue** dans le panneau **Catégories**, afin que vous puissiez les identifier rapidement. Le texte **Etendue** reste affiché jusqu'à ce que vous étendiez à nouveau la catégorie, que vous la modifiez d'une autre manière, ou la supprimiez via le menu contextuel.

Figure 6-13
Boîte de dialogue *Etendre les catégories*



Chaque technique disponible lors de la création ou l'extension de catégories convient à certains types de données et de situations. Cependant, il est souvent judicieux de combiner plusieurs techniques dans la même analyse afin de capturer le maximum d'enregistrements. Les concepts et les types qui ont été regroupés dans une catégorie seront toujours disponibles la prochaine fois que vous créerez des catégories. Aussi est-il possible de voir un concept figurer dans plusieurs catégories ou de rencontrer des catégories redondantes.

Entrée de catégorie. Sélectionnez l'entrée à utiliser pour étendre les catégories :

- Résultats d'extraction non utilisés. Cette option permet aux catégories d'être créées à partir des résultats d'extraction qui ne sont pas encore utilisés dans une catégorie existante. Ceci réduit la tendance des enregistrements à correspondre à plusieurs catégories et limite le nombre de catégories produites.
- Tous les résultats d'extraction. Cette option permet aux catégories d'être créées à l'aide de tous les résultats d'extraction. Ceci est particulièrement utile quand aucune ou peu de catégories existent déjà.

Techniques de regroupement

Pour les descriptions courtes de chacune de ces techniques, reportez-vous à la rubrique [Paramètres linguistiques avancés](#) sur p. 116. Ces techniques comprennent :

- Dérivation des racines de concept (*non disponible pour le japonais*)
- Réseau sémantique (*Anglais uniquement*)
- Inclusion de concepts
- Cooccurrence et sous-option Nombre minimal de documents.

Plusieurs types sont définitivement exclus de la technique de réseau sémantique car ils ne renverront pas de résultats pertinents. Ils comprennent <Positive>, <Negative>, <IP>, d'autres types non-linguistiques, etc.

Distance de recherche maximale. Sélectionnez jusqu'où vous souhaitez que les techniques effectuent la recherche avant de créer des catégories. Plus la valeur est basse, moins vous obtenez de résultats. Cependant, ces résultats produisent moins de parasites, et sont davantage susceptibles d'être liés ou associés de façon significative. Plus la valeur est élevée, plus vous obtenez de résultats. Toutefois, ces résultats risquent d'être moins fiables ou moins pertinents. Bien que cette option soit généralement appliquée à toutes les techniques, son effet est maximal sur les cooccurrences et les réseaux sémantiques.

Éviter l'appariement de concepts spécifiques. Cochez cette case pour arrêter le processus de regroupement ou d'appariement de deux concepts dans les résultats. Pour créer ou gérer des paires de concepts, cliquez sur Gérer les paires... [Pour plus d'informations, reportez-vous à la section Gestion des paires d'exceptions de liens sur p. 120.](#)

Lorsque cela est possible : choisissez s'il faut simplement effectuer une extension, généraliser les descripteurs avec des caractères génériques ou effectuer ces deux opérations.

- **Étendre et généraliser.** Cette option permet d'étendre les catégories sélectionnées et de généraliser les descripteurs. Lorsque vous choisissez de généraliser, le produit crée des règles de catégorie génériques dans les catégories à l'aide du caractère générique astérisque. Par exemple, au lieu de produire plusieurs descripteurs tels que [tarte aux pommes + .] et [tarte aux fraises + .], l'utilisation de caractères génériques peut donner [tarte * + .]. Si vous généralisez avec des caractères génériques, vous obtenez souvent exactement le même nombre d'enregistrements que précédemment. Toutefois, cette option a l'avantage de réduire le nombre de descripteurs de catégorie et de les simplifier. De plus, cette option augmente les possibilités de catégoriser davantage d'enregistrements en utilisant ces catégories sur de nouvelles données textuelles (par exemple dans les enquêtes longitudinales/par vagues).

- **Étendre uniquement.** Cette option étend vos catégories sans généralisation. Il peut être utile de choisir d'abord l'option *Étendre uniquement* pour les catégories créées manuellement puis d'étendre de nouveau ces mêmes catégories à l'aide de l'option *Étendre et généraliser*.
- **Généraliser uniquement.** Cette option généralisera les descripteurs sans étendre vos catégories de quelque autre façon.

Autres options d'extension de catégories

En plus de sélectionner les techniques de regroupement à appliquer, vous pouvez éditer les options suivantes :

Nombre maximum d'éléments pour étendre un descripteur. Lors de l'extension d'un descripteur avec des éléments (concepts, types et autres expressions), définissez le nombre maximum d'éléments pouvant être ajoutés à un seul descripteur. Si vous fixez cette limite à 10, un maximum de 10 éléments supplémentaires peuvent être ajoutés à un descripteur existant. Si plus de 10 éléments doivent être ajoutés, les techniques arrêteront d'ajouter de nouveaux éléments après l'ajout du dixième. Ceci peut raccourcir la liste de descripteur mais ne garantit pas que les éléments les plus intéressants aient été utilisés en premier. Vous préférez peut-être réduire la taille de l'extension sans pénaliser la qualité en utilisant l'option *Généraliser* avec des caractères génériques lorsque cela est possible. Cette option ne s'applique qu'aux descripteurs qui contiennent les booléens & (AND) ou ! (NOT).

Étendre également les sous-catégories. Cette option étendra également toutes les sous-catégories sous les catégories sélectionnées.

Étendre les catégories avec des descripteurs en fonction des noms de catégories. Cette option tente de créer automatiquement des descripteurs pour chaque catégorie en fonction des mots constituant le nom de la catégorie. Le nom de catégorie est analysé pour voir si les mots du nom correspondent à des concepts extraits. Si un concept est reconnu, il est utilisé pour trouver les patrons de concept correspondants et ceux-ci sont utilisés pour générer des descripteurs pour la catégorie. Cette option produit les meilleurs résultats lorsque les noms de catégories sont à la fois longs et descriptifs. C'est une méthode rapide pour générer des descripteurs de catégories, qui à leur tour permettent à la catégorie de capturer les enregistrements contenant ces descripteurs. Cette option est particulièrement utile quand vous importez des catégories d'ailleurs ou quand vous créez des catégories manuellement avec de longs noms descriptifs. Cette méthode s'applique uniquement aux catégories vides qui ont 0 descripteurs. Si une catégorie contient déjà des descripteurs, elle ne sera pas étendue de cette manière.

Générer des descripteurs en tant que. Cette option s'applique uniquement si l'option précédente est sélectionnée.

- **Concepts.** Choisissez cette option pour produire les résultats des descripteurs sous la forme de concepts, qu'ils aient été extraits ou non du texte source.
- **Patrons.** Choisissez cette option pour produire les résultats des descripteurs sous la forme de patrons, que les résultats des patrons ou tout autre patron aient été extraits ou non.

Création de catégories manuellement

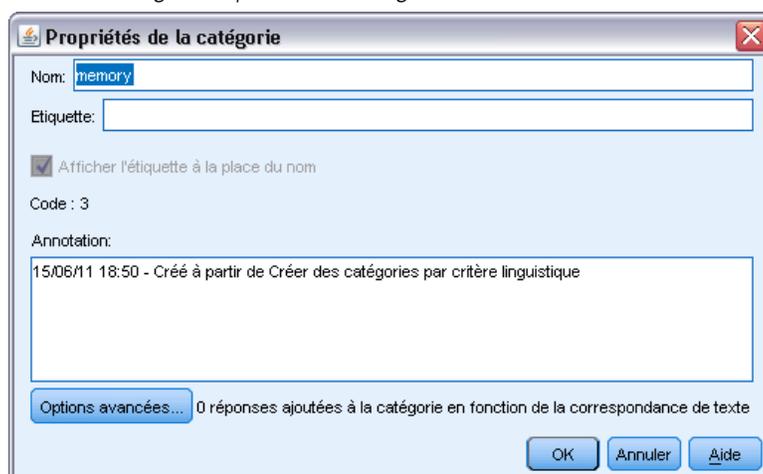
En plus de créer des catégories à l'aide des techniques de création de catégories automatiques que sont le plan de codage et l'éditeur de règles, vous pouvez également créer des catégories manuellement. Les méthodes manuelles suivantes existent :

- Création d'une catégorie vide dans laquelle vous ajouterez des éléments un par un. [Pour plus d'informations, reportez-vous à la section Création de catégories ou attribution d'un nouveau nom aux catégories sur p. 133.](#)
- Déplacement de termes, de types et de patrons dans le panneau des catégories. [Pour plus d'informations, reportez-vous à la section Création de catégories par la méthode Glisser-déposer sur p. 134.](#)

Création de catégories ou attribution d'un nouveau nom aux catégories

Vous pouvez créer des catégories vides en vue d'y ajouter des concepts et des types. Vous pouvez également renommer vos catégories.

Figure 6-14
Boîte de dialogue Propriétés de catégorie



Pour créer une catégorie vide

- ▶ Accédez au panneau Catégories.
- ▶ Dans les menus, choisissez Catégories > Créer une catégorie vide. La boîte de dialogue apparaît.
- ▶ Entrez un nom dans le champ Nom pour cette catégorie.

- ▶ Cliquez sur OK pour valider ce nom et fermer la boîte de dialogue. La boîte de dialogue se ferme et un nouveau nom de catégorie apparaît dans le panneau.

Vous pouvez maintenant commencer à ajouter des éléments à cette catégorie. [Pour plus d'informations, reportez-vous à la section Ajout de descripteurs aux catégories sur p. 160.](#)

Pour renommer une catégorie

- ▶ Sélectionnez une catégorie, puis choisissez Catégories > Renommer la catégorie. La boîte de dialogue apparaît.
- ▶ Entrez un nouveau nom dans le champ Nom pour cette catégorie.
- ▶ Cliquez sur OK pour valider ce nom et fermer la boîte de dialogue. La boîte de dialogue se ferme et un nouveau nom de catégorie apparaît dans le panneau.

Création de catégories par la méthode Glisser-déposer

La technique du glisser-déposer est manuelle et n'est pas basée sur des algorithmes. Vous pouvez créer des catégories dans le panneau Catégories en glissant-déposant :

- des concepts, des types ou des patrons extraits du panneau Résultats d'extraction vers le panneau Catégories.
- des concepts extraits du panneau Données vers le panneau Catégories.
- des lignes entières du panneau Données vers le panneau Catégories. Cela permettra de créer une catégorie composée de tous les concepts et les patrons extraits contenus dans cette ligne.

Remarque : Le panneau Résultats d'extraction prend en charge la sélection multiple pour faciliter le glisser-déposer de différents éléments.

Important ! Vous ne pouvez pas faire glisser et déposer des concepts provenant du panneau Données et qui n'ont pas été extraits du texte. Si vous souhaitez forcer l'extraction d'un concept trouvé dans les données, vous devez l'ajouter à un type. Puis réexécutez l'extraction. Les nouveaux résultats de l'extraction contiendront le concept que vous venez d'ajouter. Vous pouvez l'utiliser dans votre catégorie. [Pour plus d'informations, reportez-vous à la section Ajout de concepts à des types dans le chapitre 5 sur p. 91.](#)

Pour créer des catégories avec la méthode de glisser-déposer :

- ▶ Dans le panneau Résultats d'extraction ou le panneau Données, sélectionnez un ou plusieurs concepts, patrons, types, enregistrements ou enregistrements partiels.
- ▶ Tout en maintenant le bouton de la souris appuyé, faites glisser l'élément vers une catégorie existante ou vers la zone du panneau pour créer une nouvelle catégorie.
- ▶ Lorsque vous atteignez la zone où vous souhaitez déposer cet élément, relâchez le bouton de la souris. Cet élément est ajouté au panneau Catégories. Les catégories modifiées apparaissent avec une couleur d'arrière-plan spécifique. Cette couleur s'appelle l'arrière-plan des commentaires de catégorie. [Pour plus d'informations, reportez-vous à la section Définition des options dans le chapitre 2 sur p. 19.](#)

Remarque : La catégorie résultante a été automatiquement nommée. Si vous le souhaitez, vous pouvez modifier un nom. [Pour plus d'informations, reportez-vous à la section Modification des propriétés de catégories sur p. 159.](#)

Si vous souhaitez voir quels enregistrements sont affectés à une catégorie, sélectionnez cette catégorie dans le panneau Catégories. Le panneau des données est automatiquement actualisé et affiche tous les enregistrements pour cette catégorie. Pour afficher l'ensemble entier des réponses à une question, sélectionnez le noeud Tous les enregistrements en haut de l'arborescence des catégories.

Import et export de catégories prédéfinies

Si vos propres catégories sont stockées dans un fichier Microsoft Excel (*.xls, *.xlsx), vous pouvez les importer dans IBM® SPSS® Text Analytics for Surveys.

Vous pouvez également exporter vers un fichier Microsoft Excel (*.xls, *.xlsx), les catégories que vous avez dans un projet ouvert. Lorsque vous exportez vos catégories, vous pouvez choisir d'inclure ou d'exclure des informations supplémentaires telles que les descripteurs et les scores. [Pour plus d'informations, reportez-vous à la section Exporter des catégories sur p. 144.](#)

Important : A partir de SPSS Text Analytics for Surveys Version 4.0.1, les catégories prédéfinies ont en grande partie remplacé l'utilisation des plans de codage. Par exemple, l'assistant Importer un plan de codage a été remplacé par l'assistant Importer des catégories prédéfinies. Toutefois, ce nouvel assistant vous permet toujours d'importer vos plans de codage existants. En outre, le gestionnaire de plan de codage n'est plus pris en charge ; pour modifier des valeurs de code, sélectionnez Afficher > Codes de catégories dans les menus pour afficher la colonne Code dans le panneau Catégorie et modifier les codes comme nécessaire.

Si vos catégories prédéfinies n'ont pas de code ou si vous souhaitez de nouveaux codes, vous pouvez générer automatiquement un nouvel ensemble de codes pour l'ensemble de catégories, dans le panneau des catégories en choisissant Catégories > Gestion des catégories > Générer automatiquement des codes à partir des menus. Tous les codes existants seront supprimés et renumérotés automatiquement.

Import de catégories prédéfinies

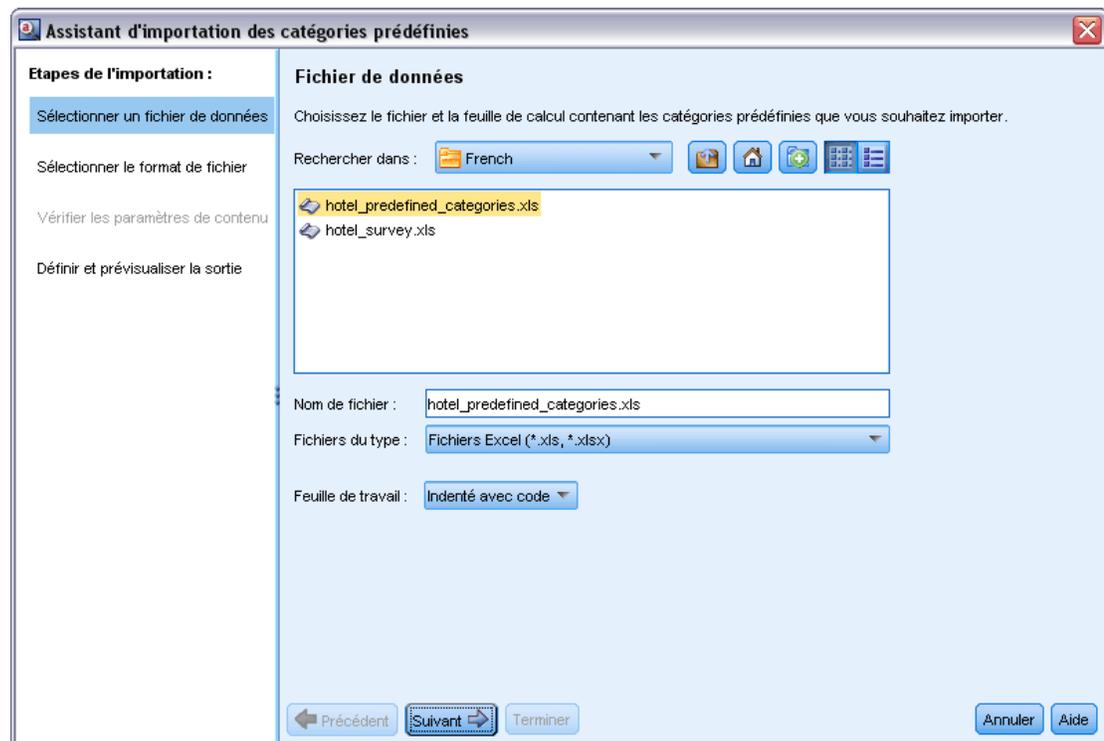
Vous pouvez importer vos catégories prédéfinies dans IBM® SPSS® Text Analytics for Surveys. Avant l'importation, assurez-vous que le fichier de la catégorie prédéfinie est un fichier Microsoft Excel (*.xls, *.xlsx) et qu'il est structuré dans un format pris en charge. Vous pouvez également choisir de laisser le produit détecter le format pour vous. Les formats suivants sont pris en charge :

- Format liste plate : [Pour plus d'informations, reportez-vous à la section Format liste plate sur p. 140.](#)
- Format compact : [Pour plus d'informations, reportez-vous à la section Format compact sur p. 141.](#)
- Format indenté : [Pour plus d'informations, reportez-vous à la section Format indenté sur p. 142.](#)

Pour importer des catégories prédéfinies

- Dans les menus, choisissez Catégories > Gestion des catégories > Importer des catégories prédéfinies. Un assistant d'importation des catégories prédéfinies apparaît.

Figure 6-15
Assistant Import des catégories prédéfinies

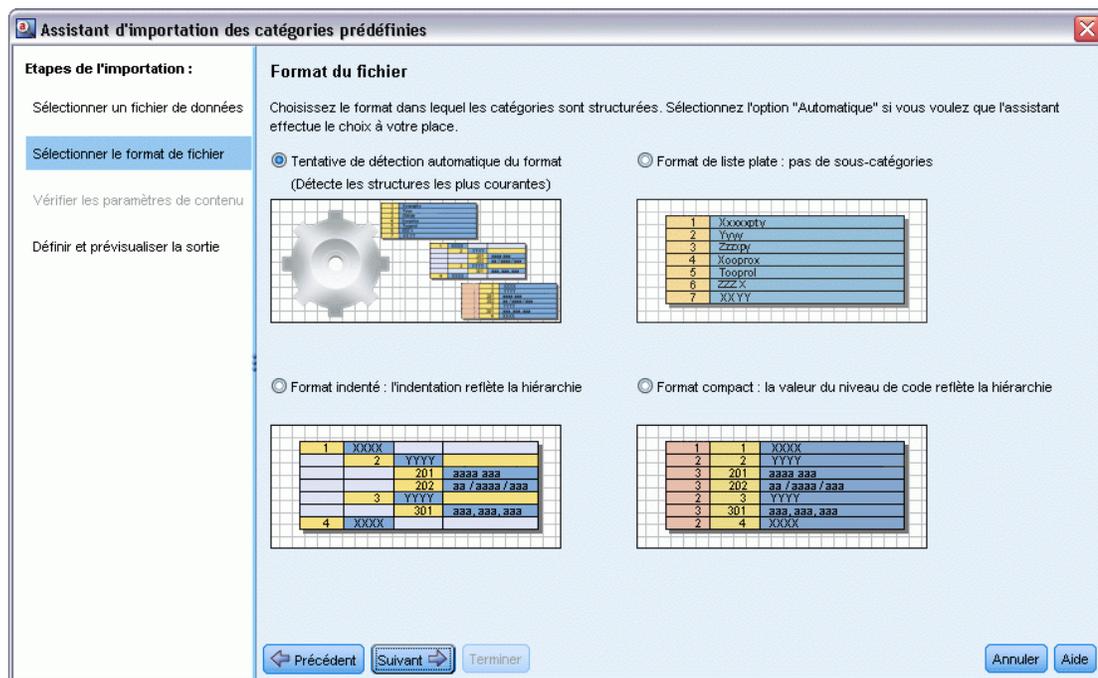


- Dans la liste déroulante Rechercher dans, sélectionnez le pilote et le dossier dans lesquels se trouve le fichier.
- Sélectionnez le fichier dans la liste. Le nom du fichier apparaît dans la zone de texte Nom de fichier.

- ▶ Sélectionnez dans la liste, la feuille de calcul contenant les catégories prédéfinies. Le nom de la feuille de calcul apparaît dans le champ Feuille de calcul.
- ▶ Cliquez sur Suivant pour commencer à choisir le format de données.

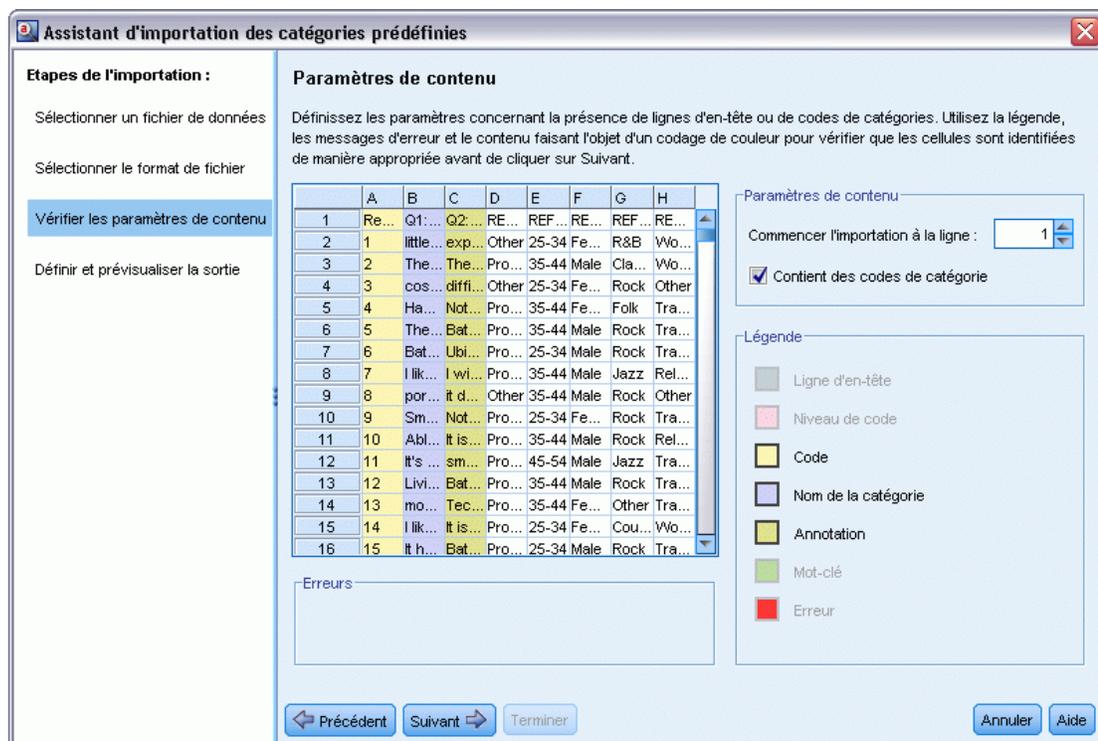
Figure 6-16

Boîte de dialogue Importer des catégories prédéfinies, étape Format de données



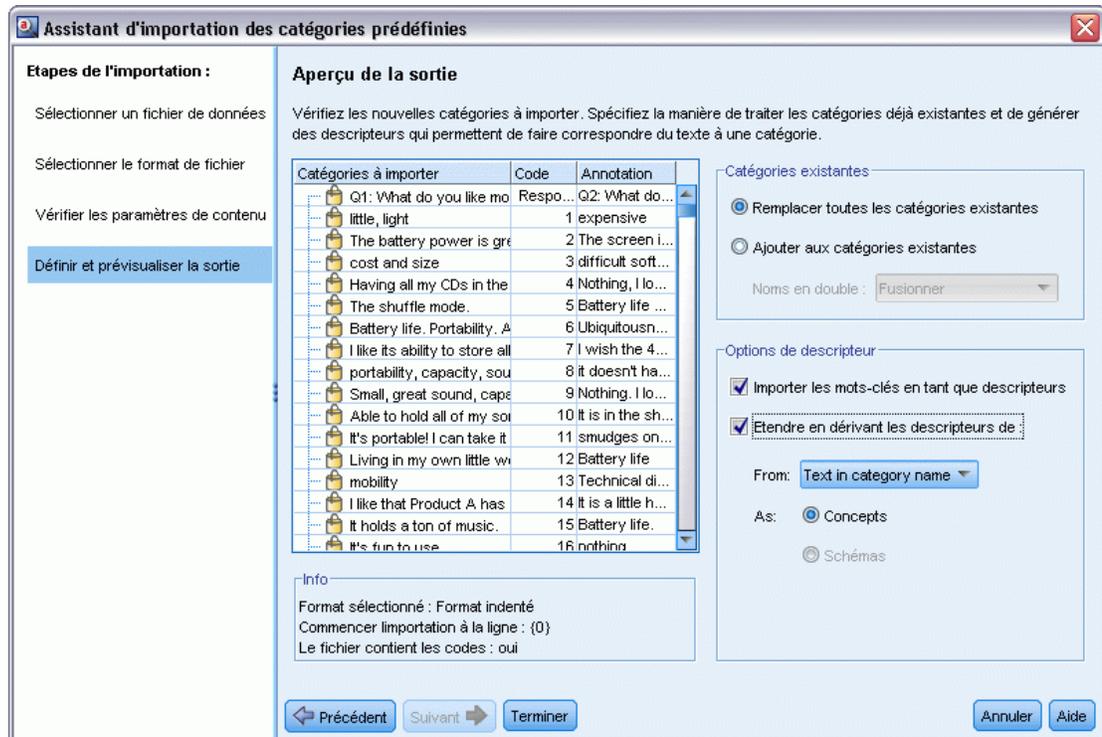
- ▶ Choisissez le format de votre fichier, ou choisissez l'option permettant au produit de détecter automatiquement le format. La détection automatique est la plus efficace sur les formats les plus courants.
 - Format liste plate : [Pour plus d'informations, reportez-vous à la section Format liste plate sur p. 140.](#)
 - Format compact : [Pour plus d'informations, reportez-vous à la section Format compact sur p. 141.](#)
 - Format indenté : [Pour plus d'informations, reportez-vous à la section Format indenté sur p. 142.](#)
- ▶ Cliquez sur Suivant pour définir les options supplémentaires d'import. Si vous avez choisi de détecter automatiquement le format, vous serez orienté directement vers l'étape finale.

Figure 6-17
 Importer des catégories prédéfinies, étape Options d'import



- ▶ Si une ou plusieurs lignes contiennent des titres de colonnes ou d'autres types d'information, choisissez le numéro de la ligne à partir de laquelle vous voulez commencer l'importation dans l'option Commencer l'import à la ligne. Par exemple : si vos noms de catégories commencent à la ligne 7, vous devez entrer le chiffre 7 dans cette option pour importer correctement votre fichier.
- ▶ Si votre fichier contient des codes de catégories, choisissez l'option Contient des codes de catégories. Ainsi, il sera plus facile à l'assistant de reconnaître correctement vos données.
- ▶ Consultez les cellules de codage couleur et la légende afin de vous assurer que les données ont été correctement identifiées. Toute erreur détectée dans le fichier est affichée en rouge et référencée sous le tableau d'aperçu de format. Si le format sélectionné n'est pas le bon, revenez en arrière et choisissez-en un autre. Si vous devez apporter des corrections à votre fichier, effectuez vos changements et redémarrez l'assistant en sélectionnant de nouveau votre fichier. Vous devez corriger toutes les erreurs avant de fermer l'assistant.
- ▶ Cliquez sur Suivant pour consulter l'ensemble des catégories et des sous-catégories qui seront importées et pour définir comment créer des descripteurs pour ces catégories.

Figure 6-18
Boîte de dialogue Importer des catégories prédéfinies, étape Prévisualisation



- ▶ Consultez l'ensemble des catégories qui seront importées dans le tableau. Si les mots clés que vous avez définis comme descripteurs n'apparaissent pas, il se peut qu'ils n'aient pas été reconnus durant l'import. Assurez-vous qu'ils ont été correctement préfixés et qu'ils apparaissent dans la bonne cellule.
- ▶ Choisissez la manière dont vous voulez traiter les catégories pré-existantes dans votre projet.
 - Remplacer toutes les catégories existantes. Cette option purge toutes les catégories existantes et les catégories nouvellement importées sont alors utilisées seules à leur place.
 - Ajouter aux catégories existantes. Cette option vous permettra d'importer les catégories et de les fusionner avec les catégories déjà existantes. En ajoutant des catégories à des catégories déjà existantes, vous devez déterminer comment vous voulez que les doublons soient traités. Un des choix (option : Fusionner) est de fusionner toute catégorie importée avec les catégories existantes si elles ont un nom de catégorie commun. Une autre possibilité (option : Exclure de l'import) est d'interdire l'import de catégories si une catégorie portant le même nom existe déjà.
- ▶ Import de mot-clés comme descripteurs est une option de descripteur, qui importe les mots-clés identifiés dans vos données en tant que descripteurs pour la catégorie associée.
- ▶ Étendre les catégories à partir des descripteurs est une option générant des descripteurs à partir de mots qui représentent le nom de la catégorie ou de la sous-catégorie, et/ou à partir de mots qui forment l'annotation. Si ces mots correspondent aux résultats extraits, alors ils seront ajoutés en tant que descripteurs de la catégorie. Cette option produit de meilleurs résultats lorsque les noms de catégories sont à la fois longs et descriptifs. C'est une méthode rapide pour générer des

descripteurs de catégories, qui à leur tour permettent à la catégorie de capturer les enregistrements contenant ces descripteurs.

- Le champ À partir de vous permet de choisir depuis quel texte les descripteurs seront choisis, les noms ou les catégories ou sous-catégories, les mots dans les annotations, ou les deux.
- Le champ En tant que vous permet de créer ces descripteurs sous la forme de concepts ou de patrons TLA. Si l'extraction TLA n'a pas eu lieu, les options de patrons de cet assistant sont désactivés.

► Cliquez sur Terminer pour importer les catégories prédéfinies dans le panneau Catégories.

Format liste plate

Dans ce format liste plate, il n'y a qu'un niveau de catégories, sans aucune hiérarchie, ce qui signifie qu'il n'y a pas de sous-catégories ou de sous-réseaux. Les noms de catégories sont dans une seule colonne.

Figure 6-19

Exemple de format liste plate

	A	B	C
1	1	1 Logement	
2	2	10 Cadre	tout commentaire ayant trait à la décoration
3	2	11 Propreté	
4	2	12 Taille	tout commentaire ayant trait aux grandes pièces
5	2	13 Tranquillité	tout commentaire ayant trait au fait que c'est calme
6	2	14 Confort	tout commentaire indiquant que l'endroit est confortable, correctement équipé
7		climatisation	
8	1	2 Emplacement	tout commentaire ayant trait à la proximité au centre-ville, l'aéroport... ou indiquant que l'er
9		bien placé	
10	1	3 Service	tout commentaire ayant trait au personnel
11		employés	
12	2	30 Accueil	
13		accueillant	
14		avenant	
15		chaleureux	
16		sympathique	
17		poli	
18	2	31 Compétence	

Les informations suivantes peuvent être contenues dans un fichier de ce format :

- La colonne des **codes** optionnels contient des valeurs numériques qui identifient chaque catégorie de façon unique. Si vous précisez que les fichiers de données contiennent des codes (option Contient des codes de catégories dans l'étape Paramètres de contenu), alors une colonne contenant des codes uniques pour chaque catégorie doit exister dans la cellule située à gauche du nom de la catégorie. Si vos données ne contiennent pas de codes, mais que vous voulez créer des codes plus tard, vous pourrez toujours le faire en utilisant l'option Catégories > Gestion des catégories > Générer automatiquement des codes. Vous pouvez modifier des codes ultérieurement en choisissant Afficher > Code de catégorie ; les codes sont affichés dans une colonne Code du panneau Catégorie où vous pouvez les modifier manuellement.

- Une colonne *obligatoire* **nom des catégories** contient tous les noms de catégories. Cette colonne est nécessaire pour importer en utilisant ce format.
- Des **annotations** optionnelles sont disponibles dans la cellule située à droite du nom de catégorie. Cette annotation consiste en un texte décrivant vos catégories ou sous-catégories.
- Des **mots clés** optionnels peuvent être importés en tant que descripteurs de catégories. Afin qu'ils soient reconnus, ces mots clés doivent être placés dans la cellule située directement sous le nom de la catégorie/sous-catégorie associée. La liste des mots clés doit également être préfixée d'un caractère de soulignement () par exemple « _armesàfeu, armes / revolvers ». La cellule de mot clé peut contenir un ou plusieurs mots pour décrire chaque catégorie. Ces mots seront importés en tant que descripteurs, ou ignorés, en fonction de ce que vous aurez spécifié durant la dernière étape de l'assistant. Plus tard, les descripteurs sont comparés aux résultats extraits du texte. Si une correspondance est détectée, le document ou l'enregistrement est attribué à la catégorie contenant ce descripteur.

Table 6-3
Format liste plate avec codes, mots clés et annotations

Colonne A	Colonne B	Colonne C
Code de catégorie (<i>facultatif</i>)	Nom de catégorie	Annotation
	<u> </u> Descripteur/liste des mots-clés (<i>facultatif</i>)	

Format compact

Ce format compact est similaire au format liste plate, mis à part le fait qu'il est utilisé avec les catégories hiérarchiques. Par conséquent, une colonne de niveau de code est nécessaire afin de déterminer le niveau hiérarchique de chaque catégorie et sous-catégorie.

Figure 6-20
Exemple de fichier de catégorie prédéfinie compact sous Microsoft Excel

	A	B	C
1	1	Logement	
2	10	Logement/Cadre	tout commentaire ayant trait à la décoration
3	11	Logement/Propreté	
4	12	Logement/Taille	tout commentaire ayant trait aux grandes pièces
5	13	Logement/Tranquilité	tout commentaire ayant trait au fait que c'est calme
6	14	Logement/Confort	tout commentaire indiquant que l'endroit est confortable, correctement équipé
7		<u>_</u> climatisation	
8	2	Emplacement	tout commentaire ayant trait à la proximité au centre-ville, l'aéroport...ou indiquant que l'endroit est bien placé
9		<u>_</u> bien placé	
10	3	Service	tout commentaire ayant trait au personnel
11		<u>_</u> employés	
12	30	Service/Accueil	
13		<u>_</u> accueillant	
14		<u>_</u> avenant	
15		<u>_</u> chaleureux	
16		<u>_</u> sympathique	
17		<u>_</u> poli	
18	31	Service/Compétence	

Les informations suivantes peuvent être contenues dans un fichier de ce format :

- Une colonne *obligatoire* de **niveau de code** contient les nombres indiquant la position hiérarchique des informations ultérieures pour cette ligne. Par exemple; si les valeurs 1, 2 ou 3 sont spécifiées et que vous avez à la fois des catégories et des sous-catégories, alors 1 correspond aux catégories, 2 aux sous-catégories et 3 aux sous-sous-catégories. Si vous avez uniquement des catégories et des sous-catégories, 1 correspond aux catégories et 2 correspond aux sous-catégories. Et ainsi de suite, jusqu'à la profondeur de catégories souhaitée.
- La colonne optionnelle des **codes** contient des valeurs numériques qui identifient chaque catégorie de façon unique. Si vous précisez que les fichiers de données contiennent des codes (option Contient des codes de catégories dans l'étape Paramètres de contenu), alors une colonne contenant des codes uniques pour chaque catégorie doit exister dans la cellule située à gauche du nom de la catégorie. Si vos données ne contiennent pas de codes, mais que vous voulez créer des codes plus tard, vous pourrez toujours le faire en utilisant l'option Catégories > Gestion des catégories > Générer automatiquement des codes. Vous pouvez modifier des codes ultérieurement en choisissant Afficher > Code de catégorie ; les codes sont affichés dans une colonne Code du panneau Catégorie où vous pouvez les modifier manuellement.
- Une colonne *obligatoire* **nom des catégories** contient tous les noms de catégories et de sous-catégories. Cette colonne est nécessaire pour importer en utilisant ce format.
- Des **annotations** optionnelles sont disponibles dans la cellule située à droite du nom de catégorie. Cette annotation consiste en un texte décrivant vos catégories ou sous-catégories.
- Des **mots clés** optionnels peuvent être importés en tant que descripteurs de catégories. Afin qu'ils soient reconnus, ces mots clés doivent être placés dans la cellule située directement sous le nom de la catégorie/sous-catégorie associée. La liste des mots clés doit également être préfixée d'un caractère de soulignement () par exemple « _armesàfeu, armes / revolvers ». La cellule de mot clé peut contenir un ou plusieurs mots pour décrire chaque catégorie. Ces mots seront importés en tant que descripteurs, ou ignorés, en fonction de ce que vous aurez spécifié durant la dernière étape de l'assistant. Plus tard, les descripteurs sont comparés aux résultats extraits du texte. Si une correspondance est détectée, le document ou l'enregistrement est attribué à la catégorie contenant ce descripteur.

Table 6-4

Exemple de format compact avec codes

Colonne A	Colonne B	Colonne C
Niveau de code hiérarchique	Code de catégorie (<i>facultatif</i>)	Nom de la catégorie
Niveau de code hiérarchique	Code de sous-catégorie (<i>facultatif</i>)	Nom de sous-catégorie

Table 6-5

Exemple de format compact sans codes

Colonne A	Colonne B
Niveau de code hiérarchique	Nom de la catégorie
Niveau de code hiérarchique	Nom de sous-catégorie

Format indenté

Dans le format de fichier indenté, le contenu est organisé de façon hiérarchique ; c'est-à-dire qu'il contient des catégories et un ou plusieurs niveaux de sous-catégories. De plus, sa structure est indentée pour refléter cette hiérarchie. Chaque ligne du fichier contient soit une catégorie,

soit une sous-catégorie. Les sous-catégories sont indentées à partir des catégories, et toute sous-sous-catégorie est indentée à partir des sous-catégories, etc. Vous pouvez créer manuellement cette structure dans Microsoft Excel ou utiliser une structure exportée à partir d'un autre produit et enregistrée sous un format Microsoft Excel.

Figure 6-21
Exemple d'une catégorie indentée dans Microsoft Excel

	A	B	C	D
1	1	Logement		
2			10 Cadre	tout commentaire ayant trait à la décoration
3			11 Propreté	
4			12 Taille	tout commentaire ayant trait aux grandes pièces
5			13 Tranquillité	tout commentaire ayant trait au fait que c'est calme
6			14 Confort	tout commentaire indiquant que l'endroit est confortable, correctement équip
7			_climatisation	
8	2	Emplacement		tout commentaire ayant trait à la proximité au centre-ville, l'aéroport...ou indiquant que l'endroit est b
9		bien placé		
10	3	Service		tout commentaire ayant trait au personnel
11			_employés	
12			30 Accueil	
13			_accueillant	
14			_avenant	
15			_chaleureux	
16			_sympathique	
17			_poli	
18			31 Compétence	

- Les **Codes et les noms de catégories de niveau supérieur** occupent respectivement les colonnes A et B. Ou, si aucun code n'est présent, alors le nom de catégorie occupe la colonne A.
- Les **codes de sous-catégories et les noms de sous-catégories** occupent respectivement les colonnes B et C. Ou, si aucun code n'est présent, les noms de sous-catégories occupent la colonne B. La sous-catégorie est membre d'une catégorie. Vous ne pouvez pas avoir de sous-catégories si vous n'avez pas de catégories.

Table 6-6
Structure indentée avec des codes

Colonne A	Colonne B	Colonne C	Colonne D
Code de catégorie (facultatif)	Nom de catégorie		
	Code de sous-catégorie (facultatif)	Nom de sous-catégorie	
		Code de sous-sous-catégorie (facultatif)	Nom de sous-sous-catégorie

Table 6-7
Structure indentée sans codes

Colonne A	Colonne B	Colonne C
Nom de la catégorie		
	Nom de sous-catégorie	
		Nom de sous-sous-catégorie

Les informations suivantes peuvent être contenues dans un fichier de ce format :

- Les **codes** optionnels doivent être des valeurs identifiant de manière unique chaque catégorie ou sous-catégorie. Si vous précisez que les fichiers de données contiennent des codes (option Contient des codes de catégories dans l'étape Paramètres de contenu), alors un code unique pour chaque catégorie ou sous-catégorie doit exister dans la cellule située à gauche du nom de la catégorie/sous-catégorie. Si vos données ne contiennent pas de codes, mais que vous voulez créer des codes plus tard, vous pourrez toujours le faire en utilisant l'option Catégories > Gestion des catégories > Générer automatiquement des codes. Vous pouvez modifier des codes ultérieurement en choisissant Afficher > Code de catégorie ; les codes sont affichés dans une colonne Code du panneau Catégorie où vous pouvez les modifier manuellement.
- Un **nom** est *obligatoire* pour chaque catégorie et sous-catégorie. Les sous-catégories doivent être indentées des catégories d'une cellule vers la droite et sur une ligne différente.
- Des **annotations** optionnelles sont disponibles dans la cellule située à droite du nom de catégorie. Cette annotation consiste en un texte décrivant vos catégories ou sous-catégories.
- Des **mots clés** optionnels peuvent être importés en tant que descripteurs de catégories. Afin qu'ils soient reconnus, ces mots clés doivent être placés dans la cellule située directement sous le nom de la catégorie/sous-catégorie associée. La liste des mots clés doit également être préfixée d'un caractère de soulignement () par exemple « _armesàfeu, armes / revolvers ». La cellule de mot clé peut contenir un ou plusieurs mots pour décrire chaque catégorie. Ces mots seront importés en tant que descripteurs, ou ignorés, en fonction de ce que vous aurez spécifié durant la dernière étape de l'assistant. Plus tard, les descripteurs sont comparés aux résultats extraits du texte. Si une correspondance est détectée, le document ou l'enregistrement est attribué à la catégorie contenant ce descripteur.

Important ! Si vous utilisez un code à un niveau, vous devez inclure un code pour chaque catégorie et sous-catégorie. Si cela n'est pas le cas, votre import échouera.

Exporter des catégories

Vous pouvez également exporter vers un fichier Microsoft Excel (*.xls, *.xlsx), les catégories que vous avez dans un projet ouvert. Les données qui seront exportées viennent essentiellement du contenu actuel du panneau des catégories ou des propriétés des catégories. Toutefois, nous recommandons de scorer de nouveau, si vous prévoyez d'exporter également la valeur de score de Docs..

Toujours exporté..

- S'ils sont présents, les codes de catégories
- Noms de catégories (et de sous-catégories)
- S'ils sont présents, les niveaux de codes (Format *Plat/Compact*)
- En-têtes de colonnes (Format *Plat/Compact*)

Exporté de manière optionnelle...

- Scores de Docs.
- Annotations de catégories
- Noms des descripteurs
- Nombres de descripteurs

Important ! Lorsque vous exportez des descripteurs, ils sont convertis en chaînes de texte et un caractère de soulignement leur est ajouté comme préfixe. Si vous effectuez de nouveau une importation dans ce produit, la capacité de distinction entre les descripteurs qui sont des patrons, des règles de catégorie ou des concepts bruts est perdue. Si vous souhaitez réutiliser ces catégories

dans ce produit, nous vous recommandons fortement de créer un fichier de package d'analyse de texte (TAP), car ce format permet de conserver tous les descripteurs tels qu'ils sont définis, ainsi que toutes vos catégories, vos codes et les ressources linguistiques utilisées. Les fichiers TAP peuvent être utilisés à la fois dans IBM® SPSS® Text Analytics et dans IBM® SPSS® Text Analytics for Surveys. [Pour plus d'informations, reportez-vous à la section Utilisation des packages d'analyse de texte dans le chapitre 3 sur p. 43.](#)

Pour exporter des catégories prédéfinies

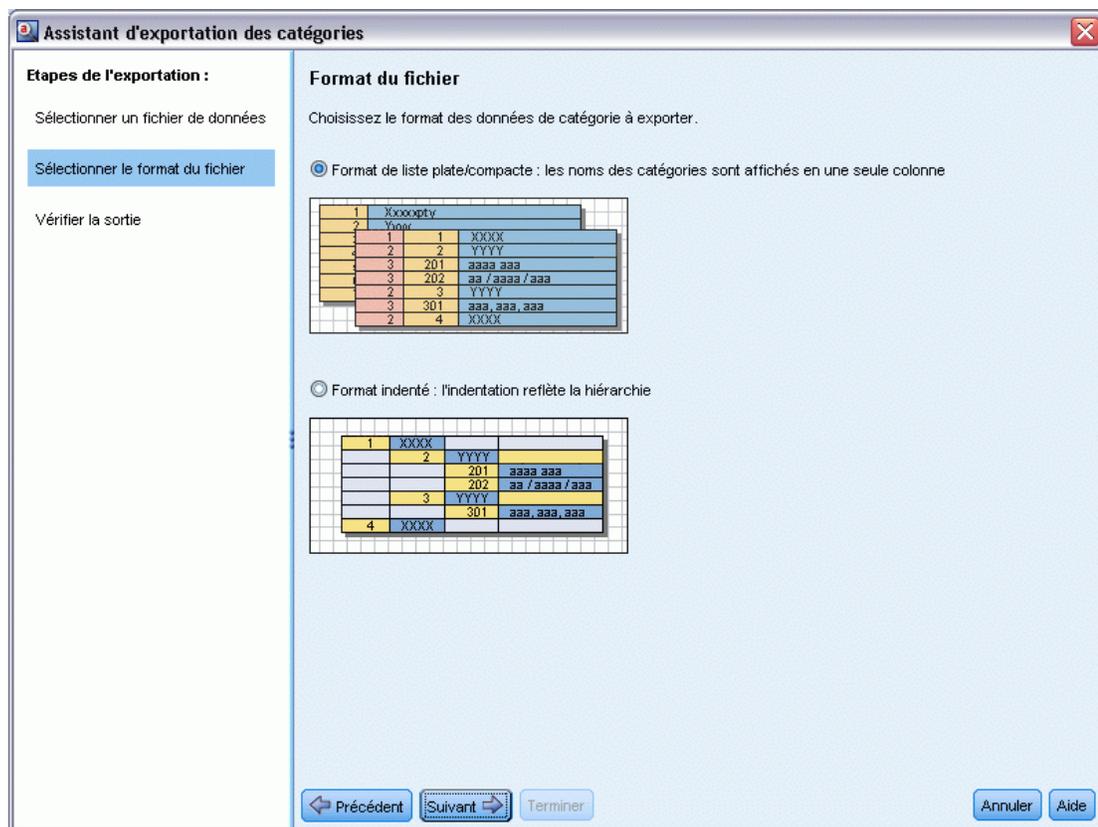
- Dans les menus, choisissez Catégories > Gestion des catégories > Exporter des catégories. Un assistant d'exportation des catégories apparaît.

Figure 6-22
Assistant d'export de catégories, étape 1



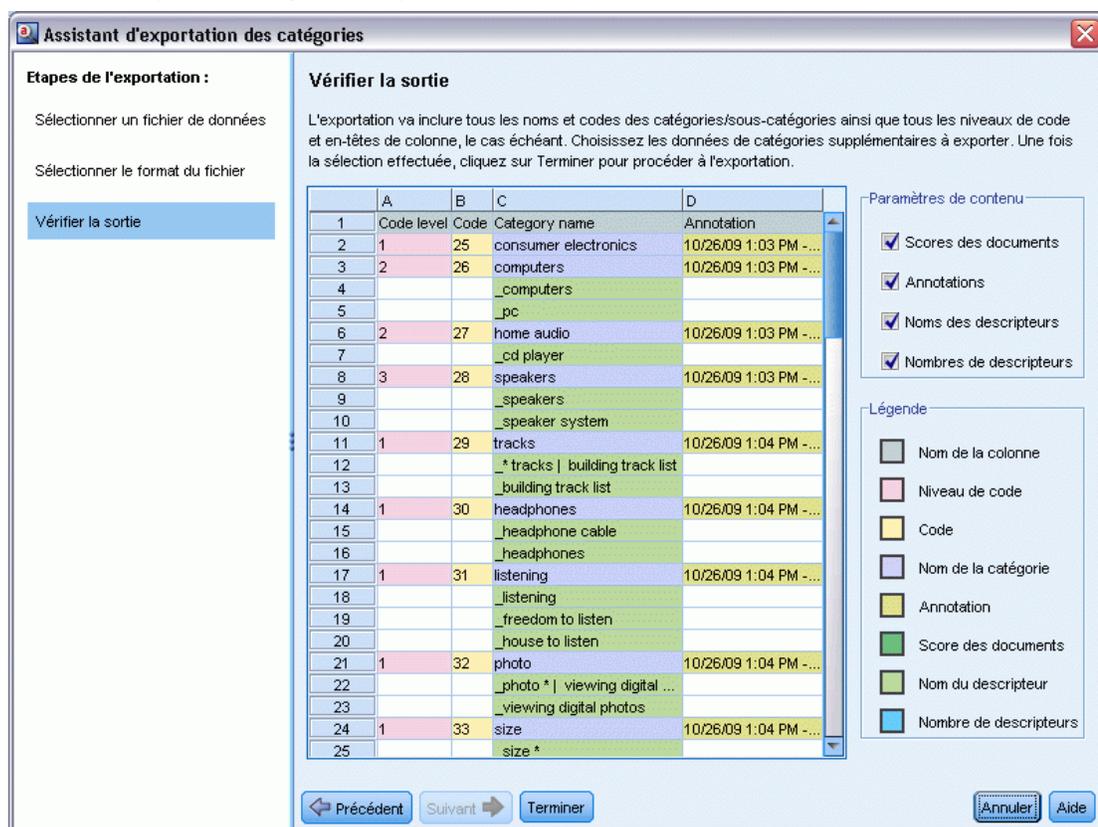
- Sélectionnez un emplacement et saisissez le nom du fichier à exporter.
- Entrez le nom du fichier de sortie dans la zone de texte Nom du Fichier.
- Cliquez sur Suivant pour choisir le format dans lequel vous allez exporter vos données de catégorie.

Figure 6-23
 Assistant d'export de catégories, étape 2



- ▶ Choisissez le format parmi l'une des options suivantes :
 - Format liste plate ou compacte : [Pour plus d'informations, reportez-vous à la section Format liste plate sur p. 140.](#) La liste plate ne contient aucune sous-catégorie. [Pour plus d'informations, reportez-vous à la section Format compact sur p. 141.](#) Le format de liste compacte contient des données hiérarchiques.
 - Format indenté : [Pour plus d'informations, reportez-vous à la section Format indenté sur p. 142.](#)
- ▶ Cliquez sur Suivant pour commencer à choisir le contenu à exporter et pour consulter les données proposées.

Figure 6-24
Assistant d'export de catégories, étape 3



- ▶ Consulter le contenu du fichier exporté.
- ▶ Sélectionnez ou désélectionnez les paramètres de contenu supplémentaire à exporter tels que les annotations ou les noms de descripteur.
- ▶ Cliquez sur Terminer pour exporter les catégories.

Utilisation des règles de catégorie

Vous pouvez créer des catégories de différentes façons. Une de ces façons est de définir des règles de catégorie permettant d'exprimer des idées. Les règles de catégorie sont des instructions qui classifient automatiquement les enregistrements en une catégorie basée sur une expression logique à l'aide de concepts, de types et de patrons extraits ou d'opérateurs booléens. Par exemple, vous pouvez créer une expression qui signifie *inclure tous les enregistrements contenant le concept extrait ambassade mais ne pas inclure argentinedans cette catégorie*.

Alors que certaines règles de catégorie sont produites automatiquement lors de la création des catégories à l'aide de techniques de regroupement, telles que les règles de *cooccurrence* et de *dérivation des racines de concept* (Catégories > Configurer les paramètres > Paramètres avancés : Linguistique), vous pouvez également créer des règles de catégorie manuellement dans l'éditeur de règle en utilisant votre propre compréhension des données et du contexte. Chaque règle est

rattachée à une seule catégorie afin que chaque enregistrement correspondant à la règle soit ensuite scoré dans cette catégorie.

Les règles de catégorie permettent d'améliorer la qualité et la productivité de vos résultats de Text Mining et d'autres analyses quantitatives en vous aidant à classer les réponses dans des catégories plus spécifiques. Votre expérience et vos connaissances du marché peuvent vous offrir une compréhension particulière de vos données et du contexte. Cette compréhension peut vous permettre de traduire vos connaissances en règles de catégorie afin de classer vos enregistrements de manière plus efficace et plus précise en combinant des éléments extraits avec la logique booléenne.

La possibilité de créer ces règles améliore la précision, l'efficacité et la productivité du codage, tout en vous permettant d'incorporer vos connaissances du marché à la technologie d'extraction de produit.

Remarque : pour des exemples de correspondances entre les règles et le texte, voir [Exemples de règles de catégorie sur p. 154](#)

Syntaxe des règles de catégorie

Alors que certaines règles de catégorie sont produites automatiquement lors de la création des catégories à l'aide de techniques de regroupement, telles que les règles de *cooccurrence* et de *dérivation des racines de concept* (Catégories > Configurer les paramètres > Paramètres avancés : Linguistique), vous pouvez également créer des règles de catégorie manuellement dans l'éditeur de règle. Chaque règle est le descripteur d'une seule catégorie afin que chaque enregistrement correspondant à la règle soit automatiquement scoré dans cette catégorie.

Remarque : pour des exemples de correspondances entre les règles et le texte, voir [Exemples de règles de catégorie sur p. 154](#)

Lorsque vous créez ou modifiez une règle, elle doit être ouverte dans le panneau de l'éditeur de règles. Vous pouvez ajouter des concepts, des types ou des patrons et utiliser des caractères génériques pour étendre les correspondances. Lorsque vous utilisez des concepts, des types ou des patrons extraits, vous pouvez bénéficier de la recherche de tous les concepts associés.

Important ! Pour éviter les erreurs courantes, nous vous recommandons de glisser-déposer directement les concepts du panneau Résultats d'extraction, des panneaux Analyse des liens du texte ou du panneau Données dans l'éditeur de règles ou de les ajouter à l'aide des menus contextuels lorsque cela est possible.

Lorsque les concepts, les types et les patrons sont reconnus, une icône apparaît à côté du texte.



Concept extrait



Type extrait



Patron extrait

Syntaxe et opérateurs de règle

Le tableau suivant contient les caractères avec lesquels vous définissez la syntaxe de votre règle. Utilisez ces caractères avec les concepts, les types et les patrons pour créer votre règle.

Table 6-8
Syntaxe prise en charge

Caractère	Description
&	Le “et” booléen. Par exemple, a & b contient à la fois <i>aetb</i> comme dans : - invasion & états-unis - 2016 & jeux olympiques - bonne & pomme
	Le “ou” booléen est inclusif ce qui signifie que si un ou tous les éléments sont trouvés, une correspondance est établie. Par exemple, a b contient soit <i>asoitb</i> comme dans : - attaque france - condominium appartement
!()	Le “pas” booléen. Par exemple, !(a) ne contient pas a comme dans : !(bon & hôtel)
*	Un caractère générique représentant toute expression, d’un caractère unique à un mot entier selon la façon dont il est utilisé. Pour plus d’informations, reportez-vous à la section Utilisation de caractères génériques dans les règles de catégorie sur p. 152.
()	Un délimiteur d’expression. Toute expression entre parenthèses est évaluée en premier.
+	Le connecteur du patron utilisé pour former un patron avec un ordre spécifique. Vous devez utiliser les crochets lorsqu’ils sont disponibles. Pour plus d’informations, reportez-vous à la section Utilisation des patrons TLA dans les règles de catégorie sur p. 150.
[]	Le délimiteur de patrons est nécessaire si vous cherchez à effectuer des correspondances en fonction d’un patron TLA extrait à l’intérieur d’une règle de catégorie. Le contenu entre parenthèses fait référence aux patrons TLA et ne correspondra jamais à des concepts ou des types basés sur une simple cooccurrence. Si vous n’avez pas extrait ce patron TLA, aucune correspondance ne sera possible. Pour plus d’informations, reportez-vous à la section Utilisation des patrons TLA dans les règles de catégorie sur p. 150. N’utilisez pas de crochets si vous cherchez à mettre en correspondance des concepts et des types au lieu des patrons. <i>Remarque</i> : Dans les versions précédentes, les règles de cooccurrence et de synonymes générées par les techniques de création de catégories étaient généralement entre crochets. Dans toutes les nouvelles versions, les crochets indiquent désormais la présence d’un patron TLA. Les règles produites par une technique de cooccurrence et des synonymes sont maintenant entourées de parenthèses, comme dans (<i>enceintes acoustiques enceintes</i>).

Les opérateurs & et | sont commutatifs par conséquent $a \& b = b \& a$ et $a | b = b | a$.

Ignorer les caractères avec une barre oblique inverse

Si un concept contient un caractère qui est également un caractère de syntaxe, vous devez placer une barre oblique inverse devant ce caractère afin que la règle soit correctement interprétée. La barre oblique inverse (\) permet d’ignorer des caractères qui autrement auraient une signification particulière. Lorsque vous le faites glisser dans l’éditeur, la barre oblique inverse est automatiquement ajoutée.

Les caractères de syntaxe de règle suivants doivent être précédés d’une barre oblique inverse si vous souhaitez qu’ils soient traités tels qu’ils sont plutôt que comme une syntaxe de règle :

& ! | + < > () [] *

Par exemple, le concept *r&d* contenant l’opérateur « et » (&), la barre oblique inverse est requise lorsque ce concept est saisi dans l’éditeur de règle, sous la forme : *r\d&d*.

Utilisation des patrons TLA dans les règles de catégorie

Les patrons d'analyse des liens du texte peuvent être précisément définis en règles de catégorie afin d'obtenir des résultats encore plus précis et contextuels. Lorsque vous définissez un patron dans une règle de catégorie, vous ignorez les résultats d'extraction de concept les plus simples et vous utilisez uniquement les documents et les enregistrements correspondants basés sur les résultats des patrons d'analyse des liens du texte extraits.

Délimitation par crochets. Un patron TLA doit être entouré de crochets [] si vous l'utilisez à l'intérieur d'une règle de catégorie. Le délimiteur de patrons est nécessaire si vous cherchez à effectuer des correspondances en fonction d'un patron TLA extrait. Puisque les catégories peuvent contenir des types, des concepts ou des patrons, les crochets expliquent à la règle que le contenu entre crochets fait référence au patron TLA extrait. Si vous n'avez pas extrait ce patron TLA, aucune correspondance ne sera possible. Si vous voyez un patron sans crochets, comme pomme + bonne dans le panneau Catégories, cela signifie que le patron a été ajouté directement à la catégorie en-dehors de l'éditeur de règle de catégorie. Par exemple, si vous ajoutez un patron de concept directement à une catégorie du panneau des résultats d'extraction, il n'apparaîtra pas avec des crochets. Mais, lorsqu'un patron est utilisé à l'intérieur d'une règle de catégorie, vous devez placer le patron entre crochets dans la règle de catégorie, comme dans [banane + !(bon)].

Utilisation du signe + dans les patrons. Dans IBM® SPSS® Text Analytics for Surveys, vous pouvez avoir des patrons en deux parties. Lorsque vous souhaitez indiquer que l'ordre est important, utilisez le signe + pour connecter les éléments. Si l'ordre n'a pas d'importance, vous pouvez utiliser le booléen & à la place. Dans les textes suivants, la position du terme *meilleur* est importante : “les *brocolis*, c'est *meilleur*” et “c'est *meilleur* que les *brocolis*”.

Par exemple, disons que vous aviez les deux échantillons de texte suivants : “J'aime l'ananas” et “Je déteste l'ananas. Néanmoins, j'aime les fraises”. L'expression aime & l'ananas correspondrait aux deux textes car il s'agit de l'expression d'un concept et non pas d'une règle des liens du texte (elle n'est pas entre parenthèses). L'expression ananas + aime correspond uniquement à “J'aime l'ananas” car dans le deuxième texte, le terme *aime* est associé à *fraises* à la place.

Regroupement des patrons. Vous pouvez simplifier les règles avec vos propres patrons. Imaginons que vous souhaitez capturer les trois expressions suivantes, poivre de cayenne + aime, poivre blanc + aime, et poivre + aime. Vous pouvez les regrouper en une seule règle de catégorie comme [poivre * & aime]. Si vous avez une autre expression poivre noir + bon, vous pouvez regrouper les quatre en une règle comme [poivre * + <Positive>].

Ordre des patrons. Afin de mieux organiser les résultats, les règles d'analyse des liens du texte fournies dans les modèles installés avec votre produits tentent de générer des patrons de base dans le même ordre quel que soit l'ordre des mots dans une phrase. Par exemple, si vous avez un enregistrement contenant le texte, “Bonnes présentations.” et un autre contenant “les présentations étaient bonnes”, la même règle correspond à ces deux textes qui sont générés dans le même ordre que présentation + bonne dans les résultats des patrons de concept plutôt que présentation + bonne et que bonne + présentation. Et dans les patrons à deux propriétés comme dans ceux de cet exemple, les concepts affectés aux types dans la

bibliothèque Opinions seront présentés en dernier dans les résultats par défaut comme dans pomme + mauvaise.

Table 6-9

Syntaxe de patrons et utilisation booléenne

Expression	Correspond à un enregistrement qui
[]	Contient tous les patrons TLA. Le délimiteur de patrons est nécessaire <i>dans les règles de catégorie</i> si vous cherchez à effectuer des correspondances en fonction d'un patron TLA extrait. Le contenu entre crochets fait référence aux patrons TLA et pas à des concepts ou types simples. Si vous n'avez pas extrait ce patron TLA, aucune correspondance ne sera possible. Si vous voulez créer une règle qui ne comprend pas de patrons, vous pouvez utiliser !([]).
[a]	Contient un patron dont au moins un des éléments est a quelle que soit sa position dans le patron. Par exemple, [affaire] peut renvoyer [affaire + bonne] ou simplement [affaire + .]
[a + b]	Contient un patron de concept. Par exemple, [affaire + bonne]. <i>Remarque</i> : Si vous souhaitez capturer uniquement ce patron sans ajouter d'autres éléments, nous vous recommandons d'ajouter ce patron directement à votre catégorie plutôt que d'en faire une règle.
[<A> +]	Contient tout patron avec le type <A> comme première propriété et le type comme deuxième propriété et il y a exactement deux propriétés. Le signe + indique que l'ordre des éléments correspondants est important. Par exemple, [<Budget> + <Negative>]. <i>Remarque</i> : Si vous souhaitez capturer uniquement ce patron sans ajouter d'autres éléments, nous vous recommandons d'ajouter ce patron directement à votre catégorie plutôt que d'en faire une règle.
[<A> &]	Contient tout patron de type avec type <A> et type . Par exemple, [<Budget> & <Negative>]. Ce patron TLA ne sera jamais extrait, mais lorsqu'il est écrit ainsi, il est vraiment égal à [<Budget> + <Négatif>] [<Négatif> + <Budget>]. L'ordre des éléments mis en correspondance n'est pas important.
[a + .]	Contient un patron où a est le seul concept et où les autres propriétés de ce patron sont vides. Par exemple : [affaire + .] correspond au patron de concept dans lequel le seul résultat est le concept affaire. Si vous avez ajouté le concept affaire comme descripteur de catégorie, vous obtiendrez tous les enregistrements avec le concept affaire, y compris les instructions positives sur une affaire. Mais, l'utilisation de [affaire + .] ne renverra que les résultats de patrons d'enregistrements qui représentent affaire et aucune autre relation ou opinion et ne renverra pas affaire + fantastique. <i>Remarque</i> : Si vous souhaitez capturer uniquement ce patron sans ajouter d'autres éléments, nous vous recommandons d'ajouter ce patron directement à votre catégorie plutôt que d'en faire une règle.
[<A> + <>]	Contient un patron où <A> est le seul type. Par exemple, [<Budget> + <>] correspond au patron dans lequel le seul résultat est un concept de type <Budget>. <i>Remarque</i> : Vous pouvez utiliser l'opérateur <> pour indiquer un type vide uniquement lorsque vous le placez derrière le symbole + dans un patron de type tel que [<Budget> + <>] mais pas [prix + <>]. <i>Remarque</i> : Si vous souhaitez capturer uniquement ce patron sans ajouter d'autres éléments, nous vous recommandons d'ajouter ce patron directement à votre catégorie plutôt que d'en faire une règle.
[a + !(b)]	Contient au moins un patron qui comprend le concept a mais ne comprend pas le concept b. Doit contenir au moins un patron. Par exemple, [prix + !(élevé)] ou pour les types, [!(<Fruits> <Légumes>) + <Positif>]
!([<A> &])	Ne contient pas de patrons spécifiques. Par exemple, !([<Budget> & <Negative>]).

Remarque : pour des exemples de correspondances entre les règles et le texte, voir [Exemples de règles de catégorie sur p. 154](#)

Utilisation de caractères génériques dans les règles de catégorie

Les caractères génériques peuvent être ajoutés dans les règles afin d'étendre les capacités de mise en correspondance. Le caractère générique * (astérisque) peut être placé avant et/ou après un mot pour indiquer la façon dont les concepts peuvent être mis en correspondance. Deux types de caractères génériques sont disponibles :

- **Les caractères génériques affixes.** Ces caractères génériques sont placés en suffixe ou en préfixe sans espace entre la chaîne et l'astérisque. Par exemple, *opérat** pourrait renvoyer *opérateur, opération, opérations, opérationnel, opérationnelles, etc.*
- **Les caractères génériques nominaux.** Ces caractères génériques sont placés en suffixe ou en préfixe un concept avec un espace entre le concept et l'astérisque. Par exemple, *opération ** pourrait renvoyer *opération, opération chirurgicale, opération mathématique, etc.* De plus, un caractère générique nominal peut être utilisé avec un caractère générique affixe, comme par exemple ** opérat* **, qui pourrait renvoyer *opération, opération chirurgicale, opérateur téléphonique, post opératoire, etc.* Comme cet exemple l'illustre, nous vous recommandons d'utiliser les caractères génériques avec prudence afin que la recherche ne soit pas trop large et ne capture pas de correspondances indésirables.

Exceptions !

- Un caractère générique ne peut jamais être utilisé seul. Par exemple, (*pomme | **) n'est pas accepté.
- Un caractère générique ne peut jamais être utilisé pour mettre en correspondance des noms de type. *<Négatif*>* ne sera pas mis en correspondance avec des noms de type.
- Vous ne pouvez pas éviter que certains types soient mis en correspondance avec des concepts à l'aide de caractères génériques. Le type auquel le concept est affecté est automatiquement utilisé.
- Un caractère générique ne peut pas être placé au milieu d'une séquence de mots, qu'il soit ajouté à la fin d'un mot (*ouv* compte*) ou qu'il soit utilisé seul (*ouvrir * compte*). Vous ne pouvez pas non plus utiliser les caractères génériques dans les noms de type. Par exemple, *mot* mot*, tel que *pot* recette*, ne correspondra pas à une recette de potage ni à rien d'autre. Toutefois, *pomme* ** correspondra à *pomme de terre, pomme vapeur, pomme* etc. Dans un autre exemple, *mot * mot*, comme *gâteau * pomme*, ne correspondra pas à *gâteau cannelle pomme* ni à rien d'autre car l'astérisque apparaît entre deux mots. However, *pomme ** correspondra à *pomme de terre, pomme, pomme vapeur* etc.

Table 6-10
Utilisation des caractères génériques

Expression	Correspond à un enregistrement qui
*pin	Contient un concept qui se termine par une lettre mais qui peut contenir n'importe quel nombre de lettres comme préfixe. Par exemple : *pin se termine par les lettres <i>pin</i> mais peut avoir comme préfixe : - pommes - lapin - rupin
pin*	Contient un concept qui commence par des lettres mais qui peut contenir n'importe quel nombre de lettres comme suffixe. Par exemple : pin* commence par les lettres <i>pin</i> mais peut être suivi d'un suffixe ou d'aucun suffixe comme : - pommes - pincement - pinceau Par exemple, pin* & !(arbre* pomme), qui contient un concept qui débute par les lettres <i>pin</i> mais pas un concept qui commence par les lettres <i>arbre</i> ni le concept <i>pomme</i> ne correspondrait PAS : pin & pomme mais correspondrait à : - pincement - pin & forêt
vers	Contient un concept qui comprend les lettres vers, mais qui peut avoir n'importe quel nombre de lettres comme préfixe, suffixe ou les deux. Par exemple : *vers* pourrait renvoyer : - inverse - pervers - versatile
* prêt	Contient un concept qui comprend le mot prêt mais peut être un composé avec un autre mot placé devant. Par exemple, * prêt pourrait renvoyer : - prêt - banque de prêt - remboursement de prêt Par exemple, [* argent + <Negative>] contient un concept qui se termine par le mot argent et contient un type <Negative> et pourrait renvoyer les patrons de concept suivants : - transfert d'argent + lent - remboursement d'argent + retard
voiture *	Contient un concept qui comprend le mot voiture mais qui peut être un composé suivi d'un autre mot. Par exemple, voiture * pourrait renvoyer : - voiture - voiture de location - voiture de tourisme
* pomme *	Contient un concept qui peut commencer par n'importe quel mot suivi du mot pomme et suivi d'un autre mot. * signifie 0 ou n caractères, donc cela correspond aussi au mot pomme. Par exemple, * pomme * pourrait renvoyer : - pomme de pin - tarte aux pommes caramélisée - Gratin aux pommes de terre - pommes Par exemple, [* réservation * * + <Positive>], qui contient un concept commençant par le mot réservation (sans tenir compte de sa place dans le concept) en première position et contient un type <Positive> en deuxième position pourrait renvoyer les patrons de concept : - système de réservation + bon - réservation en ligne + bon

Remarque : pour des exemples de correspondances entre les règles et le texte, voir [Exemples de règles de catégorie sur p. 154](#)

Exemples de règles de catégorie

Pour mieux comprendre en quoi les correspondances entre les règles et les enregistrements sont différentes selon la syntaxe utilisée pour les exprimer, étudiez l'exemple suivant.

Exemples d'enregistrements

Imaginons deux enregistrements :

- **Enregistrement A** : “lorsque j’ai regardé dans mon porte-monnaie, j’ai vu qu’il y manquait 5 dollars.”
- **Enregistrement B** : “on a trouvé les 5\$ sur l’aire de pique-nique mais la couverture n’était plus là.”

Les deux tableaux suivants montrent ce qui peut être extrait pour les concepts et les types ainsi que les patrons de concept et les patron de type.

Concepts et types extraits de l'exemple

Table 6-11
Exemple de concepts et de types extraits

Concept extrait	Concepts saisis comme
porte-monnaie	<Inconnu>
manquer	<Négatif>
5 dollars	<Devise>
couverture	<Inconnu>
aire de pique-nique	<Inconnu>

Patrons TLA extraits de l'exemple

Table 6-12
Exemple de sortie de patron TLA extraite

Patrons de concepts extraits	Patrons de types extraits	De l'enregistrement
aire de pique-nique + .	<Inconnu> + <>	Enregistrement B
porte-monnaie + .	<Inconnu> + <>	Enregistrement A
couverture + manquer	<Inconnu> + <Négatif>	Enregistrement B
5 dollars + .	<Devise> + <>	Enregistrement B
5 dollars + manquer	<Devise> + <Négatif>	Enregistrement A

Comment mettre en correspondance les règles de catégorie possibles

Le tableau suivant contient une syntaxe qui pourrait être entrée dans l’éditeur de règle de catégorie. Toutes les règles qui se trouvent ici ne fonctionnent pas et toutes ne correspondent pas aux mêmes enregistrements. Regardez la façon dont les différentes syntaxes ont un impact sur les enregistrements mis en correspondance.

Table 6-13
Règles concernant les échantillons

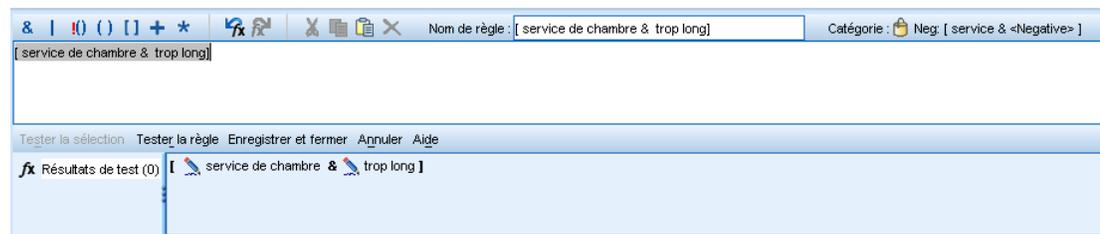
Syntaxe de règle	Résultat
5 dollars & manquer	Correspond à la fois à A et à B car ils contiennent tous les deux le concept extrait manquer et le concept extrait 5 dollars. Cela revient à : (5 dollars & manquer)
manquer & 5 dollars	Correspond à la fois à A et à B car ils contiennent tous les deux le concept extrait manquer et le concept extrait 5 dollars. Cela revient à : (manquer & 5 dollars)
manquer & <Devise>	Correspond à la fois à A et à B car ils contiennent tous les deux le concept extrait manquer et un concept correspondant au type <Devise>. Cela revient à : (manquer & <Devise>)
<Devise> & manquer	Correspond à la fois à A et à B car ils contiennent tous les deux le concept extrait manquer et un concept correspondant au type <Devise>. Cela revient à : (<Devise> & manquer)
[5 dollars + manquer]	Correspond à A mais pas à B car l'enregistrement B n'a pas produit de sortie de patron B contenant 5 dollars + manquer (voir le tableau précédent). Cela revient à la sortie de patron TLA : 5 dollars + manquer
[manquer + 5 dollars]	Ne correspond ni à l'enregistrement A ni au B car aucun patron TLA extrait (voir tableau précédent) ne correspond à l'ordre exprimé ici avec manquer en première position. Cela revient à la sortie de patron TLA : 5 dollars + manquer
[manquer & 5 dollars]	Correspond à A mais pas à B car aucun patron TLA correspondant n'a été extrait de l'enregistrement B. Utiliser le caractère & indique que l'ordre n'a pas d'importance lors de la mise en correspondance ; par conséquent, cette règle recherche une correspondance de patron avec [manquer + 5 dollars] ou [5 dollars + manquer]. Seul [5 dollars + manquer] de l'enregistrement A possède une correspondance.
[manquer + <Devise>]	Ne correspond ni à l'enregistrement A ni à l'enregistrement B car aucun patron TLA extrait ne correspondait à cet ordre. Cela n'a pas d'équivalent car une sortie TLA n'est basée que sur les termes (5 dollars + manquer) ou sur les types (<Devise> + <Négatif>), mais ne mélange pas les types et les concepts.
[<Devise> + <Négatif>]	Correspond à l'enregistrement A mais pas à l'enregistrement B car aucun patron TLA n'a été extrait de l'enregistrement B. Cela équivaut à la sortie TLA : <Devise> + <Négatif>
[<Négatif> + <Devise>]	Ne correspond ni à l'enregistrement A ni à l'enregistrement B car aucun patron TLA extrait ne correspondait à cet ordre. Par défaut, dans le modèle Opinions lorsqu'une rubrique est trouvée avec une opinion, cette rubrique (<Devise>) occupe la première position et opinion (<Négatif>) occupe la deuxième.

Création de règles de catégorie

Lorsque vous créez ou modifiez une règle de catégorie, elle doit être ouverte dans le panneau de l'éditeur de règles. Vous pouvez ajouter des concepts, des types ou des patrons et utiliser des caractères génériques pour étendre les correspondances. Lorsque vous utilisez des concepts, des types ou des patrons reconnus, vous pouvez bénéficier de la recherche de tous les concepts associés. Par exemple, lorsque vous utilisez un concept, tous ses termes associés, ses formes plurielles et ses synonymes sont également associés à cette règle. De même, lorsque vous utilisez un type, tous ses concepts sont également capturés par la règle.

Vous pouvez ouvrir l'éditeur de règles en modifiant une règle existante ou en faisant un clic droit sur le nom de la catégorie et en choisissant Créer une règle.

Figure 6-25
Panneau de l'éditeur de règles



Vous pouvez utiliser les menus contextuels, la méthode glisser-déposer ou saisir manuellement les concepts, les types et les patrons dans l'éditeur. Combinez-les ensuite avec les opérateurs booléens (&, ! (), |) et les parenthèses pour former vos expressions de règle. Pour éviter les erreurs communes, nous vous recommandons de déplacer les concepts en les faisant glisser directement du panneau Résultats d'extraction ou du panneau Données vers l'éditeur de règle. Soyez attentif à la syntaxe des règles afin d'éviter les erreurs. [Pour plus d'informations, reportez-vous à la section Syntaxe des règles de catégorie sur p. 148.](#)

Remarque : pour des exemples de correspondances entre les règles et le texte, voir [Exemples de règles de catégorie sur p. 154](#)

Pour créer une règle

- ▶ Si vous n'avez pas encore extrait de données ou que votre extraction a expiré, faites-le maintenant. [Pour plus d'informations, reportez-vous à la section Extraction de données dans le chapitre 5 sur p. 85.](#)
- ▶ Dans le panneau Catégories, sélectionnez la catégorie dans laquelle vous voulez ajouter votre règle.
- ▶ Dans les menus, choisissez Catégories > Créer une règle. Le panneau de l'éditeur des règles de catégorie apparaît dans la fenêtre.
- ▶ Dans le champ Nom de règle, entrez un nom pour votre règle. Si vous ne donnez pas de nom à la règle, l'expression sera automatiquement utilisée comme nom. Vous pourrez renommer cette règle ultérieurement.

- ▶ Dans le plus grand champ de texte de l'expression, vous pouvez effectuer les opérations suivantes :
 - Saisir directement le texte dans le champ ou le faire glisser depuis un autre panneau. Utilisez uniquement des concepts, types et patrons extraits. Par exemple, si vous entrez le mot `chats`, mais que seule la forme au singulier, `chat`, apparaît dans le panneau Résultats d'extraction, l'éditeur ne sera pas en mesure de reconnaître `chats`. Dans ce cas, la forme singulière pourra inclure automatiquement la forme plurielle, ou vous pourrez utiliser un caractère générique. [Pour plus d'informations, reportez-vous à la section Syntaxe des règles de catégorie sur p. 148.](#)
 - Sélectionner les concepts, types ou patrons à ajouter aux règles et utiliser les menus.
 - Ajouter des opérateurs booléens pour relier les éléments de votre règle. Utilisez les boutons de la barre d'outils pour ajouter le booléen "et" `&`, le booléen "ou" `|`, le booléen "pas" `!`, les parenthèses `()`, et les crochets pour les patrons `[]` à votre règle.
- ▶ Cliquez sur le bouton Tester la règle pour vérifier que votre règle est bien constituée. [Pour plus d'informations, reportez-vous à la section Syntaxe des règles de catégorie sur p. 148.](#) Le nombre d'enregistrements trouvés apparaît entre parenthèses en regard du texte Résultats de test. Les éléments de la règle qui ont été reconnus ou les messages d'erreur éventuels apparaissent à droite de ce texte. Si le graphique à côté du type, du patron ou du concept apparaît avec un point d'interrogation rouge, ceci indique que l'élément ne correspond à aucune extraction connue. S'il n'y a pas de correspondance, la règle ne trouvera aucun enregistrement.
- ▶ Pour tester une partie de votre règle, sélectionnez-la, puis cliquez sur Tester la sélection.
- ▶ En cas de problème, apportez toutes les modifications nécessaires et testez à nouveau la règle.
- ▶ Lorsque vous avez terminé, cliquez sur Enregistrer & Fermer pour enregistrer à nouveau la règle et fermer l'éditeur. Le nouveau nom de la règle apparaît dans la catégorie.

Figure 6-26
Règle dans le panneau Catégories

Catégorie	Nombre ...	Nombre ...
friends+<Contextual>	1	2
interface+<Positive>	1	4
cd collection+<Positive>	1	2
volume+<Contextual>	1	2
unit+<Positive>	1	1
music	12	28
sound	7	23
listening	2	6
capacity	4	12
capacity+<Contextual>	3	10
fx [<Budget> & (<Positive> low small)]	5	
storage capacity+<Contextual>	2	5
capacity+<Positive>	1	2
size	3	14
cds	2	5

Modification et suppression des règles

Vous pouvez éditer à tout moment une règle que vous avez créée et enregistrée. [Pour plus d'informations, reportez-vous à la section Syntaxe des règles de catégorie sur p. 148.](#)

Si une règle ne vous est plus utile, vous pouvez la supprimer.

Pour modifier des règles

- ▶ Dans le panneau des catégories, sélectionnez la règle à modifier.
- ▶ Dans les menus, choisissez Catégories > Editer la règle ou double-cliquez sur le nom de la règle. L'éditeur s'ouvre, avec la règle sélectionnée.
- ▶ Modifiez la règle à l'aide des résultats de l'extraction et des boutons de la barre d'outils.
- ▶ Testez à nouveau la règle pour vous assurer qu'elle renvoie les résultats attendus.
- ▶ Cliquez sur Enregistrer & Fermer pour enregistrer à nouveau la règle et fermer l'éditeur.

Pour supprimer une règle

- ▶ Dans le panneau des catégories, sélectionnez la règle à supprimer.
- ▶ Dans les menus, sélectionnez Edition > Supprimer. La règle est supprimée de la catégorie.

Edition et réglage des catégories

Une fois les catégories créées, vous souhaitez certainement les analyser et procéder à des ajustements. Outre le réglage des ressources linguistiques, vous devez procéder à l'examen de vos catégories en recherchant des moyens de combiner ou de nettoyer leurs définitions. Vérifiez également certains enregistrements catégorisés. Vous pouvez également examiner les enregistrements d'une catégorie et y opérer des ajustements, de sorte que les catégories puissent collecter les nuances et les distinctions.

Vous pouvez utiliser les techniques intégrées de création de catégories automatisées pour créer vos catégories. Cependant, vous voudrez probablement apporter quelques modifications à ces catégories. Lorsque vous utilisez une ou plusieurs techniques, un certain nombre de nouvelles catégories apparaissent dans la fenêtre. Vous pouvez alors examiner les données d'une catégorie et les ajuster jusqu'à ce que vos définitions de catégorie vous conviennent. [Pour plus d'informations, reportez-vous à la section A propos des catégories sur p. 110.](#)

Voici quelques options pour affiner vos catégories, la plupart étant décrites dans les pages suivantes :

- Modification des propriétés de la catégorie (changement de nom, ajout d'étiquettes et d'annotations)
- Ajout de descripteurs à vos catégories
- Modification de catégories
- Déplacement de catégories
- Aplatissement des catégories hiérarchiques

- Fusion de catégories
- Ajout d'une correspondance de texte
- Forcer les réponses à entrer dans les catégories
- Copie et réutilisation de catégories
- Suppression de catégories
- Modification et réextraction des ressources linguistiques
- Visualisation de l'interaction de vos catégories et ajustements. [Pour plus d'informations, reportez-vous à la section Visualisation des graphiques dans le chapitre 7 sur p. 170.](#)

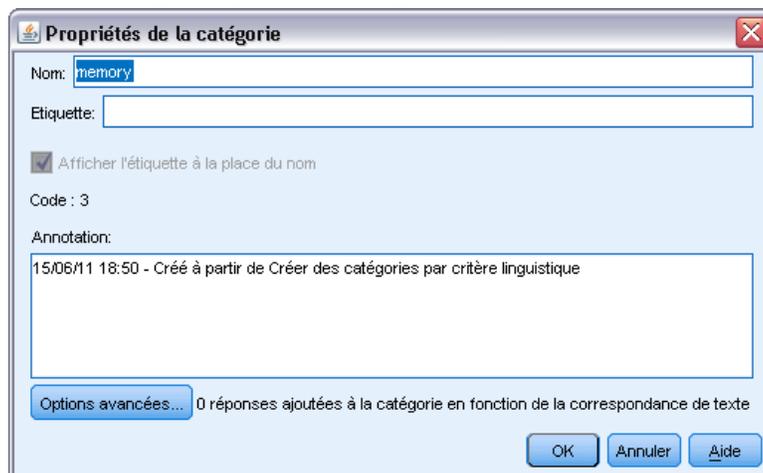
Modification des propriétés de catégories

Comme de nombreux autres éléments dans IBM® SPSS® Text Analytics for Surveys, vous pouvez éditer les propriétés de vos catégories : nom, étiquette, annotations et correspondances de texte avancées. [Pour plus d'informations, reportez-vous à la section Propriétés de la catégorie sur p. 111.](#) En plus de modifier les propriétés, vous pouvez voir le nombre d'éléments inclus dans la définition de catégorie, c'est-à-dire le nombre de termes, de types, de patrons TLA ou de règles de catégorie constituant cette catégorie. Le numéro de code s'affiche également et correspond à la valeur du code dans le Plan de codage.

Pour modifier les propriétés de catégorie :

- ▶ Dans les menus, sélectionnez Catégories > Propriétés de catégorie. La boîte de dialogue Propriétés de catégorie apparaît.

Figure 6-27
Boîte de dialogue Propriétés de catégorie



- ▶ Vous pouvez renommer la catégorie en saisissant un nouveau nom dans le champ Nom.
- ▶ Modifier le nom ou l'étiquette de la catégorie.
- ▶ Pour utiliser l'étiquette au lieu du nom de la catégorie dans l'interface, par exemple dans le panneau des catégories, sélectionnez Afficher l'étiquette à la place du nom.

- ▶ Si vous le souhaitez, vous pouvez entrer une annotation dans le champ Annotation.
- ▶ Pour forcer un mot ou une phrase à entrer dans la définition de catégorie, cliquez sur Avancé et entrez vos correspondances de texte dans le tableau. [Pour plus d'informations, reportez-vous à la section Correspondance de texte dans les catégories sur p. 165.](#)
- ▶ Cliquez sur OK pour appliquer les modifications.

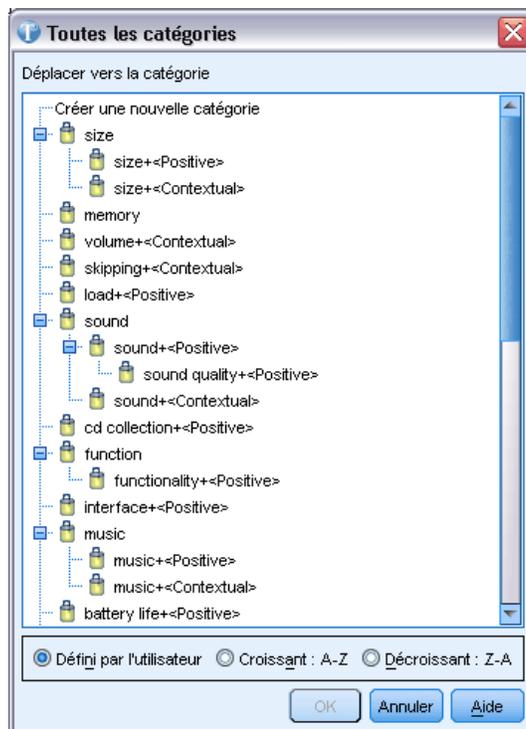
Ajout de descripteurs aux catégories

Après avoir utilisé des techniques automatisées, il est possible que vous disposiez encore de résultats d'extraction qui n'ont été utilisés dans aucune des définitions de catégorie. Consultez cette liste dans le panneau des Résultats de l'extraction. Si vous trouvez des éléments que vous souhaitez déplacer vers une catégorie, vous pouvez les ajouter dans une catégorie existante ou nouvelle.

Pour ajouter un concept ou un type à une catégorie

- ▶ Dans les panneaux Résultats d'extraction et Données, sélectionnez les éléments que vous souhaitez ajouter à une catégorie nouvelle ou existante.
- ▶ Dans les menus, sélectionnez Catégories > Ajouter à la catégorie. La boîte de dialogue Toutes les catégories apparaît pour présenter l'ensemble des catégories. Sélectionnez la catégorie à laquelle vous voulez ajouter les éléments sélectionnés. Si vous voulez ajouter les éléments à une nouvelle catégorie, sélectionnez Nouvelle catégorie. Une nouvelle catégorie apparaît dans le panneau Catégories, sous le nom du premier élément sélectionné.

Figure 6-28
Boîte de dialogue Toutes les catégories



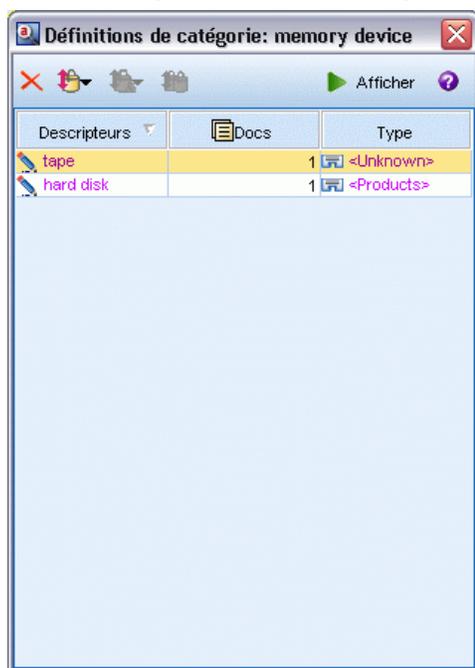
Modification des descripteurs de catégorie

Dès que vous avez créé des catégories, vous pouvez ouvrir chacune d'entre elles pour visualiser l'ensemble des descripteurs qui constituent sa définition. Dans la boîte de dialogue Définitions de catégorie, vous pouvez apporter un certain nombre de modifications à vos descripteurs de catégorie. De plus, si des catégories apparaissent dans l'arborescence des catégories, vous pouvez également les utiliser.

Pour éditer une catégorie

- ▶ Sélectionnez la catégorie à éditer dans le panneau Catégories.
- ▶ Dans les menus, sélectionnez Affichage > Définitions de catégorie. La boîte de dialogue Définitions de catégorie apparaît.

Figure 6-29
Boîte de dialogue Définitions de catégorie



- Sélectionnez le descripteur que vous souhaitez éditer, puis cliquez sur le bouton correspondant dans la barre d'outils.

Le tableau ci-dessous décrit tous les boutons de la barre d'outils qui permettent de modifier vos définitions de catégorie.

Table 6-14
Boutons de la barre d'outils et descriptions

Icônes	Description
	Supprime les descripteurs sélectionnés de la catégorie.
	Déplace les descripteurs sélectionnés vers une catégorie existante ou nouvelle.
	Déplace les descripteurs sélectionnés vers une catégorie sous la forme d'une règle de catégorie &. Pour plus d'informations, reportez-vous à la section Utilisation des règles de catégorie sur p. 147.
	Déplace chacun des descripteurs sélectionnés dans une nouvelle catégorie qui lui est propre.
	Met à jour l'affichage des panneaux Données et Visualisation en fonction des descripteurs sélectionnés.

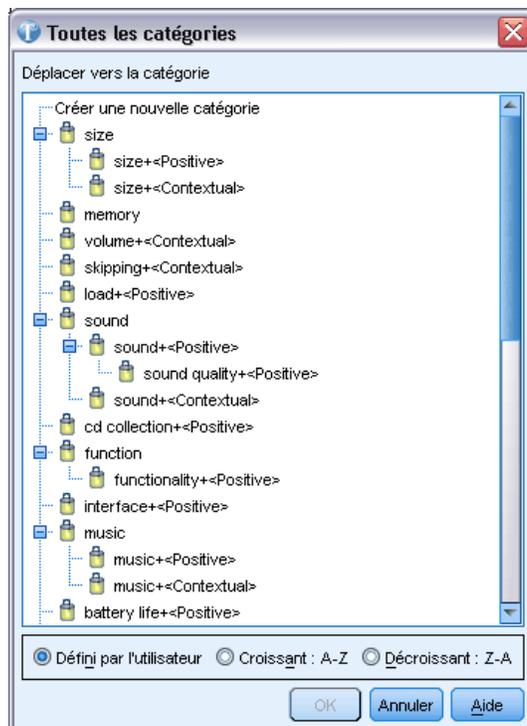
Déplacement de catégories

Si vous voulez placer une catégorie dans une autre catégorie existante ou déplacer des descripteurs dans une autre catégorie, vous pouvez les déplacer.

Pour déplacer une catégorie

- ▶ Dans le panneau Catégories, sélectionnez les catégories ou les descripteurs à déplacer dans une autre catégorie.
- ▶ Dans les menus, choisissez Catégories > Déplacer vers la catégorie. Le menu présente un ensemble de catégories, la plus récente figurant en haut de la liste. Sélectionnez le nom de la catégorie vers laquelle vous voulez déplacer les concepts sélectionnés.
 - Si vous voyez le nom que vous recherchez, sélectionnez-le pour l'ajouter à cette catégorie.
 - S'il n'apparaît pas, sélectionnez Plus pour afficher la boîte de dialogue Toutes les catégories et sélectionnez la catégorie dans la liste.

Figure 6-30
Boîte de dialogue Toutes les catégories



Fusion ou combinaison de catégories

Si vous souhaitez combiner deux catégories ou plus dans une nouvelle catégorie, vous pouvez les fusionner. Quand vous fusionnez des catégories, une nouvelle catégorie est créée avec un nom générique. Tous les concepts, types et patrons utilisés dans les descripteurs de catégorie sont déplacés dans cette nouvelle catégorie. Par la suite, vous pouvez renommer cette catégorie en éditant ses propriétés. [Pour plus d'informations, reportez-vous à la section Modification des propriétés de catégories sur p. 159.](#)

Pour fusionner tout ou partie d'une catégorie

- ▶ Dans le panneau Catégories, sélectionnez les éléments que vous souhaitez fusionner.
- ▶ Dans les menus, sélectionnez Catégories > Fusionner les catégories. La boîte de dialogue Propriétés des catégories s'affiche et vous permet d'entrer un nom pour la catégorie nouvellement créée. Les catégories sélectionnées sont fusionnées dans la nouvelle catégorie en tant que sous-catégories.

Forcer les réponses à entrer dans des catégories

Le fait de forcer les réponses à entrer et sortir des catégories vous permet de remplacer les définitions de catégorie créées par les techniques de création de catégorie automatisées sans modifier la définition de catégorie actuelle. Vous pouvez découvrir que bien que la réponse contienne des termes utilisés pour définir une catégorie particulière, la réponse elle-même ne doit pas se trouver dans cette catégorie. Dans ce cas, vous pouvez forcer la réponse à sortir de cette catégorie sans devoir supprimer les termes de la définition de catégorie.

L'ajout des termes forcés est utilisée dans des cas particuliers où une réponse correspond (ou ne correspond pas) à une catégorie mais pour une raison ou une autre (par exemple, elle contient un terme particulier), elle est attribuée (ou non attribuée) à cette catégorie. Ceci se produit généralement lorsqu'une personne interrogée utilise l'ironie dans sa réponse, par exemple "La pizza était excellente. Je suis sûr que tout le monde aime la pizza froide et brûlée." Supposons que vous avez une catégorie nommée Pos: [`<Nourriture>` + `<Positive>`] pour capturer l'opinion positive concernant la nourriture servie par un restaurant, cette réponse peut être attribuée à cette catégorie. Dans ce cas, vous souhaitez peut-être forcer cette réponse à sortir de la catégorie.

Pour forcer les réponses à entrer ou à sortir des catégories

- ▶ Dans le panneau Données, sélectionnez les réponses que vous souhaitez forcer dans ou hors d'une catégorie particulière.
- ▶ Dans les menus, sélectionnez Catégories > Forcer la réponse à entrer ou Catégories > Forcer la réponse à sortir. Un sous-menu affiche la liste des catégories que vous pouvez sélectionner.
- ▶ Sélectionnez la catégorie vers laquelle ou hors de laquelle vous souhaitez forcer cette réponse. Si vous avez créé de nombreuses catégories, certaines peuvent ne pas être visibles dans le sous-menu.
 - Dans ce cas, sélectionnez Plus au bas du sous-menu. La boîte de dialogue Toutes les catégories s'ouvre, dans laquelle vous pouvez sélectionner la catégorie et cliquer sur OK pour appliquer la modification.
 - Si vous voulez forcer la réponse à entrer dans une nouvelle catégorie, sélectionnez Nouvelle catégorie vide. Une nouvelle catégorie apparaît dans l'arborescence des catégories sous un nom générique.

Chaque fois qu'une catégorie contient une ou plusieurs réponses forcées, une pseudo-catégorie appelée soit Forcer à entrer soit Forcer à sortir est affichée sous le nom de la catégorie dans l'arborescence. Vous pouvez aussi voir les réponses forcées dans ou hors d'une catégorie en affichant les colonnes Forcée à entrer et Forcée à sortir dans le panneau Données. [Pour plus d'informations, reportez-vous à la section Le panneau Données sur p. 100.](#)

Pour supprimer un état de réponse forcée

- ▶ Dans le panneau Données, sélectionnez les réponses que vous ne souhaitez plus forcer à entrer ou sortir d'une catégorie.
- ▶ Dans les menus, sélectionnez Catégories > Forcer la réponse à entrer pour la forcer à entrer, ou sélectionnez Catégories > Forcer la réponse à sortir pour la forcer à sortir. Les catégories dans lesquelles la réponse est forcée à sortir ou entrer sont précédées d'une coche.

Figure 6-31

Forcer des réponses dans le panneau Données

	Id	Réponse	Catégories
1	1	Excellentes prestations générales. J'ai particulièrement apprécié les salons où l'on pouvait travailler ou se reposer en toute tranquillité	prestations
2	2	qualité du service, accueil très chaleureux, discrétion du personnel	accueil
3	3	propreté des lieux, accueil chaleureux, buffet extraordinaire	accueil buffet lieu
4	5		accueil
5	7		climatisation
6	8		ore urant e
7	9		ore
8	10		e
9	11		ore
10	12		il
11	13		urant ore
12	14	elle chambre, service agréable	ore e

Ajouter à la catégorie	
Ajouter à la règle	Ctrl+U
Ajouter à l'annotation	
<input checked="" type="checkbox"/> Forcer la réponse à entrer dans	
<input checked="" type="checkbox"/> Forcer la réponse à sortir de	
Marquer les réponses par	
Trier	
Sélectionner tout	Ctrl+A
Copier	Ctrl+C
Rechercher	Ctrl+F
Afficher les colonnes	

Nouvelle catégorie vide...	
<input type="checkbox"/>	internet
<input checked="" type="checkbox"/>	ambiance
<input type="checkbox"/>	buffet
<input type="checkbox"/>	cuisine
<input type="checkbox"/>	serveurs
<input type="checkbox"/>	enfants
<input type="checkbox"/>	nourriture
<input type="checkbox"/>	airport

- ▶ Sélectionnez la catégorie dans le sous-menu coché pour laquelle vous souhaitez supprimer l'imposition. La coche est supprimée et la réponse n'est plus forcée.

Pour supprimer un état de réponse forcée

Pour effacer tous les états de réponse forcée :

- ▶ Dans les menus, sélectionnez Catégories > Effacer tout > Forcées à entrer ou Catégories > Supprimer tout > Forcées à sortir. L'état forcé des réponses est effacé et elles ne sont plus forcées vers ou hors des catégories.

Correspondance de texte dans les catégories

Si vous avez essayé de forcer l'extraction d'un mot ou d'une phrase à travers les ressources linguistiques mais sans succès en raison d'autres règles de traitement linguistique, vous pouvez créer une correspondance de texte pour attribuer directement toute catégorie contenant ce texte dans une catégorie particulière sans utiliser de résultat d'extraction.

Lorsque vous ajoutez des correspondances de texte (un mot ou une phrase) dans une catégorie, IBM® SPSS® Text Analytics for Surveys attribue automatiquement toute réponse contenant le mot ou la phrase à cette catégorie. La correspondance de texte doit uniquement être utilisée si vous avez déjà essayé d'ajouter ce mot aux ressources linguistiques afin de tirer profit au mieux de cette définition. [Pour plus d'informations, reportez-vous à la section Ajout des termes forcés dans le chapitre 10 sur p. 228.](#)

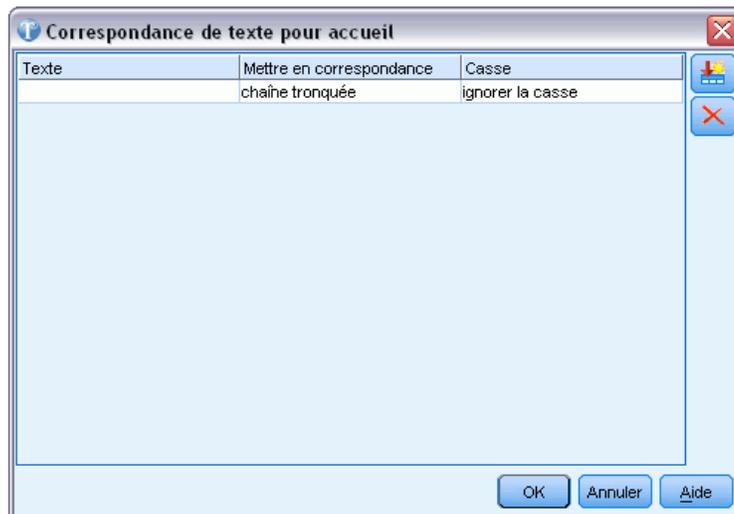
Chaque fois qu'une correspondance de texte est ajoutée à la définition d'une catégorie, une pseudo-catégorie appelée Correspondance de texte est affichée sous la catégorie dans l'arborescence des catégories. [Pour plus d'informations, reportez-vous à la section Le panneau Données sur p. 100.](#)

Vous pouvez aussi supprimer des entrées de ce tableau en sélectionnant la ou les ligne(s) concernées et en cliquant sur le bouton Supprimer.

Pour ajouter une correspondance de texte à une définition de catégorie :

- ▶ Dans le panneau Données, identifiez le mot ou la phrase que vous voulez forcer à entrer dans une définition de catégorie.
- ▶ Dans le panneau Catégories, sélectionnez la catégorie dans laquelle vous voulez forcer ce mot ou cette phrase à entrer.
- ▶ Dans les menus, sélectionnez Catégories > Propriétés de catégorie. La boîte de dialogue Propriétés de catégorie apparaît.
- ▶ Cliquez sur Avancé. La boîte de dialogue Correspondance de texte apparaît.

Figure 6-32
Boîte de dialogue Correspondance de texte



- ▶ Dans le tableau, saisissez le mot ou la phrase dans la première cellule de la colonne Texte.
- ▶ Sélectionnez la manière dont ce mot ou cette phrase doit correspondre au texte trouvé dans les réponses. Pour que le mot ou la phrase corresponde exactement à ce que vous avez saisi, sélectionnez l'ensemble du mot ou de la phrase. Pour une correspondance à de plus longues phrases, sélectionnez chaîne partielle.

- ▶ Si le mot ou la phrase que vous saisissez contient des majuscules et des minuscules, sélectionnez respecter la casse dans la colonne Casse.
- ▶ Cliquez sur OK pour enregistrer vos modifications et fermer la boîte de dialogue. Le nombre de réponses attribuées à la catégorie avec les correspondances de texte est mis à jour et s'affiche dans la boîte de dialogue.
- ▶ Cliquez sur OK pour appliquer les modifications.

Copie de catégories

Quand vous utilisez des questions identiques ou similaires dans une ou plusieurs enquêtes, vous pouvez réutiliser les définitions de catégorie pour gagner un temps précieux. Vous pouvez copier les catégories d'une question vers une autre au sein d'un même projet. Lorsque vous réutilisez des catégories, vous devez procéder à une nouvelle extraction pour faire correspondre les catégories et les données de réponse. Avant la réextraction, les catégories apparaissent dans le panneau Catégories, accompagnées d'un point d'interrogation (?) qui représente l'évaluation de fréquence.

Remarque : Pour réutiliser des catégories dans un autre projet, nous vous recommandons de créer un package d'analyse de texte avec les catégories et les ressources dans votre projet et d'utiliser ce package d'analyse de texte (TAP) lors de la création de votre nouveau projet dans l'assistant. [Pour plus d'informations, reportez-vous à la section Utilisation des packages d'analyse de texte dans le chapitre 3 sur p. 43.](#)

Pour copier des catégories au sein d'un même projet

- ▶ Accédez à la question dont vous voulez copier les catégories.
- ▶ Dans l'arborescence du panneau Catégories, sélectionnez toutes les catégories.
- ▶ Dans les menus, sélectionnez Edition > Copier pour copier les catégories.
- ▶ Accédez à la question (Afficher > Question) dans laquelle vous souhaitez coller les catégories.
- ▶ Dans les menus, sélectionnez Edition > Coller pour coller les catégories. Les catégories sont ajoutées au panneau. Les évaluations de fréquences ne sont pas connues car vous n'avez pas effectué de réextraction. Par conséquent, les évaluations apparaissent sous la forme d'un point d'interrogation (?).

Impression de catégories

Vous pouvez imprimer la vue de l'arborescence du panneau Catégories.

Pour imprimer la vue de l'arborescence de catégorie

- ▶ En fonction de ce que vous souhaitez imprimer, vous pouvez développer, réduire ou trier les éléments de l'arborescence du panneau Catégories.
- ▶ Dans les menus, choisissez Fichier > Imprimer > Imprimer les catégories. La boîte de dialogue Aperçu avant impression apparaît.

Figure 6-33
Boîte de dialogue Aperçu avant impression

Catégorie	Compteur 1	Compteur 2
Tous les enregistrements	-	405
Sans catégorie	-	282
Aucun concept extrait	-	4
songs+<Positive>	3	9
memory	2	2
fx [memory + easy to use]		1
fx [storage + able]		1
battery+<Contextual>	2	3
design+<Positive>	2	7
store+<Positive>	2	3
skipping+<Contextual>	2	2
look+<Positive>	2	3
feature+<Positive>	2	3
software+<Positive>	1	3
function	2	2
playlists+<Positive>	1	2
battery life+<Positive>	1	5
headphones+<Positive>	1	2
load+<Positive>	1	2
friends+<Contextual>	1	2
interface+<Positive>	1	4
cd collection+<Positive>	1	2
volume+<Contextual>	1	2
unit+<Positive>	1	1
music	12	28
sound	7	23
listening	2	6
capacity	4	12
size	3	14
cds	2	5

- Cliquez sur le bouton imprimer pour imprimer la vue telle qu'elle apparaît dans la boîte de dialogue.

Suppression de catégories

Si vous ne souhaitez pas conserver une catégorie, vous pouvez la supprimer. Lorsque vous supprimez une catégorie, tout concept non utilisé dans une autre catégorie est visible dans l'onglet Extractions non utilisées dans le panneau des Résultats de l'extraction.

Pour supprimer une catégorie

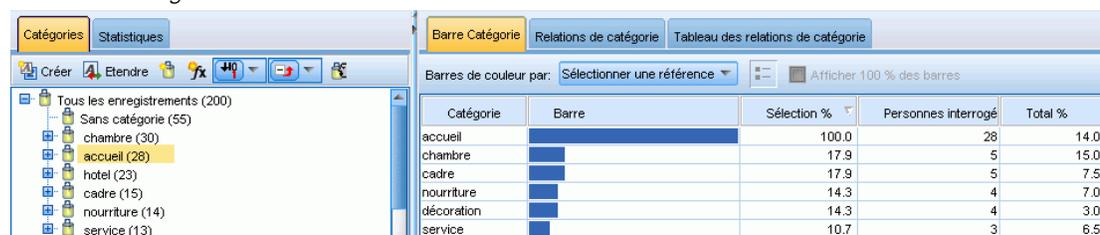
- ▶ Dans le panneau Catégories, sélectionnez la ou les catégories à supprimer.
- ▶ Dans les menus, sélectionnez Edition > Supprimer.

Visualisation des graphiques

Lors de la création des catégories, il est important de prendre le temps de passer en revue les définitions de catégorie, les réponses qu'elles contiennent, et les chevauchements de catégorie. Le panneau de visualisation offre plusieurs perspectives sur les catégories. Le panneau Visualisation est situé dans l'angle supérieur droit de la vue Question. S'il est masqué, vous pouvez y accéder à partir du menu Affichage (Affichage > Panneaux > Visualisation).

Dans cette vue, le panneau de visualisation offre trois perspectives sur les similarités de catégorisation des réponses. Vous pouvez vous appuyer sur les graphiques et diagrammes de ce panneau pour analyser les résultats de la catégorisation, et affiner les catégories ou générer des rapports. Lorsque vous affinez des catégories, vous pouvez utiliser ce panneau pour examiner les définitions de catégorie afin de découvrir les catégories trop similaires (par exemple, les catégories qui partagent plus de 75 % de leurs réponses) ou trop différentes. Si deux catégories sont trop similaires, vous pouvez décider de les combiner. Vous pouvez également décider d'affiner les définitions des catégories en supprimant certains descripteurs d'une catégorie. Vous pouvez copier, coller et imprimer les résultats dans ce panneau pour faciliter votre analyse ou à des fins de rapport.

Figure 7-1
Panneaux Catégorie et Visualisation



En fonction des éléments sélectionnés dans le panneau Résultats d'extraction ou dans le panneau Catégories, vous pouvez visualiser les interactions correspondantes entre les réponses et les catégories de chaque onglet de ce panneau. Chacun présente des informations similaires, mais d'une façon différente ou avec un niveau de détail différent. Si nécessaire, vous pouvez personnaliser les couleurs utilisées dans ces graphiques et diagrammes dans la boîte de dialogue Options. [Pour plus d'informations, reportez-vous à la section Options : Onglet Affichage dans le chapitre 2 sur p. 21.](#)

Remarque : Vous pouvez aussi générer des graphiques récapitulatifs, comme un diagramme en bâton des 5 premières catégories. Ces graphiques, qui sont exportés en HTML, peuvent ensuite être utilisés dans des présentations. [Pour plus d'informations, reportez-vous à la section Exporter des graphiques récapitulatifs dans le chapitre 4 sur p. 60.](#)

Le panneau Visualisation propose les graphiques et diagrammes suivants :

- **Diagramme Barre Catégorie.** Un tableau et un diagramme en bâton présentent le chevauchement entre les réponses correspondant à votre sélection et les catégories associées. Le diagramme en bâton présente également le rapport entre les réponses dans les catégories et le nombre total de réponses. Vous pouvez également sélectionner une variable de référence (si vous en

avez importé une) afin de comparer les valeurs de variable de référence des enregistrements dans chaque catégorie. [Pour plus d'informations, reportez-vous à la section Diagramme Barre Catégorie sur p. 171.](#)

- **Graphique Relations de catégorie.** Ce graphique présente le chevauchement des réponses pour les catégories auxquelles appartiennent les réponses, en fonction de la sélection dans les autres panneaux. [Pour plus d'informations, reportez-vous à la section Graphique Relations de catégorie sur p. 172.](#)
- **Tableau des relations de catégorie.** Cet onglet présente les mêmes informations que l'onglet Relations de catégorie, mais sous forme de tableau. Il contient trois colonnes qu'il est possible de trier en cliquant sur leur en-tête. [Pour plus d'informations, reportez-vous à la section Tableau des relations de catégorie sur p. 173.](#)

[Pour plus d'informations, reportez-vous à la section Catégorisation des données textuelles dans le chapitre 6 sur p. 96.](#)

Diagramme Barre Catégorie

Cet onglet affiche un tableau et un diagramme en bâton qui présentent le chevauchement entre les réponses correspondant à votre sélection et les catégories associées. Le diagramme en bâton présente également le rapport entre les réponses dans les catégories et le nombre total de réponses. Vous ne pouvez pas modifier la présentation de ce diagramme.

Vous pouvez utiliser les menus contextuels dans ce diagramme en bâton pour trier les colonnes, modifier les couleurs du graphique, sélectionner le contenu du diagramme, copier ce contenu ainsi qu'afficher ou masquer la légende.

Le diagramme contient les colonnes suivantes :

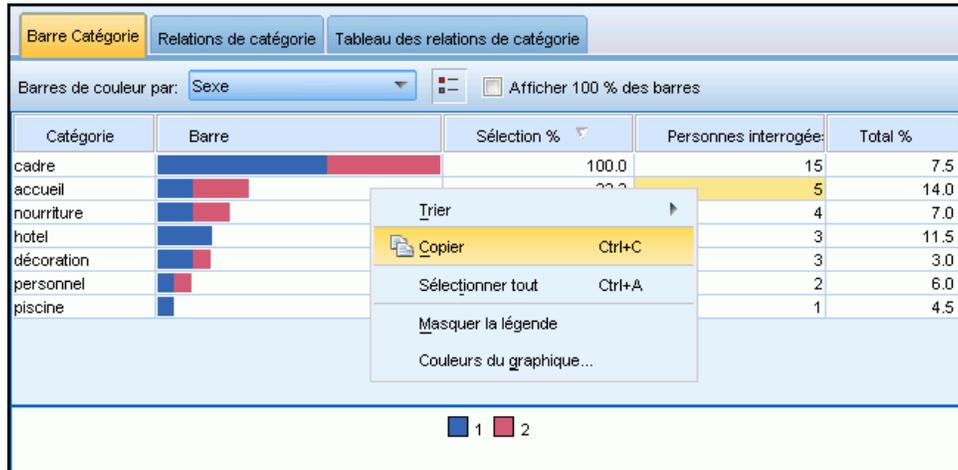
- **Catégorie.** Cette colonne présente le nom des catégories de votre sélection. Par défaut, la catégorie la plus courante de votre sélection est répertoriée en premier.
- **Barre.** Cette colonne présente, de manière visuelle, le rapport entre les enregistrements d'une catégorie donnée et le nombre total d'enregistrements.
- **Sélection %.** Cette colonne présente un pourcentage basé sur le rapport entre le nombre total d'enregistrements pour une catégorie et le nombre total d'enregistrements représentés dans la sélection.
- **Réponses** Cette colonne présente le nombre de réponses d'une sélection pour la catégorie donnée.
- **Total %.** Cette colonne présente un pourcentage basé sur le rapport entre le nombre total d'enregistrements pour une catégorie et le nombre total d'enregistrements pour cette question (pas la sélection).

Vous pouvez aussi sélectionner une variable de référence disponible dans la liste déroulante pour comparer leurs valeurs. Lorsque vous sélectionnez une variable de référence, les barres du tableau sont divisées selon un code de couleur en fonction de la valeur des variables de référence. Lorsque vous cliquez sur chaque valeur de référence colorée d'une barre, le panneau Données est mis à jour et affiche la sous-sélection des réponses, selon la valeur de la variable de référence. Pour visualiser la légende des valeurs de variable de référence, cliquez sur le bouton Légende de la barre d'outils.

Figure 7-2
Bouton Légende de la barre d'outils



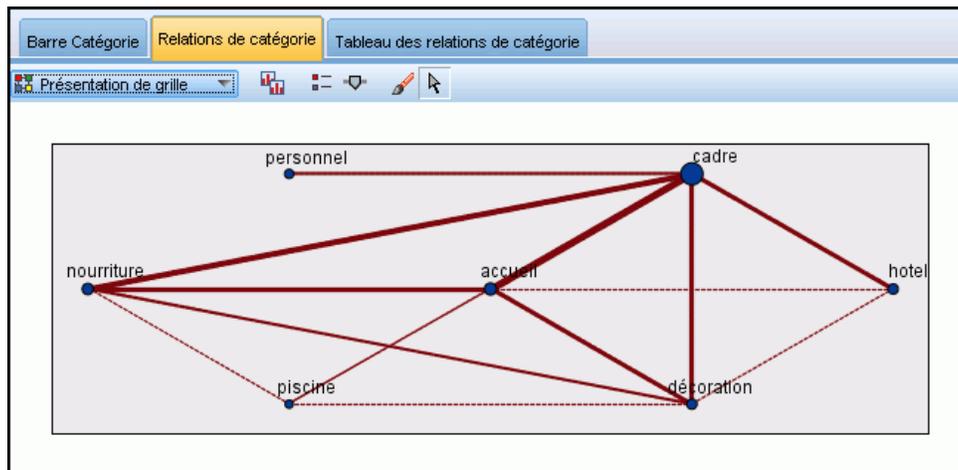
Figure 7-3
panneau Visualisation : Barre Catégorie



Graphique Relations de catégorie

Cet onglet affiche un graphique Relations de catégorie. Les relations présentent le chevauchement des réponses pour les catégories auxquelles les réponses appartiennent, en fonction de la sélection dans les autres panneaux. S'il existe des étiquettes de catégorie, celles-ci apparaissent dans le graphique. Vous pouvez sélectionner la présentation du graphique (réseau, cercle, orientée ou grille) à l'aide des boutons de la barre d'outils de ce panneau.

Figure 7-4
Graphique Relations de catégorie, présentation de grille



Dans les relations, chaque noeud représente une catégorie. A l'aide de la souris, vous pouvez sélectionner et déplacer les noeuds au sein du panneau. La taille du noeud représente la taille relative basée sur le nombre d'enregistrements pour cette catégorie dans votre sélection. L'épaisseur et la couleur de la ligne entre deux catégories représentent le nombre d'enregistrements communs qu'elles contiennent. Si vous pointez sur un nœud en mode d'interaction, une info-bulle affiche le nom (ou l'étiquette) de la catégorie et le nombre global de enregistrements dans la catégorie.

Remarque : Par défaut, le mode d'interaction est activé pour les graphiques dans lesquels vous pouvez déplacer des noeuds. Toutefois, vous pouvez passer en mode d'édition pour modifier la présentation de vos graphiques, notamment les couleurs, polices et légendes. [Pour plus d'informations, reportez-vous à la section Utilisation des palettes et des barres d'outils de graphiques sur p. 173.](#)

Tableau des relations de catégorie

Cet onglet présente les mêmes informations que l'onglet Relations de catégorie, mais sous forme de tableau. Le tableau contient trois colonnes qu'il est possible de trier en cliquant sur leur en-tête :

- **Effectif.** Cette colonne présente le nombre d'enregistrements que les deux catégories partagent ou ont en commun.
- **Catégorie 1.** Cette colonne présente le nom de la première catégorie, suivi entre parenthèses du nombre total d'enregistrements qu'elle contient.
- **Catégorie 2.** Cette colonne présente le nom de la deuxième catégorie, suivi entre parenthèses du nombre total d'enregistrements qu'elle contient.

Figure 7-5
panneau Visualisation : Tableau des relations de catégorie

Barre Catégorie	Relations de catégorie	Tableau des relations de catégorie
Nombre	Catégorie 1	Catégorie 2
4	chambre(30)	restaurant(6)
2	restaurant(6)	service(13)
1	chambre(30)	service(13)
1	hotel(23)	restaurant(6)
1	nourriture(14)	service(13)
1	restaurant(6)	nourriture(14)

Utilisation des palettes et des barres d'outils de graphiques

Le graphique Relations de catégorie a une barre d'outils qui fournit un accès rapide à certaines palettes courantes avec lesquelles vous pouvez effectuer un certain nombre d'actions sur vos graphiques. Vous pouvez choisir entre le mode d'affichage *Sélection/Interaction* ou le mode d'affichage *Edition*.

Tandis que le mode d'interaction vous permet d'explorer de manière analytique les données et les valeurs représentées par la visualisation, le mode d'édition vous permet de modifier la mise en forme et l'aspect de la visualisation. Vous pouvez par exemple modifier les polices et les couleurs pour respecter le guide de style de votre organisation. Pour sélectionner ce mode, sélectionnez Affichage > Panneau Visualisation > Mode d'édition dans les menus (ou cliquez sur l'icône de la barre d'outils).

En mode d'édition, plusieurs barres d'outils ont un impact sur différents aspects de la présentation de la visualisation. Si vous n'utilisez pas certaines barres d'outils, vous pouvez les masquer afin d'augmenter l'espace disponible dans la boîte de dialogue dans laquelle le graphique est affiché. Pour sélectionner ou désélectionner des barres d'outils, cliquez sur le nom de la barre d'outils ou de la palette souhaitée dans le menu Affichage.

Table 7-1
Boutons de la barre d'outil de Text Analytics

Bouton/liste	Description
	Active le mode d'édition. Basculez vers le mode d'édition pour modifier l'apparence du graphique, par exemple pour agrandir la police, modifier les couleurs pour respecter le guide de style de l'entreprise, ou supprimer des étiquettes et des légendes. Pour plus d'informations, reportez-vous à la section Modification des visualisations sur p. 174.
	Active le mode d'interaction. Par défaut, le mode d'interaction est activé, ce qui signifie que vous pouvez déplacer et faire glisser des noeuds dans le graphique et pointer sur des objets du graphique pour obtenir des informations supplémentaires via une info-bulle.
	Sélectionnez un type d'affichage des relations pour les graphiques. <ul style="list-style-type: none"> ■ Présentation en cercle. Présentation générale qu'il est possible d'appliquer à n'importe quel graphique. Pour créer un graphique, elle suppose que les liens ne sont pas orientés et traite tous les noeuds de la même façon. Les noeuds sont placés uniquement autour du périmètre d'un cercle. ■ Présentation en réseau. Présentation générale qu'il est possible d'appliquer à n'importe quel graphique. Pour créer un graphique, elle suppose que les liens ne sont pas orientés et traite tous les noeuds de la même façon. Les noeuds sont placés librement au sein de la présentation. ■ Présentation orientée. Il convient d'utiliser cette présentation uniquement pour les graphiques orientés. Cette présentation produit des structures ressemblant à un arbre, des noeuds racine aux noeuds feuille, et fournit une organisation par couleurs. ■ Présentation de grille. Présentation générale qu'il est possible d'appliquer à n'importe quel graphique. Pour créer un graphique, elle suppose que les liens ne sont pas orientés et traite tous les noeuds de la même façon. Les noeuds sont placés uniquement aux points de grille au sein de l'espace.
	Bouton bascule qui, lorsqu'il est pressé, affiche la légende. Lorsque le bouton est désactivé, la légende n'apparaît pas.
	Bouton bascule qui, lorsqu'il est pressé, affiche le curseur des liens sous le graphique. Vous pouvez filtrer les résultats en faisant glisser la flèche.

Modification des visualisations

Vous pouvez éditer une visualisation de plusieurs manières dans le **mode Edition** Vous pouvez :

- Editer le texte et le formater
- Changer la couleur de remplissage, la transparence, le motif des cadres et les éléments graphiques.

- Changer la couleur et les tirets des bordures et des lignes
- Changer la forme et le rapport d'aspect des points, et faire pivoter les points
- Changer la taille des éléments graphiques (par exemple, les barres et les points)
- Ajuster l'espace entourant les éléments à l'aide des marges et via l'extension
- Spécifier le format des numéros.
- Changer les paramètres d'axe et d'échelle
- Trier, exclure et réduire des catégories sur un axe catégoriel.
- Définir l'orientation des panneaux.
- Appliquer des transformations à un système de coordonnées.
- Modifier les statistiques, les types d'élément graphique et les modificateurs de collision.
- Changer la position de la légende
- Appliquer des feuilles de style de visualisation.

Les rubriques suivantes décrivent ces diverses tâches. Il est également recommandé de prendre connaissance des règles générales liées à l'édition de graphiques.

Comment basculer vers le mode d'édition

- ▶ Dans les menus, sélectionnez :
Afficher > Mode Edition

Règles générales d'édition de visualisations

Mode Edition

Toutes les modifications s'effectuent dans le mode d'édition. Pour activer le mode d'édition, sélectionnez dans les menus :

Afficher > Mode Edition

Sélection

Les options d'édition disponibles dépendent de la sélection effectuée. Différentes options de barre d'outils et de palette de propriétés sont activées selon les éléments sélectionnés. Seules les options activées s'appliquent à la sélection courante. Par exemple, si un axe est sélectionné, les onglets Echelle, Graduations principales et Graduations secondaires sont disponibles dans la palette des propriétés.

Voici quelques conseils pour sélectionner des éléments dans une visualisation :

- Cliquez sur un élément pour le sélectionner.
- Sélectionnez un élément graphique (par exemple, des points dans un diagramme de dispersion ou des barres dans un diagramme à barres), en cliquant une seule fois. Après une première sélection, cliquez de nouveau pour réduire la sélection à des groupes d'éléments graphiques ou à un seul élément graphique.
- Appuyez sur la touche Echap pour désélectionner tous les éléments.

Palettes

Lorsqu'un élément est sélectionné dans la visualisation, les différentes palettes sont mises à jour pour refléter la sélection. Les palettes contiennent des commandes pour apporter des modifications à la sélection. Les palettes peuvent être des barres d'outils ou un panneau avec plusieurs commandes et onglets. Les palettes peuvent être masquées, pour garantir que la palette nécessaire est affichée pour effectuer des modifications. Consultez le menu Affichage pour savoir quelles palettes sont actuellement affichées.

Vous pouvez repositionner les palettes en cliquant et déplaçant l'espace vide dans une palette barre d'outils ou sur le côté gauche d'autres palettes. Le contrôle visuel vous indique où vous pouvez ancrer la palette. Pour les palettes qui ne sont pas des barres d'outils, vous pouvez également cliquer sur le bouton fermer pour masquer la palette et sur le bouton détacher pour afficher la palette dans une fenêtre séparée. Cliquez sur le bouton aide pour afficher de l'aide concernant une palette en particulier.

Paramètres automatiques

Certains paramètres fournissent une option -Automatique-. Cette option indique que des valeurs automatiques sont appliquées. Les paramètres automatiques utilisés dépendent de la visualisation et des valeurs de données spécifiques. Vous pouvez entrer une valeur pour remplacer le paramètre automatique. Pour restaurer le paramètre automatique, supprimez la valeur actuelle et appuyez sur Entrée. Le paramètre affiche de nouveau -Automatique-.

Suppression/Masquage d'éléments

Vous pouvez supprimer/masquer divers éléments dans une visualisation. Par exemple, vous pouvez masquer la légende ou l'étiquette d'axe. Pour supprimer un élément, sélectionnez-le et appuyez sur Supprimer. Si la suppression de l'élément n'est pas autorisée, rien ne se produit. Si vous supprimez un élément par erreur, appuyez sur Ctrl+Z pour annuler la suppression.

Etat

Certaines barres d'outils reflètent l'état de la sélection actuelle, contrairement à d'autres. La palette des propriétés reflète toujours cet état. Si une barre d'outils ne reflète *pas* l'état, ceci est mentionné dans la rubrique décrivant cette barre d'outils.

Edition et formatage de texte

Vous pouvez éditer le texte en place et changer le formatage d'un bloc de texte entier. Vous ne pouvez pas éditer un texte directement lié à des valeurs de données. Par exemple, vous ne pouvez pas éditer une étiquette de graduation car le contenu de l'étiquette provient des données sous-jacentes. Cependant, vous pouvez formater n'importe quel texte dans la visualisation.

Comment éditer le texte en place

- ▶ Double-cliquez sur le bloc de texte. Cette action sélectionne tout le texte. A ce stade, toutes les barres d'outils sont désactivées car vous ne pouvez changer aucun autre élément de la visualisation pendant que vous éditez du texte.
- ▶ Tapez le nouveau texte pour remplacer le texte existant. Vous pouvez également cliquer de nouveau sur le texte pour afficher un curseur. Placez le curseur à la position souhaitée et entrez le texte supplémentaire.

Comment formater du texte

- ▶ Sélectionnez le cadre qui contient le texte. Ne double-cliquez pas sur le texte.
- ▶ Formatez le texte à l'aide de la barre d'outils des polices. Si cette barre d'outils n'est pas activée, vérifiez que seul le *cadre* contenant le texte est sélectionné. Si le texte lui-même est sélectionné, la barre d'outils est désactivée.

Figure 7-6
Barre d'outils des polices



Vous pouvez changer la police :

- Couleur
- Famille (par exemple, Arial ou Verdana)
- Taille (l'unité utilisée est le point, sauf si vous indiquez une unité différente telle que le pica, pc)
- Poids
- Aligement par rapport au cadre du texte

Le formatage s'applique à tout le texte figurant dans le cadre. Vous ne pouvez pas changer le formatage de certaines lettres ou de certains mots dans un bloc de texte spécifique.

Changement des couleurs, des motifs, des tirets et de la transparence

Dans une visualisation, de nombreux éléments différents comportent un remplissage et une bordure. L'exemple le plus évident est celui d'une barre dans un diagramme à barres. La couleur des barres est la couleur de remplissage. Les barres peuvent également être entourées d'une bordure unie noire.

D'autres éléments moins évidents d'une visualisation comportent également des couleurs de remplissage. Si la couleur de remplissage est transparente, le remplissage n'est pas nécessairement visible. Par exemple, considérons le texte d'une étiquette d'axe. Ce texte semble "flotter" mais figure en fait dans un cadre comportant une couleur de remplissage transparente. Le cadre est visible lorsque vous sélectionnez l'étiquette d'axe.

N'importe quel cadre de la visualisation peut comporter un style de remplissage et de bordure, y compris le cadre entourant la visualisation dans son intégralité. De plus, n'importe quel remplissage possède un niveau d'opacité/transparence associé qui peut être ajusté.

Comment changer les couleurs, les motifs, les tirets et la transparence

- ▶ Sélectionnez l'élément à formater. Par exemple, sélectionnez les barres d'un diagramme à barres ou un cadre contenant du texte. Si la visualisation est fractionnée par une variable ou un champ catégoriel, vous pouvez également sélectionner le groupe correspondant à une catégorie individuelle. Vous pouvez ainsi changer l'apparence par défaut attribuée à ce groupe. Par exemple, vous pouvez changer la couleur de l'un des groupes de superposition d'un diagramme à barres superposées.
- ▶ Pour changer la couleur de remplissage, la couleur de bordure ou le motif de remplissage, utilisez la barre d'outils des couleurs.

Figure 7-7
Barre d'outils des couleurs

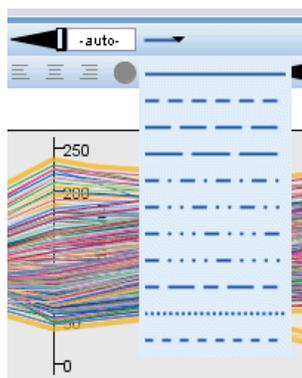


Remarque : Cette barre d'outils ne reflète pas l'état de la sélection courante.

Pour changer une couleur ou un remplissage, vous pouvez cliquer sur le bouton pour sélectionner l'option affichée ou sur la flèche de la liste déroulante pour sélectionner une autre option. Pour les couleurs, il existe une couleur paraissant blanche et traversée d'une ligne diagonale rouge. Il s'agit de la couleur transparente. Cette couleur peut être utilisée par exemple pour masquer les bordures des barres dans un histogramme.

- Le premier bouton contrôle la couleur de remplissage.
 - Le deuxième bouton contrôle la couleur de bordure.
 - Le troisième bouton contrôle le motif de remplissage. Le motif de remplissage utilise la couleur de bordure. Par conséquent, le motif de remplissage n'est visible que s'il existe une couleur de bordure visible.
 - La quatrième commande est composée d'un curseur de défilement et d'une zone de texte qui contrôlent l'opacité de la couleur et du motif de remplissage. Un pourcentage inférieur signifie moins d'opacité et plus de transparence. 100 % indique une couleur complètement opaque (aucune transparence).
- ▶ Pour changer les tirets d'une bordure ou d'une ligne, utilisez la barre d'outils de ligne.

Figure 7-8
Barre d'outils de ligne



Remarque : Cette barre d'outils ne reflète pas l'état de la sélection courante.

Comme dans le cas de l'autre barre d'outils, vous pouvez cliquer sur le bouton pour sélectionner l'option affichée ou sur la flèche de la liste déroulante pour sélectionner une autre option.

Changement de la forme et du rapport d'aspect des points et rotation des points

Vous pouvez faire pivoter des points, attribuer une forme prédéfinie différente ou changer le rapport d'aspect (le rapport entre la largeur et la hauteur).

Comment modifier des points

- ▶ Sélectionnez les points. Vous ne pouvez pas changer la forme et le rapport d'aspect de points individuels ni faire pivoter ces points.
- ▶ Utilisez la barre d'outils des symboles pour modifier les points.

Figure 7-9
Barre d'outils des symboles



- Le premier bouton permet de changer la forme des points. Cliquez sur la flèche de la liste déroulante et sélectionnez une forme prédéfinie.
- Le deuxième bouton permet de faire pivoter les points sur une position de compas spécifique. Cliquez sur la flèche de la liste déroulante, puis faites glisser l'aiguille vers la position désirée.
- Le troisième bouton permet de changer le rapport d'aspect. Cliquez sur la flèche de la liste déroulante, puis cliquez sur le rectangle qui apparaît et faites-le glisser. La forme du rectangle représente le rapport d'aspect.

Changement de la taille des éléments graphiques

Vous pouvez changer la taille des éléments graphiques figurant dans la visualisation. Ces éléments sont notamment les barres, les lignes et les points. Si la taille de l'élément graphique est déterminée par une variable ou un champ, la taille spécifiée est la taille *minimale*.

Comment changer la taille des éléments graphiques

- ▶ Sélectionnez les éléments graphiques à redimensionner.
- ▶ Utilisez le curseur ou entrez une taille spécifique pour l'option disponible dans la barre d'outils des symboles. L'unité utilisée est le pixel, sauf si vous indiquez une unité différente (vous trouverez ci-dessous une liste complète des abréviations d'unité). Vous pouvez également spécifier un pourcentage (par exemple, 30 %), ce qui signifie qu'un élément graphique utilise le pourcentage d'espace disponible spécifié. L'espace disponible dépend du type d'élément graphique et de la visualisation.

Table 7-2
Abréviations d'unités valides

Abréviation	Unité
cm	centimètre
in	pouce
mm	millimètre
pc	pica
pt	point
px	pixel

Figure 7-10
Contrôle de la taille dans la barre d'outils des symboles



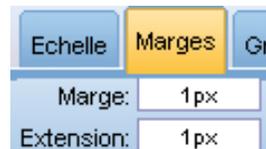
Spécification des marges et de l'extension

Si l'espace autour ou à l'intérieur d'un cadre dans une visualisation est trop important ou insuffisant, vous pouvez changer les paramètres de marge et d'extension du graphique. La **marge** est la quantité d'espace séparant le cadre des autres éléments situés autour de ce cadre. L'**extension** est la quantité d'espace situé entre la bordure et le *contenu* du cadre.

Comment spécifier les marges et l'extension

- Sélectionnez le cadre pour lequel vous souhaitez spécifier des marges et une extension. Il peut s'agir d'un cadre de texte, d'un cadre entourant une légende ou même d'un cadre de données affichant des éléments graphiques (par exemple, des barres et des points).
- Utilisez l'onglet Marges de la palette des propriétés pour spécifier les paramètres. Toutes les tailles sont exprimées en pixels, sauf si vous indiquez une unité différente (par exemple, le centimètre cm ou le pouce in).

Figure 7-11
Onglet Marges



Formatage des nombres

Vous pouvez spécifier le format des nombres figurant dans les étiquettes de graduation sur un axe continu ou dans les étiquettes de valeurs de données affichant un nombre. Vous pouvez par exemple spécifier que les nombres affichés dans les étiquettes de graduation sont affichés en milliers.

Pour spécifier les formats de nombre

- Sélectionnez les étiquettes de graduation de l'axe continu ou les étiquettes de valeur de données si elles comportent des nombres.
- Cliquez sur l'onglet Format dans la palette des propriétés.

Figure 7-12
Onglet Format



- Sélectionnez les options de formatage des nombres souhaitées :

Préfixe. Caractère à afficher devant le nombre. Par exemple, saisissez le symbole (\$) si les nombres correspondent à des salaires en dollars U.S.

Suffixe. Caractère à afficher derrière le nombre. Par exemple, saisissez le symbole du pourcentage (%) si les nombres sont des pourcentages.

Chiffres entiers min.. Nombre minimum de chiffres à afficher dans la partie entière d'une représentation décimale. Si la valeur réelle ne contient pas le nombre minimum de chiffres, la partie entière de la valeur sera remplie de zéros.

Chiffres entiers max.. Nombre maximum de chiffres à afficher dans la partie entière d'une représentation décimale. Si la valeur réelle dépasse le nombre minimum de chiffres, la partie entière de la valeur sera remplacée par des astérisques.

Chiffres décimaux min.. Nombre minimum de chiffres à afficher dans la partie décimale d'une représentation décimale ou scientifique. Si la valeur réelle ne contient pas le nombre minimum de chiffres, la partie décimale de la valeur sera remplie de zéros.

Chiffres décimaux max.. Nombre maximum de chiffres à afficher dans la partie décimale d'une représentation décimale ou scientifique. Si la valeur réelle dépasse le nombre minimum de chiffres, la décimale est arrondie au nombre de chiffres approprié.

Scientifique. Afficher les chiffres sous forme de notation scientifique. Cette notation est utile pour des nombres très grands ou très petits. -auto- permet à l'application de déterminer quand la notation scientifique est appropriée.

Mise à l'échelle. Facteur d'échelle, représentant un numéro par lequel la valeur originale est divisée. Utilisez un facteur d'échelle si vous souhaitez que l'étiquette ne s'étende pas trop pour s'ajuster aux nombres élevés. Si vous modifiez le format des nombres des étiquettes de graduation, veillez à modifier le titre de l'axe pour indiquer comment les nombres doivent être interprétés. Par exemple, si votre axe d'échelle affiche des salaires et que les étiquettes sont 30 000, 50 000 et 70 000, vous pouvez saisir un facteur d'échelle de 1 000 pour afficher 30, 50 et 70. Vous devez ensuite modifier l'axe d'échelle pour inclure le texte en milliers.

Parenthèses pour -ve. Afficher des parenthèses autour des valeurs négatives.

Regroupement. Afficher un caractère entre les groupes de chiffres. Les paramètres régionaux actuels de votre ordinateur déterminent le caractère utilisé pour le regroupement des chiffres.

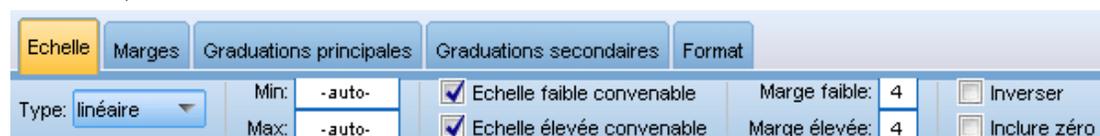
Changement des paramètres d'axe et d'échelle

Plusieurs options permettent de changer les axes et les échelles.

Comment changer les paramètres d'axe et d'échelle

- ▶ Sélectionnez une partie de l'axe (par exemple, l'étiquette d'axe ou les étiquettes de graduation).
- ▶ Utilisez les onglets Echelle, Graduations principales et Graduations secondaires de la palette des propriétés pour changer les paramètres d'axe et d'échelle.

Figure 7-13
Palette Propriétés



Onglet Echelle

Type. Indique si l'échelle est linéaire ou transformée. Les transformations d'échelle aident à comprendre les données ou à émettre les hypothèses nécessaires à la déduction statistique. Dans les diagrammes de dispersion, vous pouvez utiliser une échelle transformée si la relation entre les variables indépendantes et dépendantes ou les champs est non linéaire. Les transformations d'échelle permettent également de rendre un histogramme asymétrique plus symétrique de sorte qu'il ressemble à une distribution normale. Vous ne transformez que l'échelle à laquelle les données sont affichées et non pas les données elles-mêmes.

- **linéaire.** Indique une échelle non transformée linéaire.
- **log.** Indique une échelle transformée logarithmique décimale. Pour prendre en compte les valeurs nulles (zéro) et négatives, cette transformation utilise une version modifiée de la fonction log. Cette fonction de « log valide » est définie comme $\text{sign}(x) * \log(1 + \text{abs}(x))$. Ainsi $\text{safeLog}(-99)$ est égal à :

$$\text{sign}(-99) * \log(1 + \text{abs}(-99)) = -1 * \log(1 + 99) = -1 * 2 = -2$$
- **puissance.** Indique une échelle transformée de puissance à l'aide de l'exposant 0,5. Pour prendre en compte les valeurs négatives, cette transformation utilise une version modifiée de la fonction de puissance. Cette fonction de « log valide » est définie comme $\text{sign}(x) * \text{pow}(\text{abs}(x), 0.5)$. Ainsi $\text{safePower}(-100)$ est égal à :

$$\text{sign}(-100) * \text{pow}(\text{abs}(-100), 0.5) = -1 * \text{pow}(100, 0.5) = -1 * 10 = -10$$

Minimum/Maximum/Echelle faible convenable/Echelle élevée convenable. Indique l'intervalle de l'échelle. Le fait de sélectionner Echelle faible convenable et Echelle élevée convenable permet à l'application de sélectionner une échelle appropriée en fonction des données. Le minimum et le maximum sont convenables car ils constituent généralement des valeurs entières supérieures ou inférieures aux valeurs de données maximum et minimum. Par exemple, si l'intervalle de données va de 4 à 92, la valeur Echelle faible convenable ou Echelle élevée convenable pour l'échelle peut être 0 ou 100 respectivement plutôt que les valeurs réelles de données maximum et minimum. Veuillez à ne pas définir un intervalle trop restreint et qui risquerait de masquer des éléments

importants. En outre, vous ne pouvez pas définir de valeurs minimum et maximum explicites si l'option Inclure zéro est sélectionnée.

Marge basse/Marge haute. Crée des marges à l'extrémité inférieure et/ou supérieure de l'axe. La marge apparaît, perpendiculaire à l'axe sélectionné. L'unité utilisée est le pixel, sauf si vous indiquez une unité différente (par exemple, le centimètre, cm ou le pouce, in). Par exemple, si vous définissez la marge Marge élevée sur 5 pour l'axe vertical, une marge horizontale de 5 pixels s'étend le long de la partie supérieure du cadre de données.

Inverser. Indique si l'échelle est inversée.

Inclure le zéro. Indique que l'échelle doit inclure 0. Cette option est généralement utilisée pour les diagrammes en barres afin de s'assurer que les barres commencent à 0 plutôt qu'à une valeur proche de la hauteur de la barre la plus petite. Si cette option est sélectionnée, les options Minimum et Maximum sont désactivées car vous ne pouvez pas définir de valeur minimum et maximum personnalisée pour la plage de l'échelle.

Onglets Graduations principales/Graduations secondaires

Les **graduations** ou **marques de graduation** sont les lignes qui apparaissent sur un axe. Celles-ci indiquent des valeurs à des intervalles ou catégories spécifiques. Les **graduations principales** sont les marques de graduation avec étiquettes. Celles-ci sont également plus longues que les autres marques de graduation. Les **graduations secondaires** sont les marques de graduation qui apparaissent entre les marques de graduation principales. Certaines options sont spécifiques au type de graduation mais la plupart des options sont disponibles pour les graduations principales et secondaires.

Afficher les graduations. Indique si les graduations principales ou secondaires apparaissent sur un graphique.

Afficher les quadrillages. Indique si les quadrillages apparaissent au niveau des graduations principales ou secondaires. Les **quadrillages** sont les lignes qui traversent l'intégralité d'un graphique d'un axe à l'autre.

Position. Indique la position des marques de graduation par rapport à l'axe.

Longueur. Spécifie la longueur des marques de graduation. L'unité utilisée est le pixel, sauf si vous indiquez une unité différente (par exemple, le centimètre, cm ou le pouce, in).

Base. *S'applique uniquement aux graduations principales.* Indique la valeur à laquelle la première graduation principale apparaît.

Delta. *S'applique uniquement aux graduations principales.* Indique la différence entre les graduations principales. En d'autres termes, les graduations principales apparaîtront à chaque ne valeur, n étant la valeur delta.

Divisions. *S'applique uniquement aux graduations secondaires.* Indique le nombre de divisions de graduation secondaire entre les graduations principales. Le nombre de graduations secondaires est inférieur de une unité au nombre de divisions. Par exemple, supposons qu'il existe des graduations principales à 0 et 100. Si vous entrez 2 comme nombre de divisions de graduation secondaire, il y aura *une* graduation secondaire à 50, divisant l'intervalle 0–100 et créant *deux* divisions.

Modification de catégories

Vous pouvez modifier les catégories sur un axe catégoriel de différentes manières :

- Changer l'ordre de tri de l'affichage des catégories.
- Exclure des catégories spécifiques.
- Ajouter une catégorie qui n'apparaît pas dans l'ensemble de données.
- Fusionner/associer de petites catégories en une catégorie.

Comment changer l'ordre de tri des catégories

- ▶ Sélectionnez un axe catégoriel. La palette Catégories affiche les catégories sur l'axe.

Remarque : Si la palette n'est pas visible, vérifiez que vous l'avez bien activée.

- ▶ Dans la palette Catégories, sélectionnez une option de tri dans la liste déroulante :

Personnalisé. Trier les catégories en fonction de l'ordre dans lequel elles apparaissent dans la palette. Utilisez les flèches pour placer les catégories en haut ou en bas de la liste, ou les déplacer vers le haut ou vers le bas.

Données. Trier les catégories en fonction de l'ordre dans lequel elles apparaissent dans l'ensemble de données.

Nom. Trier les catégories dans l'ordre alphabétique, en utilisant les noms affichés dans la palette.

Valeur. Trier les catégories en fonction de la valeur de données sous-jacente, en utilisant les valeurs affichées entre parenthèses dans la palette. Seules les sources de données avec des métadonnées (telles que les fichiers de données IBM® SPSS® Statistics) prennent en charge cette option.

Statistique. Trier les catégories en fonction de la statistique calculée pour chaque catégorie. Il peut s'agir par exemple d'effectifs, de pourcentages et de moyennes. Cette option est uniquement disponible si une statistique est utilisée dans le graphique.

Ajout d'une catégorie

Par défaut, seules les catégories qui apparaissent dans l'ensemble de données sont disponibles. Le cas échéant, vous pouvez ajouter une catégorie à la visualisation.

- ▶ Sélectionnez un axe catégoriel. La palette Catégories affiche les catégories sur l'axe.

Remarque : Si la palette n'est pas visible, vérifiez que vous l'avez bien activée.

- ▶ Dans la palette Catégories, cliquez sur le bouton Ajouter une catégorie :

Figure 7-14

Bouton Ajouter une catégorie



- ▶ Dans la boîte de dialogue Ajouter une nouvelle catégorie, entrez le nom de la catégorie.
- ▶ Cliquez sur OK.

Comment exclure des catégories spécifiques

- ▶ Sélectionnez un axe catégoriel. La palette Catégories affiche les catégories sur l'axe.
Remarque : Si la palette n'est pas visible, vérifiez que vous l'avez bien activée.
- ▶ Dans la palette Catégories, sélectionnez un nom de catégorie dans la liste Inclure, puis cliquez sur le bouton X. Pour inclure de nouveau la catégorie, sélectionnez son nom dans les Exclusions, puis cliquez sur la flèche à droite de la liste.

Comment fusionner/associer des petites catégories

Vous pouvez associer des catégories que vous n'avez pas besoin d'afficher séparément du fait de leur petite taille. Vous pouvez ainsi, dans un graphique sectoriel contenant plusieurs catégories, envisager de fusionner les catégories avec un pourcentage inférieur à 10. La fusion est disponible uniquement pour les statistiques additives. Par exemple, vous ne pouvez pas ajouter des moyennes car les moyennes ne sont pas additives. Par conséquent, la fusion/association de catégories en utilisant une moyenne n'est pas disponible.

- ▶ Sélectionnez un axe catégoriel. La palette Catégories affiche les catégories sur l'axe.
Remarque : Si la palette n'est pas visible, vérifiez que vous l'avez bien activée.
- ▶ Dans la palette Catégories, sélectionnez Réduire et indiquez un pourcentage. Toutes les catégories dont le pourcentage du total est inférieur au nombre indiqué sont associées en une seule catégorie. Le pourcentage est basé sur la statistique présentée dans le tableau. La réduction est disponible uniquement pour les statistiques basées sur les effectifs ou une addition (somme).

Modification de l'orientation des panneaux

Si vous utilisez des panneaux dans votre visualisation, vous pouvez modifier leur orientation.

Comment changer l'orientation des panneaux

- ▶ Sélectionnez une partie de la visualisation.
- ▶ Cliquez sur l'onglet Panneaux dans la palette des propriétés.

Figure 7-15
Onglet Panneaux



- ▶ Sélectionnez une option dans Présentation :
Table. Dispose les panneaux comme une table : une ligne ou une colonne est attribuée à chaque valeur individuelle.

Transposé. Dispose les panneaux comme une table, mais permute également les lignes et les colonnes d'origine. Cette option est différente de l'option de transposition du graphique lui-même. Les axes x et y restent inchangés lorsque vous sélectionnez cette option.

Liste. Dispose les panneaux comme une liste : chaque cellule représente une combinaison de valeurs. Les colonnes et les lignes ne sont plus affectées à des valeurs individuelles. Cette option permet de réorganiser les panneaux si nécessaire.

Transformation du système de coordonnées

De nombreuses visualisations sont affichées dans un système de coordonnées plat et rectangulaire. Vous pouvez transformer le système de coordonnées si nécessaire. Vous pouvez par exemple appliquer une transformation polaire au système de coordonnées, ajouter des effets d'ombrage oblique, et transposer les axes. Vous pouvez également annuler ces transformations si elles sont déjà appliquées à la visualisation en cours. Par exemple, un diagramme en secteurs est représenté dans un système de coordonnées polaires. Si vous le souhaitez, vous pouvez annuler la transformation polaire et afficher le diagramme sous forme de barre empilée unique dans un système de coordonnées rectangulaires.

Comment transformer le système de coordonnées

- ▶ Sélectionnez le système de coordonnées à transformer. Vous pouvez sélectionner le système de coordonnées en sélectionnant le cadre autour du graphique individuel.
- ▶ Cliquez sur l'onglet Coordonnées dans la palette des propriétés.

Figure 7-16
Onglet Coordonnées



- ▶ Sélectionnez les transformations à appliquer au système de coordonnées. Vous pouvez également désélectionner une transformation pour l'annuler.

Transposé. L'opération consistant à changer l'orientation des axes est nommée **transposition**. Cette opération est comparable à la permutation des axes vertical et horizontal dans une visualisation en 2D.

Polaire. Une transformation polaire dessine les éléments graphiques à un angle et une distance spécifiques du centre du graphique. Un diagramme en secteurs est une visualisation en 1D avec une transformation polaire qui dessine les barres individuelles à des angles spécifiques. Un graphique radar est une visualisation en 2D avec une transformation polaire qui dessine les éléments graphiques à des angles et distances spécifiques du centre du graphique. Une visualisation en 3D inclurait également une dimension de profondeur supplémentaire.

Oblique. Une transformation oblique ajoute un effet 3D aux éléments graphiques. Cette transformation ajoute la profondeur aux éléments graphiques, mais la profondeur est purement décorative. Elle n'est pas influencée par des valeurs de données particulières.

Même rapport. L'application du même rapport indique que la même distance sur chaque échelle représente la même différence dans les valeurs de données. Par exemple, 2 cm sur les deux échelles représente une différence de 1000.

% d'insertion de pré-transformation. Si les axes sont coupés après la transformation, vous pouvez ajouter des insertions au graphique avant d'appliquer la transformation. Les insertions réduisent les dimensions d'un certain pourcentage avant que les transformations ne soient appliquées au système de coordonnées. Vous disposez d'un contrôle sur les dimensions x inférieure, x supérieure, y inférieure, et y supérieure, dans cet ordre.

% d'insertion de post-transformation. Si vous souhaitez modifier les dimensions d'un graphique, vous pouvez ajouter des insertions au graphique après avoir appliqué la transformation. Les insertions réduisent les dimensions d'un certain pourcentage après l'application de transformations au système de coordonnées. Ces insertions peuvent également être appliquées même si aucune transformation n'est appliquée au graphique. Vous disposez d'un contrôle sur les dimensions x inférieure, x supérieure, y inférieure, et y supérieure, dans cet ordre.

Modification des statistiques et des éléments graphiques

Vous pouvez convertir un en autre type, modifier la statistique utilisée pour dessiner l'élément graphique, ou spécifier le modificateur de collision qui détermine ce qui se passe lorsque des éléments graphiques se chevauchent.

Comment convertir un élément graphique

- ▶ Sélectionnez l'élément graphique à convertir.
- ▶ Cliquez sur l'onglet Élément dans la palette des propriétés.

Figure 7-17
Onglet Élément



- ▶ Sélectionnez un nouveau type d'élément graphique dans la liste Type.

Type d'élément graphique	Description
Point	Un marqueur identifiant un point de données spécifique. Un point est utilisé dans des diagrammes de dispersion et d'autres visualisations associées.
Intervalle	Une forme rectangulaire tracée à une valeur de données spécifiques et remplissant l'espace entre une origine et une autre valeur de données. Un élément d'intervalle est utilisé dans les diagrammes en bâton et les histogrammes.
Ligne	Une ligne qui connecte les valeurs de données.
Chemin	Une ligne qui connecte les valeurs de données dans l'ordre dans lequel elles apparaissent dans l'ensemble de données.

Type d'élément graphique	Description
Aires	Une ligne qui connecte les éléments de données avec la zone entre la ligne et une origine remplie.
Polygone	Une forme à plusieurs côtés comprenant une région de données. Un polygone peut être utilisé dans un diagramme de dispersion mis en intervalles ou une carte.
Schéma	Un élément constitué d'une boîte avec des moustaches et des marqueurs indiquant les valeurs éloignées. Un élément de schéma est utilisé pour les boîtes à moustaches.

Modification de la statistique

- ▶ Sélectionnez l'élément graphique dont vous voulez modifier la statistique.
- ▶ Cliquez sur l'onglet Elément dans la palette des propriétés.
- ▶ Sélectionnez une nouvelle statistique dans la liste déroulante Récapitulatif. Notez que la sélection d'une statistique regroupe les données. Si vous voulez plutôt que la visualisation affiche des données non groupées, sélectionnez (aucune statistique) dans la liste Récapitulatif.

Statistiques récapitulatives calculées à partir d'un champ continu

- **Moyenne.** Mesure de la tendance centrale. Moyenne arithmétique ; somme divisée par le nombre d'observations.
- **Médiane.** Valeur au-dessus ou au-dessous de laquelle se trouvent la moitié des observations ; 50e centile. Si le nombre d'observations est pair, la médiane correspond à la moyenne des deux observations du milieu lorsqu'elles sont triées dans l'ordre croissant ou décroissant. La médiane est une mesure de tendance centrale et elle n'est pas, à l'inverse de la moyenne, sensible aux valeurs éloignées.
- **Mode.** Valeur qui revient le plus fréquemment. Si plusieurs valeurs partagent le plus grand effectif d'occurrence, chacune d'elles constitue un mode.
- **Minimum.** Valeur la plus petite d'une variable numérique.
- **Maximum.** Plus grande valeur d'une variable numérique.
- **Intervalle.** Indique la différence entre les valeurs minimale et maximale.
- **Milieu d'intervalle.** Le milieu de l'intervalle, c'est-à-dire la valeur dont la différence par rapport au minimum est égale à sa différence par rapport au maximum.
- **Somme.** Somme ou total des valeurs, pour toutes les observations n'ayant pas de valeur manquante.
- **Somme cumulative.** La somme cumulative des valeurs. Chaque élément graphique montre la somme d'un sous-groupe plus la somme totale de tous les groupes précédents.
- **Somme des pourcentages.** Le pourcentage dans chaque sous-groupe basé sur un champ totalisé par rapport à la somme dans tous les groupes.
- **Somme cumulative des pourcentages.** Le pourcentage cumulatif dans chaque sous-groupe basé sur un champ totalisé par rapport à la somme dans tous les groupes. Chaque élément graphique montre le pourcentage d'un sous-groupe plus le pourcentage total de tous les groupes précédents.

- **Variance.** Mesure de dispersion autour de la moyenne, égale à la somme des carrés des écarts par rapport à la moyenne, divisée par le nombre d'observations moins un. La variance se mesure en unités, qui sont égales au carré des unités de la variable.
- **Ecart-type.** Mesure de dispersion par rapport à la moyenne. Dans le cas d'une distribution normale, 68 % des observations se situent à l'intérieur d'un écart-type de la moyenne et 95 % se situent à l'intérieur de deux écarts-types. Par exemple, si la moyenne d'âge est de 45 avec un écart-type égal à 10, une distribution normale verra 95 % des observations se situer entre 25 et 65.
- **Erreur standard.** Mesure du degré de variation de la valeur d'une statistique test, d'un échantillon à l'autre. Il s'agit de l'écart-type de la distribution de l'échantillon pour une statistique. Par exemple, l'erreur standard de la moyenne est l'écart-type des moyennes d'échantillon.
- **Aplatissement.** Mesure de l'étendue du regroupement des observations autour d'un point central. Dans le cas d'une distribution normale, la valeur de la statistique d'aplatissement est égale à zéro. Un aplatissement positif indique que par rapport à une distribution normale, les observations sont plus regroupées au centre et présentent des extrémités plus fines atteignant les valeurs extrêmes de la distribution. La distribution leptokurtique présente des extrémités plus épaisses que dans le cas d'une distribution normale. Un aplatissement négatif indique que les observations sont moins regroupées au centre et présentent des extrémités plus épaisses atteignant les valeurs extrêmes de la distribution. La distribution platykurtique présente des extrémités plus fines que dans le cas d'une distribution normale.
- **Asymétrie.** Mesure de l'asymétrie d'une distribution. La distribution normale est symétrique et possède une valeur d'asymétrie égale à 0. Une distribution dont la valeur d'asymétrie est positive présente une extrémité droite allongée. Une distribution caractérisée par une importante asymétrie négative présente une extrémité gauche plus allongée. Pour simplifier, une valeur d'asymétrie deux fois supérieure à l'erreur standard correspond à une absence de symétrie.

Les statistiques de région suivantes peuvent produire plus d'un élément graphique par sous-groupe. Lorsque vous utilisez des éléments graphiques d'intervalle, de zone ou de bord, une statistique de région résulte en un élément graphique affichant l'intervalle. Tous les autres éléments graphiques résultent en deux éléments séparés, l'un montrant le début de l'intervalle et l'autre en montrant la fin.

- **Région : intervalle.** Indique l'intervalle de valeurs entre les valeurs minimale et maximale.
- **Région : 95% intervalle de confiance moyen.** Un intervalle de valeur ayant 95% de chances de comprendre la moyenne de la population.
- **Région : 95% intervalle de confiance individuel.** Un intervalle de valeur qui a 95% de chances de comprendre la valeur prévue selon le cas individuel.
- **Région : 1 écart-type au-dessus/en dessous de la moyenne.** Un intervalle de valeurs entre 1 écart-type au-dessus et en dessous de la **moyenne**.
- **Région : 1 erreur standard au-dessus/en dessous de la moyenne.** Un intervalle de valeurs entre 1 **erreur standard** au-dessus et en dessous de la **moyenne**.

Statistiques récapitulatives basées sur les effectifs

- **Effectifs.** Nombre de lignes/observations.

- **Effectifs cumulés.** Nombre cumulé de lignes/observations. Chaque élément graphique montre les effectifs d'un sous-groupe plus les effectifs totaux de tous les groupes précédents.
- **Pourcentage des effectifs.** Le pourcentage de lignes/observations dans chaque sous-groupe par rapport au nombre total de lignes/observations.
- **Pourcentage d'effectifs cumulés.** Le pourcentage cumulatif de lignes/observations dans chaque sous-groupe par rapport au nombre total de lignes/observations. Chaque élément graphique montre le pourcentage d'un sous-groupe plus le pourcentage total de tous les groupes précédents.

Comment spécifier le modificateur de collision

Le modificateur de collision détermine ce qui se passe lorsque des éléments graphiques se chevauchent.

- ▶ Sélectionnez l'élément graphique pour lequel vous souhaitez spécifier le modificateur de collision.
- ▶ Cliquez sur l'onglet Élément dans la palette des propriétés.
- ▶ Sélectionnez un modificateur de collision dans la liste déroulante Modificateur. -auto- permet à l'application de déterminer quel modificateur de collision convient au type et à la statistique de l'élément graphique.

Superposer. Représenter les éléments graphiques les uns sur les autres lorsqu'ils ont la même valeur.

Empiler. Empiler les éléments graphiques qui devraient normalement être superposés lorsqu'ils ont les mêmes valeurs de données.

Dodge. Place les éléments graphiques à côté d'autres éléments graphiques qui apparaissent à la même valeur, au lieu de les superposer. Les éléments graphiques sont disposés symétriquement. Autrement dit, les éléments graphiques sont déplacés aux opposés d'une position centrale. Le dodging est très similaire au clustering (regroupement en clusters).

Tas. Place les éléments graphiques à côté d'autres éléments graphiques qui apparaissent à la même valeur, au lieu de les superposer. Les éléments graphiques sont disposés de manière asymétrique. Autrement dit, les éléments graphiques sont empilés les uns sur les autres, l'élément graphique du bas étant positionné à une valeur spécifique sur l'échelle.

Brouillage (normal). Repositionne de façon aléatoire les éléments graphiques à la même valeur de données au moyen d'une distribution normale.

Brouillage (uniforme). Repositionne de façon aléatoire les éléments graphiques à la même valeur de données au moyen d'une distribution uniforme.

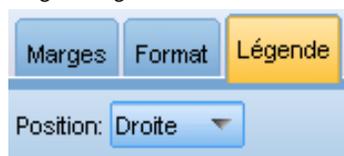
Changement de la position de la légende

Si un graphique comprend une légende, celle-ci apparaît généralement à droite du graphique. Vous pouvez changer cette position si nécessaire.

Comment changer la position de la légende

- ▶ Sélectionnez la légende.
- ▶ Cliquez sur l'onglet Légende dans la palette des propriétés.

Figure 7-18
Onglet Légende



- ▶ Sélectionnez une position.

Copie d'une visualisation et de données de visualisation

La palette générale inclut également des boutons pour la copie de la visualisation et de ses données.

Figure 7-19
Bouton Copier la visualisation



Copie de la visualisation. Cette action copie la visualisation dans le presse-papiers sous forme d'image. Plusieurs formats d'image sont disponibles. Lorsque vous collez l'image dans une autre application, vous pouvez choisir une option "collage spécial" pour sélectionner l'un des formats d'image disponibles pour le collage.

Figure 7-20
Bouton Copier les données de la visualisation



Copie des données de la visualisation. Cette action copie les données de la visualisation qui sont utilisées pour représenter la visualisation. Les données sont copiées dans le presse-papiers sous forme de texte brut ou de texte au format HTML. Lorsque vous collez les données dans une autre application, vous pouvez choisir une option "collage spécial" pour sélectionner l'un de ces formats pour le collage.

Raccourcis clavier

Table 7-3
Raccourcis clavier

Touche de raccourci	Fonction
Ctrl+Espace	Basculer entre le mode interaction et le mode d'édition
Supprimer	Supprimer un élément de visualisation

Touche de raccourci	Fonction
Ctrl+Z	Annuler
Ctrl+Y	Rétablir
F2	Afficher des conseils pour sélectionner des éléments du graphique

Partie III:
Editeur de ressources

Modèles et ressources

IBM® SPSS® Text Analytics for Surveys capture et extrait rapidement et avec précision des concepts-clés de données texte. Ce processus d'extraction repose principalement sur les ressources linguistiques afin de déterminer la façon d'extraire des informations des données textuelles. [Pour plus d'informations, reportez-vous à la section Fonctionnement de l'extraction dans le chapitre 1 sur p. 4.](#) Vous pouvez affiner ces ressources dans la vue de l'Editeur de ressources.

Lorsque vous installez le logiciel, vous obtenez aussi un ensemble de ressources spécialisées. Ces ressources fournies vous permettent de bénéficier d'années de recherche et d'un réglage pour des langues et applications spécifiques. Etant donné que les ressources fournies ne sont pas toujours parfaitement adaptées au contexte de vos données, vous pouvez modifier ces modèles de ressources, ou même créer et utiliser des bibliothèques personnalisées adaptées aux données de votre entreprise. Ces ressources sont disponibles sous diverses formes et chacune peut être utilisée dans votre projet. Les ressources se trouvent aux endroits suivants :

- **Modèles de ressources.** Les modèles sont constitués d'un ensemble de bibliothèques, de types et de ressources avancées qui forment un ensemble spécialisé de ressources adapté à un domaine ou contexte particulier comme les opinions sur des produits.
- **Packages d'analyses de texte (TAP).** En plus des ressources stockées dans un modèle, les TAP regroupent également un ou plusieurs ensembles de catégories générés à l'aide de ces ressources afin que les catégories et les ressources soient stockées ensemble et puissent être réutilisées. [Pour plus d'informations, reportez-vous à la section Utilisation des packages d'analyse de texte dans le chapitre 3 sur p. 43.](#)
- **Bibliothèques.** Les bibliothèques sont utilisées comme blocs de construction pour les TAP et les modèles. Elles peuvent aussi être ajoutées individuellement aux ressources dans votre projet. Chaque bibliothèque est composée de plusieurs déclarations utilisées pour définir et gérer les types, les synonymes et les exclusions. Alors que les bibliothèques sont également fournies individuellement, elles sont prégroupées dans les modèles et les TAP. [Pour plus d'informations, reportez-vous à la section Utilisation des bibliothèques dans le chapitre 9 sur p. 206.](#)

Remarque : Au cours de l'extraction, certaines ressources internes compilées sont aussi utilisées. Ces ressources compilées contiennent de nombreuses définitions qui complètent les types de Core Library. Ces ressources compilées ne peuvent pas être éditées.

L'Editeur de ressources permet d'accéder aux ressources utilisées pour produire les résultats de l'extraction (concepts, types et patrons). Vous pouvez effectuer de nombreuses tâches dans l'Editeur de ressources, dont :

- **Utilisation de bibliothèques**[Pour plus d'informations, reportez-vous à la section Utilisation des bibliothèques dans le chapitre 9 sur p. 206.](#)
- **Création de déclarations de types**[Pour plus d'informations, reportez-vous à la section Création de types dans le chapitre 10 sur p. 222.](#)
- **Ajout de termes à la déclaration**[Pour plus d'informations, reportez-vous à la section Ajout de termes dans le chapitre 10 sur p. 224.](#)

- **Création de synonymes** Pour plus d'informations, reportez-vous à la section Définition de synonymes dans le chapitre 10 sur p. 233.
- **Mise à jour des ressources dans les TAP.** Pour plus d'informations, reportez-vous à la section Mise à jour des Packages d'analyse de texte dans le chapitre 3 sur p. 45.
- **Création de modèles.** Pour plus d'informations, reportez-vous à la section Création et mise à jour de modèles sur p. 197.
- **Import et export des modèles** Pour plus d'informations, reportez-vous à la section Import et export des modèles sur p. 200.
- **Publication de bibliothèques.** Pour plus d'informations, reportez-vous à la section Publication de bibliothèques dans le chapitre 9 sur p. 216.

Interface de l'éditeur

Les opérations que vous effectuez dans l'Editeur de ressources concernent la gestion et l'adaptation des ressources linguistiques. Ces ressources sont stockées sous la forme de modèles et de bibliothèques. Pour plus d'informations, reportez-vous à la section Déclarations de types dans le chapitre 10 sur p. 220.

Onglet Ressources de bibliothèque

Figure 8-1

The screenshot displays the 'Ressources de bibliothèque' (Library Resources) tab in the IBM SPSS Text Analytics for Surveys software. The interface is organized into several key sections:

- Left Sidebar (1):** A tree view showing a hierarchy of libraries, including 'Opinions (English)', 'Bibliothèque locale', 'Opinions Library (English)', 'Budget Library (English)', 'Core Library (English)', 'Variations Library (English)', 'Slang Library (English)', and 'Emoticon Library (English)'.
- Main Table (2):** A table listing various terms and their associated metadata. The columns are 'Terme', 'Correspondance', 'Flexion', 'Type', and 'Bibliothèque'. The table contains numerous entries, such as '0 problem', '10 stars', '100% correct', etc., all linked to the 'Opinions Library (English)'. Large number '2' is placed over this table.
- Right Panel (3):** A list of exclusion rules, each with a checkbox and a description, such as 'any kind of prob...', 'any problems i...', 'anykind of prob...', etc. Large number '3' is placed over this panel.
- Bottom Pane (4):** A section for editing a specific resource. It shows a 'Cible' (Target) field with the text 'able to log-on' and a 'Synonymes' (Synonyms) field containing a list of alternative phrases like 'able to log-in', 'able to login', 'can always log-in', etc. Large number '4' is placed over this section.

At the bottom of the window, a status bar provides summary statistics: 7 Bibliothèques, 35 Types, 13425 Termes, 28 Exclusions, 1334 Synonymes, and 27 Optionnels.

L'interface comporte quatre parties :

1. Panneau Arborescence des bibliothèques. Situé dans l'angle supérieur gauche, ce plan présente l'arborescence des bibliothèques. Vous pouvez activer et désactiver les bibliothèques figurant dans cette arborescence. Vous pouvez également filtrer les vues des autres panneaux en sélectionnant une bibliothèque dans l'arborescence. Vous pouvez effectuer plusieurs opérations dans cette arborescence à l'aide des menus contextuels. Lorsque vous développez une bibliothèque figurant dans l'arborescence, vous pouvez consulter l'ensemble des types qu'elle comporte. Vous pouvez également filtrer cette liste avec le menu Affichage si vous souhaitez vous concentrer sur une bibliothèque spécifique.

2. Listes des termes du panneau Déclarations de types. Ce panneau, situé à droite de l'arborescence des bibliothèques, affiche les listes de termes des déclarations de types correspondant aux bibliothèques sélectionnées dans l'arborescence des bibliothèques. Une **déclaration de types** est un ensemble de termes devant être regroupés sous une même étiquette ou sous un même nom de type. Lorsque le moteur du programme d'extraction lit vos données textuelles, il compare les mots trouvés dans le texte aux termes figurant dans les déclarations de types. Si un concept extrait figure sous la forme d'un terme dans une déclaration de types, le nom du type en question lui est alors affecté. Vous pouvez considérer la déclaration de types comme étant un dictionnaire distinct qui regroupe les termes présentant des points communs. Par exemple, le type <Location> qui figure dans Core Library comporte des concepts tels que `new orleans`, `great britain` et `new york`. Ces termes désignent tous des lieux géographiques. Une bibliothèque peut comporter une ou plusieurs déclarations de types. [Pour plus d'informations, reportez-vous à la section Déclarations de types dans le chapitre 10 sur p. 220.](#)

3. Panneau Dictionnaire d'exclusions. Ce panneau, situé sur le côté droit, affiche la collection de termes qui seront exclus des résultats finaux de l'extraction. Les termes apparaissant dans ce dictionnaire d'exclusions n'apparaissent pas dans le panneau Résultats d'extraction. Les termes exclus peuvent être stockés dans la bibliothèque de votre choix. Mais, le panneau Dictionnaire d'exclusions affiche tous les termes exclus de toutes les bibliothèques visibles dans l'arborescence des bibliothèques. [Pour plus d'informations, reportez-vous à la section Dictionnaires d'exclusions dans le chapitre 10 sur p. 236.](#)

4. Panneau Dictionnaire de substitutions. Ce panneau, situé en bas à gauche, affiche les synonymes et les éléments optionnels, chacun dans leur propre onglet. Les synonymes et les éléments optionnels permettent de grouper les termes similaires sous un concept principal ou cible dans les résultats d'extraction finaux. Ce dictionnaire peut comporter des synonymes connus, des synonymes et des éléments définis par l'utilisateur, ainsi que les fautes d'orthographe courantes associées à leur correction. Les définitions des synonymes et les éléments optionnels peuvent être stockés dans la bibliothèque de votre choix. Mais, le panneau Dictionnaire de substitutions affiche tous les contenus de toutes les bibliothèques visibles dans l'arborescence des bibliothèques. Pendant que ce panneau affiche tous les synonymes et éléments facultatifs de toutes les bibliothèques, les substitutions de toutes les bibliothèques de l'arborescence apparaissent ensemble dans ce panneau. Une bibliothèque ne peut comporter qu'un seul dictionnaire de substitutions. [Pour plus d'informations, reportez-vous à la section Dictionnaires des substitutions/synonymes dans le chapitre 10 sur p. 231.](#)

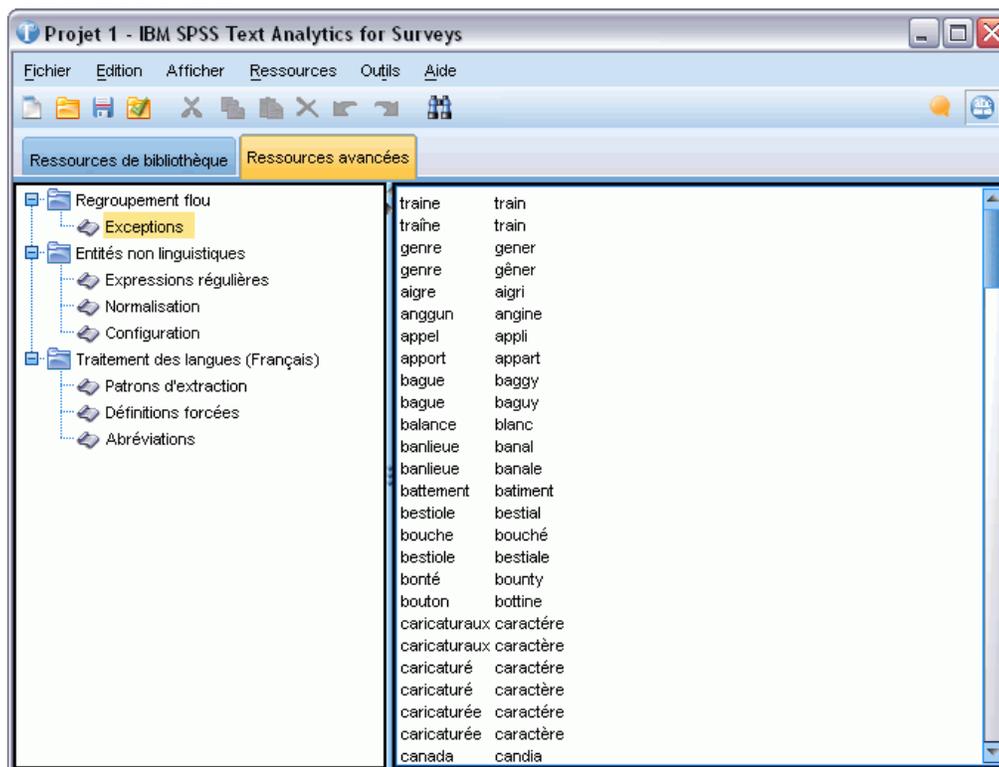
Remarque :

- pour procéder à un filtrage visant à n'afficher que les informations propres à une seule bibliothèque, vous pouvez modifier l'affichage de la bibliothèque à l'aide de la liste déroulante figurant dans la barre d'outils. Elle comporte une entrée de niveau supérieur appelée Toutes les bibliothèques, ainsi qu'une autre entrée supplémentaire pour chacune des bibliothèques. [Pour plus d'informations, reportez-vous à la section Affichage des bibliothèques dans le chapitre 9 sur p. 210.](#)

Onglet Ressources avancées

Les ressources avancées sont disponibles dans le deuxième onglet de la vue de l'éditeur. Vous pouvez consulter et modifier les ressources avancées dans cet onglet. [Pour plus d'informations, reportez-vous à la section À propos des ressources avancées dans le chapitre 11 sur p. 239.](#)

Figure 8-2
Ressources avancées

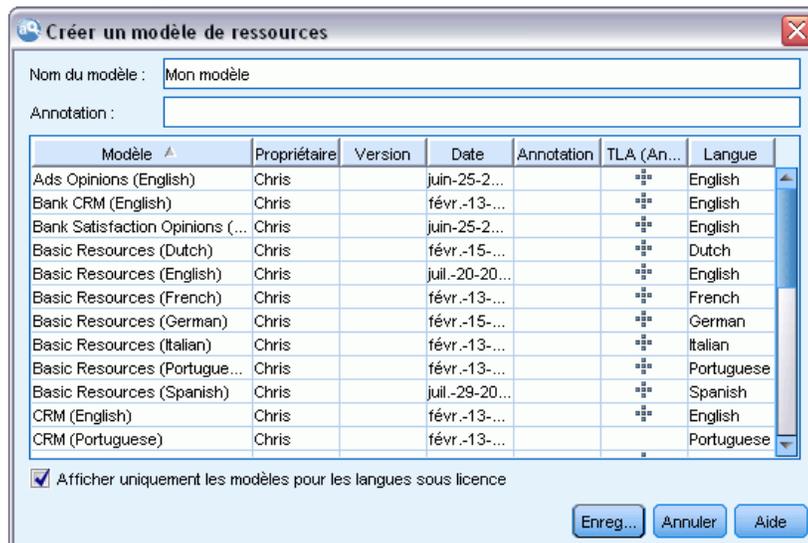


Création et mise à jour de modèles

Chaque fois que vous apportez des modifications à vos ressources et que vous souhaitez les réutiliser ultérieurement, enregistrez-les en tant que modèle. Lorsque vous effectuez cette opération, vous pouvez choisir d'enregistrer vos ressources en utilisant le nom d'un modèle existant ou en indiquant un nouveau nom. Ensuite, chaque fois que vous chargerez ce modèle, vous obtiendrez les mêmes ressources.

Remarque : Vous avez également la possibilité de publier et de partager vos bibliothèques. [Pour plus d'informations, reportez-vous à la section Partage de bibliothèques dans le chapitre 9 sur p. 214.](#)

Figure 8-3
Boîte de dialogue Créer un modèle de ressource



Pour créer (ou mettre à jour) un modèle

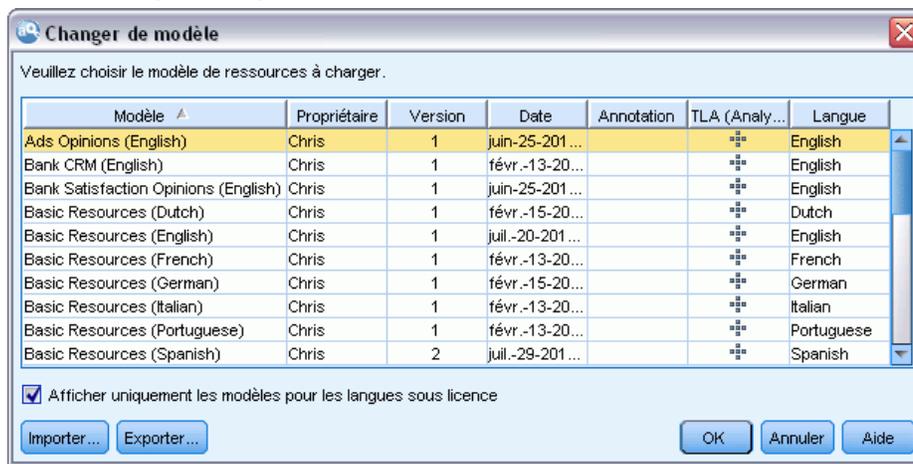
- ▶ Dans les menus de la vue Editeur de ressources, choisissez Ressources > Créer un modèle de ressource. La boîte de dialogue Créer un modèle de ressource apparaît.
- ▶ Entrez un nouveau nom dans le champ Nom du modèle, si vous créez un modèle. Sélectionnez un modèle dans le tableau si vous souhaitez remplacer un modèle existant par les ressources chargées.
- ▶ Cliquez sur Enregistrer pour créer le modèle.

Changement des modèles de ressources

Si vous voulez remplacer les ressources chargées par une copie de celles d'un autre modèle, vous pouvez changer ces modèles. Cette action remplacera toutes les ressources chargées.

Vous pouvez sélectionner le modèle dont vous souhaitez copier le contenu dans l'Editeur de ressources puis cliquer sur OK. Vous remplacez ainsi les ressources contenues dans le projet de

Figure 8-4
Boîte de dialogue Changer de modèle



Pour Changer de modèle

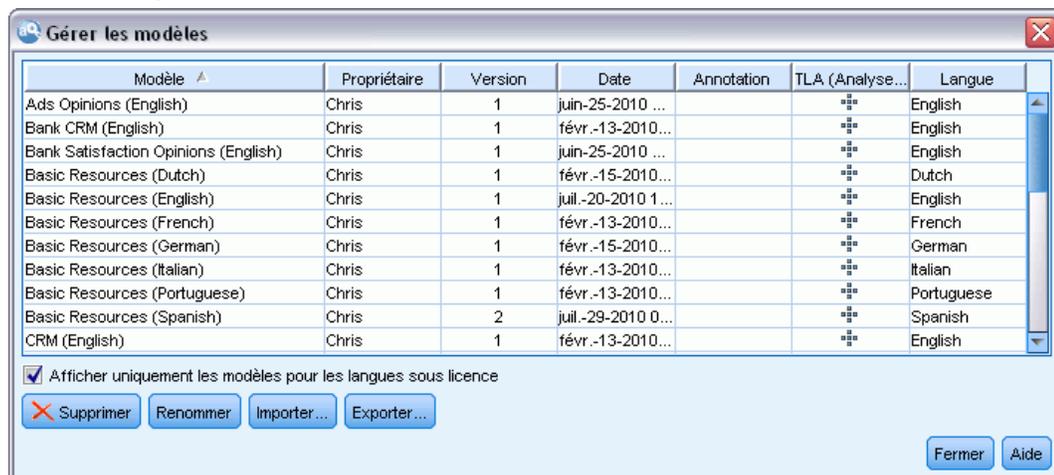
- ▶ Dans les menus de la vue Editeur de ressources, choisissez Ressources > Changer de modèles de ressource. La boîte de dialogue Changer de modèles de ressource apparaît.
- ▶ Sélectionnez le modèle à utiliser parmi ceux répertoriés dans le tableau.
- ▶ Cliquez sur OK pour abandonner les ressources chargées et charger une copie de celles contenues dans le modèle sélectionné à la place. Si vous avez apporté des modifications à vos ressources et souhaitez enregistrer vos bibliothèques pour un usage ultérieur, vous pouvez les publier, les mettre à jour et les partager avant de passer à un autre modèle. [Pour plus d'informations, reportez-vous à la section Partage de bibliothèques dans le chapitre 9 sur p. 214.](#)

Gestion des modèles

Il existe également certaines tâches de gestion de base que vous pouvez effectuer occasionnellement sur vos modèles telles que l'attribution de nouveau nom, l'import et l'export, ou la suppression des modèles obsolètes. Ces tâches sont réalisées dans la boîte de dialogue Gérer les modèles. L'import et l'export des modèles permet de les partager avec d'autres utilisateurs. [Pour plus d'informations, reportez-vous à la section Import et export des modèles sur p. 200.](#)

Remarque : Vous ne pouvez pas renommer ou supprimer les modèles installés (ou fournis) avec ce produit. Si vous voulez renommer, vous pouvez ouvrir le modèle installé et en créer un nouveau avec le nom de votre choix. Vous pouvez supprimer vos modèles personnalisés ; néanmoins, si vous essayez de supprimer un modèle fourni, il sera réinitialisé à la version installée à l'origine.

Figure 8-5
Boîte de dialogue Gérer les modèles



Pour renommer un modèle

- ▶ Dans les menus, sélectionnez Ressources > Gérer les modèles de ressources. La boîte de dialogue Gérer les modèles apparaît.
- ▶ Sélectionnez le modèle à renommer et cliquez sur Renommer. La zone de nom devient un champ modifiable dans le tableau.
- ▶ Tapez un nouveau nom et appuyez sur la touche Entrée. Un message de confirmation apparaît.
- ▶ Si le nouveau nom vous satisfait, cliquez sur Oui. Sinon, cliquez sur Non.

Pour supprimer un modèle

- ▶ Dans les menus, sélectionnez Ressources > Gérer les modèles de ressources. La boîte de dialogue Gérer les modèles apparaît.
- ▶ Dans la boîte de dialogue Gérer les modèles, sélectionnez le modèle à supprimer.
- ▶ Choisissez Supprimer. Un message de confirmation apparaît.
- ▶ Cliquez sur Oui pour supprimer ou sur Non pour annuler la demande de suppression. Si vous cliquez sur Oui, le modèle est supprimé.

Import et export des modèles

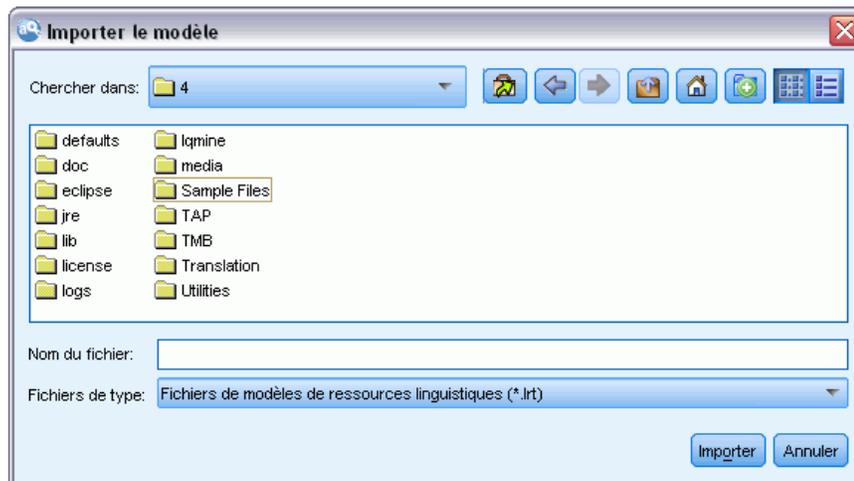
Vous pouvez partager les modèles avec d'autres utilisateurs ou ordinateurs en les important et en les exportant. Les modèles sont stockés dans une base de données interne, mais ils ne peuvent pas être exportés sous forme de fichiers *.lrt sur votre disque dur. Vous pouvez importer et exporter des modèles dans la boîte de dialogue Gérer les modèles dans l'Editeur de ressources.

Pour importer un modèle

- Dans la boîte de dialogue, cliquez sur Importer. La boîte de dialogue Importer un modèle apparaît.

Figure 8-6

Boîte de dialogue Importer un modèle



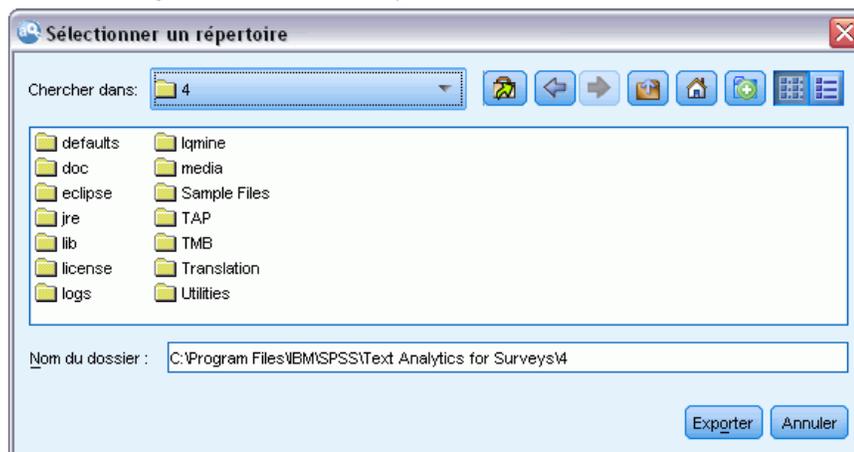
- Sélectionnez le fichier du modèle de ressources (*.lrt) à importer et cliquez sur Importer. Enregistrez le modèle importé en lui attribuant un nouveau nom ou en remplaçant le modèle existant. La boîte de dialogue se ferme et le modèle apparaît dans le tableau.

Pour exporter un modèle

- Dans la boîte de dialogue, sélectionnez le modèle à exporter et cliquez sur Exporter. La boîte de dialogue Sélectionner un répertoire apparaît.

Figure 8-7

Boîte de dialogue Sélectionner un répertoire



- Sélectionnez le répertoire vers lequel vous souhaitez exporter et cliquez sur Exporter. Cette boîte de dialogue se ferme et le modèle est exporté et porte l'extension de fichier (*.lrt).

Sauvegarde des ressources

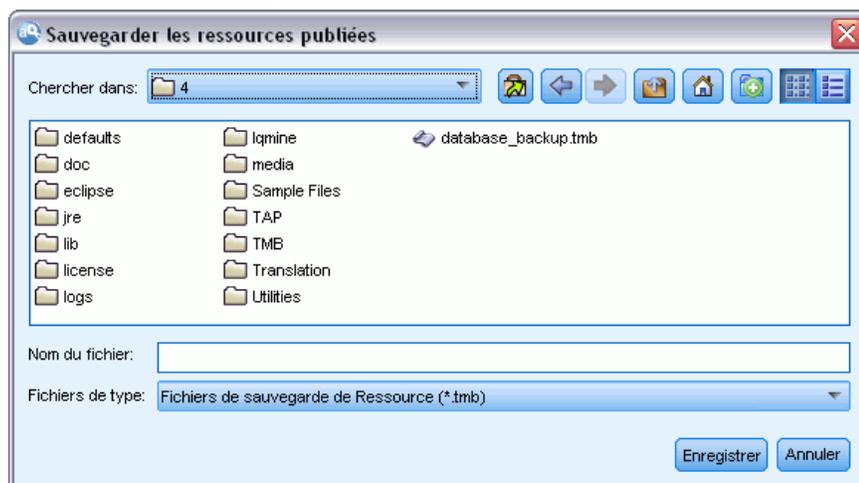
Il se peut que vous ayez besoin, de temps à autre, de sauvegarder vos ressources par mesure de sécurité.

Important ! Lorsque vous procédez à une restauration, le contenu intégral de vos ressources est supprimé et seul le contenu du fichier de sauvegarde est accessible dans le produit. Ceci inclut tout travail en cours.

Pour sauvegarder les ressources

- Dans les menus, sélectionnez Ressources > Outils de sauvegarde > Sauvegarder les ressources. La boîte de dialogue Sauvegarder apparaît.

Figure 8-8
Boîte de dialogue Sauvegarder les ressources

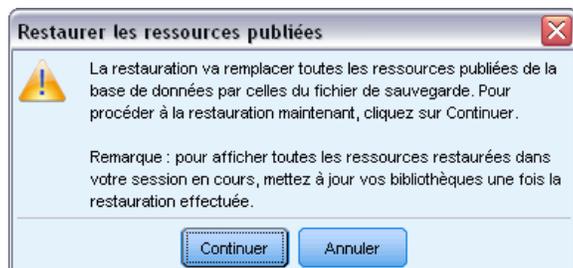


- Nommez votre fichier de sauvegarde et cliquez sur Enregistrer. La boîte de dialogue se ferme et le fichier de sauvegarde est créé.

Pour restaurer les ressources

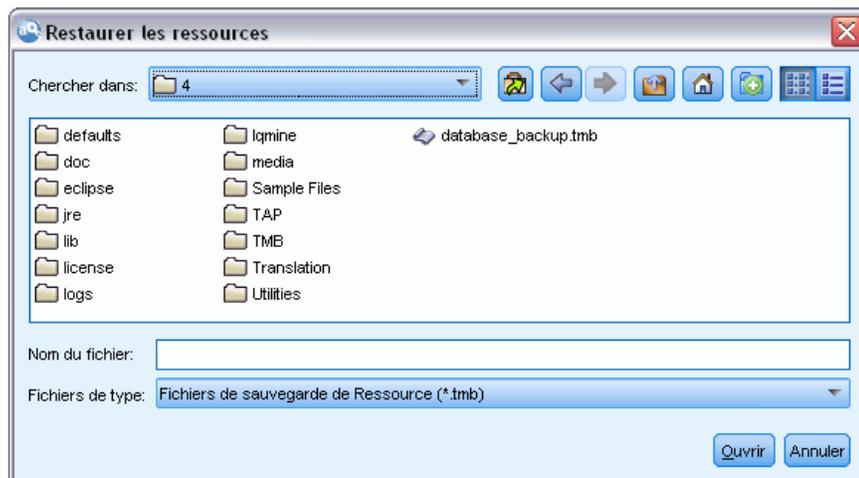
- Dans les menus, sélectionnez Ressources > Outils de sauvegarde > Restaurer les ressources. Une alerte vous avertit que la restauration va écraser le contenu actuel de votre base de données.

Figure 8-9
Message d'alerte d'écrasement



- Cliquez sur Oui pour continuer. Si un projet est ouvert, il sera conservé, car il est en mémoire ; néanmoins, vous devez le sauvegarder de nouveau pour le conserver dans la base de données nouvellement restaurée. La boîte de dialogue apparaît.

Figure 8-10
Boîte de dialogue Restaurer les ressources



- Sélectionnez le fichier de sauvegarde à restaurer et cliquez sur Ouvrir. La boîte de dialogue se ferme et les ressources sont restaurées dans l'application.

Important ! Lorsque vous procédez à une restauration, le contenu intégral de vos ressources est supprimé et seul le contenu du fichier de sauvegarde est accessible dans le produit. Ceci inclut tout travail en cours.

Import des fichiers de ressources

Si vous avez effectué des modifications directement dans les fichiers de ressources en-dehors de ce produit, vous pouvez les importer dans une bibliothèque donnée, en sélectionnant cette bibliothèque et en procédant à l'importation. Lorsque vous importez un répertoire, vous pouvez également importer l'ensemble des fichiers pris en charge dans une bibliothèque ouverte spécifique. Vous ne pouvez importer que les fichiers **.txt*.

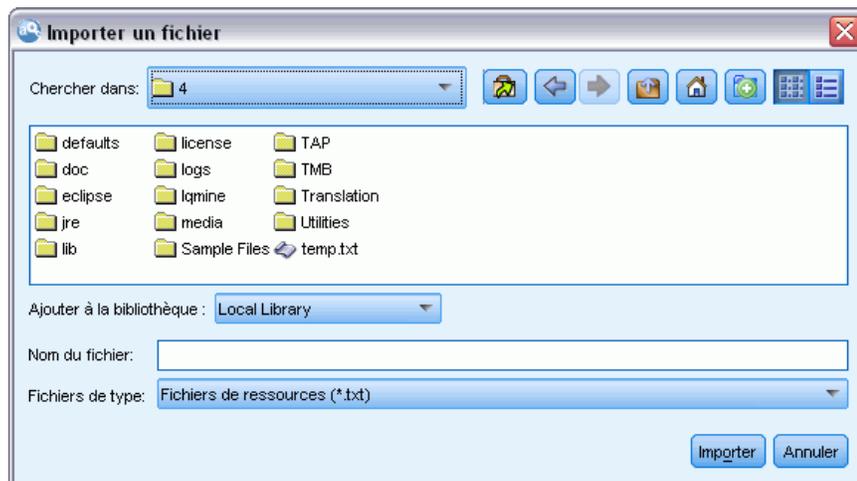
Chaque fichier importé doit contenir seulement une entrée par fichier, et si le contenu est structuré comme :

- Une liste de mots ou d'expressions (une par ligne). Le fichier est importé comme liste de termes d'une déclaration de types, où la déclaration de types prend le nom du fichier sans l'extension.
- Une liste d'entrées telle que *terme1<TAB>terme2*, est ensuite importée comme liste de synonymes, avec *terme1* comme ensemble des termes sous-jacents et *terme2* comme terme cible.

Pour importer un unique fichier de ressources

- Dans les menus, choisissez Ressources > Importer des fichiers > Importer un fichier. La boîte de dialogue Importer un fichier apparaît.

Figure 8-11
Boîte de dialogue Importer un fichier

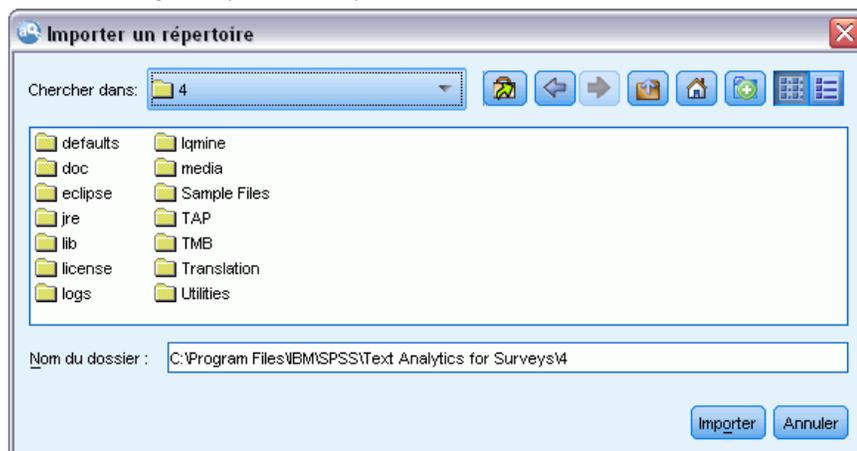


- Sélectionnez le fichier que vous souhaitez importer et cliquez sur Importer. Un format interne est appliqué au contenu du fichier qui est ensuite ajouté à votre bibliothèque.

Pour importer l'ensemble des fichiers d'un répertoire

- Dans les menus, choisissez Ressources > Importer des fichiers > Importer un répertoire entier. La boîte de dialogue Importer un répertoire apparaît.

Figure 8-12
Boîte de dialogue Importer un répertoire



- Dans la liste Importer, sélectionnez la bibliothèque dans laquelle vous souhaitez importer l'ensemble des fichiers de ressources. Si vous sélectionnez l'option Par défaut, une bibliothèque est créée. Elle porte le nom du répertoire.

- ▶ Sélectionnez le répertoire à partir duquel les fichiers doivent être importés. Les sous-répertoires ne seront pas lus.
- ▶ Cliquez sur Importer. La boîte de dialogue se ferme et le contenu des fichiers de ressources importés apparaît dans l'éditeur sous forme de dictionnaires et de fichiers de ressources avancés.

Utilisation des bibliothèques

Les ressources utilisées par le moteur d'extraction pour extraire et regrouper les termes des données textuelles contiennent une ou plusieurs bibliothèques. Vous pouvez visualiser l'ensemble des bibliothèques dans l'arborescence des bibliothèques située dans la partie supérieure gauche de l'Editeur de ressources. Les bibliothèques comportent trois types de dictionnaires : Types, Substitutions et Exclusions. [Pour plus d'informations, reportez-vous à la section A propos des dictionnaires de bibliothèque dans le chapitre 10 sur p. 220.](#)

Le modèle de ressources ou les ressources du TAP choisi comprennent plusieurs bibliothèques afin de vous permettre de procéder immédiatement à l'extraction des concepts de vos données textuelles. Cependant, vous pouvez également créer vos propres bibliothèques et les publier pour pouvoir les réutiliser. [Pour plus d'informations, reportez-vous à la section Publication de bibliothèques sur p. 216.](#)

Supposons, par exemple, que vous utilisez fréquemment des données textuelles relatives au secteur automobile. Après avoir analysé vos données, vous décidez de créer des ressources personnalisées pour gérer les termes propres à ce secteur. A l'aide d'Editeur de ressources, vous pouvez créer un nouveau modèle et, dans ce modèle, une bibliothèque permettant d'extraire et de regrouper les termes correspondant au secteur automobile. Etant donné que vous aurez à nouveau besoin de ces informations dans cette bibliothèque, publiez votre bibliothèque dans un référentiel central, accessible à partir de la boîte de dialogue **Gérer les bibliothèques**, afin qu'elle puisse être réutilisée indépendamment dans différents projets.

Supposons que vous souhaitez également regrouper les termes propres à différents sous-secteurs, tels que les dispositifs électroniques, les moteurs, les systèmes de refroidissement, voire un fabricant ou un marché particulier. Vous pouvez créer une bibliothèque pour chaque groupe, puis publier ces bibliothèques de manière à ce qu'elles puissent être utilisées avec plusieurs ensembles de données textuelles. De cette manière, vous pouvez ajouter les bibliothèques qui correspondent le mieux au contenu de vos données textuelles.

Remarque : les ressources supplémentaires peuvent être configurées et gérées dans l'onglet Ressources avancées. Certaines s'appliquent à toutes les bibliothèques et gèrent les entités non linguistiques, les exceptions de regroupement flou, etc. [Pour plus d'informations, reportez-vous à la section À propos des ressources avancées dans le chapitre 11 sur p. 239.](#)

Bibliothèques fournies

Par défaut, plusieurs bibliothèques sont installées avec IBM® SPSS® Text Analytics for Surveys. Vous pouvez utiliser ces bibliothèques préformatées pour accéder à des milliers de termes et synonymes prédéfinis, ainsi qu'à plusieurs types différents. Ces bibliothèques fournies sont affinées pour être utilisées dans plusieurs domaines différents et sont disponibles dans plusieurs langues.

De nombreuses bibliothèques existent mais les plus couramment utilisées sont les suivantes :

- **Bibliothèque locale.** Utilisée pour stocker les dictionnaires définis par l'utilisateur. Il s'agit d'une bibliothèque vide ajoutée par défaut à toutes les ressources. Elle contient également une déclaration de types vide. Elle est particulièrement utile pour effectuer des modifications ou des améliorations aux ressources directement (comme ajouter un mot à un type) depuis la fenêtre d'analyse de texte. Dans ce cas, ces modifications et ces améliorations sont automatiquement stockées dans la première bibliothèque répertoriée dans l'arborescence de bibliothèques d'Editeur de ressources ; par défaut, il s'agit de la *bibliothèque locale*. Vous ne pouvez pas publier cette bibliothèque car elle est spécifique aux données du session. Pour publier son contenu, renommez tout d'abord la bibliothèque.
- **Core library.** Utilisée dans la plupart des cas, puisqu'elle comprend les cinq types de base intégrés représentant les personnes, les lieux, les organisations, les produits et les éléments inconnus. Bien que seuls quelques termes soient répertoriés dans l'une des déclarations de types, les types représentés dans Core library viennent en fait compléter les types fiables détectés dans les ressources internes, compilées et livrées avec votre produit de Text Mining. Ces ressources internes et compilées contiennent des milliers de termes pour chaque type. Pour cette raison, même si vous ne voyez pas un terme dans une liste de termes des déclarations de types, il peut quand même être extrait et saisi avec un type Core. Cela explique la façon dont les noms comme *George* peuvent être extraits et recevoir le type <Person>, alors que seul *John* apparaît dans la déclaration de types <Person> de Core library. De même, si vous n'incluez pas Core Library, il se peut que vous aperceviez toujours ces types dans vos résultats d'extraction puisque les ressources compilées qui les contiennent sont toujours utilisées par le moteur du programme d'extraction.
- **Opinions Library.** Utilisé le plus souvent pour extraire des opinions et des sentiments des données textuelles. Cette bibliothèque comprend des milliers de mots qui représentent des attitudes, des qualificatifs et des préférences et qui — lorsqu'ils sont utilisés avec d'autres termes — indiquent une opinion sur un sujet. Cette bibliothèque contient de nombreux types intégrés, des synonymes et des exclusions. Elle comprend également un grand nombre de règles de patrons utilisées pour l'analyse des liens du texte.
- **Budget Library.** Utilisée pour extraire les termes faisant référence au coût d'un objet ou d'un service. La bibliothèque comprend plusieurs mots et expressions représentant des adjectifs, des qualificatifs et des jugements concernant le prix ou la qualité d'un objet ou d'un service.
- **Variations Library.** Utilisée pour inclure les observations dans lesquelles certaines variations de langue requièrent des définitions de synonyme pour être correctement regroupées. Cette bibliothèque ne comporte que des définitions de synonyme.

Bien que certaines bibliothèques fournies en dehors des modèles ressemblent au contenu de certains modèles, les modèles ont été spécifiquement adaptés à certaines applications et contiennent des ressources avancées supplémentaires. Nous vous recommandons d'essayer d'utiliser un modèle conçu pour le genre de données textuelles que vous utilisez et d'effectuer les modifications sur ces ressources plutôt que de simplement ajouter des bibliothèques individuelles à un modèle plus générique.

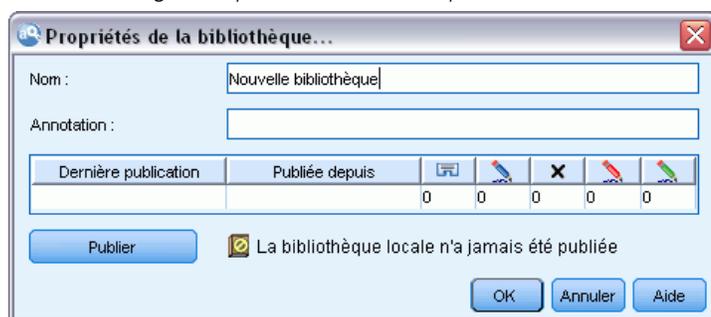
Les ressources compilées sont également fournies avec SPSS Text Analytics for Surveys. Elles sont systématiquement utilisées au cours du processus d'extraction et contiennent un grand nombre de définitions complémentaires dans les déclarations de types intégrées aux bibliothèques par défaut. Etant donné que ces ressources sont compilées, il est impossible de les visualiser ou de

les modifier. Vous pouvez toutefois forcer n'importe quel autre dictionnaire à accepter un terme classé par type par les ressources compilées. [Pour plus d'informations, reportez-vous à la section Ajout des termes forcés dans le chapitre 10 sur p. 228.](#)

Création de bibliothèques

Vous pouvez créer autant de bibliothèques que vous le souhaitez. Après avoir créé une nouvelle bibliothèque, vous pouvez commencer à y créer des déclarations de types et insérer des termes, des synonymes et des exclusions.

Figure 9-1
Boîte de dialogue Propriétés de bibliothèque



Pour créer une bibliothèque

- ▶ A partir des menus, sélectionnez Ressources > Nouvelle bibliothèque. La boîte de dialogue Ajouter une bibliothèque au projet s'ouvre.
- ▶ Nommez la bibliothèque dans le champ Nom.
- ▶ Si vous le souhaitez, insérez un commentaire dans le champ Annotation.
- ▶ Cliquez sur Publier si vous souhaitez publier cette bibliothèque maintenant, sans y insérer quoi que ce soit. [Pour plus d'informations, reportez-vous à la section Partage de bibliothèques sur p. 214.](#) Vous pouvez également la publier plus tard à n'importe quel moment.
- ▶ Cliquez sur OK pour créer la bibliothèque. La boîte de dialogue se ferme et la bibliothèque apparaît dans l'arborescence. Si vous développez la bibliothèque dans l'arborescence, vous vous apercevez qu'une déclaration de types vide y a été incluse automatiquement. Vous pouvez ajouter des termes immédiatement. [Pour plus d'informations, reportez-vous à la section Ajout de termes dans le chapitre 10 sur p. 224.](#)

Ajout de bibliothèques publiques

Si vous souhaitez réutiliser une bibliothèque contenant les données d'un autre projet, vous pouvez l'ajouter à vos ressources en cours à condition qu'il s'agisse d'une bibliothèque publique. Une **bibliothèque publique** est une bibliothèque qui a été publiée. [Pour plus d'informations, reportez-vous à la section Publication de bibliothèques sur p. 216.](#)

Lorsque vous ajoutez une bibliothèque publique, une copie **locale** est incorporée à vos données de projet. Il est possible d'apporter des modifications à cette bibliothèque, cependant, vous êtes tenu de publier à nouveau sa version publique si vous souhaitez partager les modifications effectuées.

Lorsque vous ajoutez une bibliothèque publique, une boîte de dialogue Résoudre les conflits peut apparaître s'il existe des conflits entre les termes et les types d'une bibliothèque et les autres bibliothèques locales. Il est nécessaire de résoudre ces conflits ou d'accepter les résolutions proposées afin de pouvoir réaliser cette opération. [Pour plus d'informations, reportez-vous à la section Résolution des conflits sur p. 218.](#)

Remarque : Si vous mettez systématiquement à jour vos bibliothèques lorsque vous ouvrez ou publiez une bibliothèque lorsque vous fermez un projet, vous avez moins de risques d'avoir des bibliothèques qui ne sont pas synchronisées. [Pour plus d'informations, reportez-vous à la section Partage de bibliothèques sur p. 214.](#)

Figure 9-2
Boîte de dialogue Ajouter une bibliothèque



Pour ajouter une bibliothèque

- ▶ A partir des menus, sélectionnez Ressources > Ajouter une bibliothèque. La boîte de dialogue Ajouter une bibliothèque apparaît.
- ▶ Sélectionnez la ou les bibliothèques dans la liste.
- ▶ Cliquez sur Ajouter. En cas de conflits entre les bibliothèques qui viennent d'être ajoutées et les anciennes, vous êtes invité à vérifier les résolutions de conflit ou à les modifier avant de réaliser l'opération. [Pour plus d'informations, reportez-vous à la section Résolution des conflits sur p. 218.](#)

Recherche de termes et de types

Vous pouvez rechercher dans les nombreux panneaux de l'éditeur à l'aide de la fonction Rechercher. Dans l'éditeur, sélectionnez Edition > Rechercher dans les menus. La barre d'outils Rechercher apparaît. Utilisez cette barre d'outils pour rechercher une occurrence à la fois. En cliquant à nouveau sur Rechercher, vous pouvez rechercher les occurrences suivantes du terme recherché.

Lors de la recherche, l'éditeur effectue la recherche uniquement dans les bibliothèques répertoriées dans la liste déroulante de la barre d'outils Rechercher. Si l'option Toutes les bibliothèques est sélectionnée, le programme effectue la recherche dans tous les supports dans l'éditeur.

Lorsque vous démarrez une recherche, elle débute dans la zone active. La recherche continue dans chaque section, formant une boucle jusqu'à revenir à la cellule active. Il est possible d'inverser l'ordre de la recherche en utilisant les flèches de direction. Vous pouvez également choisir si votre recherche est sensible ou non à la casse.

Pour rechercher des chaînes dans la vue

- ▶ Dans les menus, sélectionnez Edition > Rechercher. La barre d'outils Rechercher apparaît.
- ▶ Insérez la chaîne de caractères que vous souhaitez rechercher.
- ▶ Cliquez sur le bouton Rechercher pour commencer la recherche. La prochaine occurrence du terme ou du type est ensuite mise en surbrillance.
- ▶ Cliquez à nouveau sur le bouton pour passer d'une occurrence à une autre.

Affichage des bibliothèques

Vous pouvez afficher le contenu d'une bibliothèque donnée ou celui de toutes les bibliothèques. Cela peut s'avérer utile lorsque vous travaillez avec plusieurs bibliothèques ou lorsque vous souhaitez passer en revue le contenu d'une bibliothèque spécifique avant sa publication. Le fait de modifier l'affichage n'a un impact que sur ce que vous voyez dans l'onglet Ressources de bibliothèque mais cette modification n'empêche pas l'utilisation des bibliothèques au cours de l'extraction. [Pour plus d'informations, reportez-vous à la section Désactivation des bibliothèques locales sur p. 211.](#)

L'affichage par défaut est Toutes les bibliothèques. Cette option affiche toutes les bibliothèques de l'arborescence et leur contenu dans d'autres panneaux. Vous pouvez modifier cette sélection en utilisant la liste déroulante de la barre d'outils ou via une sélection de menus (Affichage > Bibliothèques). Si une seule bibliothèque est affichée, tous les éléments des autres bibliothèques disparaissent de la vue mais continuent d'être lus pendant l'extraction.

Pour afficher la vue Bibliothèque

- ▶ Dans les menus de l'onglet Ressources de bibliothèque, choisissez Vue > Bibliothèques. Un menu comprenant toutes les bibliothèques locales apparaît.
- ▶ Sélectionnez la bibliothèque que vous souhaitez afficher ou sélectionnez l'option Toutes les bibliothèques pour afficher le contenu de toutes les bibliothèques. Le contenu de la vue est filtré en fonction de votre sélection.

Gestion des bibliothèques locales

Les bibliothèques locales sont les bibliothèques qui se trouvent dans votre projet de ou dans un modèle, contrairement aux bibliothèques publiques. [Pour plus d'informations, reportez-vous à la section Gestion des bibliothèques publiques sur p. 212.](#) Vous souhaitez peut-être exécuter certaines tâches de gestion de bibliothèques locales, notamment : renommer, désactiver ou supprimer une bibliothèque locale.

Attribution d'un nouveau nom à une bibliothèque locale

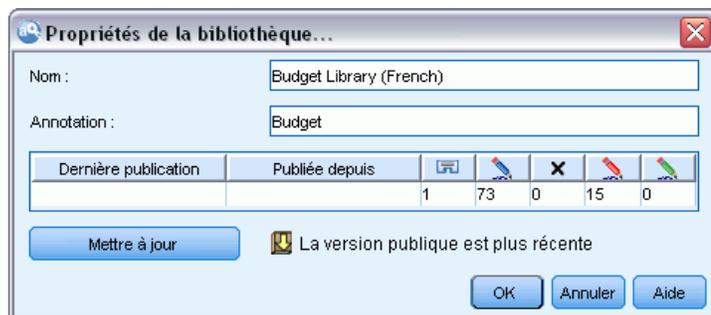
Vous pouvez renommer les bibliothèques locales. Lorsque vous renommez une bibliothèque locale, vous la dissociez de la version publique éventuelle. Cela signifie que les modifications suivantes ne peuvent plus être partagées avec la version publique. Vous pouvez republier cette bibliothèque locale sous son nouveau nom. Cela signifie également que vous ne pouvez pas incorporer les modifications apportées à cette version locale dans la version publique originale.

Remarque : Il est impossible de renommer une bibliothèque publique.

- Dans les menus, sélectionnez Edition > Propriétés de bibliothèque. La boîte de dialogue Propriétés de bibliothèque apparaît.

Figure 9-3

Boîte de dialogue Propriétés de bibliothèque



Pour renommer une bibliothèque locale

- Dans l'arborescence, sélectionnez la bibliothèque à renommer.
- Attribuez un nouveau nom à la bibliothèque dans le champ Nom.
- Cliquez sur OK pour accepter le nouveau nom de la bibliothèque. La boîte de dialogue se ferme et son nom apparaît dans l'arborescence.

Désactivation des bibliothèques locales

Si vous souhaitez exclure temporairement une bibliothèque du processus d'extraction, désélectionnez la case à cocher à gauche du nom de la bibliothèque dans l'arborescence. Vous indiquez ainsi que vous souhaitez conserver la bibliothèque, mais en ignorer le contenu pendant la recherche de conflits et le processus d'extraction.

Pour désactiver une bibliothèque

- ▶ Dans l'arborescence de bibliothèques, sélectionnez la bibliothèque à désactiver.
- ▶ Cliquez sur la barre d'espace. La coche disparaît de la case figurant à gauche du nom.

Suppression des bibliothèques locales

Vous pouvez supprimer une bibliothèque sans supprimer la version publique de la bibliothèque et inversement. La suppression d'une bibliothèque locale supprime la bibliothèque et tout son contenu du projet uniquement. Le fait de supprimer la version locale d'une bibliothèque ne supprime pas la bibliothèque des autres projets et ne supprime pas non plus la version publique. [Pour plus d'informations, reportez-vous à la section Gestion des bibliothèques publiques sur p. 212.](#)

Pour supprimer une bibliothèque locale

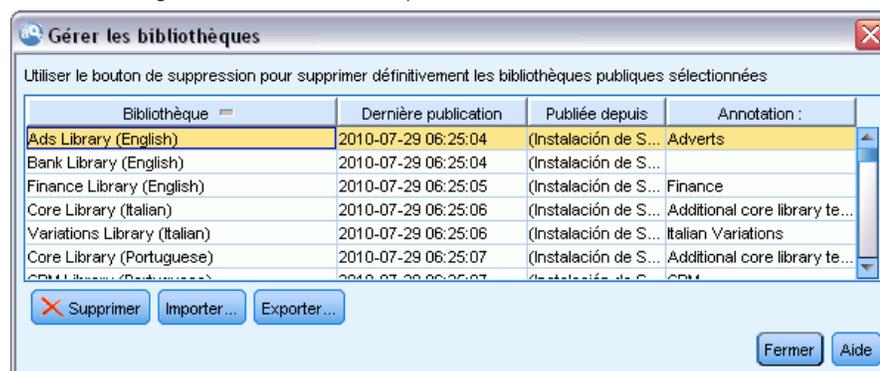
- ▶ Dans l'arborescence, sélectionnez la bibliothèque à supprimer.
- ▶ Dans les menus, choisissez Edition > Supprimer pour supprimer la bibliothèque. La bibliothèque est supprimée.
- ▶ Si vous n'avez jamais publié cette bibliothèque auparavant, un message vous demande si vous souhaitez supprimer ou conserver la bibliothèque. Cliquez sur Supprimer pour poursuivre ou sur Conserver pour conserver cette bibliothèque.

Remarque : Vous devez toujours conserver au moins une bibliothèque.

Gestion des bibliothèques publiques

Afin de réutiliser les bibliothèques locales, vous pouvez les publier et ensuite les utiliser à partir de la boîte de dialogue Gérer les bibliothèques (Ressources > Gérer les bibliothèques). [Pour plus d'informations, reportez-vous à la section Partage de bibliothèques sur p. 214.](#) Vous pouvez effectuer des tâches de base de gestion des bibliothèques publiques : import, export ou suppression d'une bibliothèque publique. Il est impossible de renommer une bibliothèque publique.

Figure 9-4
Boîte de dialogue Gérer les bibliothèques

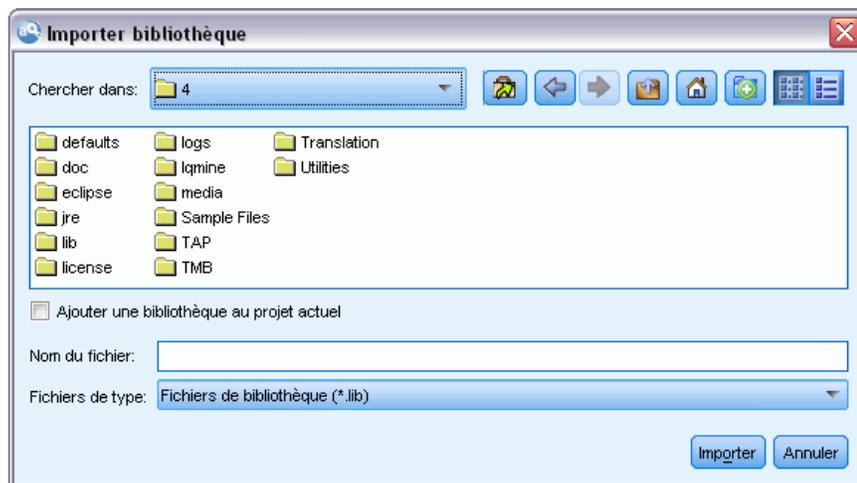


Import de bibliothèques publiques

- ▶ Dans la boîte de dialogue Gérer les bibliothèques, cliquez sur Importer. La boîte de dialogue Importer une bibliothèque apparaît.

Figure 9-5

Boîte de dialogue Importer une bibliothèque



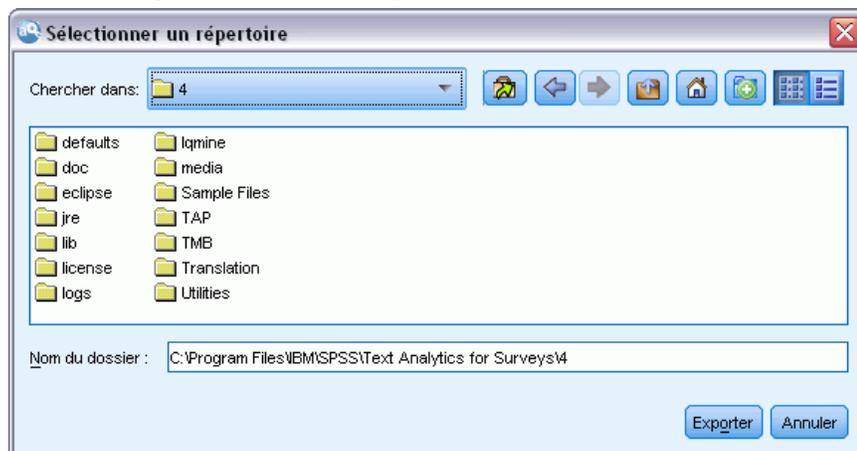
- ▶ Sélectionnez le fichier de bibliothèque (*.lib) à importer et, si vous souhaitez également ajouter cette bibliothèque localement, sélectionnez l'option Ajouter une bibliothèque au projet actuel.
- ▶ Cliquez sur Importer. La boîte de dialogue se ferme. Si une bibliothèque publique de même nom existe déjà, vous êtes invité à renommer la bibliothèque en cours d'import ou à remplacer la bibliothèque publique en cours.

Export de bibliothèques publiques

Vous pouvez exporter les bibliothèques publiques au format *.lib* afin de pouvoir les partager.

- ▶ Dans la boîte de dialogue Gérer les bibliothèques, sélectionnez dans la liste la bibliothèque à exporter.
- ▶ Cliquez sur Export. La boîte de dialogue Sélectionner un répertoire apparaît.

Figure 9-6
Boîte de dialogue Sélectionner un répertoire



- Sélectionnez le répertoire vers lequel vous souhaitez exporter et cliquez sur Exporter. La boîte de dialogue disparaît et le fichier de bibliothèque (*.lib) est exporté.

Suppression de bibliothèques publiques

Vous pouvez supprimer une bibliothèque locale sans supprimer la version publique de la bibliothèque et inversement. Cependant, si la bibliothèque est supprimée de cette boîte de dialogue, elle ne peut plus être ajoutée aux projets jusqu'à ce qu'une version locale soit à nouveau publiée.

Si vous supprimez une bibliothèque qui a été installée avec le produit, la version originale installée est restaurée.

- Dans la boîte de dialogue Gérer les bibliothèques, sélectionnez la bibliothèque à supprimer. Vous pouvez trier la liste en cliquant sur l'en-tête approprié.
- Cliquez sur Supprimer pour supprimer la bibliothèque. IBM® SPSS® Text Analytics for Surveys vérifie si la version locale de la bibliothèque est la même que la bibliothèque publique. Si c'est le cas, la bibliothèque est supprimée sans message d'alerte. Si les versions de la bibliothèque diffèrent, un message d'alerte apparaît pour vous demander si vous souhaitez conserver ou supprimer la version publique.

Partage de bibliothèques

Les bibliothèques permettent d'utiliser des ressources de manière à pouvoir les partager facilement parmi plusieurs projets. Les bibliothèques existent en deux états ou deux versions. Les bibliothèques associées à un projet donné s'appellent des **bibliothèques locales**. Lorsque vous travaillez dans un projet, vous pouvez effectuer de nombreux changements, par exemple dans la bibliothèque *Vegetables* (Légumes). Si vos changements peuvent s'avérer utiles avec d'autres données, vous pouvez rendre ces ressources disponibles en créant une **version publique** de la bibliothèque *Vegetables* (Légumes). Une bibliothèque publique, comme son nom l'indique, est mise à la disposition de n'importe quel autre projet.

Vous pouvez visualiser les bibliothèques publiques dans la boîte de dialogue Gérer les bibliothèques. Une fois que cette version de bibliothèque publique existe, vous pouvez l'ajouter aux ressources dans d'autres contextes de manière à ce que ces ressources linguistiques personnalisées puissent être partagées.

Les bibliothèques fournies sont initialement des bibliothèques publiques. Il est possible de modifier les ressources de ces bibliothèques avant de créer une version publique. Ces nouvelles versions sont ensuite accessibles dans d'autres nouveaux projets.

Lorsque vous utilisez vos bibliothèques et y apportez des modifications, les différentes versions ne sont plus synchronisées. Dans certains cas, une version locale peut être plus récente que la version publique et, dans d'autres cas, la version publique peut être plus récente que la version locale. La version publique peut contenir des changements que la version locale n'a pas et inversement. C'est le cas si la version publique a été mise à jour depuis un autre projet. Si les versions de vos bibliothèques ne sont plus synchronisées, vous pouvez rétablir la synchronisation. La synchronisation des versions de bibliothèque consiste à publier une nouvelle fois et/ou mettre à jour les bibliothèques locales.

Chaque fois que vous ouvrez ou fermez un projet, vous êtes invité à synchroniser toutes les bibliothèques qui doivent être mises à jour ou republiées. De plus, vous pouvez facilement identifier l'état de synchronisation de votre bibliothèque locale à l'aide de l'icône qui apparaît à côté du nom de la bibliothèque dans l'arborescence ou en affichant la boîte de dialogue Propriétés de bibliothèque. Vous pouvez également effectuer cette opération à tout moment en sélectionnant les options de menu correspondantes. Le tableau ci-dessous décrit les cinq états possibles et leurs icônes associées.

Table 9-1
Etats de synchronisation des bibliothèques locales

icône	Description de l'état des bibliothèques locales
	Non publiée — La bibliothèque locale n'a jamais été publiée.
	Synchronisé — Les versions locale et publique de la bibliothèque sont identiques. Cela s'applique également à la <i>bibliothèque locale</i> , qui ne peut pas être publiée car elle est censée contenir uniquement des ressources propres au projet.
	Obsolète — La version publique de la bibliothèque est plus récente que la version locale. Vous pouvez mettre à jour votre version locale avec les modifications.
	Plus récent — La version locale de la bibliothèque est plus récente que la version publique. Vous pouvez republier votre version locale pour que la version publique soit actualisée.
	Désynchronisé — Les bibliothèques locale et publique contiennent chacune des modifications que l'autre bibliothèque n'a pas. Vous devez choisir soit de mettre à jour votre bibliothèque locale, soit de la publier. Si vous la mettez à jour, vous perdez les modifications que vous avez effectuées depuis votre dernière mise à jour ou publication. Si vous optez pour la publication, vous supprimez les changements effectués dans la version publique.

Remarque : Si vous mettez systématiquement à jour vos bibliothèques lorsque vous ouvrez ou publiez une bibliothèque lorsque vous fermez un projet, vous avez moins de risques d'avoir des bibliothèques qui ne sont pas synchronisées.

Vous pouvez republier une bibliothèque si vous pensez que les changements apportés peuvent être utiles à d'autres projets pouvant également contenir cette bibliothèque. Par conséquent, si vos changements peuvent être utiles à d'autres projets, vous pouvez mettre à jour les versions locales dans ces projets. De cette manière, vous pouvez créer des projets pour chaque contexte

ou domaine qui s'applique à vos données : vous pouvez créer de nouvelles bibliothèques et/ou ajouter des bibliothèques publiques à vos ressources.

Lorsqu'une version publique de bibliothèque est partagée, les risques de différence entre les versions locale et publique sont plus importants. Chaque fois que vous ouvrez ou publiez lorsque vous fermez un projet, un message apparaît pour vous permettre de publier et/ou de mettre à jour les bibliothèques dont la version n'est pas synchronisée avec celles contenues dans la boîte de dialogue Gérer les bibliothèques. Si la version publique de la bibliothèque est plus récente que la version locale, une boîte de dialogue vous demande si vous souhaitez procéder à une mise à jour. Vous pouvez choisir de conserver la version locale en l'état au lieu de la mettre à jour pour qu'elle contienne les changements de la version publique. Vous pouvez également décider de la mettre à jour pour qu'elle reflète les changements insérés dans la version publique.

Publication de bibliothèques

Si vous n'avez jamais publié de bibliothèques, sachez que la publication consiste à créer une copie publique de votre bibliothèque locale dans la base de données. Si vous republiez une bibliothèque, le contenu de la bibliothèque locale remplace le contenu de la version publique existante. Après une nouvelle publication, vous pouvez mettre à jour cette bibliothèque dans n'importe quel projet de manière à ce que leurs versions locales soient synchronisées avec la version publique. Même si vous pouvez publier une bibliothèque, une version locale est toujours stockée dans le projet.

Important ! Si vous apportez des modifications à votre bibliothèque locale et si, dans le même temps, la version publique de la bibliothèque est également modifiée, votre bibliothèque est considérée comme désynchronisée. Nous vous recommandons de commencer par mettre à jour la version locale avec les modifications de la version publique, d'effectuer tous les changements nécessaires et ensuite, de publier à nouveau la version locale pour que les deux versions soient identiques. Si vous effectuez des changements et que vous commencez par une publication, vous remplacerez les changements apportés à la version publique.

Figure 9-7
Boîte de dialogue Publier les bibliothèques



Pour publier des bibliothèques locales dans la base de données

- ▶ A partir des menus, sélectionnez Ressources > Publier les bibliothèques. La boîte de dialogue Publier les bibliothèques apparaît. Toutes les bibliothèques devant être publiées sont sélectionnées par défaut.
- ▶ Cochez la case située à gauche de chaque bibliothèque à publier ou republier.
- ▶ Cliquez sur Publier pour publier les bibliothèques dans la base de données Gérer les bibliothèques.

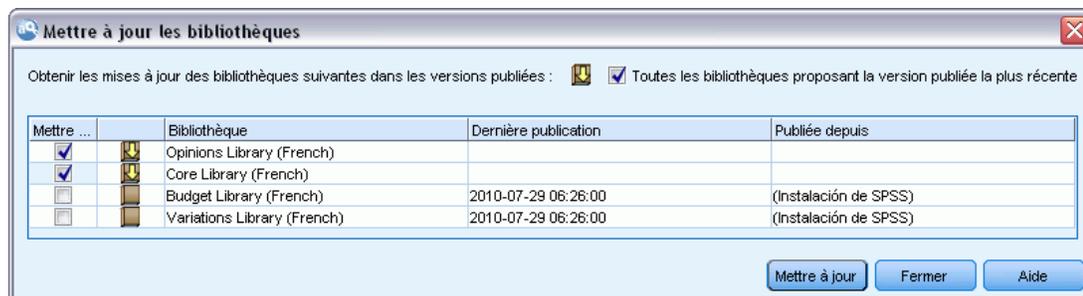
Mise à jour des bibliothèques

Chaque fois que vous ouvrez ou publiez une bibliothèque lorsque vous fermez un projet, vous pouvez mettre à jour ou publier toutes les bibliothèques qui ne sont plus synchronisées avec les versions publiques. Si la version publique de la bibliothèque est plus récente que la version locale, une boîte de dialogue vous demande si vous souhaitez mettre à jour la bibliothèque. Vous pouvez choisir soit de conserver la version locale au lieu de la mettre à jour avec le contenu de la version publique, soit de remplacer la version locale par la publique. Si la version publique d'une bibliothèque est plus récente que la version locale, vous pouvez mettre à jour la version locale pour synchroniser son contenu avec celui de la version publique. La mise à jour consiste à incorporer dans votre version locale les changements trouvés dans la version publique.

Remarque : Si vous mettez systématiquement à jour vos bibliothèques lorsque vous ouvrez ou publiez une bibliothèque lorsque vous fermez un projet, vous avez moins de risques d'avoir des bibliothèques qui ne sont pas synchronisées. [Pour plus d'informations, reportez-vous à la section Partage de bibliothèques sur p. 214.](#)

Figure 9-8

Boîte de dialogue Mettre à jour les bibliothèques



Pour mettre à jour les bibliothèques locales

- ▶ A partir des menus, sélectionnez Ressources > Mettre à jour les bibliothèques. La boîte de dialogue Mettre à jour les bibliothèques apparaît. Toutes les bibliothèques devant être mises à jour sont sélectionnées par défaut.
- ▶ Cochez la case située à gauche de chaque bibliothèque à publier ou republier.
- ▶ Cliquez sur Mettre à jour pour mettre à jour les bibliothèques locales.

Résolution des conflits

Conflits entre bibliothèque locale et bibliothèque publique

Chaque fois que vous ouvrez un projet, IBM® SPSS® Text Analytics for Surveys compare les bibliothèques locales et celles répertoriées dans la boîte de dialogue Gérer les bibliothèques. Si une bibliothèque locale de votre projet n'est pas synchronisée avec les versions publiées, la boîte de dialogue Avertissement de synchronisation de bibliothèques apparaît. Vous pouvez choisir parmi les options suivantes afin de sélectionner les versions des bibliothèques que vous souhaitez utiliser :

- **Toutes les bibliothèques locales vers un fichier.** Cette option conserve toutes vos bibliothèques en l'état. Vous pouvez toujours les republier ou les mettre à jour ultérieurement.
- **Toutes les bibliothèques publiées sur cet ordinateur.** Cette option remplace les bibliothèques locales affichées par les versions de la base de données.
- **Toutes les bibliothèques les plus récentes.** Cette option remplace toutes les anciennes bibliothèques locales par les versions publiques plus récentes contenues dans la base de données.
- **Autre.** Cette option permet de sélectionner manuellement les versions de votre choix en les sélectionnant dans le tableau.

Conflits de termes forcés

Lorsque vous ajoutez une bibliothèque publique ou mettez à jour une bibliothèque locale, vous pouvez détecter des conflits et des doublons entre les termes et les types de cette bibliothèque et ceux des autres bibliothèques stockées dans vos ressources. Si cela se produit, vous êtes invité à vérifier, dans la boîte de dialogue Editer les termes forcés, les résolutions de conflit proposées ou à les modifier avant de réaliser l'opération. [Pour plus d'informations, reportez-vous à la section Ajout des termes forcés dans le chapitre 10 sur p. 228.](#)

Figure 9-9
Boîte de dialogue Editer les termes forcés



La boîte de dialogue Editer les termes forcés contient toutes les paires de termes et types contradictoires. L'alternance des couleurs de l'arrière-plan permet de distinguer chaque paire contradictoire. Ces couleurs peuvent être modifiées dans la boîte de dialogue Options. [Pour plus](#)

d'informations, reportez-vous à la section [Options : Onglet Affichage](#) dans le chapitre 2 sur p. 21. La boîte de dialogue Éditer les termes forcés contient deux onglets :

- **Doublons.** Cet onglet contient les doublons rencontrés dans les bibliothèques. Si une icône représentant une punaise apparaît après un terme, cela signifie que l'occurrence du terme a été forcée. Si une icône X apparaît en noir, cela signifie que cette occurrence du terme sera ignorée pendant l'extraction, car elle a été imposée ailleurs.
- **Défini par l'utilisateur.** Cet onglet contient la liste de tous les termes qui ont été forcés manuellement dans le panneau des termes de la déclaration de types et non pas via les conflits.

Remarque : La boîte de dialogue Editer les termes forcés apparaît après l'ajout ou la mise à jour d'une bibliothèque. Si vous quittez la boîte de dialogue, vous n'annulez pas la mise à jour ou l'ajout de la bibliothèque.

Pour résoudre des conflits

- ▶ Dans la boîte de dialogue Editer les termes forcés, sélectionnez le bouton radio de la colonne Utiliser du terme que vous souhaitez forcer.
- ▶ Lorsque vous avez terminé, cliquez sur OK pour appliquer les termes forcés et fermer la boîte de dialogue. Si vous cliquez sur Annuler, vous annulez les changements effectués dans cette boîte de dialogue.

A propos des dictionnaires de bibliothèque

Les ressources servant à extraire des données textuelles sont stockées sous la forme de modèles et de bibliothèques. Une bibliothèque peut être constituée de trois dictionnaires.

- La **déclaration de types** comprend un ensemble de termes regroupés sous une même étiquette ou sous un même nom de type. Lorsque le moteur d'extraction lit vos données textuelles, il compare les mots rencontrés dans le texte aux termes définis dans vos déclarations de types. Au cours de l'extraction, les formes fléchies des termes et synonymes d'un type sont regroupées sous un terme cible nommé concept. Les concepts extraits sont affectés à la déclaration de types dans laquelle ils apparaissent en tant que termes. Vous pouvez gérer vos déclarations de types dans les panneaux situés en haut à gauche et au centre de l'éditeur — l'arborescence de la bibliothèque et le panneau des termes. [Pour plus d'informations, reportez-vous à la section Déclarations de types sur p. 220.](#)
- Le **dictionnaire de substitutions** contient un ensemble de mots définis comme synonymes ou comme éléments optionnels qui sont utilisés pour regrouper des termes similaires sous un seul terme cible, appelé concept dans les résultats d'extraction finaux. Vous pouvez gérer vos dictionnaires de substitutions dans le panneau inférieur gauche de l'éditeur à l'aide des onglets Synonymes et Optionnels. [Pour plus d'informations, reportez-vous à la section Dictionnaires des substitutions/synonymes sur p. 231.](#)
- Le **dictionnaire d'exclusions** comprend un ensemble de termes et de types qui seront supprimés des résultats finaux de l'extraction. Vous pouvez gérer vos dictionnaires d'exclusion dans le panneau le plus à droite de l'éditeur. [Pour plus d'informations, reportez-vous à la section Dictionnaires d'exclusions sur p. 236.](#)

[Pour plus d'informations, reportez-vous à la section Utilisation des bibliothèques dans le chapitre 9 sur p. 206.](#)

Déclarations de types

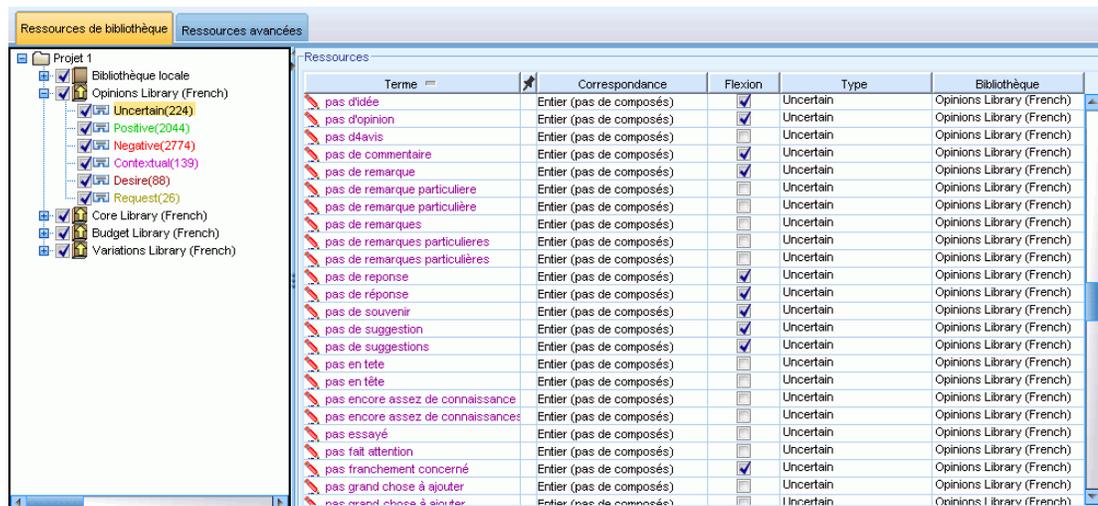
Une **déclaration de types** est constituée d'un nom de type, d'une étiquette et d'une liste de termes. La gestion des déclarations de types s'effectue dans les panneaux supérieurs gauche et central de l'onglet Ressources de bibliothèque dans l'éditeur. Vous pouvez accéder à cette vue avec *Vue > Editeur de ressources* dans les menus.

Lorsque le moteur du programme d'extraction lit vos données textuelles, il compare les mots rencontrés dans le texte aux termes définis dans vos déclarations de types. Les termes sont des mots ou des expressions dans les déclarations de types dans vos ressources linguistiques.

Lorsqu'un mot correspond à un terme, il est attribué au nom de type de ce terme. Lorsque les ressources sont lues au cours de l'extraction, les termes trouvés dans le texte traversent ensuite plusieurs étapes de traitement avant de devenir des concepts dans le panneau Résultats d'extraction. Si plusieurs termes appartenant à la même déclaration de types sont déterminés comme étant des synonymes par le moteur du programme d'extraction, ils sont regroupés sous le

terme le plus fréquent et nommé un **concept** dans le panneau Résultats d'extraction. Par exemple, si les termes *question* et *requête* peuvent apparaître dans le nom de concept *question* à la fin.

Figure 10-1
Arborescence de bibliothèques et panneau de termes



La liste des déclarations de types est affichée dans le panneau qui contient l'arborescence de bibliothèques, à gauche. Le contenu de chaque déclaration de types apparaît dans le panneau central. Les déclarations de types offrent bien plus qu'une simple liste de termes. La manière dont les mots de vos données textuelles sont mis en correspondance avec les concepts affectés aux déclarations de types est déterminée par l'option de mise en correspondance. Une **option de mise en correspondance** définit le positionnement d'un terme par rapport à une expression ou un mot candidat dans les données textuelles. [Pour plus d'informations, reportez-vous à la section Ajout de termes sur p. 224.](#)

Par ailleurs, vous pouvez étendre les termes de votre déclaration de types en indiquant si vous souhaitez générer et ajouter automatiquement des formes fléchies aux termes de la déclaration. En générant les formes fléchies, vous ajoutez automatiquement les formes plurielles de termes au singulier, les formes au singulier de termes au pluriel et les adjectifs dans la déclaration de types. [Pour plus d'informations, reportez-vous à la section Ajout de termes sur p. 224.](#)

Remarque : Les concepts qui ne se trouvent dans aucune déclaration de type mais sont extraits du texte sont automatiquement dotés du type <Unknown>.

Types intégrés

IBM® SPSS® Text Analytics for Surveys est fourni avec un ensemble de ressources linguistiques sous la forme de bibliothèques et de ressources compilées. Les bibliothèques fournies contiennent un ensemble de déclarations de types intégrées parmi lesquelles <Location>, <Organization>, <Person> et <Product>.

Ces déclarations de types sont utilisées par le moteur du programme d'extraction pour affecter des types aux concepts qu'il extrait (par exemple, le type <Location> est affecté au concept `paris`). Bien qu'un grand nombre de termes aient été définis dans les déclarations de types intégrées, ceux-ci ne couvrent pas toutes les possibilités. Par conséquent, vous pouvez les compléter ou en créer qui vous soient propres. Pour obtenir la description du contenu d'une déclaration de types fournie en particulier, reportez-vous à l'annotation dans la boîte de dialogue Propriétés de type. Sélectionnez le type dans l'arborescence et choisissez Edition > Propriétés dans le menu contextuel.

Remarque : Outre les bibliothèques fournies, les ressources compilées (également utilisées par le moteur d'extraction) contiennent un grand nombre de définitions complémentaires aux déclarations de types intégrées, mais leur contenu n'est pas visible dans le produit. Vous pouvez toutefois forcer n'importe quel autre dictionnaire à accepter un terme classé par type par les déclarations compilées. [Pour plus d'informations, reportez-vous à la section Ajout des termes forcés sur p. 228.](#)

Création de types

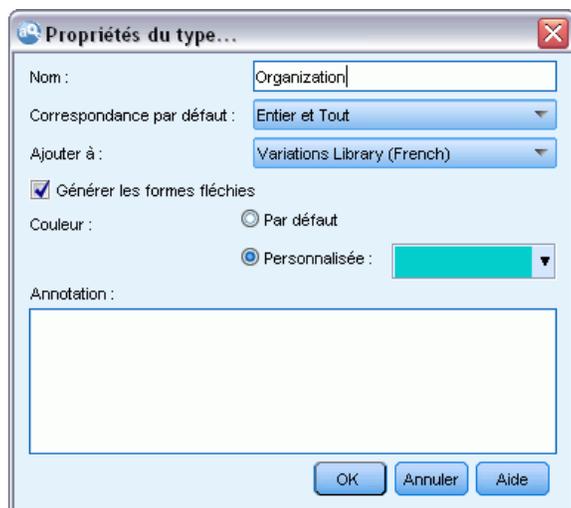
Vous pouvez créer des déclarations de types pour faciliter le regroupement de termes similaires. Quand des termes qui figurent dans cette déclaration sont découverts au cours du processus d'extraction, ils sont affectés à ce nom de type et extraits sous un nom de concept. Lorsque vous créez une bibliothèque, une bibliothèque de types vide est toujours incluse, afin de vous permettre de commencer immédiatement à entrer des termes.

Si vous analysez un texte concernant des aliments et souhaitez regrouper les termes relatifs aux légumes, vous pouvez créer votre propre déclaration de types <Légumes>. Vous pouvez ensuite y ajouter des termes tels que `carotte`, `brocoli` et `épinard` si vous estimez qu'il s'agit de termes importants qui vont apparaître dans le texte. Ensuite, au cours de l'extraction, si l'un de ces termes est identifié, il est extrait en tant que concept et affecté au type <Vegetables> (Légumes).

Il n'est pas nécessaire de définir toutes les formes d'un mot ou d'une expression, car vous pouvez choisir de générer toutes les formes fléchies des termes. Lorsque vous choisissez cette option, le moteur d'extraction reconnaît automatiquement les formes au singulier ou au pluriel des termes parmi les autres formes et les associe à ce type. Cette option s'avère particulièrement utile quand votre type contient principalement des noms, dans la mesure où il est peu probable que vous cherchiez à obtenir les formes fléchies de verbes ou d'adjectifs.

Important ! Nous vous recommandons vivement de ne pas créer de nouveaux types dans Opinions Library, car ils ne seront pas pris en compte au cours du traitement. Le contenu d'Opinions Library est traité différemment des autres bibliothèques, car il est utilisé pour produire des patrons. Travaillez plutôt dans les types existant déjà dans cette bibliothèque ou ajoutez de nouveaux types à une autre bibliothèque dans votre projet.

Figure 10-2
Boîte de dialogue Propriétés de type



Nom. Nom que vous attribuez à la déclaration de types que vous créez. Nous vous recommandons de ne pas utiliser d'espaces dans les noms de types, en particulier si deux noms de types ou plus commencent par le même mot.

Mise en correspondance par défaut. L'attribut Mise en correspondance par défaut indique au moteur d'extraction la manière dont il doit mettre en correspondance ce terme et les données textuelles. Dès que vous ajoutez un terme à cette déclaration de types, cet attribut de mise en correspondance lui est associé automatiquement. Vous avez toujours la possibilité de modifier manuellement le choix de correspondance dans la liste des termes. Vous avez le choix entre les options suivantes : Terme entier, Début, Fin, Tout, Début ou Fin, Entier et Début, Entier et Fin et Entier et (Début ou Fin) et Entier (pas de composés). [Pour plus d'informations, reportez-vous à la section Ajout de termes sur p. 224.](#)

Ajouter à. Ce champ indique la bibliothèque dans laquelle vous allez créer votre nouvelle déclaration de types.

Générer les formes fléchies. Cette option indique au moteur d'extraction d'utiliser la fonction de morphologie grammaticale pour collecter et regrouper les formes similaires des termes que vous ajoutez à cette déclaration (le singulier et le pluriel d'un terme, par exemple). Cette fonction est particulièrement utile lorsque votre type contient surtout des noms. Quand vous sélectionnez cette option, elle est appliquée par défaut à tous les nouveaux termes ajoutés au type, mais vous pouvez la modifier manuellement dans la liste.

Couleur. Ce champ vous permet de distinguer les résultats de ce type dans l'interface. Si vous sélectionnez Par défaut, la couleur par défaut du type est également utilisée pour cette déclaration de types. La couleur par défaut se configure dans la boîte de dialogue des options. [Pour plus d'informations, reportez-vous à la section Options : Onglet Affichage dans le chapitre 2 sur p. 21.](#) Si vous sélectionnez Personnalisé, sélectionnez une couleur dans la liste déroulante.

Annotations. Ce champ est facultatif et sert à entrer des commentaires ou des descriptions.

Pour créer une déclaration de types

- ▶ Sélectionnez la bibliothèque dans laquelle vous souhaitez créer une déclaration de types.
- ▶ A partir des menus, sélectionnez Outils > Nouveau type. La boîte de dialogue Propriétés de type apparaît.
- ▶ Entrez le nom de votre déclaration de types dans la zone de texte Nom et choisissez les options souhaitées.
- ▶ Cliquez sur OK pour créer la déclaration de types. Le nouveau type apparaît dans l'arborescence de bibliothèques et s'affiche dans le panneau central. Vous pouvez commencer immédiatement à ajouter des termes. Pour plus d'informations, reportez-vous à [Ajout de termes](#).

Remarque : Ces instructions indiquent la façon d'effectuer des modifications dans la vue Editeur de ressources. N'oubliez pas que vous pouvez également réaliser ce type de réglage directement à partir du panneau Résultats d'extraction ou du panneau Données. [Pour plus d'informations, reportez-vous à la section Affinage des résultats de l'extraction dans le chapitre 5 sur p. 88.](#)

Ajout de termes

L'arborescence de bibliothèques contient les bibliothèques. Il peut être développé pour présenter les déclarations de types qu'elles contiennent. Dans le panneau central, une liste affiche les termes de la bibliothèque ou de la déclaration de types sélectionnée, en fonction de l'élément sélectionné dans l'arborescence.

Figure 10-3
Panneau des termes

Terme	Correspondance	Flexion	Type	Bibliothèque
paramétrable	Entier (pas de composés)	<input checked="" type="checkbox"/>	Positive	Opinions Library (French)
paramétrable	Entier (pas de composés)	<input checked="" type="checkbox"/>	Positive	Opinions Library (French)
parasiaque	Entier (pas de composés)	<input checked="" type="checkbox"/>	Positive	Opinions Library (French)
pardonnable	Entier (pas de composés)	<input checked="" type="checkbox"/>	Positive	Opinions Library (French)
parfait état	Entier (pas de composés)	<input type="checkbox"/>	Positive	Opinions Library (French)
parfait état	Entier (pas de composés)	<input type="checkbox"/>	Positive	Opinions Library (French)
parfait	Entier (pas de composés)	<input checked="" type="checkbox"/>	Positive	Opinions Library (French)
parfait état	Entier (pas de composés)	<input type="checkbox"/>	Positive	Opinions Library (French)
parfait sur tous les points	Entier (pas de composés)	<input type="checkbox"/>	Positive	Opinions Library (French)
parfait état	Entier (pas de composés)	<input type="checkbox"/>	Positive	Opinions Library (French)
parfaite état	Entier (pas de composés)	<input type="checkbox"/>	Positive	Opinions Library (French)
parfaite état	Entier (pas de composés)	<input type="checkbox"/>	Positive	Opinions Library (French)
parfaitement	Début	<input type="checkbox"/>	Positive	Opinions Library (French)
partant	Entier (pas de composés)	<input checked="" type="checkbox"/>	Positive	Opinions Library (French)
partisan	Entier (pas de composés)	<input checked="" type="checkbox"/>	Positive	Opinions Library (French)
partisant	Entier (pas de composés)	<input checked="" type="checkbox"/>	Positive	Opinions Library (French)
pas a m en plaindre	Entier (pas de composés)	<input checked="" type="checkbox"/>	Positive	Opinions Library (French)
pas a m'en plaindre	Entier (pas de composés)	<input checked="" type="checkbox"/>	Positive	Opinions Library (French)
pas a me plaindre	Entier (pas de composés)	<input checked="" type="checkbox"/>	Positive	Opinions Library (French)
pas cher	Entier (pas de composés)	<input checked="" type="checkbox"/>	Positive	Opinions Library (French)
pas de blems	Entier (pas de composés)	<input checked="" type="checkbox"/>	Positive	Opinions Library (French)
pas de crainte	Entier (pas de composés)	<input type="checkbox"/>	Positive	Opinions Library (French)
pas de pb	Entier (pas de composés)	<input checked="" type="checkbox"/>	Positive	Opinions Library (French)
pas de proble	Entier (pas de composés)	<input type="checkbox"/>	Positive	Opinions Library (French)
pas de probleme	Entier (pas de composés)	<input checked="" type="checkbox"/>	Positive	Opinions Library (French)
pas de problemes	Entier (pas de composés)	<input type="checkbox"/>	Positive	Opinions Library (French)
pas de problème	Entier (pas de composés)	<input type="checkbox"/>	Positive	Opinions Library (French)

Dans Editeur de ressources, vous pouvez ajouter des termes directement à une déclaration de types : soit directement dans le panneau de termes, soit par l'intermédiaire de la boîte de dialogue Ajouter des nouveaux termes. Les termes que vous ajoutez peuvent être des mots simples ou composés. La ligne vide, située en haut de la liste, vous permet d'ajouter de nouveaux termes à tout moment.

Remarque : Ces instructions indiquent la façon d'effectuer des modifications dans la vue Editeur de ressources. N'oubliez pas que vous pouvez également réaliser ce type de réglage directement à partir du panneau Résultats d'extraction ou du panneau Données. [Pour plus d'informations, reportez-vous à la section Affinage des résultats de l'extraction dans le chapitre 5 sur p. 88.](#)

Colonne Terme

Dans cette colonne, entrez des mots simples ou composés dans la cellule. La couleur d'affichage du terme dépend de la couleur du type dans lequel le terme est enregistré ou a été imposé. Vous pouvez changer les couleurs des types dans la boîte de dialogue Propriétés de type. [Pour plus d'informations, reportez-vous à la section Création de types sur p. 222.](#)

Colonne Forcer

Dans cette colonne, sélectionnez cette cellule pour afficher une icône de punaise qui indique au moteur d'extraction d'ignorer les autres occurrences de ce même terme dans les autres bibliothèques. [Pour plus d'informations, reportez-vous à la section Ajout des termes forcés sur p. 228.](#)

Colonne Correspondance

Dans cette colonne, sélectionnez une option de mise en correspondance pour indiquer au moteur d'extraction comment il doit faire correspondre ce terme aux données textuelles. Consultez le tableau pour voir des exemples. Pour changer la valeur par défaut, éditez les propriétés du type. [Pour plus d'informations, reportez-vous à la section Création de types sur p. 222.](#) Dans les menus, choisissez Edition > Modifier la mise en correspondance. Ci-dessous figurent les options de mise en correspondance de base, car des combinaisons de celles-ci sont également possibles :

- Début. Si le terme dans la déclaration correspond au début d'un concept extrait du texte, ce type est attribué. Par exemple, si vous entrez *tarte*, le terme *tarte aux pommes* est renvoyé.
- Fin. Si le terme dans la déclaration correspond à la fin d'un concept extrait du texte, ce type est attribué. Par exemple, si vous entrez *tarte*, *moule à tarte* est renvoyé.
- N'importe. Si le terme dans la déclaration correspond à toute partie d'un concept extrait du texte, ce type est attribué. Par exemple, si vous entrez *tarte*, l'option *Tout classe tarte aux pommes, moule à tarte et moule pour tarte aux pommes* sous un même type.
- Terme entier. Ce type est attribué si l'intégralité du concept extrait du texte correspond exactement avec le terme de la déclaration. L'ajout d'un terme comme *Terme entier*, *Entier et début*, *Entier et fin*, *Entier et n'importe* ou *Entier* (pas de composés) force l'extraction d'un terme.

De plus, comme le type <Person> extrait uniquement les noms en deux parties, comme *edith piaf* ou *mohandas gandhi*, vous voudrez peut-être ajouter explicitement les prénoms à cette déclaration de types si vous essayez d'extraire un prénom lorsqu'aucun nom de famille n'est

mentionné. Par exemple, si vous voulez extraire toutes les instances du nom *edith*, vous devez ajouter *edith* au type `<Person>` à l'aide des options Terme entier ou Entier et début.

- Entier (pas de composés). Si le concept complet extrait du texte correspond au terme exact dans la déclaration, ce type est attribué et l'extraction s'arrête afin d'empêcher l'extraction d'associer le terme aux composés plus longs. Par exemple, si vous saisissez `pomme`, l'option Entier (pas de composé) produira `pomme` et n'extraira pas le composé `jus de pomme`, à moins que ce ne soit forcé ailleurs.

Dans le tableau suivant, on peut supposer que le terme `pomme` est dans une déclaration de types. En fonction de l'option de mise en correspondance, ce tableau indique les concepts qui seraient extraits et entrés s'ils étaient trouvés dans le texte.

Table 10-1
Exemples de correspondance

Options de mise en correspondance pour le terme :  pommes	Concepts extraits			
	pommes	tarte pommes	pommesmûres	tarte aux pommes maison
Terme entier	✓			
Démarrer		✓		
Fin			✓	
Début ou fin		✓	✓	
Entier et début	✓	✓		
Entier et fin	✓		✓	
Entier et (début ou fin)	✓	✓	✓	
N'importe		✓	✓	✓
Entier et n'importe	✓	✓	✓	✓
Entier (pas de composés)	✓	<i>jamais extrait</i>	<i>jamais extrait</i>	<i>jamais extrait</i>

Colonne Flexion

Dans cette colonne, sélectionnez si le moteur du programme d'extraction doit générer les formes fléchies de ce terme au cours de l'extraction afin qu'elles soient regroupées. La valeur par défaut de cette colonne est définie dans les propriétés du type, mais vous pouvez modifier cette option au cas par cas directement dans la colonne. Dans les menus, sélectionnez Edition > Modifier la flexion.

Colonne Type

Dans cette colonne, sélectionnez une déclaration de types dans la liste déroulante. La liste de types est filtrée en fonction de votre sélection dans l'arborescence de la bibliothèque. Le premier type de la liste est toujours celui sélectionné par défaut dans l'arborescence de la bibliothèque. Dans les menus, sélectionnez Edition > Modifier le type.

Colonne Bibliothèque

Dans cette colonne, la bibliothèque dans laquelle votre terme est stocké s'affiche. Pour transférer un terme dans un autre type de l'arborescence, faites-le glisser et déposez-le sur l'autre bibliothèque.

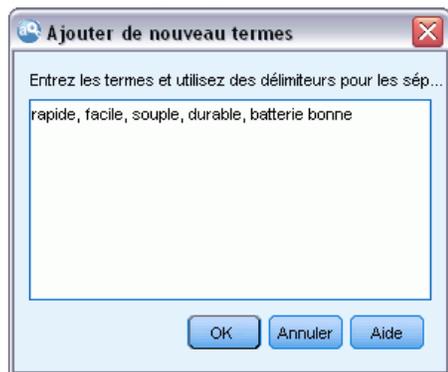
Pour ajouter un terme simple à une déclaration de types

- ▶ Dans l'arborescence de la bibliothèque, sélectionnez la déclaration de types à laquelle vous voulez ajouter le terme.
- ▶ Dans la liste de termes du panneau central, entrez votre terme dans la première cellule vide disponible et définissez les options souhaitées pour ce terme.

Pour ajouter plusieurs termes à une déclaration de types

- ▶ Dans l'arborescence de la bibliothèque, sélectionnez la déclaration de types à laquelle vous voulez ajouter des termes.
- ▶ A partir des menus, sélectionnez Outils> Nouveaux termes. La boîte de dialogue Ajouter des nouveaux termes apparaît.

Figure 10-4
Boîte de dialogue Ajouter des nouveaux termes



- ▶ Entrez les termes que vous souhaitez ajouter à la déclaration de types sélectionnée. Pour ce faire, saisissez les termes au clavier, ou copiez et collez un ensemble de termes. Si vous entrez plusieurs termes, séparez-les au moyen du séparateur défini dans la boîte de dialogue Options ou ajoutez un terme par nouvelle ligne. [Pour plus d'informations, reportez-vous à la section Définition des options dans le chapitre 2 sur p. 19.](#)
- ▶ Cliquez sur OK pour ajouter les termes à la déclaration. L'option de mise en correspondance est définie automatiquement sur la valeur par défaut de cette bibliothèque de types. La boîte de dialogue se ferme et les nouveaux termes apparaissent dans la déclaration.

Ajout des termes forcés

Si vous souhaitez affecter un terme à un type particulier, vous pouvez l'ajouter à la déclaration de types correspondante. Toutefois, si plusieurs termes ont le même nom, le moteur d'extraction doit savoir quel type utiliser. Par conséquent, vous êtes invité à sélectionner le type à utiliser. Cette opération est appelée **ajout des termes forcés** un terme dans un type. Cette option est particulièrement utile lorsque vous remplacez le type attribué d'une déclaration compilée (interne, non modifiable). En général, nous recommandons d'éviter les termes doubles.

L'ajout des termes forcés ne *supprime* pas les autres occurrences du terme, mais elles sont ignorées par le moteur d'extraction. Vous pouvez ensuite désigner l'occurrence qui doit être utilisée en activant ou en désactivant l'ajout des termes forcés. Vous pouvez également imposer un terme dans une déclaration de types lorsque vous ajoutez une bibliothèque publique ou la mettez à jour.

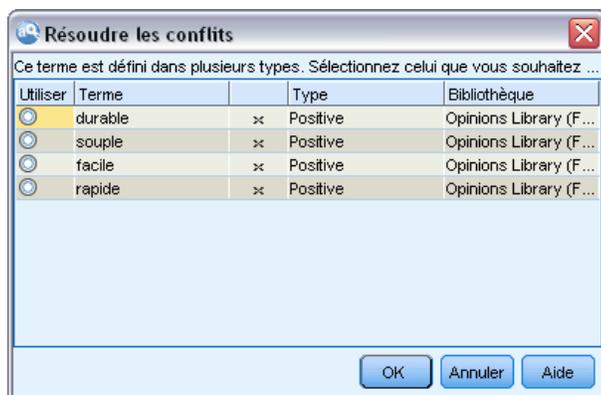
Figure 10-5
Icônes de statut d'imposition

Terme	✦	Correspondance	Flexion	Type	Bibliothèque
demitarif	✦	Terme entier	<input checked="" type="checkbox"/>	Budget	Budget Library (French)
facture	✦	Entier et (début ou fin)	<input checked="" type="checkbox"/>	Budget	Budget Library (French)
moitié prix	✦	Terme entier	<input checked="" type="checkbox"/>	Budget	Budget Library (French)
monnaie	✦	Entier et (début ou fin)	<input checked="" type="checkbox"/>	Budget	Budget Library (French)
moyens de paiement	✦	Terme entier	<input checked="" type="checkbox"/>	Budget	Budget Library (French)
paie	✦	Entier et (début ou fin)	<input checked="" type="checkbox"/>	Budget	Budget Library (French)
paiement	✦	Entier et (début ou fin)	<input checked="" type="checkbox"/>	Budget	Budget Library (French)
paiement	✦	Entier et (début ou fin)	<input checked="" type="checkbox"/>	Budget	Budget Library (French)
paye	✦	Entier et (début ou fin)	<input checked="" type="checkbox"/>	Budget	Budget Library (French)
payement	✦	Entier et (début ou fin)	<input checked="" type="checkbox"/>	Budget	Budget Library (French)
prime	✦	Entier et (début ou fin)	<input checked="" type="checkbox"/>	Budget	Budget Library (French)
prix	✦	Entier et (début ou fin)	<input type="checkbox"/>	Budget	Budget Library (French)
promotion	✦	Entier et (début ou fin)	<input checked="" type="checkbox"/>	Budget	Budget Library (French)
qualité-prix	✦	× Entier et (début ou fin)	<input type="checkbox"/>	Budget	Budget Library (French)
qualité/prix	✦	Entier et (début ou fin)	<input type="checkbox"/>	Budget	Budget Library (French)
rabais	✦	× Entier et (début ou fin)	<input type="checkbox"/>	Budget	Budget Library (French)

Vous pouvez voir les termes qui sont imposés ou ignorés dans la colonne Forcer, en deuxième position dans le panneau des termes. Si une icône de punaise apparaît, cela signifie que cette occurrence du terme a été imposée. Si une icône X apparaît en noir, cela signifie que cette occurrence du terme sera ignorée pendant l'extraction, car elle a été imposée ailleurs. Par ailleurs, quand vous imposez un terme, celui-ci apparaît dans la couleur du type dans lequel il a été imposé. Cela signifie que si vous avez imposé dans Type 1 un terme présent à la fois dans Type 1 et Type 2, il apparaît dans la fenêtre dans la couleur définie pour Type 1.

Vous pouvez double-cliquer sur l'icône pour modifier le statut. Si le terme apparaît ailleurs, la boîte de dialogue Résoudre les conflits qui apparaît vous permet de sélectionner l'occurrence à utiliser.

Figure 10-6
Boîte de dialogue Résoudre les conflits



Attribution de nouveaux noms aux types

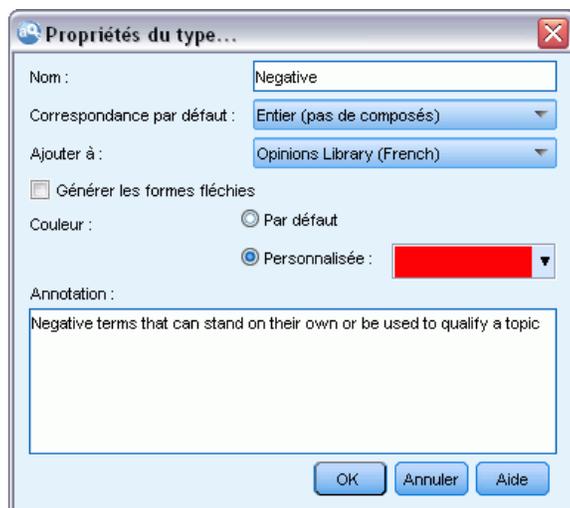
Vous pouvez renommer une déclaration de types ou configurer ses autres paramètres en éditant les propriétés du type.

Important ! Nous vous recommandons de ne pas utiliser d'espaces dans les noms de types, en particulier si deux noms de types ou plus commencent par le même mot. Nous vous recommandons également de ne pas renommer les types dans les bibliothèques Principale ou Opinions ni de modifier les attributs de correspondance par défaut.

Pour renommer un type

- ▶ Dans l'arborescence de bibliothèques, sélectionnez la déclaration de types que vous souhaitez renommer.
- ▶ Cliquez avec le bouton droit de la souris, puis choisissez Propriétés de type dans le menu contextuel. La boîte de dialogue Propriétés de type apparaît.

Figure 10-7
Boîte de dialogue Propriétés de type



- ▶ Entrez le nouveau nom de votre déclaration de types dans la zone de texte Nom.
- ▶ Cliquez sur OK pour accepter le nouveau nom. Le nouveau nom du type apparaît dans l'arborescence de bibliothèques.

Déplacement de types

Vous pouvez faire glisser une déclaration de types vers un autre emplacement d'une même bibliothèque ou vers une autre bibliothèque de l'arborescence.

Remarque : Nous vous recommandons de ne pas déplacer les types intégrés.

Pour réorganiser un type au sein d'une même bibliothèque

- ▶ Dans l'arborescence de bibliothèques, sélectionnez la déclaration de types à déplacer.
- ▶ Dans les menus, choisissez Edition > Monter d'un niveau pour remonter la déclaration de types d'une position dans l'arborescence de la bibliothèque ou Edition > Descendre d'un niveau pour le descendre d'une position.

Pour déplacer un type vers une autre bibliothèque

- ▶ Dans l'arborescence de bibliothèques, sélectionnez la déclaration de types à déplacer.
- ▶ Cliquez avec le bouton droit de la souris, puis choisissez Propriétés de type dans le menu contextuel. La boîte de dialogue Propriétés de type apparaît. (Vous pouvez également faire glisser et déposer le type dans une autre bibliothèque).
- ▶ Dans la zone de liste Ajouter à, sélectionnez la bibliothèque dans laquelle vous voulez déplacer la déclaration de types.

- ▶ Cliquez sur OK. La boîte de dialogue se referme et le type figure maintenant dans la bibliothèque sélectionnée.

Désactiver et supprimer des types

Si vous souhaitez supprimer temporairement une déclaration de types, vous pouvez la désactiver en décochant la case située à gauche de son nom dans l'arborescence de bibliothèques. Vous indiquez ainsi que vous souhaitez conserver la déclaration dans votre bibliothèque, mais en ignorer le contenu pendant la recherche de conflits et le processus d'extraction.

Vous pouvez aussi supprimer définitivement des déclarations de types d'une bibliothèque.

Remarque : Nous vous recommandons de ne pas supprimer les types intégrés dans les bibliothèques Principale ou Opinions. Nous vous recommandons plutôt de les désactiver.

Pour désactiver une déclaration de types

- ▶ Dans l'arborescence de bibliothèques, sélectionnez la déclaration de types à désactiver.
- ▶ Cliquez sur la barre d'espace. La coche disparaît de la case figurant à gauche du nom du type.

Pour supprimer une déclaration de types

- ▶ Dans l'arborescence de bibliothèques, sélectionnez la déclaration de types à supprimer.
- ▶ Dans les menus, choisissez Edition > Supprimer pour supprimer la déclaration de types.

Dictionnaires des substitutions/synonymes

Un **dictionnaire de substitutions** est un ensemble de termes qui permet de regrouper des termes similaires sous un terme cible. Les dictionnaires de substitutions sont gérés dans le panneau du bas de l'onglet Ressources de bibliothèque. Vous pouvez accéder à cette vue avec Vue > Editeur de ressources dans les menus.

Vous pouvez définir deux formes de substitution dans ce dictionnaire : **synonymes** et **éléments optionnels**. Vous pouvez cliquer sur les onglets de ce panneau pour passer de l'un à l'autre.

Après avoir exécuté une extraction sur vos données textuelles, vous pouvez trouver plusieurs concepts qui sont des synonymes ou des formes fléchies d'autres concepts. En identifiant les éléments optionnels et les synonymes, vous pouvez forcer le moteur du programme d'extraction à les faire correspondre à un terme cible unique.

La substitution à l'aide de synonymes et d'éléments optionnels réduit le nombre de concepts dans le panneau Résultats d'extraction en les combinant pour former des concepts plus importants et représentatifs avec une fréquence plus élevée.

Figure 10-8
Panneau du dictionnaire de substitutions

	Cible	Synonymes	Bibliothèque
0			
1	amour	amour	Opinions Library (French)
2	lattertif	lattertif, lattertifs, lattertive, lattertives	Opinions Library (French)
3	lclin d'oeil	lclin d'oeil	Opinions Library (French)
4	lconscience	lconscience, lconsciences	Opinions Library (French)
5	lconseil	lconseil, lconseils	Opinions Library (French)
6	lcourt	lcourt, lcourte, lcourtes, lcourts	Opinions Library (French)
7	lde mauvaise qualité	de basse qualité, de qualité décevante, de qualiter décevante, de qualité décevante, de très basse de gamme, movaiz kalité	Opinions Library (French)
8	ldésir	ldesir, ldesirs, ldésir, ldésirs	Opinions Library (French)
9	ldétail	ldetail, ldetais, ldétails	Opinions Library (French)

Synonymes

Les synonymes sont des mots ayant le même sens. Vous pouvez également utiliser des synonymes pour regrouper des termes et leurs abréviations, ou pour réunir les mots fréquemment mal orthographiés sous la version correcte du mot. Vous pouvez définir ces synonymes dans l'onglet Synonymes.

Une définition de synonyme est constituée de deux parties. La première est un terme cible, qui est le terme sous lequel vous souhaitez que le moteur du programme d'extraction regroupe tous les termes synonymes. Ce terme cible deviendra probablement le concept apparaissant dans le panneau Résultats d'extraction, sauf s'il est utilisé comme synonyme d'un autre terme cible ou s'il est exclu. La deuxième est la liste de synonymes qui sera regroupée sous le terme cible.

Par exemple, si vous voulez que *automobile* soit remplacé par *véhicule*, *automobile* sera le synonyme et *véhicule* le terme cible.

Vous pouvez entrer n'importe quel mot dans la colonne Synonyme, mais si le mot n'est pas trouvé au cours de l'extraction et si le terme avait une option de mise en correspondance de Entier, aucune substitution ne peut être effectuée. En revanche, il n'est pas nécessaire que le terme cible soit extrait pour que les synonymes soient regroupés sous ce terme.

Figure 10-9
Dictionnaire de substitutions, onglet Synonymes

	Cible	Synonymes	Bibliothèque
1	vehicule		Budget Library (French)
2	look	look, lookin, the way it looks	Opinions Library (French)
3	advertisement	ad, advert, advertasing, advertise, advertising, advertisement	Opinions Library (French)
4	aftertaste	after taste, after-taste	Opinions Library (French)
5	anti-spam	anti spam, antispam	Opinions Library (French)
6	appearance	appearance	Product Satisfaction Library (English)
7	authorisation	authorise, authorising, authorization, authorize, authorizing	Product Satisfaction Library (English)
8	battery	abtery, batery, batt, battery life, batttery	Product Satisfaction Library (English)
9	call-waiting	call waiting	Product Satisfaction Library (English)
10	characteristic	attribute, charatceristic, properties	Product Satisfaction Library (English)
11	comfort	confort	Product Satisfaction Library (English)
12	communication	^ communicate, ^ communicate, amount of mail, commuication, communciation, communicated, communicating, communicated, communication	Product Satisfaction Library (English)

Éléments optionnels

Les éléments optionnels désignent les mots optionnels d'un terme composé qui peuvent être ignorés pendant l'extraction afin de conserver un regroupement de termes similaires, même s'ils apparaissent légèrement différents dans le texte. Les éléments optionnels sont des mots simples dont la suppression d'un terme composé peut créer une correspondance avec un autre terme. Ces

mots simples peuvent apparaître à n'importe quel endroit du terme composé, à savoir au début, au milieu ou à la fin. Vous pouvez définir les éléments optionnels dans l'onglet Optionnels.

Par exemple, pour regrouper les termes `ibm` et `ibm corp`, vous devez déclarer que `corp` doit ici être traité comme élément optionnel. Dans un autre exemple, si vous signalez que le terme `accès` est un élément optionnel et que l'extraction renvoie `vitesse d'accès internet` et `vitesse internet`, ils sont regroupés sous le terme qui revient le plus fréquemment.

Figure 10-10
Dictionnaire de substitutions - Onglet Optionnels

Éléments optionnels	Bibliothèque
<input checked="" type="checkbox"/>	Local Library
<input checked="" type="checkbox"/>	Product Satisfaction Library (English)
<input checked="" type="checkbox"/>	Opinions Library (English)
<input checked="" type="checkbox"/>	Budget Library (English)
<input checked="" type="checkbox"/> a.g., a.g., ag, co, co., corp, corp., corporation, gbh, gmbh, inc, inc., incorporated, kgaa, llc, llc., llc, ltd, ltd., org, plc, s.a., s.a., s.c.a., s.c.a., sa, sca	Core Library (English)
<input checked="" type="checkbox"/> pour	Variations Library (English)
<input checked="" type="checkbox"/>	Opinions Library (French)
<input checked="" type="checkbox"/>	Budget Library (French)

Synonymes Éléments optionnels

Définition de synonymes

Dans l'onglet Synonymes, vous pouvez entrer une définition de synonyme dans la ligne vide qui figure en haut du tableau. Commencez par définir le terme cible et ses synonymes. Vous pouvez également sélectionner la bibliothèque dans laquelle vous souhaitez stocker la définition. Toutes les occurrences des synonymes sont ensuite regroupées sous le terme cible lors de l'extraction finale. [Pour plus d'informations, reportez-vous à la section Ajout de termes sur p. 224.](#)

Par exemple, si les données textuelles fournissent diverses informations sur les télécommunications, vous pouvez rencontrer ces termes : `téléphone cellulaire`, `téléphone portable` et `téléphone mobile`. Vous pouvez alors décider d'indiquer que `cellulaire` et `mobile` sont synonymes de `portable`. Si vous définissez ces synonymes, chaque occurrence extraite de `téléphone cellulaire` et `téléphone mobile` sera considérée comme le même terme que `téléphone portable` et apparaîtra avec elle dans la liste des termes.

Quand vous créez vos déclarations de types, vous pouvez entrer un terme puis penser à trois ou quatre synonymes de ce terme. Dans ce cas, vous pouvez saisir tous les termes puis votre terme cible dans le dictionnaire de substitution avant de faire glisser les synonymes.

La substitution de synonymes est également appliquée aux formes fléchies (le pluriel, par exemple) du synonyme. En fonction du contexte, il est conseillé de définir des limites régissant le processus de substitution des termes. Des caractères sont alors utilisés pour restreindre le traitement des synonymes :

- **point d'exclamation (!)** Lorsque le point d'exclamation précède directement le synonyme `!synonyme`, cela indique qu'aucune forme fléchie du synonyme ne sera substituée par le terme cible. Néanmoins, un point d'exclamation placé directement avant le terme cible `!target-term` signale qu'aucune partie du mot composé cible ou les variantes ne doit faire l'objet d'une substitution supplémentaire.

- **Astérisque (*).** Un astérisque placé directement après un synonyme, comme `synonyme*`, signifie que vous souhaitez remplacer ce mot par le terme cible. Par exemple, si vous définissez `manage*` comme synonyme et `management` comme terme cible, le mot `managers` est remplacé par le terme cible `management`. Vous pouvez également ajouter un espace et un astérisque après le mot (`synonyme *`) comme `internet *`. Si vous définissez le terme cible `internet` et les synonymes `internet *` et `web *`, les termes `internet haut débit` et `web indépendant` sont alors remplacés par `internet`. Vous ne pouvez pas utiliser le caractère générique de l'astérisque au début d'un mot ou d'une chaîne dans ce dictionnaire.
- **Caret (^).** Un caret et un espace placés avant le synonyme, comme `^ synonyme`, indiquent que le regroupement de synonymes ne s'applique que lorsque le terme commence par le synonyme. Par exemple, si vous définissez `^ salaire` comme synonyme et `revenus` comme cible et que ces deux termes sont extraits, ils sont regroupés sous le terme `revenus`. En revanche, si «bulletin de salaire» et «revenus» sont extraits, ils ne sont pas réunis car «bulletin de salaire» ne commence pas par «salaire». Un espace doit séparer ce symbole du synonyme.
- **Symbole dollar (\$).** Un espace et un symbole dollar placés après le synonyme (`synonyme $`, par exemple) signalent que le regroupement de synonymes ne s'applique que lorsque le synonyme figure à la fin du terme. Par exemple, si vous définissez le synonyme `capital $` et le terme cible `argent` et que ces deux termes sont extraits, ils sont tous deux regroupés sous le terme `argent`. En revanche, si `capital monétaire` et `argent` sont extraits, ils ne sont pas réunis car `capital monétaire` ne se termine pas par `capital`. Un espace doit séparer ce symbole du synonyme.
- **Caret (^) et symbole dollar (\$).** Si le caret et le dollar sont utilisés ensemble (par exemple `^ synonym $`), seules les concordances exactes sont mises en correspondance avec le synonyme. Cela signifie qu'aucun mot ne peut apparaître avant ou après le synonyme dans le terme extrait pour que le regroupement s'effectue. Par exemple, vous pouvez définir `^ van $` comme synonyme et `fourgon` comme cible pour que `van` soit regroupé avec `fourgon`, mais pas `ludwig van beethoven`. Par ailleurs, dès lors que vous définissez un synonyme à l'aide des symboles caret et dollar et que ce mot apparaît n'importe où dans le texte source, le synonyme est automatiquement extrait.

Figure 10-11
Dictionnaire de substitutions, onglet Synonymes avec exemple

	Cible ==	Synonymes	Bibliothèque
1	<input checked="" type="checkbox"/> vehicle		Budget Library (French)
2	<input checked="" type="checkbox"/> look	look, lookin, the way it looks	Opinions Library (French)
3	<input checked="" type="checkbox"/> advertisement	ad, advert, advertasing, advertise, advertising, advertisement	Opinions Library (French)
4	<input checked="" type="checkbox"/> aftertaste	after taste, after-taste	Opinions Library (French)
5	<input checked="" type="checkbox"/> anti-spam	anti spam, antispam	Opinions Library (French)
6	<input checked="" type="checkbox"/> appearance	appearance	Product Satisfaction Library (English)
7	<input checked="" type="checkbox"/> authorisation	authorise, authorising, authorization, authorize, authorizing	Product Satisfaction Library (English)
8	<input checked="" type="checkbox"/> battery	abbtery, betery, batt, battery life, batttery	Product Satisfaction Library (English)
9	<input checked="" type="checkbox"/> call-waiting	call waiting	Product Satisfaction Library (English)
10	<input checked="" type="checkbox"/> characteristic	attribute, charatceristic, properties	Product Satisfaction Library (English)
11	<input checked="" type="checkbox"/> comfort	confort	Product Satisfaction Library (English)
12	<input checked="" type="checkbox"/> communication	^communicate, ^communicate, amount of mail, communication, communciation, communicated, communicating, communicated, communication	Product Satisfaction Library (English)

Synonymes Éléments optionnels

Pour ajouter une entrée de synonyme

- Affichez le panneau de substitution et cliquez sur l'onglet Synonymes dans l'angle inférieur gauche.

- ▶ Dans la ligne vide en haut du tableau, entrez le terme cible dans la colonne Cible. Le terme cible que vous entrez apparaît en couleur. Cette couleur représente le type dans lequel le terme apparaît ou a été imposé, le cas échéant. Si le terme apparaît en noir, il n'est alors inclus dans aucune déclaration de types.
- ▶ Cliquez dans la deuxième cellule à droite de la cible et entrez l'ensemble de synonymes. Séparez chaque entrée à l'aide du séparateur global défini dans la boîte de dialogue Options. [Pour plus d'informations, reportez-vous à la section Définition des options dans le chapitre 2 sur p. 19.](#) Les termes que vous entrez apparaissent en couleur. Cette couleur représente le type dans lequel le terme apparaît. Si le terme apparaît en noir, il n'est alors inclus dans aucune déclaration de types.
- ▶ Cliquez dans la dernière cellule pour sélectionner la bibliothèque dans laquelle vous voulez stocker cette définition de synonyme.

Remarque : Ces instructions indiquent la façon d'effectuer des modifications dans la vue Editeur de ressources. N'oubliez pas que vous pouvez également réaliser ce type de réglage directement à partir du panneau Résultats d'extraction ou du panneau Données. [Pour plus d'informations, reportez-vous à la section Affinage des résultats de l'extraction dans le chapitre 5 sur p. 88.](#)

Définition des éléments optionnels

Dans l'onglet Optionnels, vous pouvez définir des éléments optionnels pour n'importe quelle bibliothèque. Ces entrées sont regroupées pour chaque bibliothèque. Dès que vous ajoutez une bibliothèque dans l'arborescence de bibliothèques, une ligne vide d'élément optionnel est ajoutée dans l'onglet Optionnels.

Toutes les entrées sont transformées automatiquement en mots en minuscules. Le moteur d'extraction met indifféremment en correspondance les mots du texte en majuscules et en minuscules.

Figure 10-12
Dictionnaire de substitutions - Onglet Optionnels

Éléments optionnels	Bibliothèque
<input checked="" type="checkbox"/>	Local Library
<input checked="" type="checkbox"/>	Product Satisfaction Library (English)
<input checked="" type="checkbox"/>	Opinions Library (English)
<input checked="" type="checkbox"/>	Budget Library (English)
<input checked="" type="checkbox"/> a.g., a.g., ag, co, co., corp, corp., corporation, gbh, gmbh, inc, inc., incorporated, kga, llc, llc., llc, ltd, ltd., org, plc, s.a, s.a., s.c.a., s.c.a., sa, sca	Core Library (English)
<input checked="" type="checkbox"/> pour	Variations Library (English)
	Opinions Library (French)
	Budget Library (French)

Remarque : Les termes sont délimités à l'aide du séparateur défini dans la boîte de dialogue Options. [Pour plus d'informations, reportez-vous à la section Définition des options dans le chapitre 2 sur p. 19.](#) Si l'élément optionnel que vous entrez contient le même séparateur qu'une partie du terme, faites-le précéder d'une barre oblique inversée.

Pour ajouter une entrée

- ▶ Le panneau de substitution étant affiché, cliquez sur l'onglet Optionnels dans l'angle inférieur gauche de l'éditeur.

- ▶ Cliquez dans la cellule de la colonne Éléments optionnels de la bibliothèque à laquelle vous souhaitez ajouter cette entrée.
- ▶ Entrez l'élément optionnel. Séparez chaque entrée à l'aide du séparateur global défini dans la boîte de dialogue Options. [Pour plus d'informations, reportez-vous à la section Définition des options dans le chapitre 2 sur p. 19.](#)

Désactiver et supprimer des substitutions

Vous pouvez supprimer une entrée de façon temporaire. Pour cela, vous devez la désactiver dans votre dictionnaire. Lorsque vous désactivez une entrée, celle-ci est ignorée au cours des extractions.

Vous pouvez aussi supprimer toutes les entrées obsolètes de votre dictionnaire de substitutions.

Pour désactiver une entrée

- ▶ Dans votre dictionnaire, sélectionnez l'entrée à désactiver.
- ▶ Cliquez sur la barre d'espace. La coche disparaît de la case figurant à gauche de l'entrée.
Remarque : vous pouvez également désactiver la case à cocher située à gauche de l'entrée.

Pour supprimer une entrée de synonyme

- ▶ Dans votre dictionnaire, sélectionnez l'entrée à supprimer.
- ▶ Dans le menu, sélectionnez Edition > Supprimer ou appuyez sur la touche Suppr sur votre clavier. L'entrée ne figure plus dans le dictionnaire.

Pour supprimer une entrée d'élément optionnel

- ▶ Dans votre dictionnaire, double-cliquez sur l'entrée à supprimer.
- ▶ Supprimez manuellement le terme.
- ▶ Appuyez sur Entrée pour appliquer la modification.

Dictionnaires d'exclusions

Un **dictionnaire d'exclusions** est une liste de mots, de phrases ou de chaînes partielles. Tout terme correspondant à ou contenant une entrée dans le dictionnaire d'exclusions sera ignoré ou exclu de l'extraction. La gestion des dictionnaires d'exclusions s'effectue dans le panneau de droite de l'éditeur. En règle générale, vous entrez dans cette liste les mots ou expressions de liaison utilisés pour la continuité du texte, mais qui ne lui apportent rien d'important et risquent d'encombrer les résultats de l'extraction. En ajoutant ces termes au dictionnaire d'exclusions, vous êtes assuré qu'ils ne seront jamais extraits.

La gestion des dictionnaires d'exclusion s'effectue dans le panneau supérieur droit de l'onglet Ressources de bibliothèque dans l'éditeur. Vous pouvez accéder à cette vue avec Vue > Editeur de ressources dans les menus.

Figure 10-13
Panneau Dictionnaire d'exclusions

	Exclusions	Bibliothèque
0	<input type="checkbox"/>	
1	<input checked="" type="checkbox"/> any kind of problem	Opinions Library (English)
2	<input checked="" type="checkbox"/> any problems i have	Opinions Library (English)
3	<input checked="" type="checkbox"/> anykind of problem	Opinions Library (English)
4	<input checked="" type="checkbox"/> as usual	Opinions Library (English)
5	<input checked="" type="checkbox"/> can't wait	Opinions Library (English)
6	<input checked="" type="checkbox"/> 1 conseil	Opinions Library (French)
7	<input checked="" type="checkbox"/> bien souvent	Opinions Library (French)
8	<input checked="" type="checkbox"/> en interet	Opinions Library (French)
9	<input checked="" type="checkbox"/> grand h	Opinions Library (French)
10	<input checked="" type="checkbox"/> mais qualité	Opinions Library (French)
11		

Vous pouvez entrer un mot, une expression ou une chaîne partielle dans la ligne vide figurant dans le haut du tableau du dictionnaire d'exclusions. Vous pouvez ajouter des chaînes de caractères à votre dictionnaire d'exclusions sous la forme d'un ou de plusieurs mots, voire de mots partiels à l'aide du caractère générique astérisque. Les entrées déclarées dans le dictionnaire d'exclusions servent à empêcher l'extraction de ces concepts. Si une entrée est également déclarée ailleurs dans l'interface, par exemple dans une déclaration de types, elle apparaît barrée dans les autres dictionnaires, ce qui indique qu'elle est actuellement exclue. Il n'est pas nécessaire que cette chaîne apparaisse dans les données textuelles ou qu'elle fasse partie d'une déclaration de types pour être appliquée.

Remarque : Si un concept ajouté au dictionnaire d'exclusions sert également de cible dans une entrée de synonyme, la cible et tous ses synonymes sont également exclus. [Pour plus d'informations, reportez-vous à la section Définition de synonymes sur p. 233.](#)

Utilisation de caractères génériques (*)

Vous pouvez utiliser le caractère générique astérisque pour signaler que vous souhaitez que l'entrée d'exclusion soit considérée comme une chaîne partielle. Tous les mots détectés par le moteur du programme d'extraction, et qui commencent ou finissent par l'une des chaînes entrées ici, sont exclus de l'extraction finale. Toutefois, il est interdit d'utiliser les caractères génériques dans deux cas de figure :

- Tiret (-) précédé d'un astérisque, comme *-
- Apostrophe (') précédée d'un astérisque, comme *'

Table 10-2
Exemples d'entrées d'exclusion

Entrée	Exemple	Résultats
mot	<i>suivant</i>	Aucun concept (ou ses termes) n'est extrait s'il contient le mot <i>suivant</i> .
expression	<i>par exemple</i>	Aucun concept (ou ses termes) n'est extrait s'il contient l'expression <i>par exemple</i> .

Entrée	Exemple	Résultats
partiel	<i>final*</i>	Exclut tous les concepts (ou leurs termes) concordants ou contenant des variations du mot <i>final</i> , par exemple <i>finale</i> , <i>finaliste</i> , <i>finale</i> ment ou encore <i>finale</i> 2010.
partiel	<i>*piste</i>	Exclut tous les concepts (ou leurs termes) concordants ou contenant des variations du mot <i>piste</i> , tels que <i>trappiste</i> , <i>monotypiste</i> , <i>copiste</i> , <i>endoscopiste</i> , <i>alpiste</i> ou <i>avant-piste</i> .

Pour ajouter des entrées

- ▶ Dans la ligne vide qui figure dans le haut du tableau, entrez un terme. Le terme que vous entrez apparaît en couleur. Cette couleur représente le type dans lequel le terme apparaît. Si le terme apparaît en noir, il n'est alors inclus dans aucune déclaration de types.

Pour désactiver des entrées

Vous pouvez supprimer temporairement une entrée en la désactivant dans votre dictionnaire d'exclusions. Lorsque vous désactivez une entrée, celle-ci est ignorée au cours des extractions.

- ▶ Dans votre dictionnaire d'exclusions, sélectionnez l'entrée à désactiver.
 - ▶ Cliquez sur la barre d'espace. La coche disparaît de la case figurant à gauche de l'entrée.
- Remarque* : vous pouvez également désactiver la case à cocher située à gauche de l'entrée.

Pour supprimer des entrées

Vous pouvez supprimer les entrées inutiles de votre dictionnaire d'exclusions.

- ▶ Dans votre dictionnaire d'exclusions, sélectionnez l'entrée à supprimer.
- ▶ Dans les menus, sélectionnez Edition > Supprimer. L'entrée ne figure plus dans le dictionnaire.

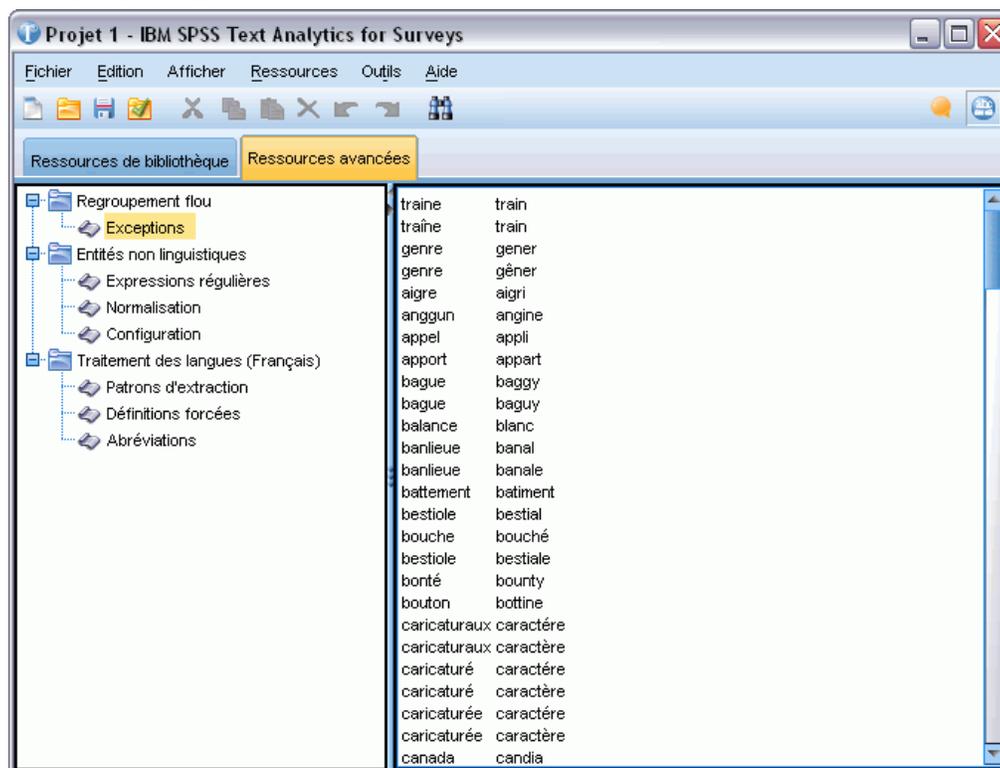
À propos des ressources avancées

Outre les déclarations de types, les dictionnaires d'exclusions et de substitutions, vous pouvez également utiliser divers paramètres de ressources avancées tels que les paramètres de regroupements flous ou les définitions de type non linguistiques. Vous pouvez trouver ces ressources dans l'onglet Ressources avancées dans la vue Editeur de ressources. Vous pouvez aussi enregistrer vos modifications comme défaut pour tous les projets ou revenir au contenu d'origine.

Lorsque vous allez dans l'onglet Ressources avancées, vous pouvez modifier les informations suivantes :

- **Regroupement flou (Exceptions).** Permet d'exclure des paires de mots de l'algorithme de regroupement flou (correction des fautes d'orthographe). [Pour plus d'informations, reportez-vous à la section Regroupement flou sur p. 242.](#)
- **Entités non linguistiques.** Permet d'activer et de désactiver les entités non linguistiques pouvant être extraites, ainsi que les expressions régulières et les règles de normalisation qui sont appliquées lors de leur extraction. [Pour plus d'informations, reportez-vous à la section Entités non linguistiques sur p. 243.](#)
- **Gestion des langues.** Permet de déclarer les méthodes spéciales de structuration des phrases (patrons d'extraction et définitions forcées) et d'utilisation des abréviations pour la langue sélectionnée. [Pour plus d'informations, reportez-vous à la section Traitement des langues sur p. 248.](#)

Figure 11-1
Ressources avancées



Remarque : vous pouvez utiliser la barre d'outils Rechercher/Remplacer pour rechercher des informations rapidement ou pour apporter des modifications globales à une section. [Pour plus d'informations, reportez-vous à la section Remplacement sur p. 241.](#)

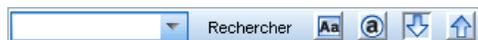
Pour modifier les ressources avancées

- ▶ Localisez et sélectionnez la section de ressources dans laquelle vous souhaitez effectuer l'édition. Le contenu apparaît dans le panneau de droite.
- ▶ Si nécessaire, à l'aide du menu ou des boutons de la barre d'outils, coupez, copiez ou collez des éléments.
- ▶ Editez les fichiers que vous souhaitez modifier en utilisant les règles de formatage de la section. Les modifications sont enregistrées dès que vous les apportez. Utilisez la flèche d'annulation ou de rétablissement de la barre d'outils pour rétablir vos précédentes modifications.

Recherche

Il peut s'avérer nécessaire de localiser rapidement des informations dans une section particulière. La fonction Rechercher vous permet de rechercher rapidement une règle spécifique. Pour rechercher des informations dans une section, utilisez la barre d'outils Rechercher.

Figure 11-2
Barre d'outils Rechercher



Pour utiliser la fonction de recherche

- ▶ Localisez et sélectionnez la section de ressources dans laquelle vous souhaitez effectuer la recherche. Son contenu apparaît dans le panneau de droite de l'éditeur.
- ▶ Dans les menus, sélectionnez Edition > Rechercher. La barre d'outils Rechercher apparaît dans la partie supérieure droite de la boîte de dialogue Editer les ressources avancées.
- ▶ Saisissez la chaîne de mots que vous recherchez dans la zone de texte. Vous pouvez contrôler la casse, les correspondances partielles et le sens de la recherche à l'aide des boutons de la barre d'outils.
- ▶ Cliquez sur Rechercher pour lancer la recherche. Si une correspondance est détectée, le texte est mis en surbrillance dans la fenêtre.
- ▶ Cliquez de nouveau sur Rechercher pour rechercher la correspondance suivante.

Remplacement

Dans certains cas, il peut s'avérer nécessaire d'effectuer des mises à jour globales à vos ressources avancées. La fonction Remplacer peut vous aider à effectuer des mises à jour homogènes sur vos informations.

Pour utiliser la fonction de remplacement

- ▶ Localisez et sélectionnez la section de ressources dans laquelle vous souhaitez effectuer la recherche et le remplacement. Son contenu apparaît dans le panneau de droite de l'éditeur.
- ▶ Dans les menus, sélectionnez Edition > Remplacer. La boîte de dialogue Remplacer apparaît.

Figure 11-3
Boîte de dialogue Remplacer



- ▶ Dans la zone de texte Rechercher, saisissez la chaîne de mots que vous souhaitez rechercher.
- ▶ Dans la zone de texte Remplacer par, saisissez la chaîne que vous souhaitez utiliser à la place du texte trouvé.
- ▶ Sélectionnez Rechercher uniquement les mots entiers si vous ne souhaitez rechercher ou remplacer que des mots complets.

- ▶ Sélectionnez **Respecter la casse** si vous ne souhaitez rechercher ou remplacer que les mots dont la casse correspond exactement.
- ▶ Cliquez sur **Suivant** pour rechercher une correspondance. Si une correspondance est détectée, le texte est mis en surbrillance dans la fenêtre. Si vous ne voulez pas remplacer cette correspondance, cliquez de nouveau sur **Suivant** jusqu'à ce que vous ayez trouvé une correspondance que vous souhaitez remplacer.
- ▶ Cliquez sur **Remplacer** pour remplacer la correspondance sélectionnée.
- ▶ Cliquez sur **Remplacer** pour remplacer toutes les correspondances de la section. Un message indique le nombre de remplacements effectués.
- ▶ Lorsque les remplacements sont terminés, cliquez sur **Fermer**. La boîte de dialogue se ferme.

Remarque : En cas d'erreur de remplacement, vous pouvez annuler l'opération. Pour cela, fermez la boîte de dialogue et choisissez **Edition > Annuler** dans les menus. Vous devez suivre cette procédure pour chaque modification à annuler.

Regroupement flou

Dans la boîte de dialogue Paramètres d'extraction, si vous sélectionnez **Traitement des fautes de frappe**. Nombre de caractères minimum requis :, vous avez activé l'algorithme de regroupement flou.

Le regroupement flou permet de regrouper les mots dont l'orthographe est souvent incorrecte ou proche en supprimant temporairement toutes les voyelles (à l'exception de la première voyelle) et les consonnes doubles ou triples des mots extraits, et en les comparant ensuite afin de déterminer s'ils sont identiques. Pendant le processus d'extraction, la fonction de regroupement flou est appliquée aux termes extraits et les résultats sont comparés afin de déterminer s'il existe des correspondances. Si tel est le cas, les termes initiaux sont regroupés dans la liste d'extraction finale. Ils sont placés sous le terme qui compte le plus d'occurrences dans les données.

Remarque : Si les deux termes comparés sont attribués à différents types, à l'exclusion du type <Unknown>, la méthode de regroupement flou ne doit pas être appliquée à cette paire. Autrement dit, les termes doivent appartenir au même type ou au type <Unknown> afin d'appliquer la méthode.

Si vous avez activé cette fonction et remarqué que deux mots orthographiés de façon similaire étaient regroupés de façon erronée, vous pouvez les exclure du regroupement flou. Pour cela, entrez les paires dont la mise en correspondance est incorrecte dans la section **Exceptions** de l'onglet **Ressources avancées**. [Pour plus d'informations, reportez-vous à la section À propos des ressources avancées sur p. 239.](#)

L'exemple suivant montre le fonctionnement du regroupement flou. Si le regroupement flou est activé, ces mots apparaissent similaires et sont mis en correspondance comme suit :

```

color -> colr          mountain -> montn
colour -> colr         montana -> montn

modeling -> modlng     furniture -> furntr
modelling -> modlng    furnature -> furntr

```

Dans l'exemple précédent, vous souhaitez vraisemblablement que les termes `paysage` et `passage` ne soient pas regroupés. Par conséquent, vous pouvez les entrer dans la section Exceptions comme suit :

```
paysage    passage
```

Important ! Dans certains cas, les exceptions de regroupement flou n'empêchent pas 2 mots d'être mis en paire car certaines règles de synonymes s'appliquent. Dans ce cas, vous pouvez essayer d'entrer des synonymes à l'aide du caractère générique du point d'exclamation (!) pour que les mots ne soient pas des synonymes dans les résultats. [Pour plus d'informations, reportez-vous à la section Définition de synonymes dans le chapitre 10 sur p. 233.](#)

Règles de formatage pour le Regroupement flou (Exceptions)

- Définissez une seule paire d'exception par ligne.
- Utilisez des mots simples ou composés.
- N'utilisez que des minuscules dans les mots. Les mots en majuscules seront ignorés.
- Utilisez le caractère TAB pour séparer les deux mots de chaque paire.

Entités non linguistiques

Lorsque vous travaillez avec certains types de données, vous pouvez souhaiter extraire des dates, des numéros de Sécurité sociale, des pourcentages ou toute autre entité non linguistique. Ces entités sont déclarées explicitement dans le fichier de configuration, dans lequel vous pouvez activer ou désactiver les entités. [Pour plus d'informations, reportez-vous à la section Configuration sur p. 247.](#) Pour optimiser la sortie dans le moteur du programme d'extraction, l'entrée issue du traitement non linguistique est normalisée pour regrouper les entités semblables en fonction de formats prédéfinis. [Pour plus d'informations, reportez-vous à la section Normalisation sur p. 246.](#)

Remarque : vous pouvez activer et désactiver l'extraction des entités non linguistiques dans les paramètres d'extraction.

Entités non linguistiques disponibles

Les entités non linguistiques du tableau suivant peuvent être extraites. Le nom de type est entre parenthèses.

Adresses (<Address>)	Organisations (<Organization>)
Acides aminés (<Aminoacid>)	Pourcentages (<Percent>)
Devises (<Currency>)	Products (<Product>)
Dates (<Date>)	Protéines (<Gene>)
Délai (<Delay>)	Numéros de téléphone (<PhoneNumber>)
Chiffres (<Digit>)	Temps (<Time>)
Adresses électroniques (<email>)	Sécurité sociale (Etats-Unis) (<SocialSecurityNumber>)
Adresses HTTP/URL (<url>)	Poids et mesures (<Weights-Measures>)
Adresse IP (<IP>)	

Nettoyage du texte pour traitement

Avant l'extraction des entités non linguistiques, le texte d'entrée est nettoyé. Durant cette étape, les modifications temporaires suivantes sont effectuées afin que les entités non linguistiques puissent être identifiées et extraites ainsi :

- Toute séquence de deux espaces ou plus est remplacée par un seul espace.
- Les tabulations sont remplacées par un espace.
- Les caractères de fin de ligne uniques ou les caractères de séquence sont remplacés par un espace, tandis que les séquences de fin de ligne multiples sont marquées comme la fin d'un paragraphe. Une fin de ligne peut être indiquée par des retours chariot (CR) et des sauts de ligne (LF) ou les deux à la fois.
- Les balises HTML et XML sont temporairement supprimées et ignorées.

Expressions régulières

Lors de l'extraction d'entités non linguistiques, vous pouvez modifier ou ajouter des définitions d'expressions régulières à utiliser pour identifier des expressions régulières. Pour cela, rendez-vous dans la section Expressions régulières dans l'onglet Ressources avancées. [Pour plus d'informations, reportez-vous à la section À propos des ressources avancées sur p. 239.](#)

Le fichier est divisé en plusieurs sections distinctes. La première section s'intitule `[macros]`. Une autre section peut venir s'y ajouter, pour chaque entité non linguistique. Vous pouvez ajouter des sections à ce fichier. Dans chaque section, les règles sont numérotées (*regexpl*, *regexp2*, etc.). Ces règles doivent être numérotées de manière séquentielle de 1 à *n*. Toute interruption dans la numérotation entraîne la suspension du traitement de ce fichier.

Dans certains cas, une entité varie en fonction de la langue. Une entité est considérée comme variant en fonction de la langue si son paramètre de langue a une valeur autre que 0 dans le fichier de configuration. [Pour plus d'informations, reportez-vous à la section Configuration sur p. 247.](#) Lorsqu'une entité varie en fonction de la langue, la langue doit être utilisée comme préfixe du nom de la section (par exemple, `[english/PhoneNumber]`). Si l'entité `PhoneNumber` adopte la valeur de langue 2, cette section contient des règles applicables uniquement aux numéros de téléphone anglais.

Important ! Si vous modifiez ce fichier (ou un autre) dans l'éditeur et que le moteur du programme d'extraction ne fonctionne plus comme vous le souhaitez, sélectionnez l'option Rétablir les valeurs d'origine dans la barre d'outils pour rétablir le fichier d'origine livré. Ce fichier requiert un certain niveau de connaissance des expressions régulières. Pour obtenir une assistance dans ce domaine, contactez IBM Corp..

Caractères spéciaux. [] {} () \ * + ? | ^ \$

Tous les caractères s'auto-correspondent à l'exception des caractères spéciaux suivants, qui sont utilisés dans un but spécifique dans les expressions : `. [{ () \ * + ? | ^ $` Pour pouvoir être utilisés tels quels, ces caractères doivent être précédés d'une barre oblique inverse (`\`) dans la définition.

Par exemple, si vous essayez d'extraire des adresses Web, le point est très important pour l'entité, par conséquent, vous devez le faire précéder d'une barre oblique inverse :

```
www\[a-z]+\.[a-z]+
```

Opérateurs de répétition et quantifieurs ? + * {}

Pour rendre les définitions plus flexibles, vous pouvez utiliser plusieurs caractères génériques qui sont standard dans les expressions régulières. Il s'agit de * ? +

- *L'astérisque ** indique qu'il y a *zéro occurrence ou plus* de la chaîne précédente. Par exemple, `ab*c` correspond à « *ac* », « *abc* », « *abbbc* », etc.
- *Le signe plus +* indique qu'il y a *une ou plusieurs occurrences* de la chaîne précédente. Par exemple, `ab+c` correspond à « *abc* », « *abbc* », « *abbbc* », mais pas à « *ac* ».
- *Le point d'interrogation ?* indique qu'il y a *zéro ou une occurrence* de la chaîne précédente. Par exemple, `modell?ing` correspond à la fois à « *modeling* » et à « *modeling* ».
- *Limiter la répétition avec des accolades {}* indique les limites de la répétition. Par exemple :
 - ▶ `[0-9]{n}` correspond à un chiffre répété exactement *n* fois. Par exemple, `[0-9]{4}` correspondra à « *1998* », mais ni à « *33* » ni à « *19983* ».
 - ▶ `[0-9]{n,}` correspond à un chiffre répété *n fois ou plus*. Par exemple, `[0-9]{3,}` correspondra à « *199* » ou à « *1998* », mais pas à « *19* ».
 - ▶ `[0-9]{n,m}` correspond à un chiffre répété entre *n et m fois inclus*. Par exemple, `[0-9]{3,5}` correspondra à « *199* », « *1998* » ou à « *19983* », mais pas à « *19* » ni à « *199835* ».

Espaces facultatifs et traits d'union

Dans certains cas, vous voulez inclure un espace facultatif dans une définition. Par exemple, si vous voulez extraire des devises telles que « *pesos uruguayens* », « *peso uruguayen* », « *pesos Uruguay* », « *peso Uruguay* », « *pesos* » ou « *peso* », vous devriez tenir compte du fait qu'il peut y avoir deux mots séparés par un espace. Dans ce cas, cette définition devrait être écrite sous la forme `?pesos?(uruguayens| Uruguay)`. Puisque *uruguayen* ou *Uruguay* sont précédés par un espace lorsqu'ils sont utilisés avec *pesos/peso*, l'espace facultatif doit être défini à l'intérieur de la séquence facultative `(uruguayens| Uruguay)`. Si l'espace ne figurait pas dans la séquence facultative (par exemple `?pesos ?(uruguayens|Uruguay)`), il ne trouverait pas de correspondance avec « *pesos* » ou « *peso* » puisque l'espace serait requis.

Si vous cherchez des séries d'éléments incluant des traits d'union (-) dans une liste, le trait d'union doit être défini en dernier. Par exemple, si vous cherchez une virgule (,) ou un trait d'union (-), utilisez `[, -]` et non pas `[- ,]`.

Ordre des chaînes dans les listes et les macros

Vous devez toujours définir la séquence la plus longue avant une plus courte, sinon la plus longue ne sera jamais lue puisque la correspondance se fera sur la plus courte. Par exemple, si vous cherchez des chaînes « *janvier* » ou « *janv* », vous devez définir « *janvier* » avant « *janv* ». Ainsi, par exemple `(janvier|janv)` et non `(janv|janvier)`. Cela s'applique aussi aux macros, puisque les macros sont des listes de chaînes.

Ordre des règles dans la section des définitions

Définissez une seule règle par ligne. Dans chaque section, les règles sont numérotées (*regex1*, *regex2*, etc.). Ces règles doivent être numérotées de manière séquentielle de 1 à *n*. Toute interruption dans la numérotation entraîne la suspension du traitement de ce fichier. Pour désactiver une entrée, placez le symbole # au début de la ligne utilisée pour définir l'expression régulière. Pour activer une entrée, supprimez le caractère # en début de ligne.

Dans chaque section, les règles les plus spécifiques doivent être définies avant les plus générales afin de garantir un traitement correct. Par exemple, si vous cherchiez une date au format "mois année" et au format "mois", la règle "mois année" doit être définie avant la règle "mois". Voici comment elle doit être définie :

```
#@# January 1932
regex1=$(MONTH),? [0-9]{4}

#@# January
regex2=$(MONTH)
```

et non pas

```
#@# January
regex1=$(MONTH)

#@# January 1932
regex2=$(MONTH),? [0-9]{4}
```

Utilisation des macros dans les règles

Chaque fois qu'une séquence spécifique est utilisée dans plusieurs règles, vous pouvez utiliser une macro. Ensuite, si vous devez modifier la définition de cette séquence, vous aurez à la modifier une seule fois et non dans toutes les règles y faisant référence. Par exemple, en supposant que vous ayez la macro suivante :

```
MOIS=((janvier|février|mars|avril|juin|juillet|août|septembre|octobre|
novembre|décembre)|(jan|fév|mar|avr|mai|juin|juil|aoû|sep|oct|nov|déc) (\.)?)
```

Chaque fois que vous faites référence au nom de la macro, il doit être mis au format \$(), comme dans : `regex1=$(MOIS)`

Toutes les macros doivent être définies dans la section [macros].

Normalisation

Lors de l'extraction d'entités non linguistiques, les entités rencontrées sont normalisées pour regrouper les entités semblables en fonction de formats prédéfinis. Par exemple, les symboles de devise et leur équivalent en toutes lettres sont considérés comme étant identiques. Les entrées de normalisation sont stockées dans la section Normalisation de l'onglet Ressources avancées. [Pour plus d'informations, reportez-vous à la section À propos des ressources avancées sur p. 239.](#) Le fichier est divisé en plusieurs sections distinctes.

Important ! Ce fichier s'adresse à des utilisateurs avertis uniquement. Il est peu probable que vous ayez à modifier ce fichier. Pour obtenir une assistance dans ce domaine, contactez IBM Corp..

Règles de formatage pour Normalisation

- Ajoutez uniquement une entrée de normalisation par ligne.
- Respectez bien les sections de ce fichier. Vous ne pouvez pas ajouter de nouvelles sections.
- Pour désactiver une entrée, entrez le symbole # au début de la ligne concernée. Pour activer une entrée, supprimez le caractère # en début de ligne.

Configuration

Vous pouvez activer et désactiver les types d'entité non linguistique que vous souhaitez extraire dans le fichier de configuration des entités non linguistiques. Désactivez les entités dont vous n'avez pas besoin pour diminuer le temps de traitement nécessaire. Pour cela, rendez-vous dans la section Configuration de l'onglet Ressources avancées. [Pour plus d'informations, reportez-vous à la section À propos des ressources avancées sur p. 239.](#) Si l'extraction non linguistique est activée, le moteur du programme d'extraction lit ce fichier de configuration lors du processus d'extraction afin de déterminer les types d'entité non linguistique à extraire.

La syntaxe de ce fichier est la suivante :

```
#name<TAB>Langue<TAB>Code
```

Table 11-1
Syntaxe du fichier de configuration

Étiquette de colonne	Description
#name	Libellé selon lequel les entités non linguistiques seront référencées dans les deux autres fichiers nécessaires à l'extraction d'entités non linguistiques. Les mots utilisés ici distinguent les majuscules des minuscules.
Langue	Langue des enregistrements. Il est préférable de sélectionner une langue précise. Toutefois, vous pouvez également choisir l'option N'importe. Les options suivantes sont proposées : 0 = N'importe quelle langue étant utilisée chaque fois qu'une expression régulière n'est pas spécifique à une langue et pourrait être utilisée dans plusieurs modèles avec différentes langues, par exemple une adresse IP/URL/e-mail ; 1 = Français ; 2 = Anglais ; 4 = Allemand ; 5 = Espagnol ; 6 = Néerlandais ; 8 = Portugais ; 10 = Italien.
Code	Code de catégories grammaticales. La plupart des entités prennent la valeur "s", sauf rares exceptions. Valeurs possibles : s = mot vide ; a = adjectif ; n = nom. Lorsque cette option est activée, les entités non linguistiques sont d'abord extraites, puis les patrons d'extraction sont appliqués afin d'identifier le rôle dans un contexte plus large. Par exemple, les pourcentages se voient attribuer la valeur "a." Imaginons que la valeur 30 % soit extraite en tant qu'entité non linguistique. Elle sera identifiée comme étant un adjectif. Par conséquent, si votre texte contient les mots "30%" augmentation salaire, l'entité non linguistique "30%" est incluse dans le patron de catégorie grammaticale "ann" (adjectif nom nom).

Ordre dans la définition des entités

L'ordre dans lequel les entités sont déclarées dans ce fichier est important car il affecte leur mode d'extraction. Ces entrées sont appliquées dans l'ordre dans lequel elles sont répertoriées. Toute modification de l'ordre se répercute sur les résultats. Les entités non linguistiques les plus spécifiques doivent être définies avant les plus générales.

Par exemple, l'entité non linguistique "Acide aminé" est définie par :

```
regexp1=$(AA)-?$(NUM)
```

où \$(AA) correspond à

"(ala|arg|asn|asp|cys|gln|glu|gly|his|ile|leu|lys|met|phe|pro|ser)", qui sont des séquences de trois lettres spécifiques correspondant à des acides aminés particuliers.

D'autre part, l'entité non linguistique « Gène » est plus générale et elle est définie par :

```
regexp1=p[0-9]{2,3}
regexp2=[a-z]{2,4}-?[0-9]{1,3}-?[r]
regexp3=[a-z]{2,4}-?[0-9]{1,3}-?p?
```

Si « Gène » est défini avant « Acide aminé » dans la section Configuration, alors « Acide aminé » ne sera jamais mis en correspondance, puisque regexp3 de « Gène » sera toujours mis en correspondance en premier.

Règles de formatage pour la configuration

- Utilisez le caractère TAB pour séparer les entrées d'une colonne.
- Ne supprimez aucune ligne.
- Respectez la syntaxe indiquée dans le tableau précédent.
- Pour désactiver une entrée, entrez le symbole # au début de la ligne concernée. Pour activer une entité, supprimez le caractère # en début de ligne.

Traitement des langues

Chaque langue moderne exprime des idées, structure des phrases et utilise des abréviations d'une façon particulière. Dans la section Gestion des langues, vous pouvez éditer les patrons d'extraction, forcer des définitions pour ces patrons et indiquer des abréviations pour la langue que vous avez sélectionnée dans la liste déroulante Langue.

- Patrons d'extraction
- Définitions forcées
- Abréviations

Patrons d'extraction

Lors de l'extraction d'informations à partir de vos enregistrements, le moteur du programme d'extraction applique une série de patrons d'extraction à un « ensemble » de mots dans le texte afin d'identifier ceux (mots et expressions) susceptibles d'être extraits. Vous pouvez ajouter ou modifier les patrons d'extraction.

Les catégories grammaticales incluent les éléments grammaticaux tels que substantifs, adjectifs, participes passés, déterminants, prépositions, coordonnants, prénoms, sigles et particules. Une série de ces éléments constitue un patron d'extraction. Dans les produits de Text Mining IBM Corp., chaque catégorie grammaticale est représentée par un caractère unique afin de faciliter la définition des patrons. Par exemple, un adjectif est représenté par la lettre (minuscule) *a*.

L'ensemble de codes pris en charge apparaît par défaut en haut de chaque fichier de patron d'extraction par défaut, conjointement avec une série de patrons et des exemples de chaque patron pour illustrer chaque code utilisé.

Règles de formatage applicables aux patrons d'extraction

- Un patron par ligne.
- Utilisez le caractère # en début de ligne pour désactiver un patron.

L'ordre dans lequel vous répertoriez les patrons d'extraction est très important car une séquence de mots donnée est lue une seule fois par le moteur du programme d'extraction et est affectée au premier patron d'extraction pour lequel le moteur détecte une correspondance.

Définitions forcées

Lors de l'extraction d'informations à partir de vos enregistrements, le moteur du programme d'extraction analyse le texte et identifie la catégorie grammaticale pour chaque mot rencontré. Dans certains cas, un mot peut présenter différents rôles en fonction du contexte. Pour imposer une catégorie grammaticale particulière à un mot ou pour exclure complètement le mot du traitement, vous pouvez utiliser la section Définition imposée dans l'onglet Ressources avancées. [Pour plus d'informations, reportez-vous à la section À propos des ressources avancées sur p. 239.](#)

Pour imposer une catégorie grammaticale pour un mot donné, vous devez ajouter une ligne à cette section à l'aide de la syntaxe suivante :

```
terme:code
```

Table 11-2
Description de la syntaxe

Entrée	Description
terme	Un nom de terme.
code	Code à caractère unique représentant la catégorie grammaticale. Vous pouvez répertorier jusqu'à six codes de catégorie grammaticale différents par expression uniterme. En outre, vous pouvez arrêter l'extraction d'un mot en mots ou phrases composés en utilisant le code <code>s</code> (en minuscule), par exemple : <code>additional:s</code> .

Règles de formatage applicables aux définitions imposées.

- Une ligne par mot.
- Les termes ne peuvent pas contenir le signe « deux-points ».
- Utilisez le code `s` (lettre minuscule) en tant que code de catégorie grammaticale pour arrêter définitivement l'extraction d'un mot.
- Utilisez jusqu'à six codes de catégorie grammaticale par ligne. Les codes de catégories grammaticales pris en charge sont affichés dans la section Patrons d'extraction . [Pour plus d'informations, reportez-vous à la section Patrons d'extraction sur p. 248.](#)
- Utilisez l'astérisque (*) en tant que caractère générique à la fin d'une chaîne pour les correspondances partielles. Par exemple, si vous saisissez `add*:s`, des mots tels que `addition`, `additionnel`, `addenda` et `additif` ne sont jamais extraits en tant que mot ou dans le cadre d'un mot composé. Toutefois, si une correspondance de mot est explicitement

déclarée en tant que terme dans un dictionnaire compilé ou dans les définitions imposées, le mot sera extrait. Par exemple, si vous saisissez à la fois `add* : s` et `addenda : n`, le mot `addenda` sera extrait s'il est détecté dans le texte.

Abréviations

Lorsque le moteur du programme d'extraction traite le texte, il interprète généralement tout point rencontré comme marquant la fin d'une phrase. La plupart du temps, cela s'avère correct ; toutefois, cette gestion des signes de ponctuation que sont les points ne s'applique pas lorsque le texte contient des abréviations.

Si vous extrayez des termes de votre texte et que vous vous apercevez que certaines abréviations ont été mal gérées, vous devez déclarer ces dernières de manière explicite dans cette section.

Remarque : si l'abréviation apparaît déjà dans une définition de synonyme ou est définie en tant que terme dans une déclaration de types, il n'est pas nécessaire d'ajouter l'entrée d'abréviation ici.

Règles de formatage applicables aux abréviations

- Définissez uniquement une abréviation par ligne.

Exceptions pour le texte en japonais

Bien que le texte en japonais soit traité et exploité d'une manière similaire aux autres langues prises en charge dans IBM® SPSS® Text Analytics for Surveys, il existe de nombreuses différences. Les différences minimales sont décrites avec les instructions concernant toutes les autres langues dans cette documentation. Toutefois, certaines des différences les plus importantes sont décrites dans ce chapitre de l'annexe.

Extraction et catégorisation du texte japonais

Lors de l'exploitation d'un texte en japonais, le processus est similaire aux autres langues prises en charge. [Pour plus d'informations, reportez-vous à la section A propos de Text Mining dans le chapitre 1 sur p. 3.](#) Cependant, pour le japonais, on peut noter les différences suivantes.

Fonctionnement de l'extraction

Lors de l'extraction des principaux concepts et idées de vos réponses, IBM® SPSS® Text Analytics for Surveys s'appuie sur une analyse de texte linguistique. Cette approche a la même efficacité en temps et en argent que les systèmes statistiques. Mais elle offre un plus grand degré de précision tout en ne nécessitant que peu d'intervention humaine. L'analyse de texte linguistique repose sur un domaine d'étude appelé processus de langage naturel, également connu sous le nom de linguistique computationnelle.

Pour du texte en japonais, la différence entre les approches basées sur les statistiques et celles basées sur la linguistique pendant l'extraction peut être illustrée à l'aide du mot 沈む comme exemple. À l'aide de ce mot, nous pouvons rechercher des expressions telles que 日が沈む, traduites par *le soleil se couche*, ou 気分が沈む, traduite par *avoir le blues*. Si vous utilisez uniquement les techniques de statistiques, 日 (traduit par *soleil*), 気分 (traduit par *ressentir*), et 沈む (traduit par *en bas*) sont extraits séparément. Mais si vous utilisez l'analyseur de sentiment, qui utilise des techniques linguistiques, 日, 気分, et 沈む sont extraits, mais 気分が沈む (traduit comme *avoir le blues*) est également extrait et affecté au type <悪い - 悲しみ全般>. L'utilisation de techniques basées sur la linguistique à l'aide de l'analyseur de sentiment permet d'extraire plus d'expressions significatives. L'analyse et la capture des émotions permettent de lever toute ambiguïté dans le texte, et fait de l'exploration de texte linguistique la méthode la plus fiable, par définition.

Si vous comprenez le fonctionnement du processus d'extraction, vous êtes plus à même de prendre les décisions-clés lorsque vous affinez vos ressources linguistiques (bibliothèques, types, synonymes, etc.). Les principales étapes du processus d'extraction sont les suivantes :

- Conversion des données source en un format standard
- Identification des termes susceptibles d'être extraits
- Identification des classes d'équivalence et intégration des synonymes

- Affectation d'un type
- Indexation et, si nécessaire, mise en correspondance de patrons avec un deuxième analyseur

Etape 1. Conversion des données source en un format standard

Au cours de cette première étape, les données que vous importez sont converties dans un format uniforme pouvant être utilisé pour effectuer d'autres analyses. Cette conversion, qui s'effectue en interne, ne modifie pas les données d'origine.

Etape 2. Identification des termes susceptibles d'être extraits

Il est important de comprendre le rôle des ressources linguistiques dans l'identification des termes susceptibles d'être extraits lors de l'extraction linguistique. Les ressources linguistiques sont utilisées lors de chaque exécution d'une extraction. Elles se présentent sous la forme de ressources compilées, de bibliothèques et de modèles. Les bibliothèques comportent des listes de mots, des relations et des informations complémentaires qui permettent de spécifier ou d'affiner l'extraction. Vous ne pouvez pas afficher ni éditer les ressources compilées. Toutefois, les autres ressources peuvent être modifiées dans l'Editeur de ressources.

Les ressources compilées sont des composants internes essentiels du moteur du programme d'extraction de SPSS Text Analytics for Surveys. Ces ressources comportent un dictionnaire général qui répertorie les formes de base avec un code concernant la catégorie grammaticale (nom, verbe, adjectif, etc.). Les ressources comprennent également des types intégrés et réservés qui permettent d'affecter de nombreux termes extraits aux types suivants : <地名>, <組織>, ou <人名>. [Pour plus d'informations, reportez-vous à la section Types disponibles pour du texte en japonais sur p. 261.](#)

Outre ces ressources compilées, plusieurs bibliothèques sont fournies avec le produit et peuvent être utilisées pour compléter les types et les définitions de concept figurant dans les ressources compilées, ainsi que pour proposer des synonymes. Ces bibliothèques —et toutes les bibliothèques personnalisées que vous créez— comprennent plusieurs dictionnaires : déclarations de types, dictionnaires de synonymes et dictionnaires d'exclusions. [Pour plus d'informations, reportez-vous à la section Modification des ressources pour du texte en japonais sur p. 256.](#)

Une fois les données importées et converties, le moteur du programme d'extraction commence à identifier les termes susceptibles d'être extraits. Ces termes sont des mots ou des groupes de mots qui permettent d'identifier des concepts du texte. Pendant le traitement du texte, les mots uniques (**unitermes**) et les mots composés (**multitermes**) sont identifiés à l'aide d'extracteurs de patrons de catégorie grammaticale. Par exemple, l'expression multiterme 青森りんご, qui suit le patron de catégorie grammaticale <地名> + <名詞> possède deux composants. Par conséquent, les mots-clés de sentiment susceptibles d'être extraits sont identifiés à l'aide de l'analyse des liens du texte de sentiment.

Imaginons par exemple que vous ayez le texte suivant en japonais : 写真が新鮮で良かった. Dans ce cas, le moteur d'extraction affecte le type de sentiment 良い - 褒め・賞賛, après avoir mis en correspondance (品物) + が + 良い à l'aide des règles de liens du texte de sentiment.

Remarque : les termes qui figurent dans le dictionnaire général compilé susmentionné constituent la liste de tous les mots qui risquent de s'avérer inintéressants ou de présenter une ambiguïté linguistique en tant qu'unitermes. Ces mots sont exclus de l'extraction lorsque vous identifiez les unitermes. Ils font toutefois l'objet d'une réévaluation lorsque vous déterminez les catégories

grammaticales ou que vous recherchez des mots composés (expressions multitermes) plus longs, susceptibles d'être extraits.

Etape 3. Identification des classes d'équivalence et intégration des synonymes

Une fois les expressions unitermes et multitermes susceptibles d'être extraites identifiées, le logiciel utilise un dictionnaire de normalisation afin d'identifier des classes d'équivalence. Une classe d'équivalence désigne la forme de base d'une phrase ou la forme unique de deux variantes d'une même phrase. L'affectation d'expression à des classes d'équivalence a pour objectif de veiller à ce que, par exemple, *effet secondaire* et 副作用 ne soient pas traités comme des concepts distincts. Pour déterminer quel concept utiliser pour la classe d'équivalence (c'est-à-dire si *effet secondaire* ou 副作用 est utilisé en tant que terme principal), le moteur du programme d'extraction applique les règles suivantes dans l'ordre indiqué ci-dessous :

- Forme définie par l'utilisateur dans une bibliothèque.
- La forme la plus fréquente, comme définie par les ressources précompilées.

Etape 4. Affectation d'un type

Des types sont ensuite affectés aux concepts extraits. Un type correspond à un regroupement sémantique de concepts. Les ressources compilées et les bibliothèques sont utilisées au cours de cette étape. Les types comprennent des éléments tels que des concepts de niveau supérieur, des mots positifs et négatifs, des prénoms, des lieux, des organisations, etc. [Pour plus d'informations, reportez-vous à la section Déclarations de types dans le chapitre 10 sur p. 220.](#)

Les ressources en japonais contiennent un ensemble de types distinct. [Pour plus d'informations, reportez-vous à la section Types disponibles pour du texte en japonais sur p. 261.](#)

Etape 5. Indexation et mise en correspondance des patrons avec extraction d'événements

L'indexation de l'ensemble des enregistrements s'effectue en définissant un pointeur entre une position de texte et le terme représentatif de chaque classe d'équivalence. Cela suppose que toutes les instances de forme fléchies d'un concept susceptible d'être extrait sont indexées en tant que forme de base susceptible d'être extraite. La fréquence globale est calculée pour chaque forme de base.

SPSS Text Analytics for Surveys peut non seulement détecter les types et les concepts, mais également les relations qui existent entre eux. Plusieurs algorithmes et bibliothèques sont fournis avec ce produit ; ils permettent d'extraire les patrons de relations d'analyse de liens du texte entre les types et les concepts. Ils s'avèrent particulièrement utiles lorsque vous tentez de découvrir des opinions spécifiques (par exemple, des réactions sur des produits).

Fonctionnement de l'extraction secondaire

Lorsque vous exécutez une extraction sur du texte japonais, vous obtenez automatiquement des concepts à partir des mots-clés de base et de 8 types de base, y compris 人名, 地名, 組織名, 名詞, 形容詞, 動詞, 形容動詞, et その他. Cependant, afin de profiter pleinement des ressources par défaut fournies pour le texte en japonais, vous devez sélectionner un des analyseurs secondaires suivants : Sentiment ou Dépendance.

Choisir un deuxième analyseur vous permet également d'extraire des patrons d'analyse des liens du texte et de connaître les relations entre les termes du texte.

Analyse secondaire. Lorsqu'une extraction est lancée, l'extraction des mots-clés de base est effectuée à l'aide de l'ensemble de types par défaut. [Pour plus d'informations, reportez-vous à la section Types disponibles pour du texte en japonais sur p. 261.](#) Mais lorsque vous sélectionnez un analyseur secondaire, vous pouvez obtenir des concepts plus nombreux et plus riches car l'extracteur inclura désormais des verbes à particules et des auxiliaires comme faisant partie du concept. Par exemple, supposons que nous avons une phrase 肩の荷が下りた, traduite par “*Cam'a enlevé un gros poids*”. Dans cet exemple, l'extraction des mots-clés de base peut extraire chaque concept séparément comme suit : 肩 (*poids*), 荷 (*gros*), 下りる (*a enlevé*), mais la relation entre ces mots n'est pas extraite. Cependant, si vous avez appliqué l'analyse de sentiment, vous pouvez extraire des concepts plus riches relatifs à un type de sentiment comme le concept =肩の荷が下りた, qui est traduit par “*avoir enlevé un gros poids*”, affecté au type <良い-安心>. Dans le cas d'une analyse de sentiment, un grand nombre de types supplémentaires est également inclus. De plus, choisir un analyseur secondaire vous permet également de générer des résultats d'analyse des liens du texte.

Remarque : Lorsqu'un analyseur secondaire est appelé, le processus d'extraction nécessite plus de temps. [Pour plus d'informations, reportez-vous à la section Fonctionnement de l'extraction secondaire sur p. 253.](#)

- **Analyse des dépendances** Choisir cette option génère des particules étendues pour les concepts d'extraction du type de base et de l'extraction des mots-clés. Vous pouvez également obtenir des résultats de patrons plus riches avec l'analyse de dépendance des liens du texte (TLA).
- **Analyse des sentiments.** Choisir cet analyseur génère l'extraction de concepts supplémentaires et, le cas échéant, l'extraction de résultats de patrons TLA. En plus des types de base, vous bénéficiez également de plus de 80 types de sentiments, notamment 嬉しい, 吉報, 幸運, 安心, 幸福, etc. Ces types permettent de découvrir des concepts et des patrons dans le texte grâce à l'expression des émotions, des sentiments et des opinions. Ce sont trois options qui dictent la cible de l'analyse des sentiments : Tous les sentiments, Sentiment représentatif uniquement et Conclusions uniquement.

Options de l'analyse des sentiments

Lorsque vous travaillez sur du texte en japonais et vous pouvez choisir d'extraire des concepts et des types supplémentaires à l'aide de l'analyseur de sentiments. Cet analyseur comprend plus de 80 types supplémentaires afin de vous aider à extraire les opinions, les sentiments et les émotions de vos données textuelles. En outre, lorsque vous choisissez l'analyse de sentiments comme analyseur secondaire, vous devez également sélectionner l'une des options suivantes pour indiquer au moteur d'extraction quel sentiment extraire :

- **Tous les sentiments**
- **Uniquement le sentiment représentatif**
- **Uniquement les conclusions**

Au cours de l'extraction, l'analyse de sentiments commence par diviser un enregistrement en propositions, chacune contenant un prédicat. Par exemple, le texte, « 月になったがまだ寒い。 », qui est traduit « *On est en avril, mais il fait encore froid.* », est interprété comme 2 propositions par l'analyseur en dépit du fait qu'il ne contient qu'un seul caractère d'arrêt. . Chaque proposition est alors examinée par le moteur d'extraction afin de voir si elle correspond à l'option sélectionnée.

Examinons les trois options à l'aide de l'exemple de texte : “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。”. Ce texte est traduit : “*Une domestique n'était pas aimable, mais la pièce était grande et tout à fait satisfaisante. J'ai aussi apprécié le dîner* ». Au cours de l'extraction, le texte original est scindé en propositions qui sont :

- 案内してくれた仲居さんは無愛想だったが、, qui signifie « *Une des domestiques n'était pas aimable mais* »
- 部屋は広くて申し分なかった。 , qui signifie « *La salle était grande et assez plaisante* »
- 夕食も満足。 , qui signifie « *J'ai aussi apprécié le dîner* »

Tous les sentiments

Cette option extrait tous les sentiments, opinions et émotions qui correspondent aux ressources et aux règles des liens du texte de sentiment. Dans notre exemple, les concepts suivants peuvent être extraits l'échantillon de texte.

Table A-1

Résultats possibles pour l'échantillon en utilisant l'option « Tous les sentiments »

Concept	Type
仲居さんは無愛想だった	<悪い - 対応が不親切>
部屋は広くて	<良い - 満足>
申し分なかった	<良い - 満足>
満足	<良い - 満足>

Remarque : Dans le tableau précédent, les deuxième et troisième lignes indiquent la façon dont l'extracteur peut obtenir deux concepts à partir de la même proposition.

Uniquement le sentiment représentatif

Cette option n'extrait que les opinions et les émotions exprimées les plus représentatives dans chaque proposition. S'il existe plusieurs opinions et émotions dans le texte, un algorithme est appliqué. Cet algorithme tente de déterminer l'importance des sentiments trouvés ainsi que la position des mots dans une proposition. Dans certains cas où deux mots-clés de sentiments ayant la même importance sont trouvés, le mots-clé de sentiments en dernière position dans la proposition est extrait plutôt que le premier.

部屋は広くて qui est traduit par *la salle était grande*, n'est pas extrait car le second *申し分なかった* est considéré comme ayant plus d'importance que le premier *部屋は広くて* dans cette proposition, lorsque l'algorithme interne et la position des mots sont appliqués.

Table A-2

Résultats possibles pour le texte en utilisant l'option « Uniquement les sentiments représentatifs »

Concept	Type
仲居さんは無愛想だった	<悪い - 対応が不親切>
申し分なかった	<満足>
満足	<満足>

Uniquement les conclusions

Cette option force l'extracteur à identifier et à extraire un mot clé de sentiment comme représentant la conclusion de la totalité de l'enregistrement. Tous les textes n'ont pas de conclusion, de sorte que dans certains cas, il est possible que rien ne soit extrait avec cette option pour une partie de texte donnée. En outre, plus l'enregistrement est long, plus l'analyse a du mal à identifier la conclusion principale. Bien que cela soit rare, il est possible que plusieurs conclusions soient extraites.

満足, traduit par *satisfait*, est considéré comme la conclusion principale des sentiments exprimés dans le texte.

Table A-3

Résultats possibles pour le texte en utilisant l'option « Uniquement les conclusions »

Concept	Type
満足	<満足>

Fonctionnement de la catégorisation

Vous disposez de plusieurs techniques de création de catégories. Etant donné que chaque ensemble de données est unique, le nombre de techniques et leur ordre d'application peuvent varier. Votre interprétation des résultats pouvant être différente de celle d'une autre personne, vous pouvez être amené à essayer plusieurs techniques de manière à déterminer celle qui donne les meilleurs résultats pour vos données textuelles.

Dans ce manuel, la **création de catégories** fait référence à la génération de définitions et de classification de catégories à l'aide d'une ou de plusieurs techniques intégrées et la **catégorisation** fait référence au scoring, ou à l'étiquetage, processus par lequel des identificateurs uniques (nom/ID/valeur) sont affectés aux définitions de catégorie de chaque enregistrement.

Pendant la création de catégories, les concepts et les types qui ont été extraits sont utilisés en tant que blocs de construction de vos catégories. Lorsque vous créez des catégories, les enregistrements sont automatiquement affectés aux catégories s'ils contiennent du texte qui correspond à un élément d'une définition de catégorie.

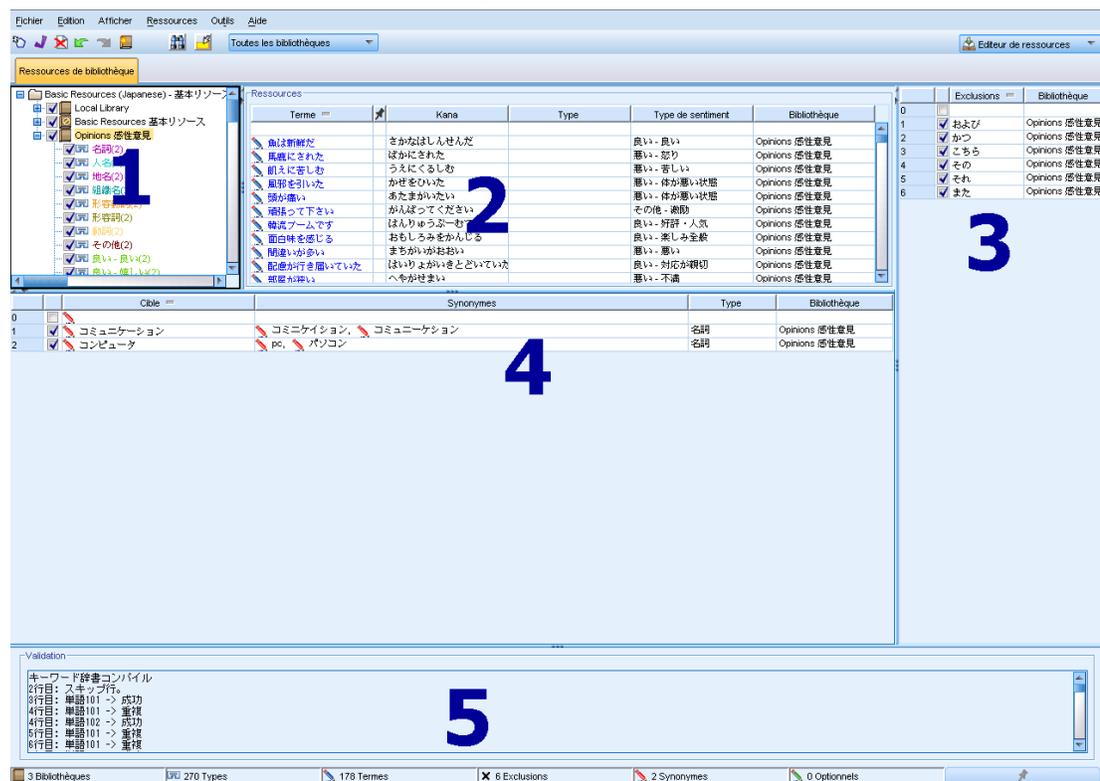
IBM® SPSS® Text Analytics for Surveys vous propose plusieurs techniques automatisées de classification supervisée qui vous permettent de catégoriser rapidement vos enregistrements. Chaque technique convient à certains types de données et de situation. Cependant, il est souvent judicieux de combiner des techniques dans la même analyse afin de capturer le maximum d'enregistrements. Aussi est-il possible de voir un concept figurer dans plusieurs catégories ou de rencontrer des catégories redondantes.

Modification des ressources pour du texte en japonais

Depuis la version 4 de IBM® SPSS® Text Analytics for Surveys, un nouveau modèle et un package d'analyse de texte (TAP) sont disponibles pour la langue japonaise. Vous pouvez effectuer les modifications des ressources en ajoutant et éditant des termes pour les personnaliser en fonction de vos données. Le package d'analyse de texte contient également une ensemble de catégories composé de catégories représentant des sentiments positifs, négatifs et des sentiments contextuels/génériques.

Vous pouvez utiliser vos ressources dans l'Editeur de ressources. Les éditeurs fonctionnent de manière similaire pour toutes les langues ; il y a toutefois des différences significatives pour le texte en japonais, comme décrit ici.

Figure A-1
Vue Editeur de ressources pour du texte en japonais



Les points suivants mettent l'accent sur les différences majeures qui existent lors de l'utilisation de ressources en japonais. Pour une description générale des quatre panneaux principaux dans l'onglet Ressources de bibliothèque, consultez [Interface de l'éditeur](#) sur p. 195.

1. Panneau des bibliothèques. Cette zone, située dans le coin supérieur gauche, fonctionne globalement comme pour les autres langues. Il existe cependant quelques différences telles que l'impossibilité de créer de nouveaux types ou d'en renommer. [Pour plus d'informations, reportez-vous à la section Utilisation des bibliothèques dans le chapitre 9 sur p. 206.](#)

2. Panneaux des termes pour les déclarations de types. Ce panneau, situé à droite du panneau de l'arborescence des bibliothèques, est différent pour le texte japonais. En plus du nom du terme, vous pouvez également ajouter le nom en kana et sélectionner un ou deux types auxquels vous souhaitez associer ce terme. Cependant, vous ne pouvez pas générer de formes fléchies des termes ni affecter des options de mise en correspondance pour des termes japonais comme vous le pourriez dans les autres langues. [Pour plus d'informations, reportez-vous à la section Panneau Arborescence des bibliothèques japonaises, panneau des types et panneau des termes sur p. 258.](#)

3. Panneau Dictionnaire de substitutions/synonymes. Dans les ressources de texte japonais, vous trouverez un onglet Synonyme dans lequel vous pouvez définir tous les synonymes pour vos ressources. Dans l'onglet Synonyme, il existe une colonne supplémentaire Type dans laquelle vous devez désigner un type pour les synonymes entrés. [Pour plus d'informations, reportez-vous à la section Utilisation du dictionnaire des synonymes pour du texte japonais sur p. 265.](#) *Remarque :* L'onglet Eléments optionnels n'est pas disponible car il ne s'applique pas au texte japonais.

4. Panneau Dictionnaire d'exclusions. Il n'y a pas de différences dans ce panneau pour les ressources en japonais en-dehors du fait que le caractère générique * n'est pas pris en charge.

5. Panneau Validation. Pour le texte japonais, il existe un panneau de validation supplémentaire servant à vérifier vos ressources avant l'extraction. Lors de l'extraction du texte japonais, le moteur d'extraction recompile automatiquement les ressources si les modifications sont détectées avant le début du processus d'extraction. Pour éviter des erreurs potentielles pendant l'extraction, vous pouvez recompiler et valider les ressources avant l'extraction afin de pouvoir corriger les erreurs rencontrées. [Pour plus d'informations, reportez-vous à la section Validation et compilation des ressources en japonais sur p. 267.](#)

Remarque : ces onglets ne sont pas disponibles car il n'existe pas de ressources avancées ou de règles des liens du texte pouvant être modifiées pour du texte japonais.

Panneau Arborescence des bibliothèques japonaises, panneau des types et panneau des termes

La façon dont vous travaillez avec les bibliothèques et les types pour les ressources en japonais est très similaire à celle des autres langues. [Pour plus d'informations, reportez-vous à la section Déclarations de types dans le chapitre 10 sur p. 220.](#)

Mais il existe quelques différences majeures, notamment :

- Les ressources en japonais contiennent un ensemble de types distinct. [Pour plus d'informations, reportez-vous à la section Types disponibles pour du texte en japonais sur p. 261.](#)
- Les types ne peuvent pas être créés ni renommés, mais leurs propriétés peuvent être modifiées. [Pour plus d'informations, reportez-vous à la section Modification des propriétés de types japonais sur p. 264.](#)
- Vous pouvez ajouter et modifier des termes, y compris la spécification d'un nom en kana pour un terme, ainsi que l'attribution à un type et à un type de sentiment secondaire. [Pour plus d'informations, reportez-vous à la section Panneau Arborescence des bibliothèques japonaises, panneau des types et panneau des termes sur p. 258.](#)

Le panneau de l'arborescence de bibliothèques affiche les bibliothèques ainsi que les déclarations de types. Si vous sélectionnez une bibliothèque ou un type dans le panneau de gauche, un panneau des termes affiche à droite les termes des bibliothèques ou des déclarations de types sélectionnées. Vous pouvez ajouter des termes directement à une déclaration de types : soit directement dans le panneau de termes, soit par l'intermédiaire de la boîte de dialogue Ajouter des termes. Les termes que vous ajoutez peuvent être des mots simples ou composés. La ligne vide, située en haut de la liste, vous permet d'ajouter de nouveaux termes à tout moment.

Lorsque vous définissez un terme dans une déclaration de types, il est considéré comme un nom par défaut et est automatiquement affecté au type <名詞>. Mais vous pouvez modifier le type en un autre type de base comme <動詞>, <形容詞>, <地名> etc. Si le moteur d'extraction trouve ce terme dans la même catégorie grammaticale que le type auquel vous l'avez affecté dans la colonne Type, il sera alors affecté à ce type et extrait. Vous pouvez également affecter ce terme à un des types de sentiment dans la colonne Type de sentiment. Ensuite, lorsque vous utilisez l'analyseur secondaire de sentiments, le texte est traité une deuxième fois afin de rechercher des termes et de les affecter aux types de sentiment. De plus, si vous définissez à la fois un type de sentiment et un type de base, et que le moteur d'extraction découvre que ce terme correspond aux deux types lorsque l'analyse secondaire de sentiment a été effectuée, alors le type de sentiment est prioritaire et est affiché dans le panneau des résultats d'extraction et dans les résultats de l'analyse des liens du texte. Par exemple, si un verbe a été extrait comme type de verbe <動詞> et est également un type positif comme « aimé » alors ce terme sera affiché comme appartenant au type positif dans l'interface car la capture de sentiments est souvent plus intéressante qu'une simple catégorie grammaticale.

Figure A-2
Panneaux Bibliothèque et Terme pour les ressources en japonais

Terme	Kana	Type	Type de sentiment	Bibliothèque
魚は新鮮だ	さかなはしんせんだ		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
頭が悪い	あたまが悪い		悪い - 体が悪い状態	Opinions 感性意見
頑張って下さい	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	おもしろみを感じる		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	はいりよかいきとどいていた		良い - 対応が親切	Opinions 感性意見
部屋が狭い	へやがせまい		悪い - 不満	Opinions 感性意見
道が混む	みちがこむ		悪い - 怒り全般	Opinions 感性意見
運勢が悪い	うんせいがわるい		悪い - 不運	Opinions 感性意見
運が良い	うんがよい		良い - 幸運	Opinions 感性意見
運路に迷っている	しんろにまよっている		悪い - 悩み	Opinions 感性意見
連絡がない	れんらくがない		悪い - 回答なし	Opinions 感性意見
逝ってよし	いってよし		悪い - 働がらせ	Opinions 感性意見

Table A-4
Descriptions des colonnes du panneau des termes

Nom de colonne	Description de colonne
Terme	Entrez des mots simples ou composés dans la cellule. La couleur d'affichage du terme dépend de la couleur du type dans lequel le terme est enregistré ou a été imposé. Vous pouvez changer les couleurs des types dans la boîte de dialogue Propriétés de type. Pour plus d'informations, reportez-vous à la section Modification des propriétés de types japonais sur p. 264. En général le terme est écrit en kanji, mais il peut aussi comprendre des kana. Important ! La saisie de verbes utilisant des caractères Katakana n'est pas prise en charge.
Forcer	Cliquer sur une icône de punaise et la placer dans cette cellule indique au moteur d'extraction d'ignorer les autres occurrences de ce même terme dans les autres bibliothèques. Pour plus d'informations, reportez-vous à la section Ajout des termes forcés dans le chapitre 10 sur p. 228. Ceci fonctionne de la même manière pour toutes les langues.
Kana	Entrez l'orthographe kana du nom du terme Kanji.
Type	Sélectionnez le nom du type de base auquel ce terme doit être affecté. Pour plus d'informations, reportez-vous à la section Types disponibles pour du texte en japonais sur p. 261.

Nom de colonne	Description de colonne
Type de sentiment	Si une analyse secondaire est réalisée, sélectionnez le nom du type de sentiment auquel le terme doit être attribué. Pour plus d'informations, reportez-vous à la section Types disponibles pour du texte en japonais sur p. 261.
Bibliothèque	Sélectionne la bibliothèque dans laquelle votre terme est stocké. Pour transférer un terme dans un autre type de l'arborescence, faites-le glisser et déposez-le sur l'autre bibliothèque.

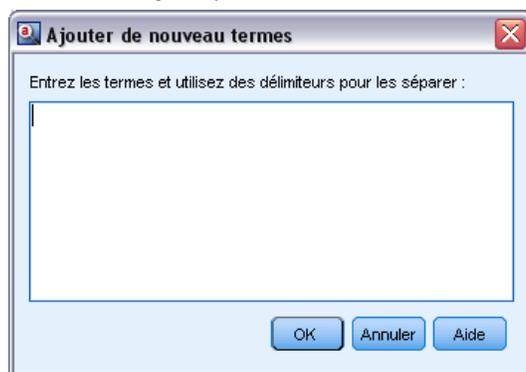
Pour ajouter un terme simple à une déclaration de types

- ▶ Dans l'arborescence de la bibliothèque, sélectionnez la déclaration de types à laquelle vous voulez ajouter le terme.
- ▶ Dans la liste de termes du panneau central, entrez votre terme dans la première cellule vide disponible et définissez les options souhaitées pour ce terme.

Pour ajouter plusieurs termes à une déclaration de types

- ▶ Dans l'arborescence de la bibliothèque, sélectionnez la déclaration de types à laquelle vous voulez ajouter des termes.
- ▶ A partir des menus, sélectionnez Outils> Nouveaux termes. La boîte de dialogue Ajouter des nouveaux termes apparaît.

Figure A-3
Boîte de dialogue Ajouter des nouveaux termes



- ▶ Entrez les termes que vous souhaitez ajouter à la déclaration de types sélectionnée en saisissant les termes ou en collant un ensemble de termes. Si vous entrez plusieurs termes, séparez-les au moyen du séparateur défini dans la boîte de dialogue Options ou ajoutez un terme par nouvelle ligne. [Pour plus d'informations, reportez-vous à la section Définition des options dans le chapitre 2 sur p. 19.](#)
- ▶ Cliquez sur OK pour ajouter les termes à la déclaration. La boîte de dialogue se ferme et les nouveaux termes apparaissent dans la déclaration.

Types disponibles pour du texte en japonais

Vous ne pouvez pas ajouter de nouveaux types pour les ressources en japonais, cependant vous pouvez ajouter ou supprimer des termes pour celles-ci. Les tableaux suivant incluent l'ensemble des types japonais actuellement disponibles.

Types pour l'extraction de base

À chaque fois qu'une extraction est lancée, les types suivants sont utilisés.

Table A-5

Types pour l'extraction de base

Types	Description
名詞	Mots faisant référence à des choses, comme « voiture » et « film ». Néanmoins, les noms de personnes, de lieux et d'organisations sont classés dans des catégories distinctes.
人名	Noms correspondant à des noms de personnes spécifiques comme « Tokugawa » et « Ieyasu ». Les combinaisons de prénoms et de noms de famille, comme « Tokugawa Ieyasu » sont également des noms propres.
地名	Noms comme « Toyko » et « Londres » faisant référence à un lieu précis.
組織名	Noms faisant référence à des entreprises et des organisations spécifiques, comme « La Fédération des Organismes Economiques ».
形容動詞	Mots comme « calme (shizuka) » qui décrivent les caractéristiques ou l'état d'une chose et qui peuvent être utilisés dans des phrases comme « pas [adjectif] (~de nai) » et « [adjectif] chose (~ na koto) ».
形容詞	Mots comme « amusant (tanoshi) » qui décrivent les caractéristiques ou l'état d'une chose et qui peuvent être utilisés dans des phrases comme « devenir [adjectif] (~ku naru) » et « une chose [adjectif] (~ i koto) ».
動詞	Les mots qui décrivent le mouvement ou l'action, y compris les verbes de type I (base verbale consonne), de type II (base verbale voyelle) et les verbes irréguliers (sagyō henkaku et kagyō henkaku).
その他	Mots comme les adverbes, les adjectifs pronominaux, les conjonctions et les interjections : par exemple, « calme », « quel que soit », « ensuite » et « merci ».

Types pour l'analyse des sentiments

À chaque fois que vous sélectionnez l'analyseur secondaire pour l'extraction des sentiments, vous obtenez un grand nombre de types en plus des 8 types de base.

Table A-6

Types pour l'analyse des sentiments

Types	Description
良い - 良い	Expressions de choses positives pouvant être classifiées comme « bonnes ».
良い - 嬉しい	Décrit un événement désirable qui produit une stimulation agréable.
良い - 吉報	Décrit un événement plaisant qui ne peut être possible qu'avec des efforts considérables.
良い - 幸運	Décrit un événement heureux qui ne peut être possible que par hasard ou en raison d'une coïncidence incroyable.
良い - 快い	Suggère que quelque chose est un stimulus ou un environnement qui déclenche une sensation physiologique agréable.

Types	Description
良い - 体が良い状態	Décrit un état dans lequel le corps n'est ni malade, ni blessé, ni fatigué ou un état dans lequel l'état physique s'améliore.
良い - 安心	Suggère que quelqu'un est calme et ne risque ni blessure ni dérangement.
良い - 幸福	Indique que quelqu'un a obtenu des conditions favorables ou un effet positif grâce à sa propre action ou aux circonstances de sa naissance.
良い - 満足	Décrit un événement désirable qui apaise l'esprit.
良い - 美味しい	Indique qu'une nourriture a bon goût.
良い - 効果が満足	Implique qu'une chose a produit l'effet désiré.
良い - 感動	Suggère que l'importance, la signification ou la valeur de quelque chose est extrêmement positive.
良い - 感謝	Suggère que quelqu'un considère les actions de quelqu'un autre de manière positive.
良い - 祝福	Exprime l'idée que la situation d'une autre personne est favorable (à un degré acceptable pour l'interlocuteur).
良い - 喜び全般	Autres événements positifs avec peu de rapport avec l'interlocuteur.
良い - 楽しい	Indique ou anticipe des activités en lien avec les relations amicales, les loisirs et le divertissement.
良い - 可笑しい	Indique que quelque chose a une qualité humoristique qui génère une stimulation plaisante.
良い - 笑い	Exprime un sourire ou un rire provoqué par quelque chose de positif et/ou de drôle.
良い - 期待	Prédit qu'un événement positif se produira dans le futur.
良い - 楽しみ全般	Autres événements agréables et/ou activités/comportements positifs avec peu de rapport avec l'interlocuteur.
良い - 金額への賞賛	Suggère que, du point de vue de l'acheteur, quelque chose a une valeur monétaire intéressante.
良い - 対応が早い	Suggère qu'un service a été fourni ou terminé à temps.
良い - 対応が親切	Suggère que l'attitude ou le comportement du fournisseur de service était plein de sollicitude.
良い - 説明が良い	Exprime l'idée que le type et/ou la quantité d'informations, et/ou le moyen de les fournir, est approprié.
良い - 対応への賞賛	Points de vue, autres que ceux cités précédemment, qui font l'éloge du fournisseur de service.
良い - 褒め・賞賛	Points de vue, autres que ceux cités précédemment, qui font l'éloge des caractéristiques, capacités et/ou fonctionnement d'un objet.
良い - 好き	Exprime le désir de posséder ou de se rapprocher de quelque chose.
良い - 入会希望	Décrit le désir de faire partie ou de rester membre d'un groupe.
良い - 買いたい	Suggère que quelqu'un veut ou prévoit d'utiliser de l'argent pour obtenir quelque chose.
良い - 好評・人気	Indique que le nombre de personnes qui souhaitent ou apprécient quelque chose a dépassé un certain niveau.
良い - 売れた	Indique la présence de personnes qui achètent quelque chose ou que le nombre ou la valeur des achats dépassent un certain niveau.
悪い - 悪い	Expressions de choses négatives pouvant être classifiées comme « mauvaises ».
悪い - 怒り	Une sensation précise de colère ressentie lorsque quelque chose ne se passe pas comme prévu.
悪い - 批判	Exprime l'idée qu'une autre personne a échoué à faire le bon choix.
悪い - お叱り	Mots ou actions qui intimident une autre personne pour pouvoir faire selon son désir.

Types	Description
悪い - 誹謗・中傷	Mots utilisés pour démontrer une opinion excessivement négative de quelqu'un d'autre.
悪い - 軽蔑	Suggère qu'une autre personne manque de caractère, de capacités et/ou d'autres qualités.
悪い - 恨み	Exprime la revanche ou le ressentiment face à un préjudice provoqué par une autre personne.
悪い - 嫌がらせ	Mots utilisés dans le but d'empêcher une communication.
悪い - 不満	Un sentiment déplaisant provoqué par l'incapacité à obtenir une chose ou un état désiré.
悪い - 不味い	Indique qu'une nourriture a mauvais goût.
悪い - 効果が不満	Suggère que quelque chose n'a pas produit l'effet escompté.
悪い - 金額が不満	Suggère que, du point de vue de l'acheteur, la valeur monétaire d'une chose n'est pas acceptable.
悪い - 対応への不満	Suggère que le fournisseur d'un service est en faute.
悪い - 対応が遅い	Suggère qu'un service n'a pas été effectué / terminé dans les temps ou qu'un service doit encore être effectué.
悪い - 対応が不親切	Indique un sentiment désagréable provoqué par l'attitude ou le comportement d'un fournisseur de service.
悪い - 説明が悪い	Exprime l'idée que le type et/ou la quantité d'informations, et/ou le moyen de les fournir, n'est pas approprié.
悪い - 返答なし	Indique que le fournisseur de service ne parvient pas à proposer la réponse désirée, même si la situation en demande une.
悪い - 不快	Suggère que quelque chose est un stimulus ou un environnement qui déclenche une sensation physiologique négative.
悪い - 怒り全般	Sentiments de colère autres que ceux cités précédemment. Colère générale ressentie par l'organisation ou l'entreprise de l'interlocuteur, ou description des événements provoqués par cette colère.
悪い - 悲しい	Un sentiment déplaisant spécifique ressenti lorsque quelqu'un perd ou ne peut pas obtenir quelque chose.
悪い - 凶報	Exprime l'idée qu'un certain objectif ne peut pas être atteint malgré de gros efforts.
悪い - 不運	Indique un résultat négatif provoqué par de la malchance, et qui n'est pas de sa faute.
悪い - ショック	Suggère que quelqu'un est contrarié par quelque chose d'imprévu ou de négatif et n'est pas capable de trouver une solution appropriée.
悪い - 残念	Un sentiment négatif ressenti lorsque quelque chose qui est prévu n'arrive pas.
悪い - 落胆	Un état dans lequel quelqu'un est envahi par un sentiment de malheur et de déception.
悪い - 諦め	Suggère qu'une chose négative, expérimentée par l'interlocuteur ou une autre personne ne peut pas s'améliorer.
悪い - 後悔	Exprime l'idée que dans le passé, quelqu'un n'est pas parvenu à faire le bon choix, même si ce choix était disponible.
悪い - 謝罪	Indique que l'interlocuteur sait qu'il a causé du tort à quelqu'un.
悪い - 淋しい	Exprime l'idée que le contact avec les autres est rare ou que ceux avec qui le contact est possible sont rares.
悪い - 哀れみ	Exprime l'idée que la situation de quelqu'un d'autre est bien pire que celle de l'interlocuteur.
悪い - 悩み	Indique que quelqu'un doit faire un choix mais est incapable de choisir parmi les options disponibles.

Types	Description
悪い - 困っている	Exprime l'idée qu'il n'existe pas de réponse adaptée à une situation demandant une action.
悪い - 苦しい	Exprime un état psychologique déplaisant dans lequel quelqu'un est incapable d'agir normalement en raison de causes externes ou en raison de ses propres erreurs.
悪い - 体が悪い状態	Décrit un état dans lequel le corps est malade, blessé, et/ou fatigué ou un état dans lequel l'état physique ne s'améliore pas.
悪い - 不安	Exprime l'idée que quelque chose de positif peut s'arrêter ou peut ne pas répondre aux attentes.
悪い - 恐怖	Suggère qu'une chose pourrait causer du tort ou blesser quelqu'un.
悪い - 悲しみ全般	Sentiments de tristesse autres que ceux mentionnés précédemment, comme la tristesse générale à propos de quelque chose qui n'a pas été précisé.
悪い - 嫌い	Indique que quelqu'un souhaite éloigner quelque chose ou s'en éloigner.
悪い - 退会希望	Décrit le désir de quitter un groupe ou de ne pas en devenir membre.
悪い - 買いたくない	Suggère que quelqu'un ne souhaite pas quelque chose ou n'a pas l'intention de payer pour cela.
悪い - 不評・不人気	Indique que le nombre de personnes aimant une certaine chose n'a pas atteint un certain niveau ou qu'il y a trop de monde ayant un sentiment négatif envers cette chose.
悪い - 売れていない	Indique l'absence de personnes qui achètent quelque chose ou que le nombre ou la valeur des achats n'a pas atteint un certain niveau.
その他 - 疑問	Expressions qui demandent des informations nécessitant un examen plus poussé ou des pensées supplémentaires de la part d'une autre personne.
その他 - 問い合わせ	Expressions qui demandent des informations qu'une autre personne possède déjà.
その他 - 要望	Expressions qui ordonnent à une autre personne (lorsque l'autre personne est directement en faute ou d'un rang inférieur à celui de l'interlocuteur) de résoudre un problème.
その他 - 提案・忠告	Expressions qui ordonnent à une autre personne (lorsque l'autre personne est directement en faute ou d'un rang inférieur à celui de l'interlocuteur) de mieux se comporter.
その他 - お願い	Expressions qui ordonnent à une autre personne, lorsque l'autre personne n'est pas en faute ou n'est pas d'un rang inférieur à celui de l'interlocuteur, de faire quelque chose.
その他 - 激励	Expressions qui encouragent une autre personne ou descriptions de comportement encourageant.
その他 - 勧誘	Expressions qui ordonnent à une autre personne de faire quelque chose avec l'interlocuteur.
その他 - 驚き	Exprime l'idée que la soudaineté ou la taille d'un événement transcende le jugement/la compréhension rationnelle.
評価なし - 評価なし	Pas d'expression de jugement.

Modification des propriétés de types japonais

Bien que vous ne puissiez pas créer de type dans les ressources en japonais, vous pouvez afficher et modifier les propriétés des types. Notez que certaines options, telles que l'option de mise en correspondance ou les formes fléchies, ne s'appliquent pas au texte en japonais.

Figure A-4
Boîte de dialogue Propriétés du type pour les ressources en japonais



Nom. Le nom de la déclaration de types.

Ajouter à. Ce champ indique la bibliothèque dans laquelle vous allez créer votre nouvelle déclaration de types.

Couleur. Ce champ vous permet de distinguer les résultats de ce type dans l'interface. Si vous sélectionnez Par défaut, la couleur par défaut du type est également utilisée pour cette déclaration de types. La couleur par défaut se configure dans la boîte de dialogue des options. [Pour plus d'informations, reportez-vous à la section Options : Onglet Affichage dans le chapitre 2 sur p. 21.](#) Si vous sélectionnez Personnalisé, sélectionnez une couleur dans la liste déroulante.

Annotations. Ce champ est facultatif et sert à entrer des commentaires ou des descriptions.

Pour afficher ou modifier les propriétés de type

- ▶ Sélectionnez le type dont vous souhaitez afficher les propriétés.
- ▶ Cliquez avec le bouton droit de la souris, puis choisissez Propriétés de type dans le menu contextuel. La boîte de dialogue Propriétés de type apparaît.
- ▶ Procédez aux modifications nécessaires.
- ▶ Cliquez sur OK pour enregistrer les modifications dans la déclaration de types.

Utilisation du dictionnaire des synonymes pour du texte japonais

Pour le texte japonais, le dictionnaire des substitutions ne contient qu'un onglet permettant de gérer vos synonymes, l'onglet Synonyme. Les synonymes sont des mots ayant le même sens. Vous pouvez également utiliser des synonymes pour regrouper des termes et leurs abréviations, ou pour réunir les mots fréquemment mal orthographiés sous la version correcte du mot.

Figure A-5
Entrées de synonymes pour le texte en japonais

	Cible	Synonymes	Type	Bibliothèque
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 感性意見
2	コンピュータ	pc, パソコン	名詞	Opinions 感性意見

Une définition de synonyme est constituée de deux parties. Le terme cible est le terme sous lequel vous souhaitez que le moteur du programme d'extraction regroupe tous les termes synonymes. A moins que ce terme cible ne soit utilisé comme synonyme d'un autre terme cible ou ne soit exclu, il deviendra probablement le concept apparaissant dans le panneau Résultats d'extraction. La liste des synonymes est la liste des termes qui seront regroupés sous le terme cible.

Dans l'onglet Synonymes, vous pouvez entrer une définition de synonyme dans la ligne vide qui figure en haut du tableau. Commencez par définir le terme cible et ses synonymes. Vous pouvez également sélectionner la bibliothèque dans laquelle vous souhaitez stocker la définition. Toutes les occurrences des synonymes sont ensuite regroupées sous le terme cible lors de l'extraction finale. [Pour plus d'informations, reportez-vous à la section Ajout de termes dans le chapitre 10 sur p. 224.](#)

Quand vous créez vos déclarations de types, vous pouvez entrer un terme puis penser à trois ou quatre synonymes de ce terme. Dans ce cas, vous pouvez saisir tous les termes puis votre terme cible dans le dictionnaire de substitution avant de faire glisser les synonymes.

Important ! Les synonymes pour le texte en japonais ne prennent pas en charge les caractères génériques et les caractères spéciaux.

Pour ajouter une entrée de synonyme

- ▶ Dans la ligne vide en haut du tableau dans l'onglet Synonyme du panneau Substitution, entrez le terme cible dans la colonne Cible. Le terme cible que vous entrez apparaît en couleur. Cette couleur représente le type dans lequel le terme apparaît ou a été imposé, le cas échéant. Si le terme apparaît en noir, il n'est alors inclus dans aucune déclaration de types.
- ▶ Cliquez dans la deuxième cellule à droite de la cible et entrez l'ensemble de synonymes. Séparez chaque entrée à l'aide du séparateur global défini dans la boîte de dialogue Options. Tous les synonymes saisis doivent être du même type. [Pour plus d'informations, reportez-vous à la section Définition des options dans le chapitre 2 sur p. 19.](#) Les termes que vous entrez apparaissent en couleur. Cette couleur représente le type dans lequel le terme apparaît. Si le terme apparaît en noir, il n'est alors inclus dans aucune déclaration de types.
- ▶ Dans la troisième colonne, la colonne Type, choisissez un type pour ces synonymes. Cependant, la cible prend le type attribué pendant l'extraction. Mais si la cible n'a pas été extraite en tant que concept, alors le type répertorié dans cette colonne est affecté à la cible dans les résultats de l'extraction.
- ▶ Cliquez dans la dernière cellule pour sélectionner la bibliothèque dans laquelle vous voulez stocker cette définition de synonyme.

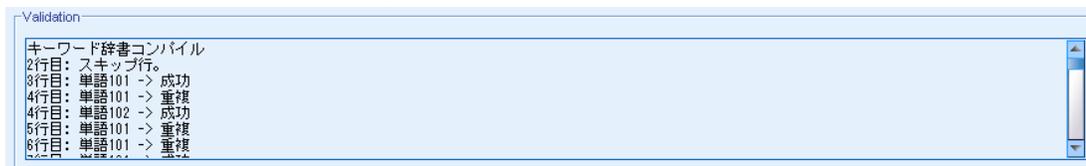
Remarque : Ces instructions indiquent la façon d'effectuer des modifications dans la vue Editeur de ressources. N'oubliez pas que vous pouvez également réaliser ce type de réglage directement à partir du panneau Résultats d'extraction ou du panneau Données. [Pour plus d'informations, reportez-vous à la section Affinage des résultats de l'extraction dans le chapitre 5 sur p. 88.](#)

Validation et compilation des ressources en japonais

Pour le texte japonais, il existe un panneau de validation supplémentaire servant à vérifier vos ressources avant l'extraction. Avant le début de l'extraction d'un texte en japonais, le moteur d'extraction recompile automatiquement les ressources lorsque des modifications sont détectées. Si une erreur est trouvée pendant l'extraction, le processus risque de ne pas s'exécuter correctement.

Pour éviter les erreurs de compilation, nous vous recommandons de valider et de compiler les ressources après avoir effectué les modifications dans l'Editeur de ressources. Si un message d'erreur s'affiche, vous pouvez procéder aux corrections et tenter une nouvelle validation.

Figure A-6
Panneau de validation pour le texte japonais



Pour valider les ressources

- ▶ A partir des menus, sélectionnez Outils> Valider des ressources. Le panneau de validation s'ouvre et affiche les messages de compilation et d'erreur.

Autres exceptions pour le japonais

Ressources internes remplaçant les ressources définies par l'utilisateur

Pour le texte en japonais, les ressources par défaut contiennent des ressources de base internes précompilées. Ces ressources internes ne peuvent pas être modifiées. C'est la raison pour laquelle vous pouvez utiliser l'Editeur de ressources pour effectuer les modifications et les réglages. La majeure partie du temps, les termes, les synonymes et les entrées de listes d'exclusion que vous définissez dans vos ressources seront prioritaires aux ressources internes précompilées. Cependant il existe plusieurs exceptions comme les exemples suivants l'indiquent.

- Il existe des cas où l'ajout de ces termes à un type particulier n'a pas d'incidence sur les résultats d'extraction. Ceci a de fortes chances de se produire lorsque les données comportent des phrases longues qui comprennent plusieurs éléments morphologiques, ponctuations ou symboles. En outre, comme les ressources pour le texte japonais comprennent un grand nombre de termes communs précompilés, il existe quelques mots communs qui seront toujours forcés dans une définition linguistique spécifique.

- Il est possible que vous ne puissiez pas exclure de termes tels que ある, いる, ou なる car le moteur d'extraction forcera toujours l'extraction de ces termes.
- Bien qu'il soit possible de modifier le type du terme 東京 de <地名> en <名詞>, le moteur d'extraction ignorera ces changements si vous essayez de modifier le type d'un terme de <地名> en <動詞> ou en <形容詞> à l'aide du dictionnaire de mots-clés (types).
- Parfois les modifications effectuées dans l'Editeur de ressources ou l' ont une incidence sur les résultats d'extraction d'une phrase mais pas d'une autre car le processus d'extraction se termine en référant les mots cooccurrents dans chaque phrase.

Problème d'affichage des katakanas de demi-largeur

Les caractères katakana de demi-largeur sont convertis en interne en caractères de pleine largeur pendant l'extraction mais apparaissent toujours dans la demi-largeur d'origine lors de l'affichage dans le panneau Données qui se trouve dans la session interactive de l'interface utilisateur . Veuillez noter que les caractères katakana de demi-largeur ne peuvent pas être mis en surbrillance dans le panneau Données. Pour éviter ce problème, convertissez tous vos enregistrements en katakanas de pleine largeur avant l'exécution.

Utilisation des minuscules et des majuscules

Les caractères alphabétiques en majuscule sont temporairement convertis en minuscules lorsqu'ils sont lus dans l'application. Mais le panneau Données affichera le texte à l'aide de la même casse que le texte d'origine. Les minuscules et les majuscules sont traitées de la même façon dans ce produit.

Avis

Ces informations sont développées pour des produits et des services proposés dans le monde entier.

Il est possible qu'IBM ne propose pas les produits, services ou fonctionnalités présentés dans ce document dans d'autres pays. Consultez votre représentant IBM local pour obtenir des informations sur les produits et services actuellement disponibles dans votre région. Toute référence à un produit, programme ou service IBM n'a pas pour but d'affirmer ou de sous-entendre que seul le produit, programme ou service IBM peut être utilisé. Tout produit, programme ou service de fonctionnalité équivalente, conforme au droit sur la propriété intellectuelle d'IBM, peut être utilisé à la place. Cependant, il est de la responsabilité de l'utilisateur d'évaluer et de vérifier le fonctionnement de tout produit, programme ou service autre qu'IBM.

IBM peut posséder des brevets ou des applications de brevet en cours qui couvrent le sujet décrit dans ce document. En aucun cas, l'obtention de ce document ne vous accorde de licence pour ces brevets. Vous pouvez envoyer vos questions sur les licences par écrit à :

IBM Director of Licensing, IBM Corporation, North Castle Drive, Armonk, NY 10504-1785, États-Unis.

Pour des informations sur le jeu de caractères à deux-octets (DBCS) des licences, contactez le service sur la propriété intellectuelle d'IBM de votre pays ou envoyez vos questions par écrit à :

Intellectual Property Licensing, Legal and Intellectual Property Law, IBM Japan Ltd., 1623-14, Shimotsuruma, Yamato-shi, Kanagawa 242-8502 Japon.

Le paragraphe suivant ne s'applique pas au Royaume-Uni ni à aucun autre pays dans lequel ces dispositions sont contraires au droit local : INTERNATIONAL BUSINESS MACHINES FOURNIT CETTE PUBLICATION « EN L'ÉTAT » SANS GARANTIE D'AUCUNE SORTE, IMPLICITE OU EXPLICITE, Y COMPRIS, NOTAMMENT, LES GARANTIES IMPLICITES DE NON-VIOLATION, DE VALEUR MARCHANDE OU D'ADÉQUATION À UN BUT PRÉCIS. Certains pays n'autorisent pas les clauses de non responsabilité des garanties explicites ou implicites de certaines transactions et par conséquent, il est possible que cette déclaration ne vous concerne pas.

Ces informations peuvent contenir des erreurs techniques ou typographiques. Ces informations sont modifiées périodiquement ; les modifications seront intégrées aux nouvelles éditions de la publication. IBM peut effectuer des améliorations et/ou des modifications sur les produits et/ou les programmes décrits dans cette publication à tout moment et sans notification préalable.

Toutes les références, contenues dans ces informations, à des sites Web autres qu'IBM sont fournies dans un but pratique uniquement et ne constituent en aucun cas une approbation de ces sites Web. Le matériel de ces sites Web ne fait pas partie du matériel de ce produit IBM et l'utilisation de ces sites Web se fait à vos propres risques.

IBM a le droit d'utiliser ou de distribuer toutes les informations que vous avez fournies, comme l'entreprise IBM le juge approprié, sans aucune obligation envers vous.

Les détenteurs de licence pour ce programme qui souhaitent obtenir des informations dans le but d'autoriser : (i) l'échange d'informations entre les programmes créés de manière indépendante et d'autres programmes (y compris celui-ci) et (ii) l'utilisation mutuelle des informations qui ont été échangées, doivent contacter :

IBM Software Group, Attention: Licensing, 233 S. Wacker Dr., Chicago, IL 60606, États-Unis.

Ces informations peuvent être disponibles conformément aux conditions générales appropriées et dans certains cas, être soumises à des frais.

Le programme sous licence décrit dans ce document et tout le matériel sous licence associé disponible sont fournis par IBM conformément à l'accord client IBM, au contrat de licence sur les programmes internationaux IBM et à tout accord équivalent entre nous.

Toutes les données sur les performances contenues ici ont été obtenues dans un environnement contrôlé. Par conséquent, les résultats obtenus dans d'autres environnements d'exploitation peuvent varier de manière significative. Il est possible que certaines mesures aient été obtenues sur des systèmes en cours de développement et il n'existe aucune garantie quant au fait que ces mesures seront les mêmes sur les systèmes publics. De plus, certaines mesures peuvent avoir été obtenues par extrapolation. Les résultats réels peuvent varier. Les utilisateurs de ce document doivent vérifier les données applicables à leur environnement spécifique.

Les informations concernant les produits autres qu'IBM ont été obtenues auprès des fournisseurs de ces produits, de leurs annonces publiées ou d'autres sources publiques disponibles. IBM n'a pas testé ces produits et ne peut pas confirmer l'exactitude de leurs performances, leur compatibilité ou toute autre information associée à des produits autres qu'IBM. Les questions sur les capacités de produits autres qu'IBM doivent être adressées aux fabricants de ces produits.

Ces informations contiennent des exemples de données et de rapports utilisés dans les opérations quotidiennes d'entreprises. Pour les illustrer aussi précisément que possible, ces exemples contiennent des noms de personnes, d'entreprises, de marques et de produits. Tous ces noms sont fictifs et toute similarité avec des noms et adresses utilisés par une entreprise réelle ne serait que pure coïncidence.

Si vous utilisez la version électronique de ces informations, il est possible que les photographies et illustrations en couleurs n'apparaissent pas.

Marques

IBM, le logo IBM, ibm.com et SPSS sont des marques d'IBM Corporation, déposées dans de nombreuses juridictions du monde entier. Une liste actuelle des marques d'IBM est disponible sur le Web à l'adresse <http://www.ibm.com/legal/copytrade.shtml>.

Adobe, le logo Adobe, PostScript et le logo PostScript sont des marques commerciales ou déposées d'Adobe Systems Incorporated aux États-Unis et/ou dans d'autres pays.

Intel, le logo Intel, Intel Inside, le logo Intel Inside, Intel Centrino, le logo Intel Centrino, Celeron, Intel Xeon, Intel SpeedStep, Itanium et Pentium sont des marques commerciales ou déposées d'Intel Corporation ou de ses filiales aux États-Unis ou dans d'autres pays.

Linux est une marque déposée de Linus Torvalds aux États-Unis, dans d'autres pays ou les deux.

Microsoft, Windows, Windows NT et le logo Windows sont des marques déposées de Microsoft Corporation aux États-Unis, dans d'autres pays ou les deux.

UNIX est une marque déposée de The Open Group aux États-Unis et dans d'autres pays.

Java et toutes les marques et logos Java sont des marques de Sun Microsystems, Inc. aux États-Unis, dans d'autres pays, ou les deux.

Les captures d'écran de produits Microsoft sont réimprimées avec l'autorisation de Microsoft Corporation.

Les autres noms de produits ou de services peuvent être des marques d'IBM ou d'autres entreprises.

Index

- ! symboles ^ * \$ dans les synonymes, 233
- abréviations, 248, 250
- acides aminés (entité non linguistique), 243
- activation des entités non linguistiques, 247
- actualisation des graphiques, 170
- adaptation
 - erreurs de ponctuation, 86
 - fautes d'orthographe, 86
- adresse électronique (entité non linguistique), 243
- adresses (entité non linguistique), 243
- adresses IP (entité non linguistique), 243
- affichage
 - bibliothèques, 210
 - catégories, 170
 - données, 51
- afficher des colonnes dans le panneau des catégories, 98
- afficher les colonnes dans le panneau Données, 100
- agitation, 190
- ajout
 - bibliothèques publiques, 208
 - concepts dans les catégories, 160
 - descripteurs, 107
 - éléments optionnels, 235
 - sons, 21, 23
 - synonymes, 89, 233
 - synonymes pour le japonais, 265
 - termes aux déclarations de types, 224
 - termes aux déclarations de types japonais, 258
 - termes aux exclusions, 237
 - types, 91
- ajout des termes forcés
 - afficher les colonnes, 100
 - extraction de concepts, 94
 - forcer à entrer, 100
 - forcer à sortir, 100
 - mots dans les catégories, 165
 - réponses, 164
 - termes, 228
- analyse de texte, 3, 8–9, 12
- analyse des dépendances, 253–254
- analyse des sentiments, 253–254
 - options, 254
- analyse secondaire
 - analyse des dépendances, 253–254
 - analyse des sentiments, 253–254
- annotations
 - pour les catégories, 111, 159
 - pour les projets, 50
- antiliens, 120
- astérisque (*)
 - dictionnaire d'exclusions, 237
 - synonymes, 233
- attribuer des booléens, 76
- attribution de nom
 - bibliothèques, 211
 - catégories, 111, 159
 - déclarations de types, 229
 - avis juridiques, 269
- barres d'état, 77
- bibliothèques, 16, 206, 220
 - affichage, 210
 - ajout, 208
 - attribution de nom, 211
 - avertissement de la synchronisation des bibliothèques, 214
 - bibliothèques locales, 214
 - bibliothèques par défaut fournies, 206
 - bibliothèques publiques, 214
 - Budget Library, 221
 - changement de nom, 211
 - Core Library, 221
 - création, 208
 - désactivation, 211
 - dictionnaires, 206
 - export, 213
 - import, 213
 - liaison, 208
 - mise à jour, 217
 - Opinions Library, 221
 - partage, 76
 - partage et publication, 214
 - publication, 53, 216
 - suppression, 212, 214
 - synchronisation, 214
- bibliothèques fournies (par défaut), 206
- bibliothèques par défaut, 206
- booléen important, 76
- booléen terminé, 76
- bouton d'affichage, 99
- bouton score, 99
- Budget Library, 221
- catégorie grammaticale, 248–249
- catégories, 27, 96–97, 110, 135–136, 158
 - ajout, 160
 - annotations, 111, 159
 - changement de nom, 133
 - copie, 167
 - création, 6, 104–105, 112, 116, 118, 121, 127, 131, 134
 - création d'une catégorie vide, 133
 - création manuelle, 133
 - déplacement, 162
 - descripteurs, 106–107, 111
 - édition, 158, 161
 - étiquettes, 111, 159
 - export, 55
 - extension, 121, 128
 - forcer les réponses, 164
 - fusion, 163

- graphique Relations, 170
- graphiques de similarités, 170
- impression, 167
- mots imposés, 165
- noms, 111, 159
- packages d'analyse de texte, 43, 45
- pertinence, 101–102
- propriétés, 111, 159
- réglage des résultats, 9, 158
- scoring, 99
- stratégies, 105
- suppression, 168
- catégories prédéfinies, 135–136, 144
 - format compact, 141
 - Format indenté, 142
 - format liste plate, 140
- catégorisation, 6, 96, 256
 - dérivation des racines de concept, 116, 121–122
 - inclusion de concepts, 116, 121, 123
 - manuellement, 133
 - méthodes, 104–105
 - règles de cooccurrence, 116, 121, 126
 - réseaux sémantiques, 116, 121, 124
 - techniques de fréquence, 127
 - techniques linguistiques, 112, 128
 - utilisation de techniques, 121
 - utilisation des techniques de regroupement, 116
- changement
 - modèles, 198
 - source de données, 63
- changement de nom
 - bibliothèques, 211
 - catégories, 133, 159
 - déclarations de types, 229
 - modèles de ressources, 199
 - projets, 53
- chiffres (entité non linguistique), 243
- cluster, 190
- colonne documents, 98–99
- combinaison de catégories, 163
- concepts
 - ajout à des catégories, 106, 111, 160
 - ajout aux types, 91
 - ajout des termes forcés dans l'extraction, 94
 - création de types, 88
 - dans les catégories, 106, 111
 - exclusion de l'extraction, 94
 - extraction, 80
 - meilleurs descripteurs, 107
- coordonnées polaires, 186
- copie
 - catégories, 167
- copie de visualisations, 191
- Core Library, 221
- correspondance de texte, 111, 159, 165
- couleur, 223, 265
- couleurs
 - définition des options de couleur, 21
 - dictionnaire d'exclusions, 238
 - pour les barres du graphique récapitulatif, 62
 - pour les types et les termes, 223, 265
 - synonymes, 235
- couleurs personnalisées, 21
- création
 - bibliothèques, 208
 - catégories, 104–105, 112, 134
 - catégories avec règles, 148
 - déclarations de types, 222, 264
 - éléments optionnels, 235
 - entrées du dictionnaire d'exclusions, 237
 - modèle à partir de ressources, 197
 - projets, 29
 - règles de catégorie, 147–148, 156
 - synonymes, 88–89, 233
 - synonymes pour le japonais, 265
 - types, 91
- création de catégories, 6, 112, 114, 116, 118, 120–128, 131, 133, 256
 - exceptions de lien de classification supervisée, 120
 - technique de dérivation des racines de concept, 6, 118, 131
 - technique de réseau sémantique, 6, 118, 131
 - technique des règles de cooccurrence, 6, 118, 131
 - technique d'inclusion de concepts, 131
 - techniques de regroupement pair/frère, 131
 - utilisation de techniques, 6, 118
- création de modèles depuis des ressources, 197
- dates (entité non linguistique), 243
- déclaration de types, 206
 - ajout de termes, 224
 - ajout de termes pour le japonais, 258
 - ajout des termes forcés, 228
 - changement de nom, 229
 - création de types, 222, 264
 - déplacement, 230
 - désactivation, 231
 - éléments optionnels, 220
 - suppression, 231
 - synonymes, 220
 - types intégrés, 221
- déclaration de types Budget, 221
- déclaration de types Location, 221
- déclaration de types Negative, 221
- déclaration de types Organization, 221
- déclaration de types Person, 221
- déclaration de types Positive, 221
- déclaration de types Product, 221
- déclaration de types Uncertain, 221
- déclaration de types Unknown, 221
- définitions, 106, 111
- définitions forcées, 248–249
- délimiteur, 19

- déplacement
 - catégories, 162
 - déclarations de types, 230
- désactivation
 - barres d'état, 77
 - bibliothèques, 211
 - déclarations de types, 231
 - dictionnaires de substitution, 236
 - dictionnaires de synonymes, 242
 - dictionnaires d'exclusions, 238
 - entités non linguistiques, 247
- désactivation des entités non linguistiques, 247
- désactivation des sons, 23
- descripteurs, 98
 - catégories, 106, 111
 - choix des meilleurs, 107
 - édition dans les catégories, 161
- devises (entité non linguistique), 243
- diagramme Barre Catégorie, 170–171
- diagrammes en bâton, 170
- dictionnaire de substitutions, 206, 231, 235–236
- dictionnaire d'exclusions, 206, 236–238
- dictionnaires, 16, 220
 - exclusions, 206, 220, 236
 - substitutions, 206, 220, 231
 - types, 206, 220
- distance de recherche maximale, 119, 125, 131
- Dodge (regroupement), 190
- données
 - actualisation, 74
 - affichage, 51
 - catégorisation, 96, 112, 133
 - changer la source de données, 63
 - création de catégories, 6, 116, 118, 121, 128
 - export, 55
 - extraction, 85
 - fichiers IBM SPSS Statistics.sav, 55
 - IBM SPSS Data Collection, 35, 56, 69
 - IBM SPSS Statistics fichiers .sav, 31, 56, 65
 - Microsoft Excel Fichiers .xls / .xlsx, 33, 55, 58, 66
 - modifier les propriétés des variables, 52
 - ODBC, 34, 68
 - réglage des résultats, 88
 - sélection de sources de données, 30, 64
 - tri, 52
- données d'enquête, 2, 27–28, 35, 69
- éditeur de ressources, 12, 197–198, 239
 - affichage de différentes ressources, 198
 - création de modèles, 197
 - mise à jour de modèles, 197
 - option séparateur global, 19
 - pour le japonais, 256
- édition
 - affinage des résultats de l'extraction, 88
 - catégories, 158–159, 161
 - propriétés, 159
 - règles de catégorie, 158
- édition de graphiques
 - taille des éléments graphiques, 179
- éléments graphiques
 - changement, 187
 - conversion, 187
 - modificateurs de collision, 190
 - types, 187
- éléments optionnels, 231
 - ajout, 235
 - définition, 232
 - suppression d'entrées, 236
 - target, 235
- enregistrement
 - enregistrement automatique de projets, 19
 - extraction, résultats, 87
 - projets, 53
 - ressources, 202
 - ressources en tant que modèles, 197
- enregistrement automatique de projets, 19
- enregistrements, 100
- entités non linguistiques, 87
 - acides aminés, 243
 - activation et désactivation, 247
 - adresses, 243
 - adresses électroniques, 243
 - adresses HTTP/URL, 243
 - adresses IP, 243
 - chiffres, 243
 - dates, 243
 - devises, 243
 - expressions régulières, *RegExp.ini*, 244
 - heures, 243
 - normalisation, *NonLingNorm.ini*, 246
 - numéro de Sécurité sociale (Etats-Unis), 243
 - numéros de téléphone, 243
 - poids et mesures, 243
 - pourcentages, 243
 - protéines, 243
- erreurs de ponctuation, 86
- étiquettes pour les catégories, 111, 159
- exceptions de lien, 120
- exclusion
 - concepts de l'extraction, 94
 - désactivation de dictionnaires, 231, 236
 - désactivation des bibliothèques, 211
 - désactivation des entrées d'exclusion, 238
 - liens de catégorie, 120
 - ressemblance, 242
- export, 55
 - bibliothèques publiques, 213
 - catégories prédéfinies, 144
 - en tant que fichiers IBM SPSS Statistics.sav, 56
 - formats de sortie, 55
 - graphiques récapitulatifs, 60
 - modèles, 200
 - pour IBM SPSS Data Collection, 56

- résultats de catégorisation, 55–56, 58
 - sous forme de fichiers Microsoft Excel *.xls* / *.xlsx*, 58
- extension de catégories, 128
- extraction, 3–4, 80, 85–86, 206, 220, 251
 - Concepts, types et patrons, 80
 - enregistrement des résultats, 87
 - entités non linguistiques, 87
 - mots imposés, 94
 - réglage des résultats, 9, 88
 - résultats, 27
 - uniternes, 5, 87, 252
- fautes d'orthographe, 28, 86, 242
- fiabilité, 9
- Fichiers *.xls* / *.xlsx*, 55, 58
- Fichiers Microsoft Excel *.xls* / *.xlsx*
 - import de catégories prédéfinies, 135
- fichiers restaurés, 19
- filtrage des bibliothèques, 210
- format compact, 141
- Format indenté, 142
- format liste plate, 140
- formes au pluriel des mots, 223
- formes fléchies, 122, 220, 222–224, 264
- fractionnement des termes en composants, 122
- fractionnement en composants, 122
- fréquence, 127
- fréquence du type, 127
- fusion de catégories, 163
- générer des formes fléchies, 220, 222–224, 264
- gestion
 - bibliothèques locales, 211
 - bibliothèques publiques, 212
 - catégories, 158
- glisser et déposer, 134
- graphiques, 170
 - actualisation, 170
 - édition, 174
 - exporter des graphiques récapitulatifs, 60
 - graphique Relations de catégorie, 170
 - Mode d'interaction, 173
 - taille des éléments graphiques, 179
- graphiques récapitulatifs, 60
- graphiques Relations, 170
- heures (entité non linguistique), 243
- HTTP/URL (entités non linguistiques), 243
- IBM SPSS Data Collection, 35, 69
 - changer la source de données, 63
 - export, 56
- IBM SPSS Statistics fichiers *.sav*, 31, 65
 - changer la source de données, 63
 - export, 56
- format de sortie, 55
- import
 - actualiser les données, 74
 - bibliothèques publiques, 213
 - catégories prédéfinies, 136
 - données d'entrée, 35, 69
 - modèles, 200
 - ODBC, 33, 67
 - préparation des données, 28
 - réimporter les données, 63
- impression de catégories, 167
- Japonais, 251, 256
 - Editeur de ressources, 256
 - propriétés du type, 264
 - types, 261, 267
- Language Weaver, 23, 37, 72
- légende
 - position, 190
- **.lib*, 213
- marquer des réponses, 76
- marques, 270
- Microsoft Excel Fichiers *.xls* / *.xlsx*, 33, 58, 66
 - changer la source de données, 63
 - exporter des catégories prédéfinies, 144
 - format de sortie, 55
 - import de catégories prédéfinies, 136
- mise à jour
 - bibliothèques, 214, 217
 - graphiques, 170
 - modèles, 197
- mode d'édition, 174
- Mode d'interaction, 173
- modèles, 4, 194, 252
 - affichage de différents modèles, 198
 - changement de nom, 199
 - création depuis des ressources, 197
 - import et export, 200
 - mise à jour ou enregistrement, 197
 - restauration, 202
 - sauvegarde, 202
 - suppression, 199
 - TLA (Analyse des liens du texte), 198
- modèles de ressources, 4, 194, 252
- modificateurs de collision, 187
- modification des visualisations, 174
 - ajout d'effets 3D, 186
 - axes, 182
 - catégories, 184
 - combinaison de catégories, 184
 - couleurs et motifs, 177
 - échelles, 182
 - exclusion de catégories, 184
 - extension, 180

- formats des nombres, 180
 - forme de point, 179
 - marges, 180
 - panneaux, 185
 - paramètres automatiques, 175
 - position de la légende, 190
 - rapport d'aspect de point, 179
 - réduction des catégories, 184
 - règles, 175
 - rotation de point, 179
 - sélection, 175
 - text, 176
 - tirets, 177
 - transformation du système de coordonnées, 186
 - transparency (transparence), 177
 - transposer, 185–186
 - tri de catégories, 184
-
- N°de sécurité sociale (entité non linguistique), 243
 - nom de catégorie, 98
 - nombre maximal de catégories à créer., 120
 - non-prise en compte de concepts, 94
 - normalisation, 246
 - nouveautés, 1
 - nouvelles catégories, 133
 - nouvelles fonctions, 1
 - numéros de téléphone (entité non linguistique), 243
-
- opérateur de règle AND, 157
 - opérateur de règle NOT, 157
 - opérateur de règle OR, 157
 - Opérateurs booléens, 157
 - opérateurs dans les règles & | !() , 157
 - opérateurs de règle | !() &, 157
 - Opinions Library, 221
 - option de mise en correspondance, 220, 222, 224–226, 264
 - options, 19
 - affichage, 21
 - son, 23
 - système, 19
 - traduction, 23
 - options de son, 23
 - ouvrir des projets, 49
-
- packages d'analyse de texte, 39, 43–45
 - packages d'analyse de texte *.tap, 39, 43–45, 48
 - palettes
 - affichage, 176
 - déplacement, 176
 - masquage, 176
 - panneau catégories, 97
 - panneau Données, 100
 - bouton d'affichage, 99
 - panneau visualisation
 - graphique Relations de catégorie, 170
 - mise à jour des graphiques, 170
-
- paramètres, 19, 21, 23
 - paramètres -auto-, 176
 - partage de bibliothèques, 214
 - ajout de bibliothèques publiques, 208
 - mise à jour, 217
 - publication, 53, 216
 - partage de projets, 76
 - patrons d'extraction, 248
 - patterns, 80, 83
 - permutations, 87
 - pertinence des réponses et des catégories, 101–102
 - pile, 190
 - poids/mesures (entités non linguistiques), 243
 - point d'exclamation (!), 233
 - pourcentages (entité non linguistique), 243
 - préférences, 19, 21, 23
 - préparation de vos données, 28
 - projets, 27, 49
 - barre d'état, 77
 - changement de nom, 53
 - création, 29
 - enregistrement, 53
 - options des bibliothèques, 19
 - ouverture, 49
 - packages d'analyse de texte, 39
 - partage, 76
 - propriétés, 50
 - réutilisation de catégories, 167
 - sélection de catégories et de ressources, 39
 - sélection des variables, 35, 69
 - source de données, 30, 64
 - traduction, 37, 72
 - propriétés
 - catégories, 111, 159
 - pour les types japonais, 264
 - projets, 50
 - variables, 52
 - protéines (entité non linguistique), 243
 - publication, 53, 216
 - ajout de bibliothèques publiques, 208
 - bibliothèques, 214
-
- question fermée, 2
 - question ouverte, 2
-
- rapports et graphiques récapitulatifs, 60
 - recherche de termes et de types, 209
 - rechercher et remplacer (ressources avancées), 240–241
 - réglage des résultats
 - ajout de concepts à des types, 91
 - ajout de synonymes, 89
 - catégories, 9, 158
 - création de types, 91
 - exclusion de concepts, 94
 - extraction de concepts imposée, 94

- extraction, résultats, 9, 88
- règles
 - création, 156
 - édition, 158
 - Opérateurs booléens, 157
 - suppression, 158
 - syntaxe, 148
 - technique des règles de cooccurrence, 126
- règles de catégorie, 147–148, 154, 156–158
 - cooccurrence de concepts, 6, 118, 122, 126, 131
 - exemples, 154
 - mots synonymes, 6, 116, 118, 121–122, 128, 131
 - règles de cooccurrence, 116, 121, 128
 - syntaxe, 148
- Regroupement flou (Exceptions), 86, 239, 242
- réimporter les données, 63
- remplacement des ressources par un modèle, 198
- réponses, 100
 - ajout des termes forcés à entrer dans des catégories, 164
 - marquer, 76
 - marquer comme terminé, 76
- réponses vides, 28
- réponses vierges, 28
- ressources
 - affichage de différentes ressources de modèle, 198
 - bibliothèques par défaut fournies, 206
 - édition des ressources avancées, 239
 - restauration, 202
 - sauvegarde, 202
- ressources avancées, 239
 - rechercher et remplacer dans l'éditeur, 240–241
- ressources linguistiques, 27, 206
 - modèles de ressources, 194
 - packages d'analyse de texte, 43, 45
- restauration de fichiers, 19
- restauration des ressources, 202
- retour à la ligne dans une colonne, 21
- réutilisation
 - catégories, 167
- sans catégorie, 98
- sauvegarde des ressources, 202
- scoring, 99
- sections de gestion des langues, 239, 248
 - abréviations, 248, 250
 - définitions forcées, 248–249
 - patrons d'extraction, 248
- séparateur global, 19
- séparateurs, 19
- séparateurs de texte, 19
- Source de données ODBC, 33–34, 63, 67–68
- statistiques
 - descriptions, 188
 - modification dans les visualisations, 188
- stockage des résultats de l'extraction, 87
- suppression
 - bibliothèques, 212, 214
 - catégories, 168
 - déclarations de types, 231
 - désactivation des bibliothèques, 211
 - éléments optionnels, 236
 - entrées exclues, 238
 - modèles de ressources, 199
 - règles de catégorie, 158
 - synonymes, 236
- surveillance, 76
- symbole dollar (\$), 233
- symbole du caret (^), 233
- synchronisation des bibliothèques, 214, 216–217
- synonymes, 88, 231
 - ! symboles ^ * \$, 233
 - ajout, 89, 233, 265
 - couleurs, 235
 - définition, 232
 - pour les textes en japonais, 265
 - Regroupement flou (Exceptions), 86, 242
 - suppression d'entrées, 236
 - termes cibles, 233, 265
- Systèmes de coordonnées
 - transformation, 186
- tableau des relations, 170
- tableau/graphique Relations de catégorie, 170, 172–173
- technique de dérivation des racines de concept, 116, 121–122, 128, 131
- technique de réseau sémantique, 7, 116, 118, 121–122, 124, 128, 131
- technique des règles de cooccurrence, 7, 116, 119, 121–122, 126, 128, 131
- technique d'inclusion de concepts, 7, 116, 119, 121, 123, 128
- techniques
 - dérivation des racines de concept, 116, 121–122, 128
 - fréquence, 127
 - glisser et déposer, 134
 - inclusion de concepts, 116, 121, 123, 128
 - règles de cooccurrence, 116, 121, 126, 128
 - réseaux sémantiques, 116, 121, 124, 128
- techniques linguistiques, 3, 6, 9, 118
- termes
 - ajout à des types japonais, 258
 - ajout au dictionnaire d'exclusions, 237
 - ajout aux types, 224
 - ajout des termes forcés, 228
 - Couleur, 223, 265
 - formes fléchies, 220
 - imposition de mots dans les catégories, 165
 - options de mise en correspondance, 220
 - recherche dans l'éditeur, 209
- termes cibles, 235
- text mining, 3
- titres du graphique récapitulatif d'export, 62
- TLA (Analyse des liens du texte), 198
- tous les documents, 98

- tracer les réponses, 76
- traduction, 23, 74
 - boîte de dialogue options, onglet traduction, 23
 - boîte de dialogue paramètres de traduction, 74
 - paramètre de précision de traduction, 38, 73, 75
 - traduction vers l'anglais, 37, 72, 74
- tri naturel, 52
- trier les données et les variables, 52
- types, 220
 - ajout de concepts, 88
 - couleur par défaut, 21, 223, 265
 - création, 222, 264
 - dictionnaires, 206
 - extraction, 80
 - fréquence du type, 127
 - pour le japonais, 261, 264, 267
 - recherche dans l'éditeur, 209
 - types intégrés, 221

- unitermes, 87

- valeur de lien minimale, 120
- variables
 - actualisation, 74
 - changer la source de données, 63
 - export, 55
 - import, 31, 33, 65, 67
 - mise en correspondance, 71
 - modifier les propriétés, 52
 - variables de référence, 27–28, 35, 69
 - variables de texte, 27–28, 35, 69
 - variables d'ID, 27–28, 35, 69
- variables de référence, 27–28, 35, 69
- variables de texte, 27–28, 35, 69
- variables d'ID, 27–28, 35, 69
- visualisations
 - axes, 182
 - catégories, 184
 - copie, 191
 - couleurs et motifs, 177
 - échelles, 182
 - édition, 174
 - extension, 180
 - formats des nombres, 180
 - forme de point, 179
 - marges, 180
 - panneaux, 184–185
 - position de la légende, 190
 - rapport d'aspect de point, 179
 - rotation de point, 179
 - text, 176
 - tirets, 177
 - transformation du système de coordonnées, 186
 - transparency (transparence), 177
 - transposer, 184–186
- vue projet complet, 12, 15
- vue question, 12–13
- vues
 - fenêtre d'analyse de texte, 12
 - fenêtre éditeur de ressources, 16
 - projet complet, 15
 - vue question, 13