

Guía del usuario de IBM SPSS Text  
Analytics for Surveys 4.0.1



*Nota:* Antes de usar esta información y el producto al que hace referencia, lea la información general en Avisos el p. 262.

Esta edición se aplica a IBM® SPSS® Text Analytics for Surveys 4 y a todas las versiones y modificaciones subsiguientes hasta que se indique lo contrario en nuevas ediciones.

Capturas de pantalla de productos de Adobe reimpresas con permiso de Adobe Systems Incorporated.

Capturas de pantalla de productos de Microsoft reimpresas con permiso de Microsoft Corporation.

Material bajo licencia - Property of IBM

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

Derechos restringidos de usuarios del Gobierno de EE.UU - El uso, duplicación o revelación está restringido por el GSA ADP Schedule Contract con IBM Corp.

---

# Prefacio

Bienvenido a IBM® SPSS® Text Analytics for Surveys versión 4.0.1, una aplicación de codificación de texto de encuestas que proporciona un análisis significativo de las respuestas a las preguntas con final abierto. Con este producto, cualquiera que realice análisis de encuestas puede transformar rápidamente las respuestas no estructuradas en datos cuantitativos. El “desbloqueo” de estos datos de texto con final abierto pueden mejorar significativamente la calidad del análisis y la capacidad de la toma de decisiones. Esta aplicación permite importar los datos de la encuesta, extraer los conceptos clave, refinar los resultados y categorizar las respuestas. Una vez categorizados los datos, puede exportar las categorías para importarlas en herramientas analíticas cuantitativas, como el sistema IBM® SPSS® Statistics, para un análisis y generación de gráficos más exhaustivo.

SPSS Text Analytics for Surveys combina las tecnologías lingüísticas avanzadas diseñadas para extraer y clasificar con fiabilidad los conceptos clave de las respuestas de las encuestas con final abierto mediante técnicas manuales. Gracias a los sólidos algoritmos de generación de categorías y a la sencilla funcionalidad de arrastrar y soltar, puede crear categorías o “códigos” en los que categorizar las respuestas de las encuestas. Además, las categorías que se generan podrán reutilizarse para proporcionar así un resultado coherente en estudios idénticos o similares. Puesto que los datos de las respuestas con final abierto pueden variar mucho de una encuesta a otra, dos proyectos nunca serán exactamente iguales; sin embargo, puede seguir el mismo proceso básico para llevar a cabo los análisis. [Si desea obtener más información, consulte el tema El proceso típico en el capítulo 2 el p. 10.](#)

## ***Acerca de IBM Business Analytics***

El software IBM Business Analytics ofrece información completa, coherente y precisa en la que los órganos de toma de decisiones confían para mejorar el rendimiento comercial. Un conjunto integral de [inteligencia empresarial](#), [análisis predictivo](#), [rendimiento comercial y gestión de estrategias](#), así como de [aplicaciones de análisis predictivo](#), le ofrece una perspectiva clara, inmediata e interactiva del rendimiento actual y la capacidad para predecir resultados futuros. En combinación con extensas soluciones sectoriales, prácticas probadas y servicios profesionales, las organizaciones de cualquier tamaño pueden conseguir el máximo de productividad, automatizar decisiones de forma fiable y alcanzar mejores resultados.

Como parte de esta familia, el software de análisis predictivo de IBM SPSS ayuda a las organizaciones a predecir eventos futuros y actuar proactivamente según esa información para lograr mejores resultados comerciales. Los clientes comerciales, gubernamentales y académicos de todo el mundo confían en la tecnología de IBM SPSS como ventaja ante la competencia para atraer, conservar y hacer crecer los clientes, reduciendo al mismo tiempo el fraude y los riesgos. Al incorporar el software de IBM SPSS en sus operaciones diarias, las organizaciones se convierten en empresas predictivas, capaces de dirigir y automatizar decisiones para alcanzar los objetivos comerciales y lograr una ventaja considerable sobre la competencia. Para obtener más información o contactar con un representante, visite <http://www.ibm.com/spss>.

## ***Asistencia técnica***

La asistencia técnica está disponible para clientes con servicio de mantenimiento. Los clientes podrán ponerse en contacto con el servicio de asistencia técnica si desean recibir ayuda sobre la utilización de los productos de IBM Corp. o sobre la instalación en los entornos de hardware admitidos. Para contactar con el servicio de asistencia técnica, visite el sitio Web de IBM Corp. en <http://www.ibm.com/support>. Tenga preparada su identificación, la de su empresa y el acuerdo de asistencia técnica cuando solicite asistencia.

---

# Contenido

## **Parte I: Primeros pasos**

### **1 Acerca del Análisis de texto 1**

Novedades . . . . .	1
Datos de encuestas con final abierto . . . . .	2
Acerca de la minería de datos de texto . . . . .	3
Cómo funciona la extracción . . . . .	3
Funcionamiento de la categorización . . . . .	6
Preparación para el análisis del texto . . . . .	7
Fiabilidad y ajuste . . . . .	8
Refinamiento de los recursos lingüísticos . . . . .	9
Refinamiento de las definiciones de categoría . . . . .	9

### **2 Primeros pasos 10**

El proceso típico . . . . .	10
La ventana Análisis de texto . . . . .	11
La vista Pregunta . . . . .	12
La vista Proyecto completo . . . . .	14
La ventana Editor de recursos . . . . .	15
Opciones de configuración . . . . .	17
Opciones: pestaña Sistema . . . . .	18
Opciones: Pestaña Mostrar . . . . .	19
Opciones: pestaña Sonidos . . . . .	21
Opciones: pestaña Traducción . . . . .	22
Configuración de Microsoft Internet Explorer para obtener ayuda . . . . .	23

## **Parte II: Análisis de texto**

### **3 Creación de proyectos y paquetes 26**

Creación de proyectos . . . . .	26
Preparación de los datos . . . . .	27
Inicio de proyectos nuevos . . . . .	28
Selección de los orígenes de datos . . . . .	29

Selección de variables . . . . .	34
Traducción al inglés. . . . .	36
Selección de categorías y recursos. . . . .	37
Uso de los paquetes de análisis de texto. . . . .	41
Creación de paquetes de análisis de texto. . . . .	41
Actualización de los paquetes de análisis de texto. . . . .	43

## **4 Trabajo con proyectos 47**

Apertura de proyectos . . . . .	47
Edición de las propiedades del proyecto. . . . .	48
Vista de los datos del proyecto . . . . .	49
Ordenación de las variables. . . . .	50
Edición de las propiedades de variables . . . . .	50
Guardado de proyectos . . . . .	51
Exportación de los resultados de la categorización. . . . .	53
Exportación a IBM SPSS Statistics o IBM SPSS Data Collection . . . . .	54
Exportando a Microsoft Excel . . . . .	56
Exportación de gráficos de resumen. . . . .	58
Cambio de los orígenes de datos . . . . .	61
Selección de los orígenes de datos . . . . .	62
Selección de variables . . . . .	67
Variables coincidentes . . . . .	69
Traducción al inglés. . . . .	70
Actualización de los datos . . . . .	72
Traducción al inglés . . . . .	72
Cómo compartir proyectos . . . . .	74
Marcación de las respuestas . . . . .	74
Barra de estado del proyecto . . . . .	75

## **5 Extracción de datos 78**

Resultados extraídos: Conceptos, tipos y patrones . . . . .	78
Extracción de datos . . . . .	82
Cómo guardar los resultados de la extracción . . . . .	85
Refinamiento de los resultados de la extracción . . . . .	86
Adición de sinónimos. . . . .	87
Adición de conceptos a tipos. . . . .	89

Exclusión de conceptos de la extracción . . . . .	91
Forzado de palabras en la extracción . . . . .	92

## **6 Categorización de los datos de texto 93**

El panel de categorías . . . . .	94
El panel de datos . . . . .	97
Relevancia de categoría . . . . .	99
Métodos y estrategias para crear categorías . . . . .	101
Métodos para crear categorías . . . . .	101
Estrategias para crear categorías . . . . .	101
Consejos sobre la creación de categorías . . . . .	102
Selección de los mejores descriptores . . . . .	104
Acerca de las categorías . . . . .	107
Propiedades de categoría . . . . .	108
Generación de categorías . . . . .	110
Configuración avanzada: Lingüística . . . . .	113
Acerca de las técnicas lingüísticas . . . . .	118
Configuración avanzada de frecuencia . . . . .	123
Ampliación de categorías . . . . .	125
Creación manual de categorías . . . . .	129
Creación de categorías nuevas o cambio de nombre de categorías . . . . .	130
Creación de categorías mediante el método de arrastrar y soltar . . . . .	131
Importación y exportación de categorías predefinidas . . . . .	131
Importación de categorías predefinidas . . . . .	132
Exportación de categorías . . . . .	142
Uso de reglas de categoría . . . . .	145
Sintaxis de regla de categoría . . . . .	146
Uso de patrones TLA en las reglas de categoría . . . . .	148
Uso de comodines en reglas de categoría . . . . .	150
Ejemplos de reglas de categoría . . . . .	151
Creación de reglas de categoría . . . . .	153
Edición y eliminación de reglas . . . . .	155
Edición y refinamiento de categorías . . . . .	156
Edición de las propiedades de las categorías . . . . .	156
Añadir descriptores a las categorías . . . . .	157
Edición de descriptores de categoría . . . . .	158
Cómo mover categorías . . . . .	159
Fusión o combinación de categorías . . . . .	160
Forzado de respuestas dentro de categorías . . . . .	161
Coincidencia de texto en las categorías . . . . .	162

Copia de categorías . . . . .	164
Impresión de categorías . . . . .	164
Eliminación de categorías . . . . .	166

## **7 Visualización de gráficos 167**

Gráfico de barras de categorías . . . . .	168
Gráfico de malla de categorías . . . . .	169
Tabla de malla de categorías . . . . .	170
Uso de barras de herramientas y paletas de gráficos . . . . .	170
Edición de visualizaciones . . . . .	171
Reglas generales de edición de visualizaciones. . . . .	172
Edición y formato de texto . . . . .	173
Cambio de colores, tramas, estilos de guión y transparencia . . . . .	174
Rotación y cambio de la forma y la relación de aspecto de los elementos de punto . . . . .	175
Cambio de tamaño de los elementos gráficos . . . . .	176
Especificación de márgenes y relleno . . . . .	177
Formato de numeración . . . . .	177
Cambio de la configuración del eje y la escala . . . . .	178
Edición de categorías . . . . .	180
Cambio de la orientación de paneles . . . . .	182
Transformación del sistema de coordenadas . . . . .	182
Cambio de estadísticos y elementos gráficos. . . . .	183
Cambio de la posición de la leyenda . . . . .	187
Copia de una visualización y de datos de visualización . . . . .	187
Métodos abreviados de teclado. . . . .	188

## **Parte III: Editor de recursos**

### **8 Plantillas y recursos 190**

La interfaz del editor . . . . .	191
Creación y actualización de plantillas . . . . .	193
Cambio de plantillas de recursos . . . . .	194
Administración de plantillas . . . . .	195
Importación y exportación de plantillas. . . . .	196
Copia de seguridad de los recursos . . . . .	197
Importación de los archivos de recursos . . . . .	199

## **9 Trabajo con bibliotecas 202**

Bibliotecas enviadas . . . . .	202
Creación de bibliotecas . . . . .	203
Adición de bibliotecas públicas . . . . .	204
Búsqueda de términos y tipos . . . . .	205
Visión de bibliotecas . . . . .	206
Administración de las bibliotecas locales . . . . .	206
Cambio de nombre de las bibliotecas locales . . . . .	206
Desactivación de bibliotecas locales . . . . .	207
Eliminación de bibliotecas locales . . . . .	207
Administración de bibliotecas públicas . . . . .	208
Compartimiento de bibliotecas . . . . .	210
Publicación de bibliotecas . . . . .	211
Actualización de bibliotecas . . . . .	212
Resolución de conflictos . . . . .	213

## **10 Acerca de los diccionarios de biblioteca 215**

Diccionarios de tipo . . . . .	215
Tipos incorporados . . . . .	216
Creación de tipos . . . . .	217
Adición de términos . . . . .	219
Forzado de términos . . . . .	222
Cambio de nombre de los tipos . . . . .	223
Cómo mover tipos . . . . .	224
Desactivación y eliminación de tipos . . . . .	225
Diccionarios de sustitución/sinónimos . . . . .	225
Definición de sinónimos . . . . .	227
Definición de elementos opcionales . . . . .	229
Desactivación y eliminación de sustituciones . . . . .	230
Diccionarios de exclusión . . . . .	231

## **11 Acerca de los recursos avanzados 233**

Búsquedas . . . . .	234
Reemplazo . . . . .	235
Agrupación difusa . . . . .	236

Entidades no lingüísticas . . . . .	237
Definiciones de expresiones regulares . . . . .	238
Normalización . . . . .	240
Configuración . . . . .	241
Gestión de idiomas . . . . .	242
Patrones de extracción . . . . .	242
Definiciones forzadas . . . . .	243
Abreviaturas . . . . .	243

## ***Apéndices***

### ***A Excepciones de textos en japonés*** **245**

Extracción y categorización de textos en japonés . . . . .	245
Cómo funciona la extracción . . . . .	245
Cómo funciona la extracción secundaria . . . . .	247
Funcionamiento de la categorización . . . . .	250
Edición de recursos para textos en japonés . . . . .	250
Panel de términos, de tipos y del árbol de bibliotecas para el japonés . . . . .	252
Tipos disponibles para textos en japonés . . . . .	254
Edición de las propiedades de tipo del japonés . . . . .	258
Uso del diccionario de sinónimos para textos en japonés . . . . .	259
Validación y compilación de recursos del japonés . . . . .	260
Otras excepciones para el japonés . . . . .	261

### ***B Avisos*** **262**

### ***Índice*** **265**

***Parte I:***  
***Primeros pasos***



# ***Acerca del Análisis de texto***

Bienvenido a IBM® SPSS® Text Analytics for Surveys versión 4.0.1, una aplicación de codificación de texto de encuestas que proporciona un análisis significativo de las respuestas a las preguntas con final abierto. Con este producto, cualquiera que realice análisis de encuestas puede transformar rápidamente las respuestas no estructuradas en datos cuantitativos. El “desbloqueo” de estos datos de texto con final abierto pueden mejorar significativamente la calidad del análisis y la capacidad de la toma de decisiones. Esta aplicación permite importar los datos de la encuesta, extraer los conceptos clave, refinar los resultados y categorizar las respuestas. Una vez categorizados los datos, puede exportar las categorías para importarlas en herramientas analíticas cuantitativas, como el sistema IBM® SPSS® Statistics, para un análisis y generación de gráficos más exhaustivo.

SPSS Text Analytics for Surveys combina las tecnologías lingüísticas avanzadas diseñadas para extraer y clasificar con fiabilidad los conceptos clave de las respuestas de las encuestas con final abierto mediante técnicas manuales. Gracias a los sólidos algoritmos de generación de categorías y a la sencilla funcionalidad de arrastrar y soltar, puede crear categorías o “códigos” en los que categorizar las respuestas de las encuestas. Además, las categorías que se generan podrán reutilizarse para proporcionar así un resultado coherente en estudios idénticos o similares. Puesto que los datos de las respuestas con final abierto pueden variar mucho de una encuesta a otra, dos proyectos nunca serán exactamente iguales; sin embargo, puede seguir el mismo proceso básico para llevar a cabo los análisis. [Si desea obtener más información, consulte el tema El proceso típico en el capítulo 2 el p. 10.](#)

## ***Novedades***

En IBM® SPSS® Text Analytics for Surveys 4.0.1 puede encontrar las siguientes funciones nuevas:

**Categorías jerárquicas.** Ahora las categorías pueden tener una estructura jerárquica, lo que significa que pueden contener subcategorías y estas subcategorías pueden también tener sus propias subcategorías y así sucesivamente. Puede importar estructuras de categorías predefinidas, antes denominadas marcos de código, con categorías jerárquicas así como crear estas categorías jerárquicas en el producto.

De hecho, las categorías jerárquicas le permiten crear una estructura de árbol con una o más subcategorías para agrupar elementos como áreas de temas o conceptos diferentes de manera más precisa. Un ejemplo sencillo puede estar relacionado con actividades de ocio; al responder a una pregunta como *¿Qué actividad le gustaría hacer si tuviera más tiempo?*, puede tener categorías principales como *deportes*, *actividades artísticas*, *pesca*, etc.; bajando un nivel, debajo de *deportes*, puede tener subcategorías para determinar si son *juegos con pelota*, *deportes acuáticos*, etc.

**Acceso a Language Weaver.** El modo de acceso a la interfaz de traducción de Language Weaver se ha simplificado para utilizar una única URL y detalles de seguridad asociados.

## ***Datos de encuestas con final abierto***

Los cuestionarios de las encuestas suelen contener distintos tipos de preguntas, incluidas aquellas con final abierto y cerrado. Una **pregunta con final cerrado** presenta un conjunto limitado de respuestas que permite varios tipos de análisis cuantitativos. Una **pregunta con final abierto** permite que el encuestado aporte una respuesta no estructurada de longitud y grado de detalle variable.

Las palabras que la gente utiliza para responder una pregunta aportan mucha información sobre su opinión. Estas son las razones por las que a menudo se incluyen preguntas con final abierto en las encuestas: proporcionan información más variada y matizada que las preguntas con final cerrado, y a menudo ofrecen una perspectiva que el diseñador de la encuesta no pudo anticipar.

Sin embargo, el uso de un mayor número de preguntas con final abierto siempre se ha considerado económicamente muy costoso, porque implica un sobreesfuerzo a la hora de interpretar las respuestas. Además, estas respuestas largas deben codificarse de una manera estandarizada, empleando un conjunto detallado de instrucciones de codificación. Esta tarea puede resultar difícil porque no siempre los codificadores se ponen de acuerdo a la hora de categorizar respuestas específicas. Cuando los codificadores discrepan entre sí, la fiabilidad de los resultados disminuye. Por todas estas razones, la codificación de respuestas con final abierto se ha considerado siempre un factor que aumenta costes y tiempo que sobrepasa los beneficios derivados de la recopilación de datos.

IBM® SPSS® Text Analytics for Surveys ofrece una alternativa a este costoso procedimiento, puesto que puede afrontar la codificación de respuestas con final abierto en un período de tiempo menor del que se emplea manualmente. Mediante el uso de técnicas y teorías lingüísticas avanzadas, SPSS Text Analytics for Surveys analiza el texto de las respuestas con final abierto como conjuntos de frases y oraciones cuya estructura gramatical ofrece contexto para el significado de una respuesta. Después de analizar el texto, los conceptos clave y los patrones de palabra se extraen y clasifican en categorías.

Puede utilizar las técnicas incorporadas de generación de categorías para crear automáticamente categorías y técnicas manuales para ajustar los resultados. La fiabilidad de los resultados aumenta considerablemente, puesto que la extracción y categorización siempre se realiza de una forma coherente y repetitiva—la misma respuesta se categoriza cada vez en las mismas categorías, a menos que opte por ajustar las definiciones de categorías o las bibliotecas lingüísticas.

El análisis exitoso de la encuesta no depende solamente de un método. La naturaleza subjetiva de la interpretación de las respuestas con final abierto demanda el uso de múltiples técnicas. Además de estas técnicas incorporadas de extracción y generación de categorías, SPSS Text Analytics for Surveys también se basa en la forma en que el usuario se toma los objetivos del análisis del texto de cada encuesta. El análisis del texto es más potente cuando se realiza de una manera repetitiva (extraer, revisar, refinar, volver a extraer), y su grado de utilidad a menudo dependerá de la cantidad de tiempo y esfuerzo que se dedique en revisar y refinar manualmente los resultados de la extracción y las definiciones de categorías. [Si desea obtener más información, consulte el tema \*\*Fiabilidad y ajuste\*\* el p. 8.](#)

Si trabaja con preguntas idénticas o similares en encuestas recurrentes, podrá volver a utilizar las categorías en otras preguntas o proyectos. La reutilización de categorías aumenta en gran medida la coherencia en la codificación, y supone un gran ahorro de tiempo y esfuerzo.

También es posible realizar otros análisis. Las categorías que genere pueden utilizarse en diversos tipos de análisis estadísticos con las otras preguntas del cuestionario o con otros datos demográficos para tener más puntos de vista sobre los encuestados, sus opiniones y comportamientos. Después de utilizar SPSS Text Analytics for Surveys para detectar las categorías que subyacen en un conjunto de respuestas, también puede exportar las categorías para realizar otros análisis cuantitativos en otros programas, como IBM® SPSS® Statistics Base.

## ***Acerca de la minería de datos de texto***

El análisis de texto, un tipo de análisis cualitativo, es la extracción de información útil del texto (como respuestas con final abierto) de manera que las ideas o los conceptos clave que contiene el texto pueden agruparse en una serie de categorías apropiadas. El análisis del texto puede realizarse en textos de cualquier tipo y longitud, aunque el enfoque del análisis puede ser diferente.

Los registros más breves se categorizan con mayor facilidad, porque no son tan complejos y suelen tener menos palabras y respuestas ambiguas. Por ejemplo, una encuesta con preguntas breves con final abierto; si pedimos a la gente que nombre sus tres actividades vacacionales favoritas, podemos esperar muchas respuestas cortas del tipo *ir a la playa, visitar parques nacionales o no hacer nada*. Las respuestas más largas con final abierto, por otro lado, pueden ser bastante complejas y muy largas, sobre todo si los encuestados son cultos, están motivados y tienen tiempo suficiente para completar un cuestionario. Si pedimos a la gente que nos hable sobre sus creencias políticas en una encuesta, o que participen en un blog sobre política, encontraremos comentarios más largos sobre todo tipo de temas y posturas.

Los investigadores de encuestas no suelen analizar las respuestas muy largas. Las respuestas en la mayoría de las encuestas tienden a tener una extensión entre breve y mediana (desde una frase a un párrafo corto). IBM® SPSS® Text Analytics for Surveys se diseñó para poder manejar extensiones de texto como éstas pero puede analizar respuestas mucho más largas.

Existen varios métodos distintos para analizar texto. En primer lugar, está el método manual: hacer que la gente lea las respuestas de la encuesta, que anoten los contenidos, determinen los conceptos clave que contienen y les asignen códigos. Puesto que la gente suele comprender bien el texto, este método es bastante preciso. Pero se invierte mucho tiempo, supone bastante esfuerzo y, dado el cada vez mayor volumen de textos actualmente disponibles, es muy poco práctico. Este método también se basa mayoritariamente en la interpretación de cada codificador.

Otro método es emplear soluciones automáticas. Hay muchas soluciones automáticas diferentes entre las que elegir, entre las que se incluyen soluciones estadísticas y lingüísticas. SPSS Text Analytics for Surveys ofrece una combinación de técnicas lingüísticas y estadísticas automáticas para sacar el máximo partido a los resultados más fiables de cada etapa del proceso. En este producto, las técnicas basadas en la lingüística se utilizan para extraer automáticamente los conceptos clave a partir de las respuestas, y pueden utilizarse tanto técnicas lingüísticas como estadísticas para crear las definiciones (códigos) de categoría que se asignan a las respuestas.

## ***Cómo funciona la extracción***

Durante la extracción de los conceptos e ideas clave de las respuestas, IBM® SPSS® Text Analytics for Surveys se basa en el análisis de texto basado en lingüística. Este método ofrece la velocidad y la rentabilidad propia de los sistemas basados en estadísticas. Pero ofrece un grado de precisión mucho mayor, y con menos intervención humana. El análisis de texto basado en

lingüística se basa en el ámbito de estudio conocido como proceso de idioma natural, conocido también como lingüística computacional.

Para ilustrar la diferencia entre los métodos basados en estadísticas y en lingüística durante el proceso de extracción, considere cómo respondería cada uno a una pregunta sobre la reproducción de documentos. Tanto las soluciones basadas en estadísticas como en lingüística tendrían que ampliar la palabra *reproducción* para que incluyera sinónimos como *copia* y *duplicación*. De lo contrario, se pasaría por alto información relevante. Pero si la solución basada en estadísticas intenta realizar este tipo de sinonimia—buscar otros términos con el mismo significado—probablemente incluiría el término *nacimiento*, lo que generaría resultados irrelevantes. La comprensión del idioma reduce la ambigüedad del texto, lo que convierte a la minería de texto basada en lingüística, por definición, en el método más fiable.

Comprender el funcionamiento del proceso de extracción puede ayudarle a tomar decisiones clave a la hora de ajustar los recursos lingüísticos (bibliotecas, tipos, sinónimos, etc.). Estos son los pasos del proceso de extracción:

- Conversión de datos de origen en un formato estándar
- Identificar los términos candidatos
- Identificar las clases de equivalencias y la integración de sinónimos
- Asignar un tipo
- Crear índices
- Extraer patrones y eventos de relaciones

### **Paso 1. Conversión de datos de origen en un formato estándar**

En este primer paso, los datos que se importan se convierten en un formato uniforme que puede utilizarse para realizar otros análisis. Esta conversión se lleva a cabo internamente y no cambia los datos originales.

### **Paso 2. Identificar los términos candidatos**

Es importante comprender el rol de los recursos lingüísticos en la identificación de los términos candidatos durante la extracción lingüística. Los recursos lingüísticos se utilizan siempre que se ejecuta una extracción. Existen en forma de plantillas, bibliotecas y recursos compilados. Las bibliotecas incluyen listas de palabras, relaciones y otra información que se utiliza para especificar o ajustar la extracción. Estos recursos compilados no se pueden ver ni editar. Sin embargo, los recursos restantes (plantillas) sí pueden editarse en el Editor de recursos.

Los recursos compilados son los componentes internos principales del motor de extracción en IBM® SPSS® Text Analytics for Surveys. Estos recursos incluyen un diccionario general que contiene una lista de formatos base con un código de categoría léxica (sustantivo, verbo, adjetivo, adverbio, participio, conjunción, determinante o preposición). Los recursos también incluyen tipos incorporados reservados que se utilizan para asignar muchos términos extraídos a los tipos siguientes: <Location>, <Organization>, o <Person>. [Si desea obtener más información, consulte el tema Tipos incorporados en el capítulo 10 el p. 216.](#)

Además de los recursos compilados, se entregan varias bibliotecas con el producto y pueden utilizarse para complementar los tipos y las definiciones de conceptos en los recursos compilados, así como para ofrecer otros tipos y sinónimos. Estas bibliotecas—y las bibliotecas personalizadas que haya creado—se componen de varios diccionarios. Entre ellos están los diccionarios de

tipo, los diccionarios de sustitución (sinónimos y elementos opcionales) y los diccionarios de exclusión. [Si desea obtener más información, consulte el tema Trabajo con bibliotecas en el capítulo 9 el p. 202.](#)

Una vez importados y convertidos los datos, el motor de extracción empezará a identificar los términos candidatos para la extracción. Los términos candidatos son palabras o grupos de palabras que se utilizan para identificar conceptos en el texto. Durante el proceso del texto, las palabras simples (**unitérminos**) que no están en los recursos compilados se consideran como extracciones de términos candidatos. Las palabras compuestas candidatas (**multitérminos**) se identifican mediante extractores de patrones de categorías léxicas. Por ejemplo, el multitérmino `coche deportivo`, que responde al patrón de categoría léxica “sustantivo adjetivo”, tiene dos componentes. El multitérmino `coche deportivo veloz`, que responde al patrón de categoría léxica “sustantivo adjetivo adjetivo”, tiene tres componentes.

*Nota:* los términos del diccionario general compilado arriba mencionado representan una lista de todas las palabras que probablemente no resultan relevantes o que son lingüísticamente ambiguas, como los unitérminos. Estas palabras se excluyen de la extracción cuando se están identificando los unitérminos. Sin embargo, volverán a evaluarse cuando determine las categorías léxicas no cuando busque palabras compuestas candidatas más largas (multitérminos).

Por último, se utiliza un algoritmo especial para gestionar las cadenas de letras en mayúsculas, como cargos laborales, de manera que estos patrones especiales pueden extraerse.

### **Paso 3. Identificar las clases de equivalencias y la integración de sinónimos**

Una vez identificados los unitérminos y los multitérminos candidatos, el software emplea un conjunto de algoritmos para compararlos y para identificar clases de equivalencias. Una clase de equivalencia es la forma básica de una frase o una forma simple de dos variantes de la misma frase. El propósito de asignar frases a las clases de equivalencias es asegurarse de que, por ejemplo, `presidente de la compañía` y `la compañía` y `su presidente` no se consideren conceptos separados. Para determinar qué concepto se va a utilizar para la clase de equivalencia—es decir, cuál de las dos opciones, `presidente de la compañía` o bien `la compañía` y `su presidente` se utiliza como término principal, el motor de extracción aplica las reglas siguientes en el orden que aparece:

- La forma especificada por el usuario en una biblioteca.
- La forma más frecuente en el cuerpo completo del texto.
- La forma más corta en el cuerpo completo del texto (que generalmente se corresponde a la forma básica).

### **Paso 4. Asignar un tipo**

A continuación, se asignan tipos a los conceptos extraídos. Un tipo es una agrupación semántica de conceptos. En este paso se utilizan tanto los recursos compilados como las bibliotecas. Los tipos incluyen elementos como conceptos de nivel más alto, palabras positivas y negativas, nombres propios, lugares, organizaciones, etc. El usuario puede definir tipos adicionales. [Si desea obtener más información, consulte el tema Diccionarios de tipo en el capítulo 10 el p. 215.](#)

### **Paso 5. Crear índices**

Se crea el índice del conjunto completo de registros estableciendo un marca entre una posición de texto y el término representativo de cada clase de equivalencia. De esta manera se presupone que todos los casos de la forma declinada de un concepto candidato se indexa como forma básica candidata. Para cada forma básica se calcula la frecuencia global.

### **Paso 6. Extraer patrones y eventos de relaciones**

IBM SPSS Text Analytics for Surveys puede detectar no solamente tipos y conceptos, sino también las relaciones entre ellos. Hay varios algoritmos y bibliotecas disponibles en el producto que proporcionan la capacidad de extraer patrones de relaciones entre tipos y conceptos. Son particularmente útiles cuando se intentan detectar opiniones específicas (por ejemplo, reacciones ante productos) o los enlaces relacionales entre personas y objetos (por ejemplo, enlaces entre grupos políticos o genomas).

## ***Funcionamiento de la categorización***

Existen varias técnicas diferentes entre las que puede escoger para crear categorías. Puesto que cada conjunto de datos es exclusivo, el número de técnicas y el orden en el que las aplica puede cambiar con el tiempo. Puesto que su interpretación de los resultados puede ser diferente de la interpretación de otros, puede que deba experimentar con diferentes técnicas para comprobar con cuál de ellas se obtiene el mejor resultado para los datos del texto.

En esta guía, la **generación de categorías** hace referencia a la generación de definiciones de categoría y clasificación mediante el uso de una o más técnicas incorporadas, y **categorización** hace referencia al proceso de puntuación o etiquetaje por el que se asignan identificadores exclusivos (nombre/ID/valor) a las definiciones de categorías para cada registro.

Durante la generación de categorías, los conceptos y los tipos que se extrajeron se utilizan como los cimientos para las categorías. Cuando crea categorías, los registros se asignan automáticamente a categorías si contienen texto que coincida con un elemento de una definición de categoría.

IBM® SPSS® Text Analytics for Surveys ofrece varias técnicas automáticas de generación de categorías para ayudarle a categorizar los registros rápidamente.

### ***Agrupación de técnicas***

Cada una de las técnicas disponibles resulta idónea para determinados tipos de datos y situaciones, pero a menudo conviene combinar técnicas en el mismo análisis para capturar el rango completo de registros. Puede ver un concepto en diversas categorías o detectar categorías redundantes.

**Derivación de raíz de conceptos.** Esta técnica crea categorías tomando un concepto y buscando otros conceptos que estén relacionados con el primero analizando si alguno de los componentes de los conceptos están morfológicamente relacionados o comparten raíces. Esta técnica es muy útil para identificar conceptos de palabras compuestas sinónimas, puesto que los conceptos de cada categoría generada son sinónimos o tienen un significado muy similar. Funciona con datos de extensión diversa y genera un número más reducido de categorías compactas. Por ejemplo, el concepto `ocasiones de progreso` se agruparía con los conceptos `oportunidad de progresar` y `oportunidad de progresión`. [Si desea obtener más información, consulte el tema Derivación de raíz de conceptos en el capítulo 6 el p. 118.](#)

**Red semántica.** Esta técnica comienza identificando los posibles sentidos de cada concepto a partir de un amplio índice de relaciones de palabras, y luego crea categorías agrupando los conceptos relacionados. Esta técnica resulta idónea cuando los conceptos son conocidos en la red semántica y no son muy ambiguos. Es menos idónea si el texto contiene terminología específica o jerga desconocida en la red. Por ejemplo, el concepto `manzana golden` se podría agrupar con `manzana reineta` y `manzana fuji` puesto que son familia de la `golden`. En otro ejemplo, el concepto `animal` se agruparía con `gato` y `canguro` puesto que ambos son hipónimos de `animal`. En esta versión esta técnica está disponible solo para texto en inglés. [Si desea obtener más información, consulte el tema Redes semánticas en el capítulo 6 el p. 121.](#)

**Inclusión de conceptos.** Esta técnica genera categorías agrupando los conceptos multitérmino (palabras compuestas) basándose en si contienen palabras que son subconjuntos o superconjuntos de una palabra en la otra. Por ejemplo, el concepto `seguridad` estaría agrupado en `asiento de seguridad`, `cinturón de seguridad` y `silla infantil de seguridad`. [Si desea obtener más información, consulte el tema Inclusión de conceptos en el capítulo 6 el p. 120.](#)

**Co-ocurrencia.** Esta técnica crea categorías a partir de las co-ocurrencias que se encuentran en el texto. La idea radica en que cuando en los documentos y registros a menudo se encuentran conceptos o patrones de conceptos que aparecen juntos, esa co-ocurrencia refleja una relación subyacente que probablemente sea valiosa para las definiciones de categorías. Cuando la co-ocurrencia de algunas palabras es significativa, se crea una regla de co-ocurrencia que puede utilizarse como un descriptor de categoría para una nueva subcategoría. Por ejemplo, si muchos registros contienen las palabras `precio` y `disponibilidad` (pero hay pocos registros que contengan sólo una de las dos), estos conceptos se podrían agrupar en una regla de co-ocurrencia, (`precio & disponible`) y asignarse a una subcategoría de la categoría `precio` por ejemplo. [Si desea obtener más información, consulte el tema Reglas de co-ocurrencia en el capítulo 6 el p. 122.](#)

- **Número mínimo de registros** Para ayudar a determinar la relevancia de las co-ocurrencias, defina el número mínimo de registros que deben contener una co-ocurrencia determinada para que se utilice como descriptor en una categoría.

## ***Preparación para el análisis del texto***

El análisis de texto implica algo más que extraer y categorizar. Para analizar el texto satisfactoriamente, tenga en cuenta los puntos siguientes:

- Al igual que ocurre en el diseño de las encuestas, la calidad de las respuestas que se importan a IBM® SPSS® Text Analytics for Surveys afecta directamente a la calidad de las categorizaciones resultantes. En general, las preguntas vagas o confusas conllevan respuestas que pueden desviarse o divagar y que resultan muy difíciles de analizar.
- Como ocurre en el análisis estadístico, el análisis de texto debe realizarse teniendo muy claros los objetivos. Antes de empezar cualquier análisis, debe determinar y reflejar en el estudio qué es lo que pretende averiguar.

Por ejemplo, supongamos que se realizó una encuesta en una escuela pública para valorar la actitud de los padres con relación a la calidad de la enseñanza que están recibiendo sus hijos. Durante el análisis, podemos centrarnos en temas como el nombre de los profesores, los programas educativos, etc., o bien en identificar y agrupar las opiniones positivas y negativas. Asimismo, debe definirse el nivel de granularidad necesario para el análisis, como por ejemplo

agrupar todos los comentarios sobre financiación conjunta o dividir esta categoría en financiación por programa. Los códigos o categorías que creemos deberán reflejar el enfoque y los objetivos de nuestros análisis.

- Lejos de ser un mero análisis estadístico, el análisis de texto no es una ciencia exacta, puesto que no existe un resultado “correcto”. El análisis de texto se realiza teniendo presentes los objetivos, pero también es subjetivo en tanto que está influido por la interpretación que el analista hace de la opinión del encuestado—por ejemplo, cómo se identifica y se clasifican opiniones sarcásticas. En función de los objetivos y del enfoque, dos empleados competentes pueden analizar los mismos datos y alcanzar diferentes conclusiones.
- El análisis de texto es, en gran medida, un *proceso repetitivo*. A medida que se trabaja con las respuestas de la encuesta, se vuelven a extraer y a categorizar las respuestas utilizando diferentes definiciones de categoría (es decir, esquemas de codificación), diferentes definiciones de sinónimos o conceptos, y diferentes agrupaciones de respuestas.

Una vez extraídos los conceptos del texto y creadas las categorías, los resultados deben examinarse con detenimiento. Si se encuentran elementos que desee ajustar, adapte el análisis reajustando las definiciones de categoría y de biblioteca lingüística. Entonces las respuestas se volverán a categorizar automáticamente cuando las extraiga de nuevo. Puede que tenga que llevar a cabo este proceso tantas veces como sea necesario hasta que esté conforme con los resultados del análisis.

*Nota:* Para obtener más información sobre las consideraciones a tener en cuenta antes de importar datos, consulte [Preparación de los datos el p. 27](#)

## ***Fiabilidad y ajuste***

Cuando codifique los datos querrá que las categorías resultantes sean fiables. En el contexto de codificación de respuestas con final abierto, esto significa que dos codificadores independientes que utilicen las mismas reglas (marco de codificación) codificarán de forma idéntica una misma respuesta. Cuando el análisis de texto se hace manualmente, esto puede convertirse en un problema grave. Puede crearse un valioso conjunto de categorías, pero si no pueden aplicarse con fiabilidad a las respuestas, su valor disminuye considerablemente. Cuando se aplica IBM® SPSS® Text Analytics for Surveys a los mismos datos y con los mismos recursos lingüísticos, siempre reproduce a la perfección un análisis anterior. Esto significa que es un 100% fiable.

Sin embargo, esto no significa que no haya errores en el análisis, sino que el enfoque en la codificación puede cambiar por otro—*ajuste*. En el proceso de codificación humana, los codificadores leen la respuesta y pueden capturar todos los matices de una frase (aunque tengan dificultades para aplicar las categorías de codificación). SPSS Text Analytics for Surveys puede aplicar las categorías de clasificación, pero éstas deben estar definidas para que puedan captarse los matices y las diferencias. Existen dos formas de realizar el ajuste:

- Refinar los recursos lingüísticos
- Refinar las definiciones de categoría

## ***Refinamiento de los recursos lingüísticos***

IBM® SPSS® Text Analytics for Surveys crea fácilmente categorías sin necesidad de que usted intervenga, pero no capturará toda la información de las respuestas. Necesitará mejorar la base lingüística que utiliza el programa para que la creación de categorías se adapte cada vez mejor a las idiosincrasias del texto. Para mejorar esta base, puede personalizar y ajustar los recursos lingüísticos que se utilizan a la hora de realizar la extracción del texto.

El ajuste, en este caso, implica añadir palabras y frases a varios diccionarios y bibliotecas lingüísticas, especificando palabras para excluirlas del análisis, definiendo sinónimos o creando bibliotecas personalizadas teniendo presente un objetivo específico. Este objetivo radica en capturar con precisión las ideas del texto de los encuestados y evitar la ambigüedad en los resultados.

## ***Refinamiento de las definiciones de categoría***

Además de refinar los recursos lingüísticos, debe revisar las categorías y averiguar maneras de combinar o limpiar las definiciones, así como comprobar algunas de las respuestas categorizadas. Puede utilizar las técnicas automáticas de generación de categorías para crear las categorías; sin embargo, seguramente querrá realizar algunos ajustes en estas definiciones. Después de utilizar una técnica, en la ventana aparecerán una serie de categorías nuevas. Puede expandir las categorías para poder ver los conceptos que definen cada categoría. Luego puede revisar las respuestas de una categoría y realizar ajustes hasta que esté conforme con las definiciones de categoría.

Ninguna de las técnicas automáticas categorizará los datos con exactitud, por lo que se recomienda que busque y aplique una o más técnicas automáticas que se adapten bien a sus datos. Luego puede utilizar técnicas manuales para realizar pequeños ajustes, eliminar clasificaciones incorrectas o añadir registros o conceptos que pueden haberse dejado de lado.

## ***Primeros pasos***

Esta documentación presenta las tareas que puede realizar con IBM® SPSS® Text Analytics for Surveys y las técnicas que puede utilizar para categorizar sus respuestas. La información que se presenta aquí le guiará a través de su análisis inicial. Trata sobre todos los procesos necesarios para analizar sus datos en su totalidad, pero puesto que cada conjunto de datos es diferente, deberá decidir en qué momento el análisis está completo.

En este capítulo, se trata sobre el típico proceso que los usuarios realizan al llevar a cabo el análisis del texto. La interfaz también se explica desde una perspectiva de alto nivel junto con las tareas y los elementos principales con los que va a trabajar.

### ***El proceso típico***

A continuación se presenta un resumen del proceso de flujo de trabajo típico que deberá realizar cuando utilice IBM® SPSS® Text Analytics for Surveys.

- **Crear un proyecto** mediante la importación de los datos de la encuesta, incluyendo respuestas con final abierto, una variable de ID y otras variables de referencia en SPSS Text Analytics for Surveys. Los datos pueden leerse desde los archivos de datos de IBM® SPSS® Statistics, Microsoft Excel, cualquier programa de base de datos que utilice ODBC o un origen de datos de IBM® SPSS® Data Collection. Puede elegir un paquete de análisis de texto para beneficiarse de algunas categorías predefinidas y recursos especializados para comenzar a trabajar rápidamente.
- **Extraer** conceptos y patrones para cada pregunta con final abierto que importe. El motor interno de extracción identifica y recopila automáticamente los términos clave que se expresan en el texto. Estos términos se agrupan bajo un concepto principal. Los conceptos luego se agrupan en tipos, que son recopilaciones de términos similares, como organizaciones, productos u opiniones positivas. También se extraen patrones, que representan combinaciones de términos y tipos que a su vez representan opiniones y relaciones, como comentarios positivos sobre una empresa.
- **Refinar** los conceptos extraídos y ajustar las extracciones mediante la manipulación directa de una o más bibliotecas que contienen tipos de palabras, términos, sinónimos, listas de exclusión y otras construcciones lingüísticas. Tal como se mencionó anteriormente, el análisis del texto es un proceso repetitivo donde el proceso de refinar las bibliotecas y los diccionarios genera directamente resultados que se ajustan a los datos.
- **Categorizar** las respuestas mediante la creación y edición manual de categorías utilizando reglas de categorías, marcos de código o empleando automáticamente técnicas de generación de categorías. Las categorías representan conceptos de mayor nivel que capturan las ideas y las actitudes principales que expresan los encuestados.
- **Exportar** las categorías junto con la variable de ID a formatos de archivo comunes para la generación de gráficos y análisis más profundos en otras aplicaciones. El resultado puede ser un conjunto de variables de respuesta múltiple, como un archivo SPSS Statistics o Microsoft Excel.

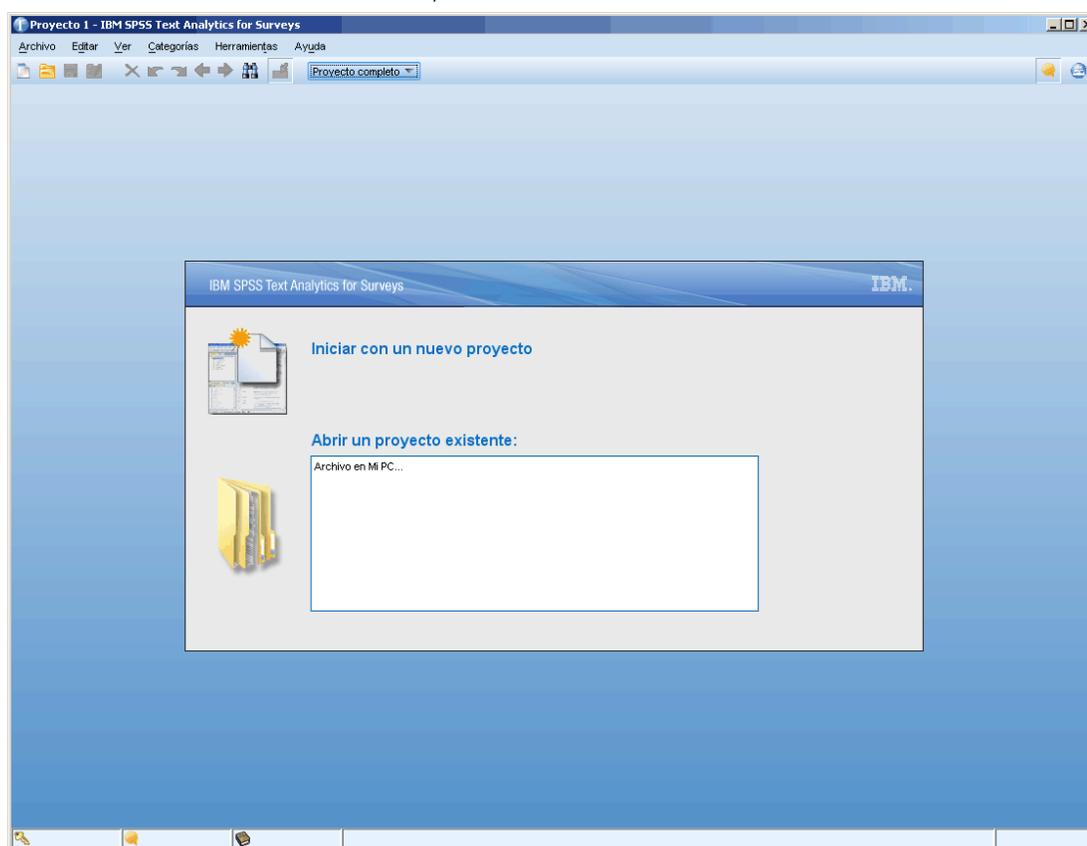
## La ventana Análisis de texto

La interfaz de la aplicación está formada por dos ventanas. La primera es la ventana del análisis de texto, donde se realiza la mayor parte del trabajo. En esta ventana, puede analizar cada pregunta en los datos. Para cada pregunta puede extraer conceptos, tipos y patrones, y luego categorizar sus respuestas.

Cuando inicia la aplicación, se presenta una pantalla donde puede abrir un proyecto existente o crear uno nuevo. Si opta por crear un proyecto nuevo, se abre un asistente que le guía a través del proceso de creación del proyecto. [Si desea obtener más información, consulte el tema Creación de proyectos en el capítulo 3 el p. 26.](#)

Figura 2-1

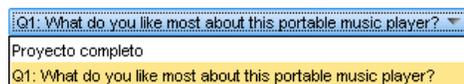
La ventana Análisis de texto al iniciar el producto



Una vez importados los datos, puede ver lo siguiente: las vistas de Pregunta o la vista Proyecto completo. Puede cambiar las vistas seleccionando una en la lista desplegable de la barra de herramientas en la ventana del análisis de texto, o seleccionando la vista desde el menú Ver. El texto que aparece en el cuadro de lista procede de la etiqueta de la variable de cada pregunta.

Figura 2-2

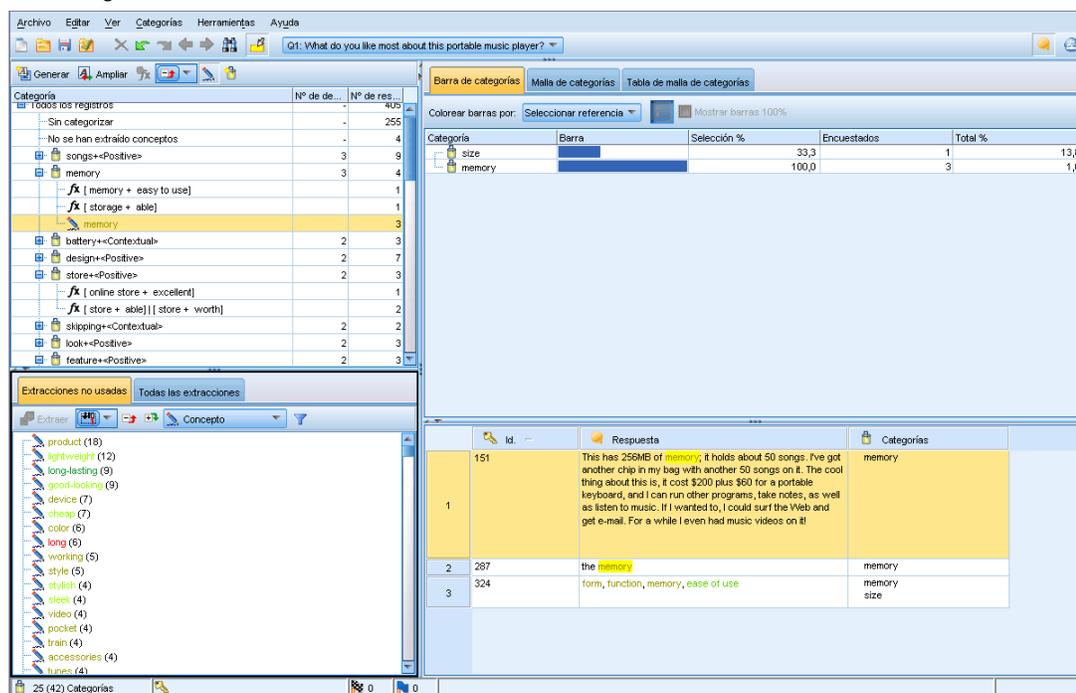
Barra de herramientas de la aplicación con la lista desplegable del selector de vistas



## La vista Pregunta

La vista Pregunta proporciona un espacio en el que puede analizar y categorizar las respuestas en una pregunta determinada. Después de crear un proyecto nuevo, la vista Pregunta aparece con las respuestas correspondientes a la primera variable de texto con final abierto que ha importado. Puede seleccionar esta vista desde la lista desplegable o elegir un nombre de pregunta desde el menú Ver > Pregunta >.

Figura 2-3  
Vista Pregunta



Las operaciones que puede realizar en la vista Pregunta se concentran en tres elementos: resultados extraídos, categorías y datos de respuesta. Para ayudarle a analizar y trabajar con cada uno de los elementos independientemente, esta ventana está dividida en cuatro paneles:

### Panel Categorías

En el ángulo superior izquierdo, este panel presenta un área en el que puede crear y administrar las categorías que genere. Después de extraer los conceptos, tipos y patrones de los datos de texto, puede empezar a generar categorías utilizando técnicas automáticas, como redes semánticas e inclusión de conceptos, o bien puede hacerlo manualmente. Puede pulsar y expandir una categoría para ver todos los descriptores que conforman su definición, como conceptos, tipos y reglas. Cuando selecciona una categoría o descriptor, puede mostrar la información sobre los registros correspondientes en los paneles Datos y Visualización. [Si desea obtener más información, consulte el tema El panel de categorías en el capítulo 6 el p. 94.](#)

Figura 2-4  
Panel Categorías: Definición expandida de categoría

Categoría	Nº de de...	Nº de res...
function	1	3
headphones	2	4
home	3	3
listening	3	27
memory device	13	29
tape		2
<b>memory</b>	<b>8</b>	<b>22</b>
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
record		1

### Panel Resultados extraídos

Situada en el ángulo inferior izquierdo, esta área presenta los resultados de la extracción. Cuando ejecuta una extracción, el motor de extracción lee los datos del texto, identifica los conceptos relevantes y asigna un tipo a cada uno. Los **Conceptos** son palabras o frases extraídas de los datos del texto. Los **Tipos** son agrupaciones semánticas de conceptos que se guardan como diccionarios de tipo. Cuando finaliza la extracción, los conceptos, tipos y patrones aparecen en el panel Resultados extraídos. Los conceptos y los tipos están codificados por colores para ayudarle a identificar a qué tipo pertenecen. Puede utilizar estos conceptos, tipos y patrones que recopile aquí para generar las categorías. [Si desea obtener más información, consulte el tema Resultados extraídos: Conceptos, tipos y patrones en el capítulo 5 el p. 78.](#)

Minería de textos es un proceso repetitivo en el que los resultados de la extracción se revisan de acuerdo con el contexto de los datos de texto, se ajustan para generar resultados nuevos y luego se reevalúan. Los resultados de la extracción pueden refinarse modificando los recursos lingüísticos. Este ajuste puede realizarse en parte directamente desde el panel Resultados extraídos o Datos, pero también directamente en la vista Editor de recursos. [Si desea obtener más información, consulte el tema La ventana Editor de recursos el p. 15.](#)

### Panel Datos

Situado en el ángulo inferior derecho de esta vista, presenta, en formato tabular, los datos de respuesta correspondientes a la pregunta con final abierto seleccionada. Por defecto, el panel Datos muestra tres columnas (ID de registro, respuestas de texto y categorías asignadas). El número de respuestas que aparecen en este panel se filtran en función de lo que haya seleccionado en otro panel. Mientras esté viendo los datos que ha importado en este panel, no podrá editar, eliminar ni agregar nada en los registros. [Si desea obtener más información, consulte el tema El panel de datos en el capítulo 6 el p. 97.](#)

### **Panel Visualización**

Situado en el ángulo superior derecho de esta vista, por defecto está oculto. Para que se muestre este panel, elija Ver > Visualización). Este panel ofrece tres vistas exclusivas sobre la forma en que las respuestas se adaptan a las categorías, y la forma en que las categorías pueden compartir respuestas (gráfico de web, gráfico de barras y tabla) según las opciones que haya seleccionado en los otros paneles. [Si desea obtener más información, consulte el tema Visualización de gráficos en el capítulo 7 el p. 167.](#)

En función de si elige la opción de extracción en el Asistente de proyecto nuevo, puede que aparezcan o no los resultados de la extracción en el ángulo inferior izquierdo del panel. Pulse en Extraer en el panel Resultados extraídos para comenzar la extracción. Después de la extracción, puede revisar los resultados para comprobar si es necesario realizar algún ajuste, como la agrupación de sinónimos bajo un nombre de concepto, o la exclusión de conceptos comunes y poco relevantes de la lista.

Una vez satisfecho con los resultados de la extracción, puede empezar a categorizar las respuestas manualmente arrastrando y soltando los conceptos como categorías, o utilizando técnicas de generación de categorías, como la inclusión de conceptos y la red semántica.

### **La vista Proyecto completo**

La vista Proyecto completo proporciona una visión general de todas las variables que ha importado al proyecto. Puede seleccionar esta vista desde la lista desplegable o desde el menú Ver > Proyecto completo. En esta vista, puede revisar los datos que ha importado, cambiar el rol de la variable (por ejemplo, de pregunta a variable de referencia) y asignar etiquetas a las variables. [Si desea obtener más información, consulte el tema Vista de los datos del proyecto en el capítulo 4 el p. 49.](#) Mientras esté viendo los datos que ha importado en esta vista, no podrá editar, corregir, eliminar ni agregar nada en los registros.

*Nota:* Para ver el contenido completo de una casilla de esta vista, pase el ratón por encima de la casilla. Aparece una etiqueta con información sobre el contenido de la casilla.

Figura 2-5  
vista Proyecto completo

	Número de serie	¿Qué fue lo que más le gustó de su estadia en nuestro hotel?	¿Qué fue lo que menos le gustó de su estadia en nuestro hotel?	1. Razones de su viaje.	Demográficos - Rango de Edad
1	1	El hotel cumple con todo lo que se le pide: bi...	no disponen de televisión vía satélite	Placer	De 31 a 40 años
2	2	la comida y la piscina cubierta	el precio, demasiado caro	Placer	De 21 a 30 años
3	3	muy centrico	parking	Placer	De 21 a 30 años
4	4	El restaurante y la vista desde las habitacio...	No tiene un área de parking para los huespe...	Negocios	De 31 a 40 años
5	5	Atención del personal	Lejos de la ciudad	Placer	De 21 a 30 años
6	6	Lo que más me gustó fue el trato de los em...	El tamaño de las habitaciones es muy peque...	Negocios	41 a 50 años
7	7	Servicio de Restaurante ofrecido por el Hotel.	La piscina no estaba lo limpia que esperaba.	Placer	De 51 a 60 años
8	8	El excelente trato del personal de todo el Ho...	La localización del hotel, está muy retirado d...	Placer	De 31 a 40 años
9	9	La limpieza de las habitaciones	El hotel es un poco pequeño, y las habitacion...	Negocios	41 a 50 años
10	10	Durante mi ESTANCIA, lo que más me gustó...	El olor a cañerías	Placer	De 21 a 30 años
11	11	Bien comunicado con el centro de la ciudad ...	No tienen una política de clientes definida, ni ...	Placer	De 21 a 30 años
12	12	Lo mejor es la ubicación. Un paisaje extraor...	Los tiempos de servicio. Específicamente en ...	Placer	41 a 50 años
13	13	Que tenía todo y era muy conveniente.	El cuarto era peque-no.	Negocios	De 31 a 40 años
14	14	La oferta que me ofrecieron era lo suficien...	El restaurante. La calidad de la comida no me...	Negocios	De 31 a 40 años
15	15	El desayuno	Las vistas de la habitación	Placer	De 21 a 30 años
16	16	El servicio que me han ofrecido	El baño no tenía bañera	Negocios	41 a 50 años
17	17	La sesión de piano en el bar por la noche	Hacía demasiado calor por la noche	Negocios	De 31 a 40 años
18	18	La localización y tranquilidad interior	Está un poco antiguo	Placer	De 21 a 30 años
19	19	limpieza de las habitaciones	no televisión vía satélite	Placer	41 a 50 años
20	20	acceso a internet gratis	aparcamiento	Negocios	41 a 50 años
21	21	limpieza de la habitacion	no disponen de televisión vía satélite	Negocios	41 a 50 años
22	22	Fantástico hotel ubicado en un elegante edif...	el mini bar es excesivamente caro	Placer	Más de 60 años
23	23	muy bien ubicado	ruido exterior	Placer	De 21 a 30 años
24	24	encantador	todo correcto	Negocios	De 31 a 40 años
25	25	Las habitaciones son un poco pequeñas pe...	ruidoso	Placer	41 a 50 años
26	26	Habitaciones decoradas con buen gusto	El buffet del desayuno un poco pobre pero a...	Placer	41 a 50 años
27	27	piscina cubierta. Genial para los niños	las habitaciones son diminutas	Negocios	Más de 60 años
28	28	acceso a internet en las habitaciones	Aislamiento sonoro entre habitaciones	Negocios	De 51 a 60 años
29	29	nada exceptional	no funcionaba el aire acondicionado, jinsopo...	Placer	Más de 60 años
30	30	ubicacion	servicio absolutamente sucio y usado	Placer	De 51 a 60 años
31	31	muy limpio	acceso difícil	Negocios	De 21 a 30 años

## La ventana Editor de recursos

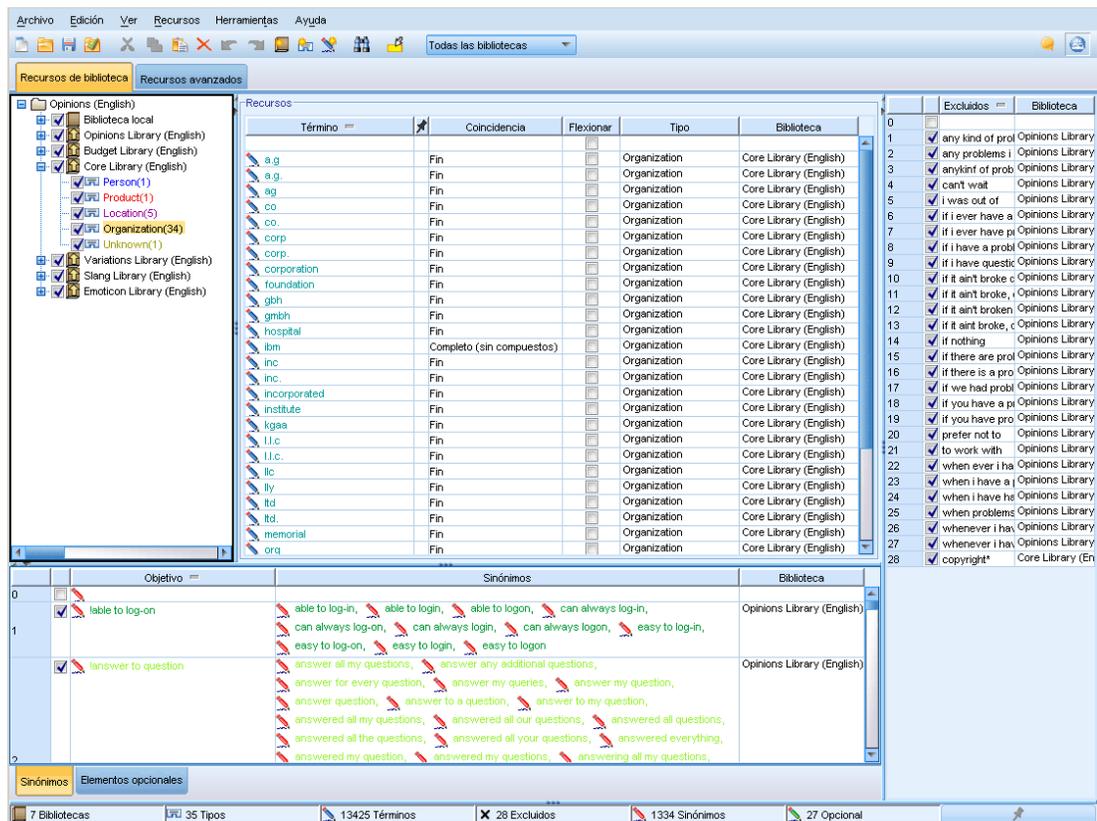
IBM® SPSS® Text Analytics for Surveys captura con rapidez y precisión los conceptos clave desde los datos de texto utilizando un motor de extracción robusto. Este motor se basa principalmente en los recursos lingüísticos para dictaminar qué cantidad de datos textuales y sin estructurar deben analizarse e interpretarse.

La vista del Editor de recursos es el lugar donde puede ver y ajustar los recursos lingüísticos que se utilizan para extraer conceptos, agruparlos bajo tipos, detectar patrones en los datos de texto y muchas cosas más. IBM® SPSS® Text Analytics for Surveys ofrece muchas plantillas de recursos preconfiguradas. Además, en algunos idiomas puede utilizar los recursos en paquetes de análisis de texto. [Si desea obtener más información, consulte el tema Uso de los paquetes de análisis de texto en el capítulo 3 el p. 41.](#)

Puesto que estos recursos no siempre se adaptan perfectamente al contexto de sus datos, puede crear, editar y administrar sus propios recursos para un contexto o dominio determinados en el Editor de recursos. [Si desea obtener más información, consulte el tema Trabajo con bibliotecas en el capítulo 9 el p. 202.](#) Es posible que esta ventana se utilice con muy poca frecuencia, pues los recursos que se entregan con el producto suelen ser suficientes. Además, gran parte del trabajo más sencillo de biblioteca puede realizarse directamente desde el panel de resultados extraídos en la ventana de análisis de texto.

Para simplificar el proceso de ajuste de los recursos lingüísticos, puede realizar las tareas de diccionario más comunes directamente en la vista de Text Analytics a través de los menús contextuales de los paneles de Datos y de Resultados extraídos. [Si desea obtener más información, consulte el tema Refinamiento de los resultados de la extracción en el capítulo 5 el p. 86.](#)

Figura 2-6  
Vista del editor de recursos



Las operaciones que realiza en el Editor de recursos se concentran en torno a la administración y ajuste de los recursos lingüísticos. Estos recursos se almacenan en forma de plantillas y bibliotecas. La vista del Editor de recursos está organizada en cuatro partes: Panel del árbol de bibliotecas, panel de diccionarios de tipo, panel del diccionario de sustitución y panel del diccionario de exclusión.

La interfaz está organizada en cuatro partes:

**1. Panel del árbol de bibliotecas.** Situada en el ángulo superior izquierdo, este plan muestra un árbol con las bibliotecas. Puede activar y desactivar las bibliotecas de este árbol y filtrar las vistas en los otros paneles seleccionando una biblioteca del árbol. Puede realizar muchas operaciones en este árbol utilizando los menús contextuales. Si expande una biblioteca en el árbol, podrá ver el conjunto de tipos que contiene. También puede filtrar esta lista a través del menú Ver si desea centrarse únicamente en una biblioteca determinada.

**2. Listas de términos del panel Diccionarios de tipo.** Situado a la derecha del árbol de bibliotecas, este panel muestra las listas de términos de los diccionarios de tipo de las bibliotecas seleccionadas en el árbol. Un **diccionario de tipo** es una recopilación de términos que se agrupan bajo una misma etiqueta, tipo o nombre. Cuando el motor de extracción lee los datos de texto, compara las palabras que encuentra en el texto con los términos definidos en los diccionarios de tipo. Si un concepto extraído aparece como término en un diccionario de tipo, se asignará ese nombre de tipo. Puede considerar que el diccionario de tipo es un diccionario específico de términos que tienen algo en común. Por ejemplo, el tipo <Location> de la biblioteca Core contiene conceptos como *nueva orleans*, *gran bretaña* y *nueva york*. Todos estos términos representan ubicaciones geográficas. Una biblioteca puede contener uno o más diccionarios de tipo. [Si desea obtener más información, consulte el tema Diccionarios de tipo en el capítulo 10 el p. 215.](#)

**3. Panel del diccionario de exclusión.** Situado en el lado derecho, este panel muestra la colección de términos que se excluirán de los resultados de extracción finales. Los términos que aparecen en este diccionario de exclusión no aparecen en el panel Resultados extraídos. Los términos excluidos pueden almacenarse en la biblioteca que usted elija. Sin embargo, el panel Diccionario de exclusión muestra todos los términos extraídos de todas las bibliotecas visibles en el árbol de bibliotecas. [Si desea obtener más información, consulte el tema Diccionarios de exclusión en el capítulo 10 el p. 231.](#)

**4. Panel del diccionario de sustitución.** Situado en la parte inferior izquierda, este panel muestra los sinónimos y los elementos opcionales, cada uno en su propia pestaña. Los sinónimos y los elementos opcionales ayudan a agrupar términos similares bajo un concepto principal u objetivo en los resultados de extracción finales. Este diccionario puede contener sinónimos conocidos y sinónimos definidos por el usuario y elementos, así como los errores ortográficos más comunes emparejados con la ortografía correcta. Las definiciones de sinónimos y los elementos opcionales pueden almacenarse en la biblioteca que elija. Sin embargo, el panel del diccionario de sustitución muestra todos los contenidos de todas las bibliotecas visibles en el árbol de bibliotecas. Mientras que este panel muestra todos los sinónimos o elementos opcionales de todas las bibliotecas, las sustituciones para todas las bibliotecas del árbol se muestran conjuntamente en este panel. Una biblioteca puede contener tan solo un diccionario de sustitución. [Si desea obtener más información, consulte el tema Diccionarios de sustitución/sinónimos en el capítulo 10 el p. 225.](#)

*Notas:*

- Si desea filtrar esta ventana de manera que solo se vea la información que pertenece a una única biblioteca, puede cambiar la vista de la biblioteca mediante la lista desplegable de la barra de herramientas. Contiene una entrada de nivel superior llamada Todas las bibliotecas así como una entrada adicional para cada biblioteca individual. [Si desea obtener más información, consulte el tema Visión de bibliotecas en el capítulo 9 el p. 206.](#)

## ***Opciones de configuración***

Puede definir opciones generales para IBM® SPSS® Text Analytics for Surveys en el cuadro de diálogo Opciones. Este cuadro de diálogo contiene las siguientes pestañas:

La pestaña **Sistema** contiene opciones para listas de bibliotecas predeterminadas, guardado automático, guardado de resultados de extracción, delimitadores y el idioma de la interfaz.

La pestaña **Mostrar** contiene opciones para los colores que se utilizan en la interfaz.

La pestaña **Sonidos** contiene opciones para las pistas de sonido.

La pestaña **Traducción** contiene opciones para las conexiones de traducción.

### **Para editar opciones**

- ▶ En los menús seleccione Herramientas > Opciones. Aparecerá el cuadro de diálogo Opciones.
- ▶ Seleccione la pestaña que contiene la información que desea cambiar.
- ▶ Cambie las opciones pertinentes.
- ▶ Pulse en Aceptar para guardar los cambios.

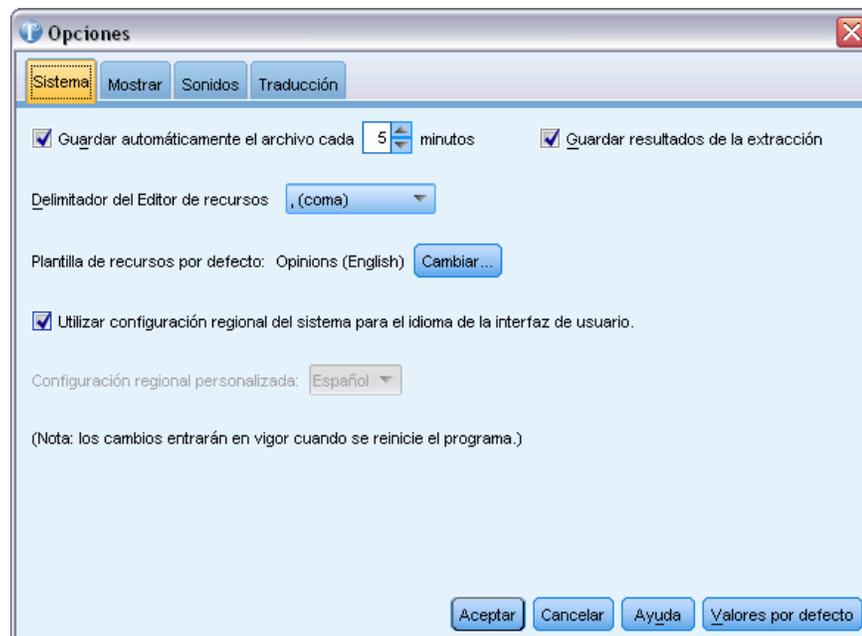
## **Opciones: pestaña Sistema**

En esta pestaña, puede definir varias opciones de proyecto, por ejemplo:

- Añadir o eliminar bibliotecas que deben aparecer en todos los proyectos nuevos por defecto
- Activar o desactivar la función de recuperación de guardado automático
- Activar o desactivar el guardado de los resultados de la extracción
- Definir el delimitador global que se utilizará en Editor de recursos para separar elementos

Figura 2-7

Cuadro de diálogo Opciones: pestaña Sistema



**Guardar automáticamente el archivo cada n minutos.** Seleccione esta opción para que el producto cree automáticamente una versión provisional guardada del archivo del proyecto abierto en caso de que se produzca un error en la máquina. Establezca también el número de minutos entre cada guardado. Si activa esta función y el producto se cierra inesperadamente, o si la máquina

presenta algún problema, la próxima vez que inicie el producto se le presentará la opción de abrir y trabajar con una versión recuperada del archivo.

**Guardar resultados de la extracción.** Seleccione esta opción para guardar los resultados de las extracciones en el proyecto. Esto puede significar un ahorro de tiempo cuando todavía esté trabajando con sus categorías. Sin embargo, puede suponer un tiempo adicional en el momento de la carga, y puede aumentar el tamaño del proyecto. Como medida de seguridad, se realiza un proceso de cifrado de dichos resultados de extracción durante el proceso de guardado, y se colocan en la base de datos. Este procedimiento dificulta la posibilidad de que otro usuario, incluso uno avanzado, vea los datos de la base de datos. Además, los resultados de la extracción nunca se presentarán en IBM® SPSS® Text Analytics for Surveys hasta que la aplicación haya localizado el origen de datos. De esta manera, si los datos están protegidos por contraseña, el usuario debe especificar el nombre de usuario y la contraseña correspondiente al origen de los datos para que los resultados de la extracción aparezcan en pantalla.

- El hecho de *guardar* es más ventajoso por la optimización del tiempo. Teniendo en cuenta que el proceso de extracción puede tardar en finalizar cuando se trabaja con conjuntos de datos voluminosos, el hecho de guardarlos le ofrece la posibilidad de acceder a los resultados de manera inmediata siempre que vuelva a abrir el proyecto. Sin embargo, es probable que advierta cierta lentitud a la hora de abrir el proyecto.
- Se opta por *no guardar* si no desea que el texto de las respuestas resida en ningún lugar que no sea el archivo de datos original, aunque cuente con medidas de seguridad.

**Editor de recursos Delimitador.** Seleccione el carácter que se utilizará como delimitador cuando introduzca elementos, como términos, sinónimos y elementos opcionales, en Editor de recursos.

**Plantilla de recursos.** Si no ha seleccionado un paquete de análisis de texto, se utilizará un conjunto de recursos por defecto. Estos recursos se guardan en una plantilla. Pulse en Cambiar para seleccionar una plantilla de recursos por defecto distinta. A continuación, en el diálogo Cambiar plantillas, seleccione la línea de la plantilla que desea utilizar y pulse en Aceptar.

**Utilice la región del sistema para el idioma de la interfaz de usuario.** Seleccione esta opción para que SPSS Text Analytics for Surveys utilice los detalles de la región de su sistema para proporcionar el idioma utilizado en la interfaz. Como alternativa, puede cancelar la selección de esta opción y, a continuación, seleccionar un idioma de interfaz diferente. Por ejemplo, puede hacer esto si está analizando información registrada en un idioma diferente al de la región de su sistema y quiere ejecutar SPSS Text Analytics for Surveys en ese idioma.

Tenga en cuenta que los cambios realizados en esta opción no surtirán efecto hasta que no reinicie SPSS Text Analytics for Surveys.

## **Opciones: Pestaña Mostrar**

En esta segunda pestaña, puede editar las opciones que inciden en el aspecto y el funcionamiento general de la aplicación, y los colores que se utilizan para distinguir los elementos.

Figura 2-8  
Cuadro de diálogo Opciones: Pestaña Mostrar



**Fuentes y colores estándar (aplicados tras el reinicio).** Las opciones de este cuadro de control se utilizan para especificar la combinación de colores y el aspecto que se muestra. Las opciones seleccionadas aquí se aplican tras cerrar y reiniciar la aplicación. Seleccione entre:

- **SPSS estándar (predeterminado)**, diseño común en los productos de la marca SPSS (parte de IBM Corp.).
- **SPSS clásico**, diseño conocido para los usuarios de versiones anteriores de esta aplicación.
- **Microsoft Windows**, diseño de Microsoft Windows que puede ser de utilidad para aumentar el contraste en lienzos y paletas de ruta.

**Colores personalizados.** Edite los colores de los elementos que aparecen en pantalla. Para cada uno de los elementos de la tabla, puede cambiar el color. Para especificar un color personalizado, pulse en el área de color a la derecha del elemento que desea cambiar y elija un color en la lista desplegable de colores.

- **Texto no extraído.** Texto de respuesta que no se extrajo y que sigue visible en el panel Datos.
- **Resaltar fondo.** Color de fondo del texto seleccionado al seleccionar elementos en los paneles o texto en el panel Datos.
- **Fondo de extracción necesaria.** Color de fondo del panel Resultados extraídos que indica que se han realizado cambios en las bibliotecas y que se necesita una extracción.
- **Fondo de comentario de categoría.** Color de fondo de la categoría que aparece después de una operación, como arrastrar y soltar respuestas y forzar respuestas desde el panel Datos al panel Categorías.
- **Tipo por defecto.** Color por defecto para los tipos y términos que aparecen en el panel Datos y en el panel Resultados extraídos. Este color aparece en la interfaz siempre que aparece el tipo `Unknown` o cualquier otro concepto asociado. Este color también se aplica a los tipos

personalizados que se crean en el Editor de recursos. Puede cambiar este color por defecto para sus diccionarios de tipo personalizado editando las propiedades de los mismos en Editor de recursos. [Si desea obtener más información, consulte el tema Creación de tipos en el capítulo 10 el p. 217.](#)

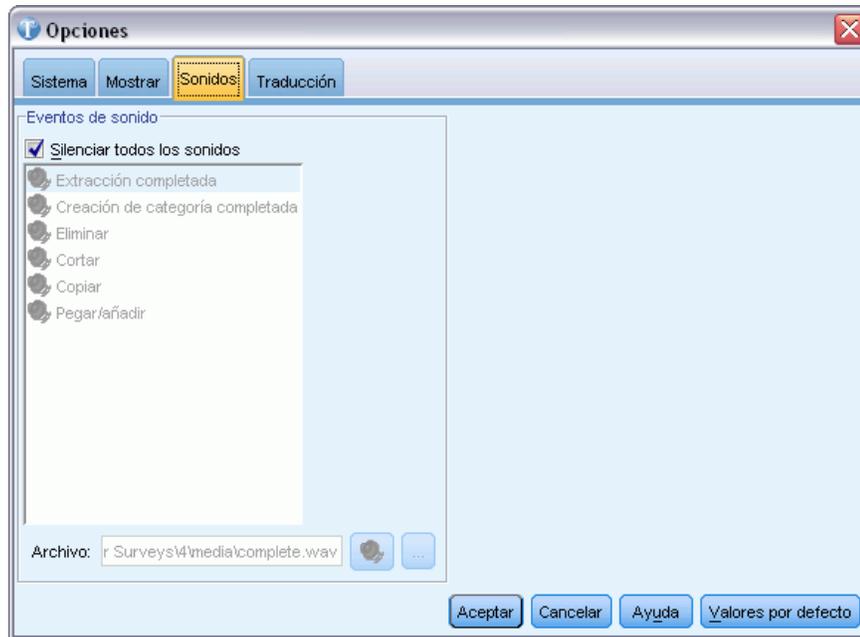
- **Tabla con bandas 1.** El primero de los dos colores utilizados de forma alternativa en la tabla del cuadro de diálogo Editar términos forzados para diferenciar cada línea.
- **Tabla con bandas 2.** El segundo de los dos colores utilizados de forma alternativa en la tabla del cuadro de diálogo Editar términos forzados para diferenciar cada línea.
- **Máscara no válida.** Color del texto de entradas duplicadas en el Administrador de marcos de código, que indica un error.
- **Fondo no válido.** Color del fondo resaltado de las entradas duplicadas en el Administrador de marcos de código, que indica un error.

**Orden de colores de visualizaciones.** Si utiliza el gráfico de barras de categoría en el panel de visualización y selecciona también una variable de referencia, puede ver cada uno de los valores posibles correspondientes a la variable de referencia en una leyenda al final del panel. Cada valor también está codificado por color para ayudarle a distinguirlo visualmente en el gráfico de barras. Puede cambiar estos colores por defecto aquí. [Si desea obtener más información, consulte el tema Visualización de gráficos en el capítulo 7 el p. 167.](#)

### ***Opciones: pestaña Sonidos***

En esta pestaña, puede editar los sonidos que se utilizan en el producto. En Eventos de sonido puede especificar que se utilice un sonido para notificar que ha ocurrido un evento. Mediante la activación y desactivación de los sonidos o la asignación de sonidos específicos, puede controlar la forma en que recibirá los avisos de determinadas operaciones del software. Por ejemplo, puede activar sonidos para eventos como el final del proceso de extracción, el final de una técnica de categorización automática, o tareas más comunes, como cortar, pegar, copiar y eliminar.

Figura 2-9  
Cuadro de diálogo Opciones: pestaña Sonidos



Hay numerosos sonidos disponibles. Utilice el botón de puntos suspensivos (...) para buscar y seleccionar un sonido. Los archivos *.wav* utilizados para crear sonidos en IBM® SPSS® Text Analytics for Surveys se almacenan en el subdirectorio */media* del directorio de instalación. Si no desea que se reproduzca ningún sonido, seleccione Silenciar todos los sonidos. Los sonidos están silenciados por defecto.

### ***Opciones: pestaña Traducción***

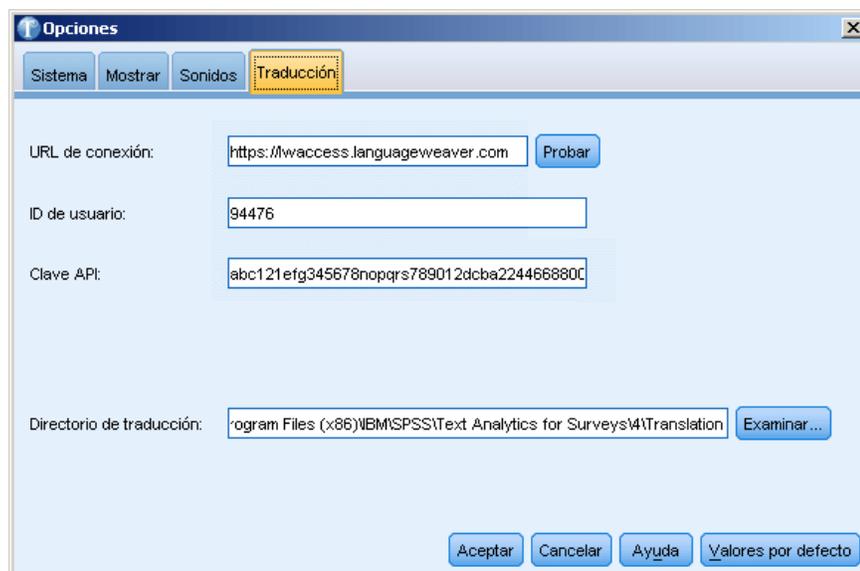
**Importante:** La traducción sólo está disponible en inglés.

En esta pestaña puede definir y gestionar la conexión del servidor de traducción Language Weaver que podrá reutilizar cuando traduzca. Una vez que defina aquí una conexión, podrá seleccionar rápidamente una conexión de par de idiomas en el momento de la traducción sin necesidad de volver a introducir toda la configuración de conexión.

Una conexión de par de idiomas identifica los idiomas de origen y de traducción así como los detalles de URL de conexión del servidor. Por ejemplo, *Chino - Inglés* significa que el texto de origen está en chino y la traducción final estará en inglés. Debe definir manualmente la conexión del servidor Language Weaver al que accede mediante los servicios en línea de Language Weaver.

Las traducciones finales se almacenan en el directorio definido en este cuadro de diálogo. Puede administrar sus archivos de traducción directamente en ese directorio y/o especificar un directorio distinto aquí.

Figura 2-10  
Cuadro de diálogo Opciones: Pestaña Traducción



**URL de conexión.** Introduzca la URL del servidor o la dirección Web en el servidor en línea Language Weaver.

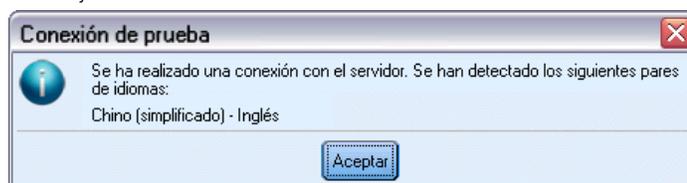
**ID de usuario.** Introduzca el ID exclusivo que le haya proporcionado Language Weaver.

**Clave de API.** Introduzca la clave que le haya proporcionado Language Weaver.

**Comprobar.** Pulse en Comprobar para verificar que la conexión se ha configurado adecuadamente y para ver los pares de idiomas que se han encontrado en esa conexión.

**Directorio de traducción.** Pulse Examinar para cambiar a un directorio distinto o escriba directamente la ruta de la carpeta en el campo.

Figura 2-11  
Mensaje de conexión correcto



## Configuración de Microsoft Internet Explorer para obtener ayuda

### Configuración de Microsoft Internet Explorer

La mayoría de las funciones de ayuda de esta aplicación utilizan tecnología basada en Microsoft Internet Explorer. Algunas versiones de Internet Explorer (incluida la versión que se incluyen con Microsoft Windows XP, Service Pack 2) bloquearán por defecto lo que se consideren “contenidos

activos” en las ventanas de Internet Explorer de su ordenador local. Esta configuración por defecto puede hacer que se bloqueen algunos contenidos de las funciones de ayuda. Para ver todos los contenidos de ayuda, puede cambiar el comportamiento por defecto de Internet Explorer.

- ▶ Elija en los menús de Internet Explorer, seleccione:  
Herramientas > Opciones de Internet...
- ▶ Pulse en la pestaña Opciones avanzadas.
- ▶ Desplácese hacia abajo hasta la sección Seguridad.
- ▶ Seleccione (marque) Permitir que el contenido activo se ejecute en archivos de Mi PC.

***Parte II:***  
***Análisis de texto***

# ***Creación de proyectos y paquetes***

## ***Creación de proyectos***

En IBM® SPSS® Text Analytics for Surveys, trabajará con datos de encuestas que podrá categorizar. Para ello, creará proyectos en los que importará datos desde el origen de datos, seleccionará algunas variables y elegirá categorías y recursos. Una vez que haya creado el proyecto, puede ajustar los recursos y las categorías hasta que esté conforme con el conjunto final de categorías. Un proyecto puede contener los elementos siguientes: datos de la encuesta, recursos lingüísticos, resultados extraídos y categorías.

### ***Datos de la encuesta***

En el proyecto se hace referencia al origen de los datos importados de la encuesta, pero dichos datos no se guardan en el proyecto. Por el contrario, cuando se abre un proyecto, los datos de la encuesta vuelven a leerse desde el origen de datos original. Mientras que las variables más importantes en este contexto son las de texto con final abierto, la variable de ID exclusiva se conserva, igual que cualquier variable de referencia (como variables demográficas) que se especifique cuando se importan los datos. Los valores de todas estas variables pueden mostrarse en el panel Datos de la vista Pregunta en la ventana de análisis de texto o en la vista Proyecto completo de la ventana de análisis de texto.

### ***Recursos lingüísticos***

Las bibliotecas de su propiedad y las personalizadas por el usuario que contienen listas de términos, sinónimos, listas de palabras excluidas y declaraciones de tipo se almacenan en un proyecto que puede modificarse. Además, se utilizan determinados recursos compilados para procesar el texto y también se guardan en el proyecto, pero en esta ocasión no puede modificarse. Las bibliotecas pueden publicarse, y pasan a ser de dominio público dentro de la base de datos para que puedan utilizarse en otros proyectos. Una biblioteca publicada puede añadirse a otros proyectos.

### ***Resultados extraídos***

Los resultados extraídos son palabras clave y frases (conceptos), sus agrupaciones semánticas (tipos) y sus relaciones (patrones) que se identifican y se extraen a partir de las respuestas de texto. Estos resultados de extracción forman parte del proyecto y constituyen la base de la creación de categorías. Por defecto, los resultados de la extracción se guardan en el proyecto, pero si considera que aumentan demasiado el tamaño del archivo del proyecto, puede desactivar esta opción, por lo que la próxima vez que abra el proyecto deberá volver a realizar la extracción. [Si desea obtener más información, consulte el tema \*Cómo guardar los resultados de la extracción en el capítulo 5\* el p. 85.](#)

## **Categorías**

Las respuestas de texto se colocan en categorías que pueden crearse automáticamente mediante las técnicas incorporadas de generación de categorías, manualmente mediante operaciones de arrastrar y soltar, y también importando archivos de definición de categorías o mediante el Administrador de marcos de código. Si opta por no guardar los resultados de la extracción, cuando se vuelve a abrir un proyecto las definiciones de categoría permanecen pero el recuento de respuestas de cualquiera de las partes de las definiciones que procedían de los resultados extraídos se mostrarán con un signo de interrogación (?) hasta que repita la extracción.

## **Preparación de los datos**

Antes de importar los datos en IBM® SPSS® Text Analytics for Surveys, revise las consideraciones siguientes:

- **Datos de entrada.** Para importar el origen de datos a SPSS Text Analytics for Surveys, debe contener determinados elementos básicos, como una variable de ID y al menos una pregunta con final abierto. La variable de ID debe contener únicamente valores exclusivos. Si existen duplicados, la importación fallará. Puede importar varias preguntas con final abierto, así como variables de referencia. [Si desea obtener más información, consulte el tema Selección de variables el p. 34.](#)
- **Errores ortográficos.** Aunque el programa arregla algunos errores ortográficos, se recomienda que los corrija antes de importar los datos en el programa. Los errores ortográficos puede ocasionar problemas en el análisis de texto, tanto si lo realizan personas como programas de software. Cuantos más errores ortográficos pueda corregir de antemano, más fiables serán las categorías resultantes. También puede crear sinónimos con la ortografía correcta de una palabra y las variaciones tipográficas incorrectas más comunes en el programa. De hecho, en la biblioteca Core ya vienen predefinidos muchos de los errores más comunes. Si no está seguro de cuánto esfuerzo debería dedicar en la verificación ortográfica, puede ejecutar algunas pruebas con ejemplos de respuestas para comprobar el porcentaje de análisis que se ve afectado por los errores ortográficos.
- **Respuestas en blanco.** No es inusual encontrar respuestas en blanco en los datos con final abierto de las encuestas. Aunque las respuestas en blanco no aportan información, pueden resultar útiles. Por ejemplo, puede ser interesante saber cuántas personas no respondieron a una pregunta, o que tipo de personas no lo hizo. Sin embargo, puesto que SPSS Text Analytics for Surveys utiliza texto para extraer términos y categorizar respuestas, estas respuestas en blanco no son fáciles de categorizar.

Una posibilidad es, antes de importar los datos, sustituir todas las respuestas en blanco por la palabra *en blanco* u otro término adecuado. A continuación, una vez importados los datos, puede crear un tipo nuevo que represente a la respuesta en blanco, con la palabra *en blanco* (o la que haya elegido) como el término que representa a ese tipo.

Otra opción sería forzar las respuestas en blanco en una categoría. Después de categorizar las respuestas, las respuestas en blanco quedarían inicialmente sin categorizar. Puede crear manualmente una nueva categoría llamada En blanco pulsando el botón derecho del ratón en

el panel Categorías. A continuación, después de seleccionar todas las respuestas en blanco, puede forzarlas a que formen parte de la nueva categoría En blanco.

- **Preguntas de respuestas múltiples.** Las preguntas con final abierto suelen ser elementos individuales; sin embargo no es este el caso. A veces las preguntas con final abierto se utilizan como conjunto de respuestas múltiples. Por ejemplo, si pregunta a un encuestado “Indíquenos tres aspectos para mejorar en el hotel” y ofrece tres espacios (variables) separados en los que responder, esto representa una pregunta de respuesta múltiple.

Puesto que SPSS Text Analytics for Surveys analiza cada variable de pregunta por separado, puede reutilizar las categorías y los recursos lingüísticos creados para analizar la primera respuesta con el fin de categorizar la segunda y la tercer. Sin embargo, puede que este no sea el método más eficaz. Existe la posibilidad de combinar las tres variables de respuesta en una sola variable antes de importar los datos en el programa. Si las combina, verifique que haya al menos un espacio entre la última palabra de una respuesta y la primera de la siguiente, o preferiblemente un punto. Puesto que se trata de una tarea que puede resultar tediosa si se trabaja con conjuntos de datos voluminosos, tiene la opción de combinar las respuestas mientras se está creando el archivo de datos en lugar de hacerlo después.

- **Ejemplos de respuestas.** Cuanto mayor es el número de respuestas y mayor su longitud de promedio, más tiempo necesitará para la extracción o categorización. Para optimizar el trabajo, si el tamaño del archivo es grande (quizá con 1.500 casos o más), puede realizar primero un muestreo aleatorio y utilizar el subconjunto de respuestas más pequeño para realizar una primera pasada del análisis.

Una muestra pequeña suele ser perfectamente adecuada para decidir cómo editar los recursos lingüísticos. Y una vez categorizados en un archivo de datos más pequeño, podrá leer el archivo completo y volver a extraer, acción que categorizará automáticamente muchas de las respuestas. A continuación podrá ver las respuestas que no se adecuaron a las categorías creadas y podrá realizar los ajustes necesarios. El tamaño del muestreo aleatorio puede variar, pero unos 300 casos suele ser adecuado.

**Importante:** Existen otras consideraciones en relación al proceso de análisis de texto en su globalidad. [Si desea obtener más información, consulte el tema Preparación para el análisis del texto en el capítulo 1 el p. 7.](#)

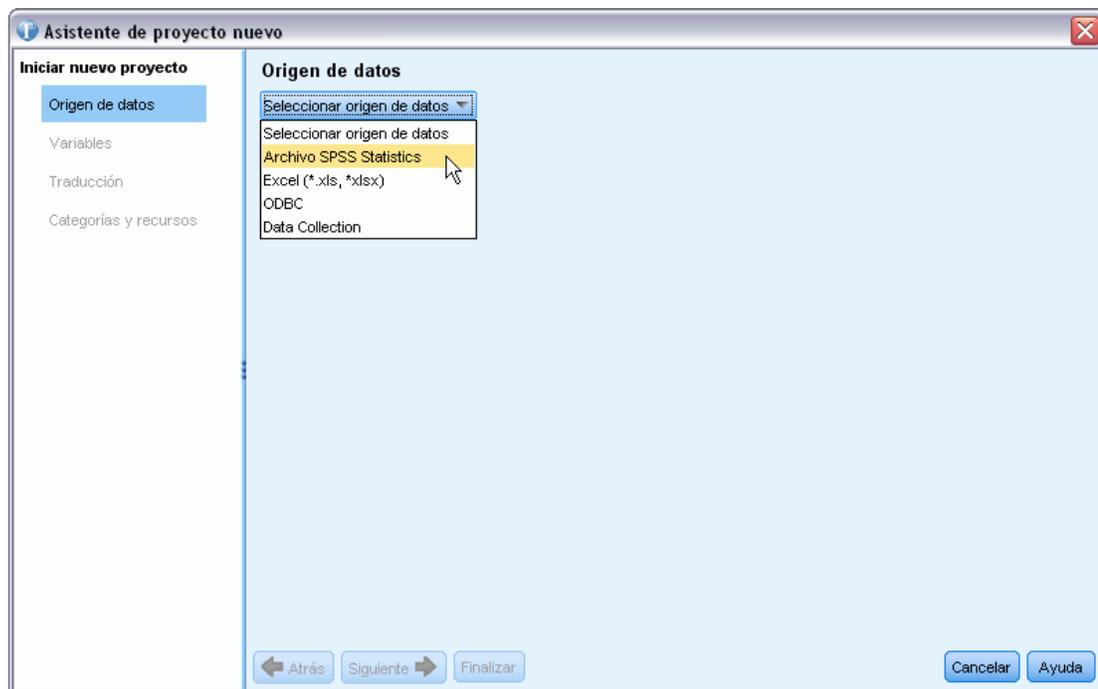
## ***Inicio de proyectos nuevos***

Para empezar a categorizar los datos de respuesta, primero debe crear un proyecto. Un asistente le guiará a través del origen de datos y de la selección de variables, especificaciones de recursos y categorías, etc. Antes de empezar a crear el proyecto, debe preparar los datos.

### ***Para iniciar un proyecto nuevo***

- En los menús elija Archivo > Nuevo proyecto. Como alternativa, pulse en Iniciar con un nuevo proyecto en la pantalla de inicio si no hay ningún proyecto abierto. Aparece el Asistente de proyecto nuevo.

Figura 3-1  
Asistente de proyecto nuevo



- ▶ Empiece seleccionando el tipo de origen de datos desde la lista desplegable Seleccionar origen de datos. Si desea obtener más información, consulte el tema [Selección de los orígenes de datos](#) el p. 29.

## ***Selección de los orígenes de datos***

Cuando se abre el asistente, debe empezar seleccionando un origen de datos. IBM® SPSS® Text Analytics for Surveys se ha optimizado para procesar conjuntos de datos de hasta 10.000 registros, aunque el rendimiento variará en función del volumen de texto contenido en estos registros. Consulte las instrucciones de instalación para conocer las estadísticas y recomendaciones sobre rendimiento.

**Importante:** Debe haber una variable de ID con un valor exclusivo para cada registro con el fin de importar los datos.

Puede elegir uno de los siguientes orígenes de datos:

- Archivos **SPSS Statistics** (\*.sav).
- Archivos **Microsoft Excel** (\*.xls / \*.xlsx).
- **Base de datos ODBC** (protocolo Open Database Connectivity de Microsoft).
- Modelo de datos **Data Collection**. Esta opción solo está disponible si tiene instalado el modelo de datos.

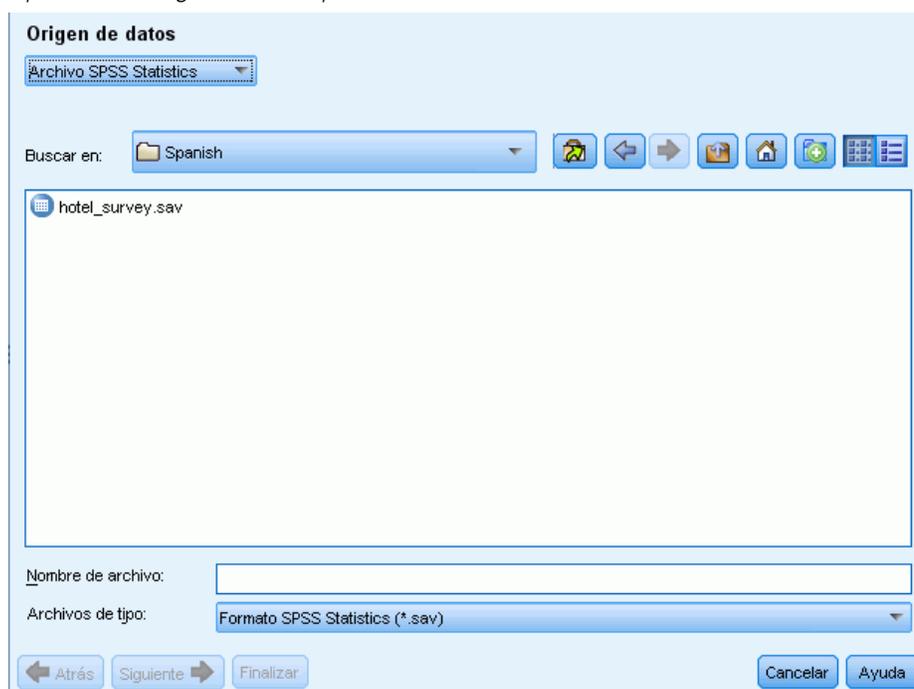
### Uso de los archivos IBM SPSS Statistics

Puede importar un archivo IBM® SPSS® Statistics (.sav) en IBM® SPSS® Text Analytics for Surveys. Debe haber una variable de ID con un valor exclusivo para cada registro con el fin de importar los datos.

**Importante:** No puede importar el archivo SPSS Statistics (.sav) con registros que sobrepasen los 4000 caracteres.

*Nota:* SPSS Text Analytics for Surveys se ha optimizado para procesar conjuntos de datos de hasta 10.000 registros, aunque el rendimiento variará en función del volumen de texto contenido en estos registros. Consulte las instrucciones de instalación para conocer las estadísticas y recomendaciones sobre rendimiento.

Figura 3-2  
Opciones de origen de datos para archivos IBM SPSS Statistics



#### Para obtener datos de IBM SPSS Statistics

- ▶ En la primera pantalla del asistente, seleccione el archivo SPSS Statistics de la lista desplegable. El asistente muestra las opciones para los archivos SPSS Statistics.
- ▶ En la lista desplegable Buscar en, seleccione la unidad y la carpeta donde se encuentra el archivo.
- ▶ Seleccione el archivo de la lista. Aparecerá el cuadro de texto Nombre de archivo.
- ▶ Pulse en Siguiente para seleccionar variables. [Si desea obtener más información, consulte el tema Selección de variables el p. 34.](#)

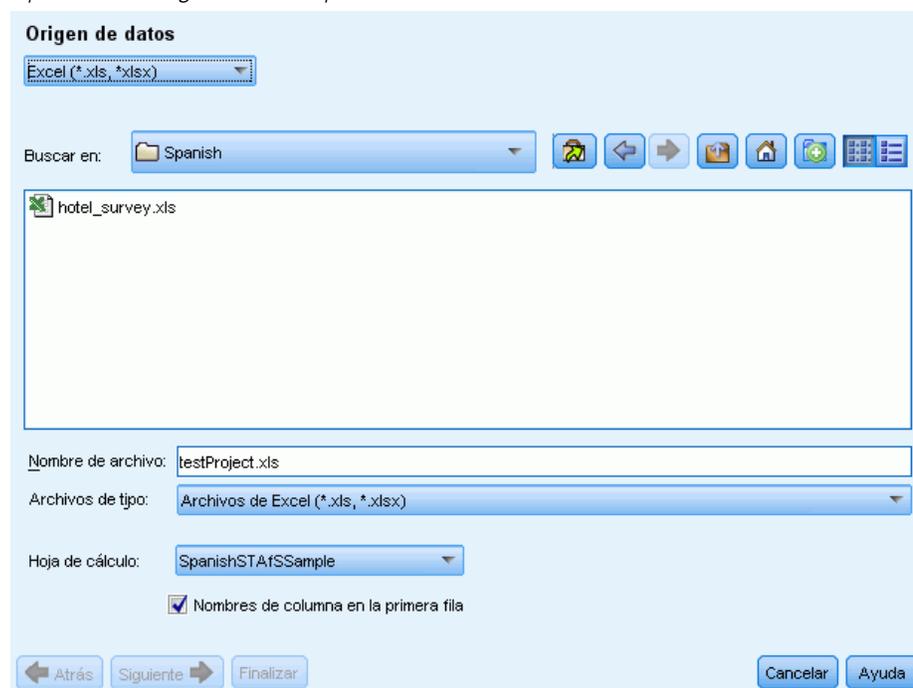
### Uso de los archivos Microsoft Excel

Puede importar un archivo Microsoft Excel (.xls / .xlsx) en IBM® SPSS® Text Analytics for Surveys. Debe haber una variable de ID con un valor exclusivo para cada registro con el fin de importar los datos.

**Importante:** Durante la importación del archivo Microsoft Excel, puede seleccionar una opción para Nombres de columna en la primera fila. Para utilizar esta opción, la primera línea de la hoja de cálculo debe contener nombres de columna—no la fila justo encima donde empiezan los datos. Por ejemplo, si los datos y los nombres de columna empiezan en la línea 7, antes de importar debe eliminar las filas 1–6 para que la importación del archivo sea correcta.

*Nota:* SPSS Text Analytics for Surveys se ha optimizado para procesar conjuntos de datos de hasta 10.000 registros, aunque el rendimiento variará en función del volumen de texto contenido en estos registros. Consulte las instrucciones de instalación para conocer las estadísticas y recomendaciones sobre rendimiento.

Figura 3-3  
Opciones de origen de datos para archivos Microsoft Excel



#### Para obtener datos de Microsoft Excel

- ▶ En la primera pantalla del asistente, seleccione Excel de la lista desplegable. El asistente muestra las opciones para los archivos Microsoft Excel.
- ▶ En la lista desplegable Buscar en, seleccione la unidad y la carpeta donde se encuentra el archivo.
- ▶ Seleccione el archivo de la lista. Aparecerá el cuadro de texto Nombre de archivo.

- ▶ Seleccione la hoja de cálculo en la lista desplegable. Solo puede importar datos de una sola hoja de cálculo. Para trabajar con los datos de varias hojas de cálculo, debe crear varios proyectos.
- ▶ Si la primera fila de esta hoja de cálculo contiene las cabeceras de columna, seleccione Nombres de columna en la primera fila. Para utilizar esta opción, la primera línea de la hoja de cálculo debe contener nombres de columna—no la fila justo encima donde empiezan los datos. Por ejemplo, si los datos y los nombres de columna empiezan en la línea 7, antes de importar debe eliminar las filas 1–6 para que la importación del archivo sea correcta. La aplicación puede utilizar estos (o bien una versión convertida si las cabeceras de columna no cumplen las convenciones de nombres de variables de IBM® SPSS® Statistics) como nombres de variable. En caso contrario, la aplicación utilizará las letras de las columnas de la hoja de cálculo como identificadores.
- ▶ Pulse en Siguiente para seleccionar variables. [Si desea obtener más información, consulte el tema Selección de variables el p. 34.](#)

### ***Uso de los datos a través de ODBC***

Los datos procedentes de orígenes de bases de datos, comúnmente de las bases de datos, se importan fácilmente en IBM® SPSS® Text Analytics for Surveys. El producto puede leer directamente las bases de datos que utilizan controladores ODBC (Open Database Connectivity) después de que los controladores apropiados se hayan instalado en la máquina en la que SPSS Text Analytics for Surveys está instalado. Debe haber una variable de ID con un valor exclusivo para cada registro con el fin de importar los datos.

*Nota:* SPSS Text Analytics for Surveys se ha optimizado para procesar conjuntos de datos de hasta 10.000 registros, aunque el rendimiento variará en función del volumen de texto contenido en estos registros. Consulte las instrucciones de instalación para conocer las estadísticas y recomendaciones sobre rendimiento.

Figura 3-4  
Opciones de origen de datos para ODBC

**Origen de datos**

ODBC

Origen ( DSN )

Nombre	Descripción
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

Usuario:

Contraseña:

Tabla

SQL

Atrás Siguiente Finalizar Cancelar Ayuda

#### Para utilizar ODBC como medio

- ▶ En la primera pantalla del asistente, seleccione ODBC de la lista desplegable. El asistente muestra las opciones para ODBC.
- ▶ Especifique el origen de datos seleccionándolo de la lista de orígenes de ODBC registrados o tecleando el nombre en el cuadro de texto Origen (DSN). Si necesita registrar nuevos orígenes de datos que no aparecen en la lista, pulse en ODBC. Así se abrirá el Administrador de orígenes de datos ODBC, que se encuentra en la mayoría de los ordenadores Microsoft Windows. Si no lo encuentra, no podrá utilizar la importación ODBC. Consulte el sistema de ayuda de Microsoft Windows para obtener más información.
- ▶ Si el origen de datos está protegido por contraseña, deberá especificar un nombre de usuario y una contraseña. Se le pedirá que lo haga cada vez que abra el proyecto, puesto que, por razones de seguridad, el nombre de usuario y la contraseña no se guardan en el proyecto.
- ▶ Puede seleccionar los datos de dos maneras distintas: Tabla o SQL. Puede seleccionar una tabla directamente, o bien utilizar comandos SQL para seleccionar los datos.
- ▶ Pulse en Siguiente para seleccionar variables. [Si desea obtener más información, consulte el tema Selección de variables el p. 34.](#)

### **Uso de datos IBM SPSS Data Collection**

#### **Para importar mediante IBM SPSS Data Collection**

- ▶ En la primera pantalla del asistente, seleccione Data Collection de la lista desplegable. La opción del modelo de datos IBM® SPSS® Data Collection solo está disponible si tiene el modelo de datos instalado con otro producto.

### **Selección de variables**

Después de seleccionar el origen de datos, el paso siguiente es especificar las variables que se van a importar. Pueden importarse tres tipos de variables en un proyecto.

#### **Variable de ID exclusiva (obligatoria)**

La variable de ID es una clave numérica exclusiva o una cadena que identifica a cada encuestado. No es necesario ordenar el archivo de datos por variable de ID exclusiva para que se lea correctamente. Después de que el programa lo haya leído, los registros pueden ordenarse según determinados criterios. [Si desea obtener más información, consulte el tema Ordenación de las variables en el capítulo 4 el p. 50.](#) Esta variable de ID es obligatoria para importar datos. Cada registro (o caso) importado debe tener un valor de ID exclusivo.

Pueden darse dos circunstancias por la que la importación falla:

- Si se detectan valores de ID duplicados
- Si hay registros con valores de ID en blanco

*Nota:* Si se detecta un ID duplicado y IBM® SPSS® Statistics está instalado en el ordenador, puede utilizar el procedimiento Identify Duplicate Cases (Identificar casos duplicados) en ese producto para identificar los duplicados y luego utilizar las opciones para indicar qué registros deben conservarse (casos principales).

#### **Variables de texto con final abierto (obligatorio)**

Las variables de texto con final abierto representan las respuestas de texto a las preguntas de la encuesta. Se requiere al menos una de estas variables para importar datos. Estas variables pueden ser variables de cadena o cadena larga en SPSS Statistics, columnas que contienen casillas de texto o generales en Microsoft Excel, o campos de notas o texto procedentes de las bases de datos. Cada variable de texto con final abierto se analizará por separado. Existe un límite de 4.000 caracteres de tamaño (extensión) para cada variable de texto importado de un archivo .SAV.

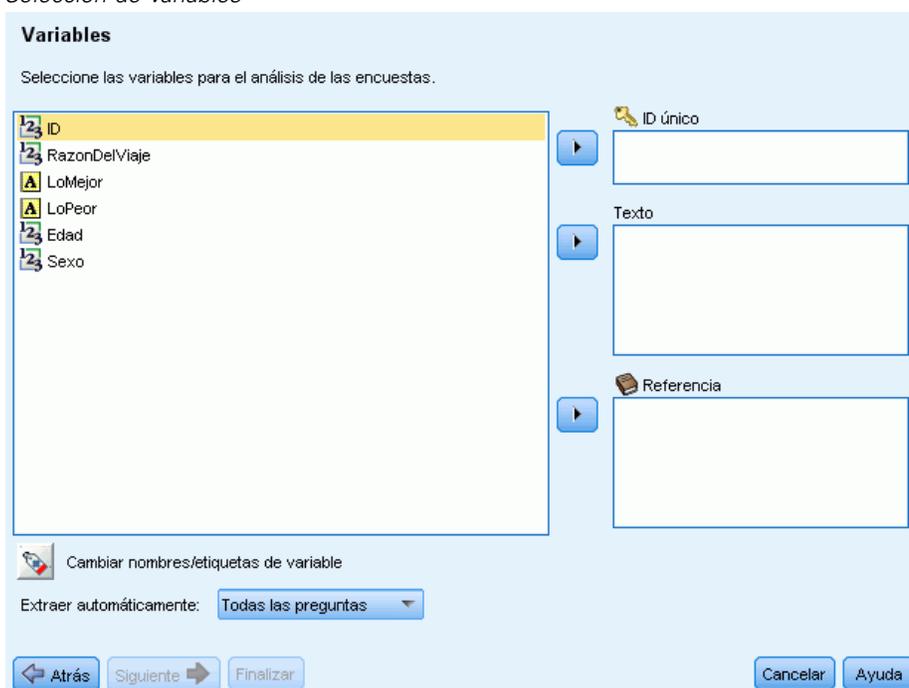
#### **Variables de referencia (opcional)**

Las variables de referencia son variables opcionales y adicionales, generalmente de categorías, que pueden importarse con fines de referencia. Las variables de referencia no se utilizan en el análisis de texto, pero proporcionan información adicional que describen al encuestado, lo que puede ayudar a entender e interpretar sus respuestas. Las variables demográficas suelen incluirse como variables de referencia, puesto que pueden contribuir a comprender qué términos o categorías se

están utilizando en cada grupo de personas. Algunos ejemplos son sexo, departamento, trabajo y nivel de estudios (para evaluaciones de formación y estudiantes). Puede ver todas las variables de referencia después de la importación en la vista Proyecto completo. También puede mostrar las variables de referencia en el panel Datos de la vista Pregunta. Además, puede seleccionar variables de referencia en el gráfico de barras del panel de visualización para poder profundizar a un subconjunto de encuestados.

*Nota:* Las variables de referencia que se leen de un archivo de datos SPSS Statistics tendrán etiquetas de variable (si se suministran) que aparecen como cabeceras de columnas, y sus etiquetas de valor correspondientes (si se suministran) se muestran en las casillas del panel Datos.

**Figura 3-5**  
Selección de variables



### **Para seleccionar variables y opciones de extracción**

- ▶ En la lista de variables disponibles, seleccione la variable que se corresponde con la variable de ID del conjunto de datos, y pulse en el botón de flecha para moverla al cuadro ID único. El ID debe ser un número o una cadena alfanumérica exclusivos para distinguir los registros entre sí. Si el conjunto de datos contiene ID duplicadas, aparecerá un mensaje de error. En tal caso, deberá limpiar los datos antes de repetir la operación.
- ▶ Desde la lista de variables disponibles, seleccione una o más variables que se correspondan con las variables de respuesta con final abierto, y pulse el botón de flecha para mover dichas variables a la lista Variables de texto con final abierto. Las variables se importarán como preguntas independientes cuyas respuestas serán analizadas y categorizadas por usted.
- ▶ Desde la lista de variables disponibles, seleccione una o más variables que se correspondan con las variables de referencia, y pulse el botón de flecha para mover dichas variables a la lista de

referencia. Las variables de referencia no se utilizan en técnicas automáticas de generación de categorías. Sin embargo, puede ver su contenido y utilizarlas para que le ayuden a tomar decisiones informadas a la hora de categorizar las respuestas.

- ▶ Para ver las etiquetas de las variables en lugar de sus nombres, pulse en el botón situado debajo de la lista de variables a la izquierda.
- ▶ Para cambiar el valor de extracción, seleccione una opción en la lista desplegable. La opción Solo la primera pregunta está seleccionada por defecto; significa que si ha seleccionado más de una variable de texto con final abierto, el proceso de extracción empezará automáticamente para la primera pregunta después de que el asistente finalice. La extracción puede tardar un tiempo en el caso de conjuntos de datos extensos. Por lo tanto, puede optar por extraer Ninguna o Todas las preguntas en función del tiempo del que disponga.
- ▶ Pulse en Siguiete > cuando haya seleccionado todas las variables.

## Traducción al inglés

Si trabaja con un texto de origen que no esté en inglés, puede conectarse a Language Weaver para traducirlo a ese idioma. La traducción sólo está disponible hacia el inglés. Debe haber configurado Language Weaver adecuadamente y con conexiones definidas para traducir. [Si desea obtener más información, consulte el tema Opciones: pestaña Traducción en el capítulo 2 el p. 22.](#)

Figura 3-6  
Opciones de traducción

**Traducción mediante Language Weaver**

Traducir texto no inglés a inglés automáticamente con Language Weaver.

Traducir al inglés

Configuración

Conexión de par de idiomas:

Precisión de traducción (1 = más rápida, 3 = máxima calidad):

Utilizar diccionario personalizado

Nombre de diccionario:

**Para traducir al inglés**

- ▶ Para traducir los datos de texto de un idioma con licencia al inglés, seleccione la casilla de verificación Traducir al inglés.
- ▶ En la lista Conexión par de idiomas, seleccione la conexión para el par de idiomas Language Weaver que desea utilizar. Si ha configurado Language Weaver en su equipo local, esos pares de idiomas aparecerán automáticamente en esta lista. Puede añadir, cambiar o comprobar la conexión de servicios en línea en la pestaña Traducción del cuadro de diálogo Opciones. [Si desea obtener más información, consulte el tema Opciones: pestaña Traducción en el capítulo 2 el p. 22.](#)
- ▶ Especifique la Precisión de la traducción que desee. Elija un valor entre 1 y 3 indicando el nivel de velocidad contra el nivel de precisión que desea. Un valor bajo hará que los resultados de traducción sean más rápidos pero que la precisión sea menor. Un valor alto producirá resultados con gran precisión pero llevará más tiempo de proceso. Para optimizar el tiempo, se recomienda comenzar con un nivel más bajo e ir aumentándolo sólo si necesita más precisión después de revisar los resultados.
- ▶ Si anteriormente había creado diccionarios personalizados en Language Weaver, puede utilizarlos en relación con la traducción. Para seleccionar un diccionario personalizado, seleccione la casilla de verificación Utilizar diccionario personalizado e introduzca el nombre del diccionario. Para utilizar más de un diccionario, separe los nombres con comas.
- ▶ En el Asistente de nuevo proyecto, pulse en Siguiete> para empezar a seleccionar las categorías y los recursos. [Si desea obtener más información, consulte el tema Selección de categorías y recursos el p. 37.](#)
- ▶ En el asistente Cambiar conjuntos de datos, pulse en Finalizar para completar el cambio de los conjuntos de datos y para iniciar el proceso de traducción.

**Para saltarse la traducción:**

- ▶ Elimine la selección de la opción Traducir al inglés.
- ▶ En el Asistente de nuevo proyecto, pulse en Siguiete> para empezar a seleccionar las categorías y los recursos. [Si desea obtener más información, consulte el tema Selección de categorías y recursos el p. 37.](#)
- ▶ En el asistente Cambiar conjuntos de datos, pulse en Finalizar para completar el cambio de los conjuntos de datos.

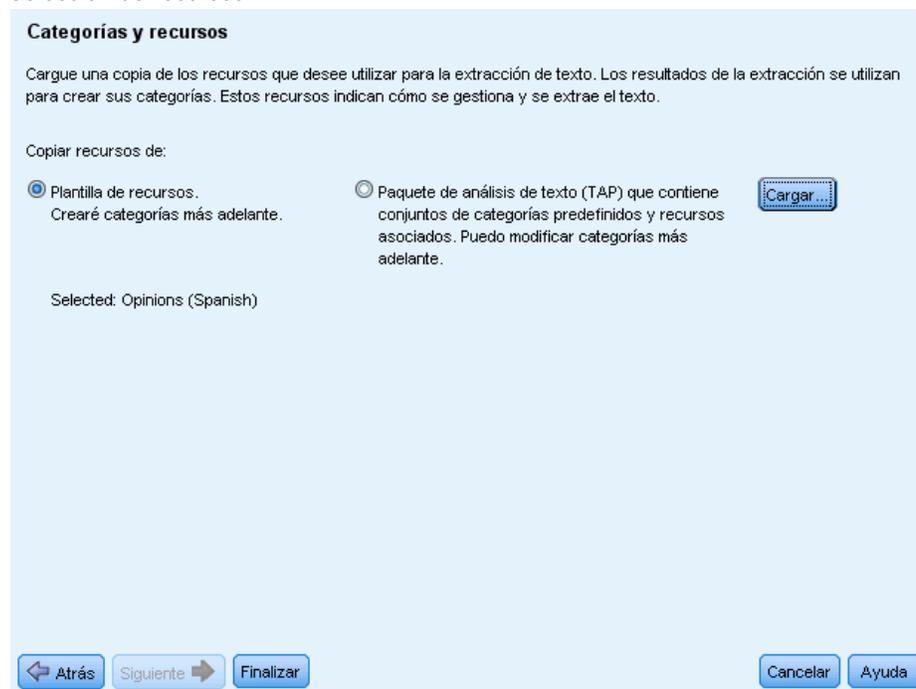
**Selección de categorías y recursos**

En este paso final, puede seleccionar los recursos lingüísticos que se utilizarán para extraer conceptos y patrones destacados del texto. Como alternativa, puede cargar un paquete de análisis de texto (TAP), que no sólo incluye los recursos lingüísticos sino también uno o más conjuntos de categorías predefinidas que representan marcos de código mejorados. [Si desea obtener más información, consulte el tema Uso de los paquetes de análisis de texto el p. 41.](#) Se ofrecen varios archivos TAP pregenerados para textos en inglés en IBM® SPSS® Text Analytics for Surveys. Cada archivo TAP incluido en este producto se ha adaptado para un tipo específico de encuesta,

como encuestas sobre empleados, productos o satisfacción de los clientes. También puede crear sus propios archivos TAP para cualquier idioma de texto compatible con el producto.

De manera predeterminada, hay una plantilla de recursos precargada. Puede cambiar la plantilla de recursos predeterminada que se propone en la primera pestaña del cuadro de diálogo Opciones. [Si desea obtener más información, consulte el tema Opciones de configuración en el capítulo 2 el p. 17.](#) Puede cargar una plantilla de recursos diferente o seleccione un archivo TAP en su lugar.

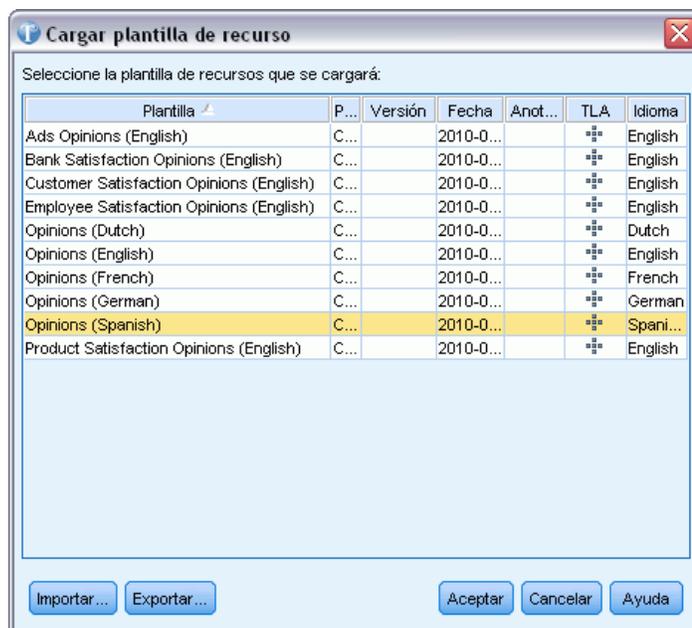
Figura 3-7  
Selección de recursos



**Para seleccionar una plantilla de recursos diferente:**

- Para cargar una plantilla de recursos diferente, asegúrese de seleccionar la opción Plantilla de recursos y pulse en Cargar. Aparecerá el cuadro de diálogo Cargar plantilla de recursos.

Figura 3-8  
Cargar plantilla de recursos



- En el cuadro de diálogo Cargar plantilla de recursos, seleccione la plantilla que quiera utilizar y pulse en Aceptar. El cuadro de diálogo se cerrará y el asistente mostrará la nueva plantilla que ha seleccionado.

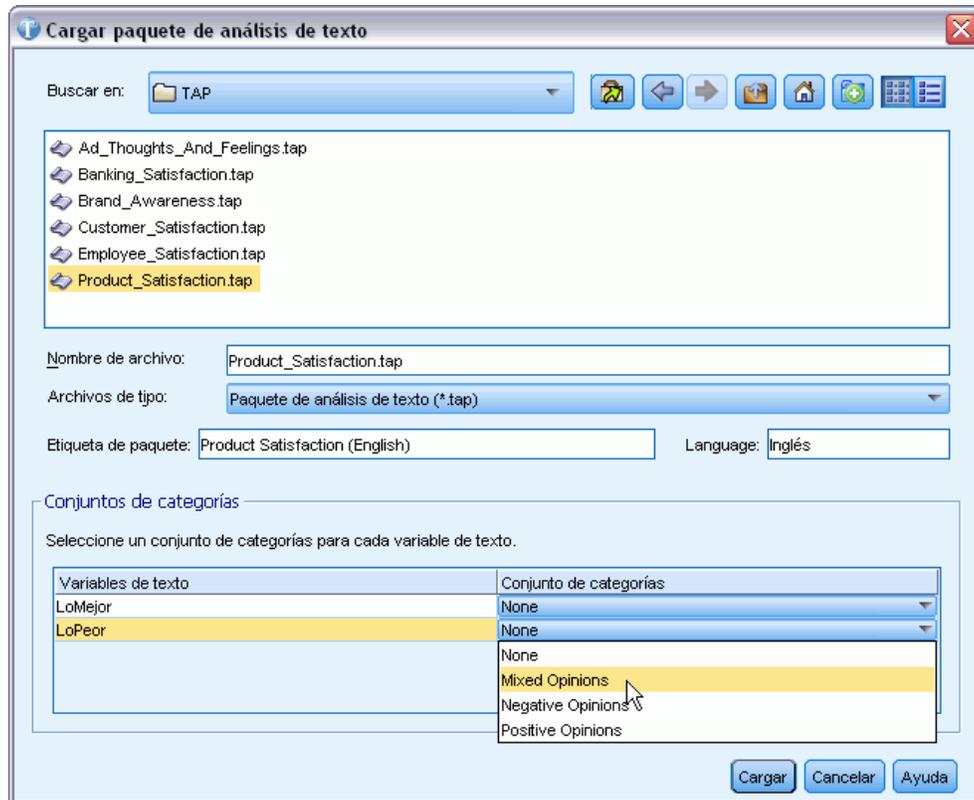
Tenga en cuenta que si tiene alguna plantilla en idiomas para los que no tiene licencia, aparecerá una casilla de verificación en la parte inferior del cuadro de diálogo para permitirle ocultar las plantillas de los idiomas sin licencia de la visualización.

- Pulse en Finalizar para empezar a importar los datos. Si opta por esta posibilidad, el proyecto resultante contendrá los recursos lingüísticos por defecto y, después de realizar extracciones, podrá generar las categorías manualmente o utilizar una técnica automática. [Si desea obtener más información, consulte el tema Categorización de los datos de texto en el capítulo 6 el p. 93.](#)

#### Para seleccionar un paquete de análisis de texto:

- Para cargar un paquete de análisis de texto, asegúrese de seleccionar la opción Paquete de análisis de texto y pulse en Cargar. Aparecerá el cuadro de diálogo Cargar Paquete de análisis de texto.

Figura 3-9  
Cargar Paquete de análisis de texto



- ▶ En el cuadro de diálogo, seleccione el archivo TAP que quiera utilizar. En la lista sólo aparecen directamente los paquetes que están almacenados en el directorio <installation\_directory>\TAP predeterminado. Los campos siguientes se actualizarán con los detalles específicos del archivo TAP seleccionado.
- ▶ En la tabla Conjuntos de categorías puede asignar un conjunto de categorías a cada una de las variables de texto. En la columna Conjunto de categorías pulse en la lista desplegable de cada casilla para elegir un conjunto de categorías disponible. Si selecciona Ninguno, no tendrá ninguna categoría para esa variable de texto hasta que no las cree en otro momento.
- ▶ Pulse en Aceptar. El cuadro de diálogo se cerrará y el asistente mostrará el nuevo archivo TAP que ha seleccionado. Después de seleccionar el TAP y los conjuntos de categorías, el asistente finaliza y poco después puede ver los registros codificados en categorías preincorporadas. A partir de ese punto, puede exportar los resultados o utilizar las categorías como punto de partida para el análisis.
- ▶ Pulse en Finalizar para cerrar el cuadro de diálogo y crear el proyecto. Cuando termine, la aplicación abre automáticamente la vista Pregunta correspondiente a la primera pregunta de texto con final abierto del proyecto. Si opta por realizar la extracción, aparecerá un diálogo de progreso de la extracción; el proceso puede tardar unos minutos en finalizar. Ahora ya puede empezar a analizar las preguntas. Para cambiar a otra pregunta, en los menús elija Ver > Pregunta.

## Uso de los paquetes de análisis de texto

Un paquete de análisis de texto, también llamado TAP, actúa como plantilla para la categorización de las respuestas del texto. Utilizar un TAP es una manera fácil de categorizar los datos de texto con una mínima intervención, ya que contiene el marco de códigos y los recursos lingüísticos necesarios para codificar un amplio número de registros rápidamente y de forma automática. Mediante el uso de los recursos lingüísticos, los datos de texto se analizan y se realiza en ellos el proceso de minería para extraer los conceptos clave. Basándose en los conceptos clave y en los patrones encontrados en el texto, los registros pueden categorizarse en el conjunto de categorías que seleccionó en el TAP. Puede realizar su propio TAP o actualizar uno.

Un TAP está compuesto por los elementos siguientes:

- **Conjuntos de categorías.** Un conjunto de categorías se compone fundamentalmente de categorías predefinidas, códigos de categorías, descriptores para cada categoría y, por último, un nombre para todo el conjunto de categorías. Los descriptores son elementos lingüísticos (conceptos, tipos, patrones y reglas) como el término *barato* o el patrón *buen precio*. Los descriptores se utilizan para definir una categoría de manera que cuando el texto coincide con un descriptor de categoría, el registro se coloca en la categoría.
- **Recursos lingüísticos.** Los recursos lingüísticos son un conjunto de bibliotecas y recursos avanzados que se ajustan para extraer patrones y conceptos clave. Estos conceptos y patrones de extracción, a su vez, se utilizan como los descriptores que permiten a los registros colocarse en una categoría del conjunto de categorías.

Puede realizar y actualizar paquetes de análisis de texto.

Después de seleccionar el TAP y de elegir un conjunto de categorías para cada variable de texto del Asistente de proyecto nuevo, IBM® SPSS® Text Analytics for Surveys puede extraer y categorizar sus registros. Desde aquí, puede exportar los resultados o continuar ajustando la categorización hasta que consiga los resultados que desea.

*Nota:* Los TAP pueden crearse y utilizarse de manera intercambiable entre SPSS Text Analytics for Surveys y IBM® SPSS® Text Analytics.

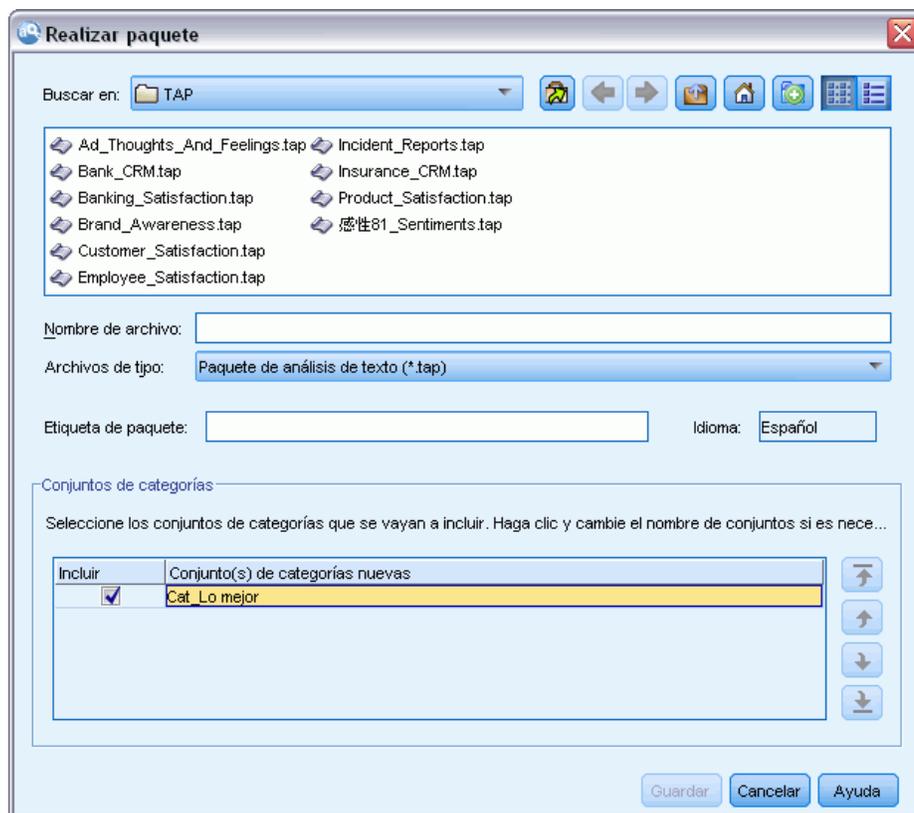
## Creación de paquetes de análisis de texto

Si tiene un proyecto con al menos una categoría y algunos recursos, puede crear un paquete de análisis de texto (TAP) a partir del contenido del proyecto abierto. El conjunto de categorías y descripciones (conceptos, tipos, reglas o resultados de patrones TLA) de cada pregunta pueden colocarse en TAP junto con todos los recursos lingüísticos abiertos en el Editor de recursos.

Puede ver el idioma para el que se crearon los recursos. El idioma se establece en la pestaña Recursos avanzados de Editor de recursos.

**Importante:** Si sus categorías contienen coincidencias de texto, registros forzados o marcas, no se guardarán en los conjuntos de categorías porque son específicos de los datos y casi nunca pueden utilizarse en otros datos. Sin embargo, los códigos de categoría y las etiquetas sí se guardan.

Figura 3-10  
Diálogo Realizar Paquete de análisis de texto

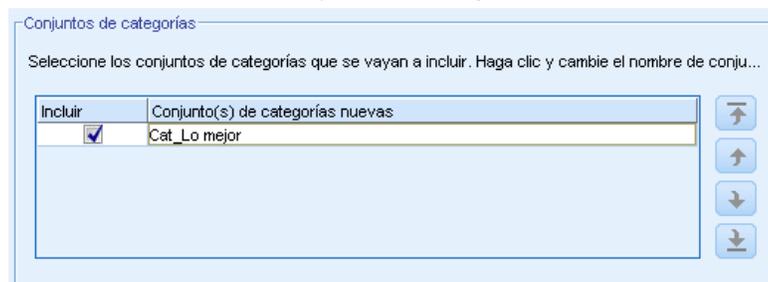


### **Para crear un paquete de análisis de texto**

- ▶ En los menús elija Archivo > Paquetes de análisis de texto > Realizar paquete. Aparecerá el diálogo Realizar paquete.
- ▶ Busque el directorio en el que se ha guardado el TAP. De manera predeterminada, los TAP se guardan en el subdirectorio \TAP del directorio de instalación del producto.
- ▶ Introduzca un nombre para el TAP en el campo Nombre de archivo.
- ▶ Introduzca una etiqueta en el campo Etiqueta de paquete. Cuando especifica un nombre de archivo, este nombre automáticamente aparece como la etiqueta, pero puede cambiarla si lo desea. Si guarda un TAP en este directorio por defecto, la etiqueta del paquete aparecerá como opción en la lista desplegable del Asistente de proyecto nuevo.
- ▶ Para excluir un conjunto de categorías del TAP, quite la marca de la casilla de verificación Incluir. Al hacerlo se asegura de que no se añadirá al paquete. Por defecto, en el TAP se incluye un conjunto de categorías por pregunta. Siempre debe haber al menos un conjunto de categorías en el TAP.
- ▶ Cambie el nombre de los conjuntos de categorías. La columna Conjunto(s) de categorías nuevas contiene nombres genéricos de manera predeterminada, que se generan al añadir el prefijo Cat\_ en

el nombre de la variable de texto. Con solo pulsar con el ratón en la casilla podrá editar el nombre. Pulse en cualquier otro lugar para que el cambio de nombre surta efecto. Si cambia el nombre de un conjunto de categorías, este nombre solo cambia en el TAP y no modifica el nombre de la variable en el proyecto abierto.

**Figura 3-11**  
*Cambie el nombre de los conjuntos de categorías*

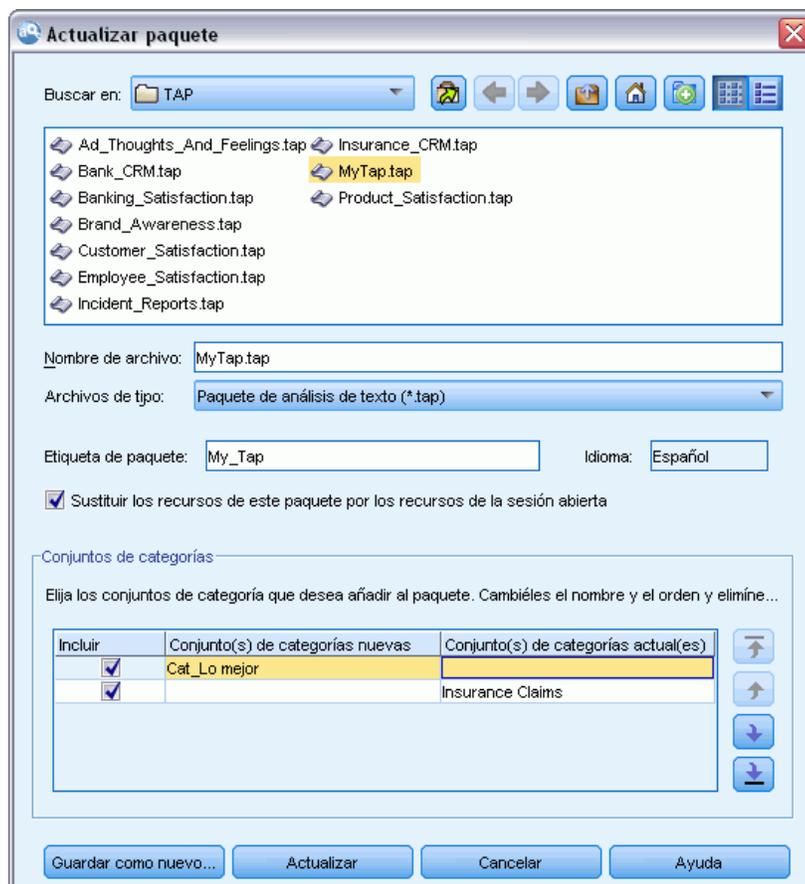


- ▶ Cambie el orden de los conjuntos de categorías mediante las teclas de flecha a la derecha de la tabla de conjuntos de categorías.
- ▶ Pulse en Guardar para crear el paquete de análisis de texto. El cuadro de diálogo se cierra.

### ***Actualización de los paquetes de análisis de texto***

Si realiza mejoras en un conjunto de categorías, en los recursos lingüísticos o crea un conjunto de categorías completamente nuevo, puede actualizar un paquete de análisis de texto (TAP) para que pueda volver a utilizar estas mejoras en otro momento. Para ello debe encontrarse en el proyecto abierto que contiene la información que desea colocar en el TAP. Al actualizar puede optar por agregar conjuntos de categorías, sustituir recursos, cambiar la etiqueta del paquete o cambiar el nombre o el orden de los conjuntos de categorías.

Figura 3-12  
Diálogo Actualizar paquete de análisis de texto



### **Para actualizar un paquete de análisis de texto**

- ▶ En los menús elija Archivo > Paquetes de análisis de texto > Actualizar paquete. Aparece el diálogo Actualizar paquete de análisis de texto.
- ▶ Busque el directorio que contiene el paquete de análisis de texto que desea actualizar.
- ▶ Introduzca un nombre para el TAP en el campo Nombre de archivo.
- ▶ Para sustituir los recursos lingüísticos que se encuentran dentro del TAP por los que están en el proyecto actual, seleccione la opción Sustituir los recursos de este paquete por los recursos de la sesión abierta. Por lo general conviene actualizar los recursos lingüísticos porque se utilizaron para extraer los conceptos clave y los patrones empleados para crear las definiciones de categorías. El hecho de tener los recursos lingüísticos más recientes, garantiza que obtenga los mejores resultados para categorizar sus registros. Si no selecciona esta opción, los recursos lingüísticos que ya estaban en el paquete se mantienen sin cambios.
- ▶ Para actualizar únicamente los recursos lingüísticos, asegúrese de que selecciona la opción Sustituir los recursos de este paquete por los recursos de la sesión abierta y elige solo los conjuntos de categorías actuales que ya se encontraban en el TAP.

- ▶ Para incluir los nuevos conjuntos de categorías desde el proyecto abierto desde la sesión abierta en el TAP, marque la casilla de verificación para cada categoría que se definió para añadirse. Puede añadir uno o varios conjuntos de categorías, o ninguno.
- ▶ Para eliminar conjuntos de categorías del TAP, quite la marca de la casilla de verificación Incluir. Puede optar por eliminar un conjunto de categorías que ya estuviera en el TAP, puesto que está añadiendo uno mejorado. Para ello, quite la marca de verificación de la casilla Incluir del conjunto de categorías correspondiente en la columna Conjunto(s) de categorías actual(es). Siempre debe haber al menos un conjunto de categorías en el TAP.
- ▶ Cambie el nombre de los conjuntos de categorías, si procede. Con solo pulsar con el ratón en la casilla podrá editar el nombre. Pulse en cualquier otro lugar para que el cambio de nombre surta efecto. Si cambia el nombre de un conjunto de categorías, este nombre solo cambia en el TAP y no modifica el nombre de la variable en el proyecto abierto. Si hay dos conjuntos de categorías con el mismo nombre, los nombres aparecerán en rojo hasta que corrija la duplicación.

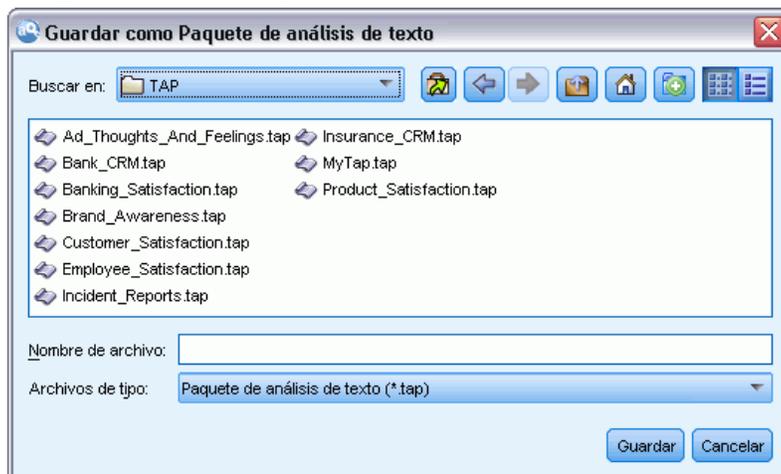
Figura 3-13

*Nombres duplicados*

Incluir	Conjunto(s) de categorías nuevas	Conjunto(s) de categorías actual(es)
<input checked="" type="checkbox"/>	Cat_Lo mejor	
<input checked="" type="checkbox"/>		Cat_Lo mejor

- ▶ Para crear un paquete nuevo con los contenidos de la sesión fusionados con los contenidos del TAP seleccionado, pulse en Guardar como nuevo. Aparece el diálogo Guardar como Paquete de análisis de texto. Consulte las instrucciones siguientes.
- ▶ Pulse en Actualizar para guardar los cambios realizados en el TAP seleccionado.

Figura 3-14

*Diálogo Guardar como Paquete de análisis de texto*

***Para guardar un paquete de análisis de texto***

- ▶ Busque el directorio en el que se ha guardado el archivo TAP. Por defecto, los archivos TAP se guardan en el subdirectorío TAP del directorío de instalación.
- ▶ Introduzca un nombre para el archivo TAP en el campo Nombre de archivo.
- ▶ Introduzca una etiqueta en el campo Etiqueta de paquete. Cuando especifica un nombre de archivo, automáticamente se utiliza este nombre también como etiqueta. Sin embargo, puede cambiar el nombre de la etiqueta. Es necesario que tenga una etiqueta. Si guarda un TAP en este directorío por defecto, la etiqueta del paquete aparecerá como opción en la lista desplegable del Asistente de proyecto nuevo.
- ▶ Pulse en Guardar para crear el nuevo paquete.

## ***Trabajo con proyectos***

En IBM® SPSS® Text Analytics for Surveys, trabajará con datos de encuestas que podrá categorizar. Para ello, deberá crear proyectos donde generará y guardará definiciones de categorías y las respuestas a las que se correspondan. Un proyecto puede contener los elementos siguientes:

- **Datos de la encuesta.** Variable de respuesta de texto para preguntas con final abierto, una variable de ID exclusiva y variables de referencia opcionales. Los datos de la encuesta no se guardan en el proyecto, sino que se leen a partir del origen de datos original cuando el proyecto se abre.
- **Recursos lingüísticos.** Plantillas de propiedad y personalizadas por el usuario y bibliotecas (sinónimos, exclusiones y diccionarios de tipo) que se utilizan cuando se extraen conceptos y patrones del texto.
- **Resultados extraídos.** Presentes después de realizar una extracción, se trata de las palabras clave y las frases identificadas y extraídas de los datos de las respuestas. Estos conceptos se utilizarán para crear las categorías.
- **Categorías.** Proceden de los conjuntos de categorías TAP, la creación manual o la técnica automática de generación de categorías. Las respuestas de la encuesta se asignan a estas categorías.

### ***Apertura de proyectos***

Para volver a un proyecto existente ábralo. Sólo se puede abrir un proyecto a la vez. Si intenta abrir un proyecto cuando hay otro abierto, se le indicará que guarde primero el otro proyecto, si procede.

Cuando el proyecto se abre, IBM® SPSS® Text Analytics for Surveys comprueba los recursos lingüísticos para determinar si hay bibliotecas públicas más recientes de las que contiene el proyecto. Si ese es el caso, se le preguntará si desea actualizar las bibliotecas. Puede optar por mantener la versión actual y no actualizar, o bien fusionar las actualizaciones en el proyecto. [Si desea obtener más información, consulte el tema Actualización de bibliotecas en el capítulo 9 el p. 212.](#)

**¡Importante!** SPSS Text Analytics for Surveys no guarda físicamente los datos de origen en sus proyectos. En su lugar, en el proyecto se guarda una referencia a dichos datos en la máquina. Si algún usuario cambia algunas de las variables originales importadas en el origen de datos, aparecerá un aviso que indica que no se han encontrado los datos. Si esto ocurre, deberá volver a importar los datos y hacer coincidir las preguntas “nuevas” con los nombres de las variables originales para poder seguir trabajando en el proyecto. En general, no se recomienda que cambie el nombre de las variables o de las cabeceras de columna de los datos de origen.

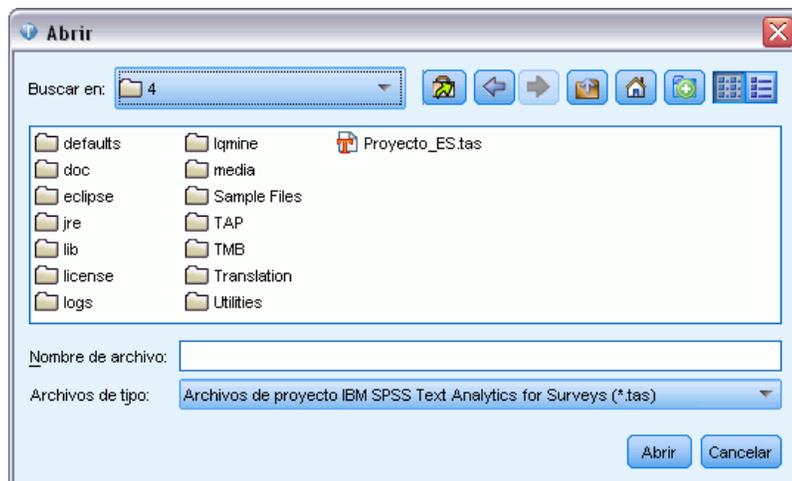
*Nota:* Los resultados de la extracción se guardan en los proyectos a menos que elija no hacerlo (Herramientas > Opciones). Cuando se cierra un proyecto, las definiciones de categoría se guardan, pero el panel Resultados extraídos quedará en blanco. Cuando abra dicho proyecto, deberá ejecutar una extracción si desea seguir categorizando las respuestas. Las definiciones de categoría

existentes muestran un signo de interrogación (?) en lugar del número de respuestas. Después de volver a realizar la extracción, aparecerá de nuevo el número de respuestas.

### **Para abrir un proyecto**

- ▶ En los menús elija Archivo > Abrir proyecto. Aparecerá el cuadro de diálogo Abrir proyecto.

Figura 4-1  
Cuadro de diálogo Abrir



- ▶ En la lista seleccione el directorio y el nombre del proyecto que desea abrir. No puede ordenar los detalles de este cuadro de diálogo, como el tamaño y la fecha del archivo.
- ▶ Pulse en Aceptar para abrir el proyecto en la ventana principal.
- ▶ Si el proyecto contiene datos procedentes de una base de datos protegida por contraseña, se le indicará que facilite la contraseña cada vez que abra el proyecto.
- ▶ Si el contenido del proyecto procede de una versión anterior del producto, se le indicará que convierta los recursos al nuevo formato. Esto implica que después de guardar el proyecto, no podrá abrirlo en una versión anterior del producto.
- ▶ Si las bibliotecas públicas del proyecto se han modificado desde la última vez que abrió el proyecto, un mensaje de alerta le notificará el cambio.

**Importante:** Siempre que abra un proyecto, se abrirá también el conjunto de datos correspondiente. Si no se encuentran los datos, aparecerá un mensaje de error. Para seguir trabajando con los datos, deberá reimportarlos. [Si desea obtener más información, consulte el tema Cambio de los orígenes de datos el p. 61.](#)

## **Edición de las propiedades del proyecto**

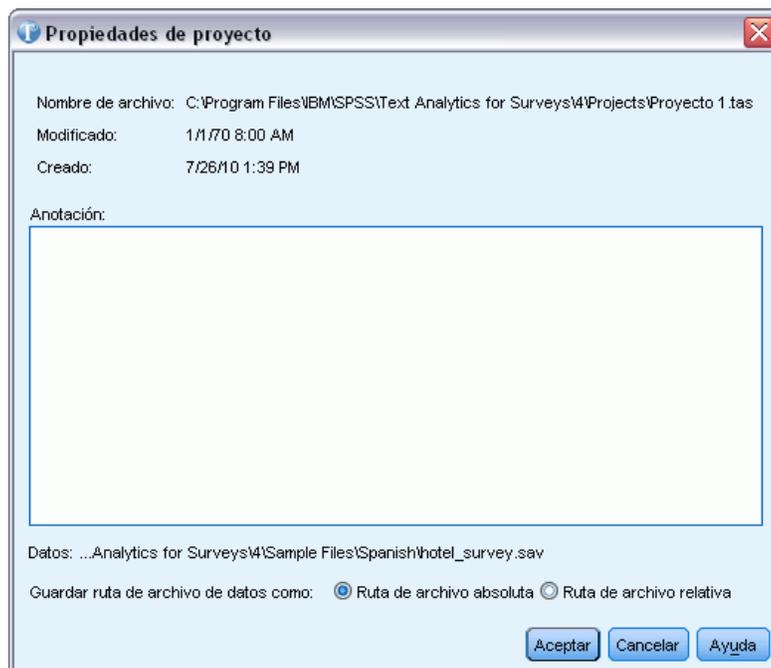
Por defecto, cada proyecto nuevo se llama *Proyecto 1*. Puede revisar las propiedades básicas del proyecto así como añadir o modificar una anotación para el proyecto.

### **Para editar las propiedades del proyecto**

- ▶ En los menús elija Archivo > Propiedades de proyecto. Aparecerá el cuadro de diálogo Propiedades de proyecto.

Figura 4-2

Cuadro de diálogo Propiedades de proyecto



- ▶ Si lo desea, escriba un comentario o descripción del proyecto en el cuadro de texto Anotación.

*Nota:* El nombre del archivo de datos se muestra en este cuadro de diálogo. Puesto que está creando un proyecto y todavía no ha importado ningún dato, el nombre del archivo es desconocido. Después de la importación, solo aparecerán los últimos 60 caracteres del nombre de archivo de los datos. Si el nombre es más largo, puede pasar el ratón por encima del mismo para verlo completo.

- ▶ Pulse en Aceptar para aceptar las nuevas propiedades. El cuadro de diálogo se cierra y las propiedades del proyecto se aplican.

## **Vista de los datos del proyecto**

Cuando haya importado el archivo de datos, aparece la vista Pregunta correspondiente a la primera pregunta de texto con final abierto del proyecto. Sin embargo, es posible que desee ver todos los datos que ha importado. Para ello vaya a la vista Proyecto completo, que ofrece una visión completa de los datos. Para acceder a esta vista, elija Ver > Proyecto completo en el menú. En esta vista podrá:

- Revisar el contenido de todas las variables importadas.
- Asignar valores y etiquetas a las variables.
- Cambiar los tipos de variables.

- Ordenar las variables.
- Copiar datos de casillas contiguas y pegarlos en otras aplicaciones.
- Cambiar el tamaño de las columnas de las variables.

**Importante:** Los datos que ha importado en el proyecto tienen carácter de solo lectura; no podrá editarlos desde IBM® SPSS® Text Analytics for Surveys.

Entonces podrá empezar a extraer conceptos de dichas respuestas, con los que se crearán las categorías. [Si desea obtener más información, consulte el tema Categorización de los datos de texto en el capítulo 6 el p. 93.](#)

Figura 4-3

Acceso a la vista Proyecto completo



## Ordenación de las variables

Puede ordenar los datos en la vista Proyecto completo alfabéticamente o por longitud de los datos.

### Para ordenar los datos en la vista Proyecto completo

- ▶ Seleccione la columna que desee ordenar y pulse el botón derecho del ratón en el título de la columna para abrir un menú contextual.
- ▶ Seleccione la opción de orden que desee entre las opciones siguientes:
  - **Clasificación natural.** Los resultados se ordenan a medida que se leen durante la importación.
  - **Orden ascendente A–Z.** Los resultados se ordenan alfabéticamente empezando por las casillas vacías, con números y luego de la A a la Z.
  - **Orden descendente: A–Z.** Los resultados se ordenan alfabéticamente de la Z a A Z, números y casillas vacías.
  - **Orden ascendente longitud.** Los resultados se ordenan por longitud; las respuestas más breves aparecen al principio.
  - **Orden descendente: longitud.** Los resultados se ordenan por longitud; las respuestas más largas aparecen al principio.

## Edición de las propiedades de variables

Mientras define los datos durante el proceso de importación, se le pide que identifique la variable que representa los ID exclusivos, las que representan las preguntas que desea analizar y, si procede, las variables de referencia que desearía incluir. Una vez importados los datos, puede añadir información a las propiedades correspondientes a dichas variables, o bien cambiar su rol en el proyecto. Por ejemplo, puede analizar una variable que haya importado como variable de referencia. Puede cambiar las siguientes propiedades de variables:

- Añadir o cambiar el nombre o la etiqueta de una variable.

- Cambiar una variable de referencia por una variable de texto con final abierto.
- Cambiar una variable con final abierto por una variable de referencia.
- Cambiar el tipo de datos de un ID o una variable de referencia.

Figura 4-4

Cuadro de diálogo Propiedades de referencia

### **Para editar las propiedades de variables**

- ▶ En la vista Proyecto completo, seleccione la columna correspondiente a la variable cuyas propiedades desea modificar, y pulse el botón derecho del ratón en el título de la columna para abrir un menú contextual.
- ▶ Elija Propiedades en el menú. Aparecerá el cuadro de diálogo Propiedades.
- ▶ Si lo desea, añada o modifique el nombre o la etiqueta de la variable.
- ▶ Para utilizar las etiquetas de la variable en lugar del nombre de la variable en el producto, seleccione la opción Mostrar etiqueta en lugar del nombre.
- ▶ Si lo desea, cambie el rol de las variables en el análisis por Referencia o Texto con final abierto. No puede cambiar el rol de una variable de ID. Si ha empezado a trabajar en una variable de texto (o pregunta) con final abierto y cambia el rol por el de una variable de referencia, el trabajo de categorización que haya realizado se perderá.
- ▶ Cambie el tipo de datos de la variable por Texto o Numérico.

## **Guardado de proyectos**

Cuando cierra un proyecto o finaliza la sesión, se le pedirá que guarde los cambios, si procede. Los proyectos se guardan en archivos con la extensión de archivo \*.tas.

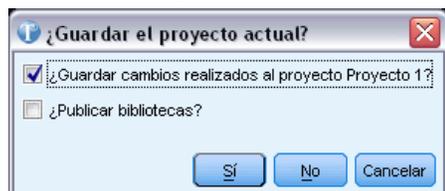
### **Para guardar los proyectos**

- ▶ En los menús elija Archivo > Guardar proyecto. El proyecto se guarda.

**Para guardar un proyecto al salir**

- ▶ Al cerrar un proyecto se abre un cuadro de diálogo que le pregunta si desea guardar los cambios realizados en el proyecto, y si desea (volver a) publicar las bibliotecas.

Figura 4-5  
Cuadro de diálogo Guardar proyecto actual

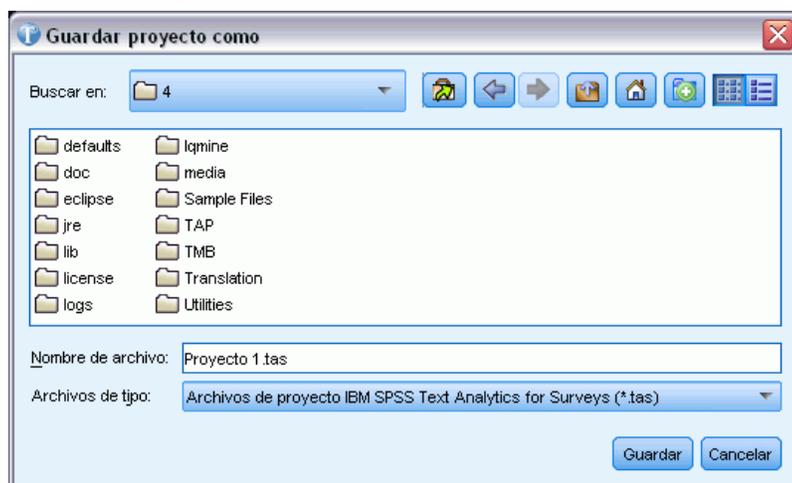


- ▶ Seleccione Guardar cambios en el proyecto.
- ▶ Si desea publicar bibliotecas para utilizarlas posteriormente, o desea actualizar las versiones públicas, seleccione también Publicar bibliotecas. Si no hay ninguna biblioteca que necesite publicación, la opción aparece desactivada. [Si desea obtener más información, consulte el tema Publicación de bibliotecas en el capítulo 9 el p. 211.](#)
- ▶ Pulse en Sí para guardar. Si eligió publicar bibliotecas, se abre otro cuadro de diálogo. [Si desea obtener más información, consulte el tema Compartimiento de bibliotecas en el capítulo 9 el p. 210.](#)

**Para guardar con otro nombre de proyecto**

Si recibe un mensaje de alerta de nombre duplicado o si elige guardar el proyecto con un nombre diferente, se abre el cuadro de diálogo Guardar proyecto como.

Figura 4-6  
Cuadro de diálogo Guardar proyecto como



- ▶ Escriba el nuevo nombre exclusivo para el proyecto en el cuadro de texto Nombre de archivo.
- ▶ Pulse en Guardar para guardar el nombre nuevo.

## **Exportación de los resultados de la categorización**

En algunos casos, la creación de categorías puede constituir el punto final del análisis. Simplemente conocer los temas principales que expresan los encuestados, y cuántos encuestados mencionan cada tema, puede ser suficiente para el propósito del análisis de texto. Sin embargo, a menudo querrá realizar más informes y análisis sobre las categorías, como crear tablas y gráficos para mostrar los resultados. También puede utilizar otras variables del cuestionario para obtener una caracterización de los encuestados más detallada en cada categoría, o utilizar las categorías para estudiar otras respuestas.

Si desea seguir trabajando con los nuevos resultados de la categorización, puede exportar las categorías en formato de texto para importarlo a una aplicación de análisis cuantitativo, como el sistema base de IBM® SPSS® Statistics. El archivo resultante contiene los ID de los encuestados y los nombres y etiquetas de las categorías, pero no contiene los valores de las variables de referencia ni de las respuestas con final abierto.

*Nota:* También puede generar gráficos de resumen, por ejemplo, un gráfico de barras Las 5 categorías principales. Estos gráficos, que se exportan en HTML, pueden utilizarse en presentaciones. [Si desea obtener más información, consulte el tema Exportación de gráficos de resumen el p. 58.](#)

### **Tipos de archivos exportados**

Cuando realiza la exportación puede utilizar uno de los tipos de archivos disponibles:

- Archivos SPSS Statistics (\*.sav). [Si desea obtener más información, consulte el tema Exportación a IBM SPSS Statistics o IBM SPSS Data Collection el p. 54.](#)
- Archivos Microsoft Excel (\*.xls / \*.xlsx). [Si desea obtener más información, consulte el tema Exportando a Microsoft Excel el p. 56.](#)
- IBM® SPSS® Data Collection. [Si desea obtener más información, consulte el tema Exportación a IBM SPSS Statistics o IBM SPSS Data Collection el p. 54.](#) Consulte también la biblioteca de desarrollo de Data Collection en *Data Collection Modelo de datos*.

### **Resultado Dicotomías frente a resultado Categorías**

Los datos de texto que se han codificado con IBM® SPSS® Text Analytics for Surveys forman un conjunto de múltiples respuestas, puesto que cada encuestado puede aportar más de una respuesta y se le puede asignar a más de una categoría por una sola pregunta. Esto significa que los datos deben codificarse en un formato especial cuando se exportan. Hay dos formatos de resultados disponibles de exportación: **dicotomías** y **categorías**.

**Dicotomías.** Los resultados se centran en las marcas de miembros de categorías para cada ID de respuesta. Para cada categoría de los datos, cada encuestado (por ID) pertenece o no a la categoría utilizando una marca binaria, que se codifica como verdadera o falsa. Los datos se estructuran en formato de tabla, con el ID en la columna de la izquierda y una columna para cada categoría. Este tipo de datos permite un número ilimitado de categorías por respuesta. Si hay 10 categorías, habrá 10 variables nuevas.

**Categorías.** Los resultados se centran en el conjunto de categorías al que pertenece una respuesta. Para cada respuesta de los datos, cada categoría a la que se ha asignado aparece como un valor (para SPSS Statistics) o como una categoría en sí misma (para Microsoft Excel). Los datos de exportación de la categoría se estructuran en formato de tabla, con el ID en la columna de la izquierda, seguido por una columna por categoría a la que al menos pertenece una respuesta. Estas columnas no representan una categoría particular, sino un espacio para registrar un código de categoría asignado. Para cada respuesta, cada código de categoría al que pertenece la respuesta se guarda en un espacio aparte. La respuesta con el número máximo de categorías asignadas determina el número de variables que debe crearse. Si hay 10 categorías pero no hay ningún encuestado codificado con más de 4 categorías, se necesitarán 4 variables para representar las categorías.

- **Para SPSS Statistics/Data Collection.** Para cada ID de respuesta de los datos, cada categoría a la que esté asignado aparece como un valor individual del 1 al  $N$ , donde  $N$  es el valor más alto del código de categoría. Si no ha asignado códigos en el Administrador de marcos de código, los códigos se asignan automáticamente cuando se crea la categoría. Si a un encuestado se le asignan menos categorías que el número máximo, las variables que quedan sin utilizar se codifican con el valor que falta del sistema SPSS Statistics (un punto).
- **Para Microsoft Excel.** Para cada ID de respuesta de los datos, cada categoría a la que se le asigna aparece como nombre de categoría o como etiqueta de categoría, dependiendo de lo que esté utilizando en la interfaz del producto. Si a un encuestado se le asignan menos categorías que el número máximo, las variables que quedan sin utilizar se codifican con un valor en blanco en Microsoft Excel.

### ***Exportación a IBM SPSS Statistics o IBM SPSS Data Collection***

Una vez categorizadas las respuestas, probablemente querrá analizar los resultados utilizando procedimientos estadísticos. IBM® SPSS® Text Analytics for Surveys permite crear un archivo de datos con un formato que puede utilizarse en diferentes productos—estas instrucciones explican cómo exportar para utilizar en IBM® SPSS® Statistics (programa de análisis estadísticos) y varios productos IBM® SPSS® Data Collection. SPSS Text Analytics for Surveys creará automáticamente la variable de respuestas múltiples en el archivo exportado. El formato exacto del archivo depende del tipo de datos que seleccione—dicotomías o categorías.

*Nota:* El archivo resultante contiene los ID de los encuestados y los nombres y etiquetas de las categorías, pero no contiene los valores de las variables de referencia ni de las respuestas con final abierto.

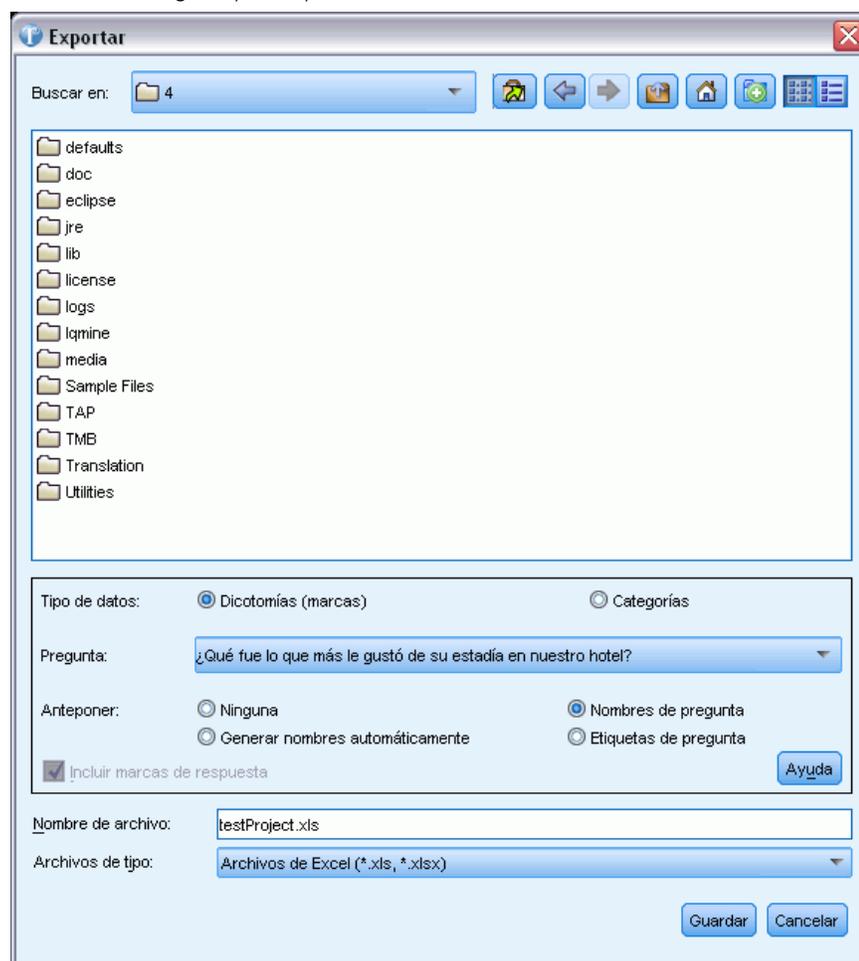
**SPSS Statistics solamente.** En relación con los resultados, si el conjunto de datos contiene datos o casos que faltan (cuando un encuestado no ha contestado a una pregunta concreta), la aplicación asigna el valor que falta del sistema a estos casos. Los archivos SPSS Statistics exportados por SPSS Text Analytics for Surveys no pueden utilizarse en versiones anteriores a la 7.5 de SPSS Statistics.

### Para exportar datos

- ▶ En el menú Archivo > Exportar resultados elija una de las siguientes opciones para abrir el cuadro de diálogo Exportar:
  - SPSS Statistics Archivo
  - Data Collection Archivo

Figura 4-7

Cuadro de diálogo Exportar para formatos de archivos IBM SPSS Statistics



- ▶ En la lista desplegable Guardar en, seleccione la unidad y la carpeta donde desea que se guarde el archivo.
- ▶ Seleccione una opción de Tipo de datos. [Si desea obtener más información, consulte el tema Exportación de los resultados de la categorización el p. 53.](#)
  - Dicotomías.
  - Categorías. Esta opción no está disponible para el archivo de datos Data Collection, y se seleccionará Dicotomías por defecto.

- ▶ En la lista desplegable Pregunta, seleccione la pregunta que desea exportar. Puede elegir si exportar los resultados de la categorización para una sola pregunta o para todo el proyecto. Si desea exportar cada pregunta por separado, deberá seleccionar y exportar cada pregunta individualmente. O puede seleccionar Proyecto completo para exportar los resultados de todas las preguntas con final abierto.
- ▶ Seleccione una opción de Anteponer para designar un prefijo al exportar los nombres de las categorías para el *proyecto completo*. Esta opción es más útil cuando se exportan los datos para múltiples preguntas. La anteposición añade un prefijo al nombre de categoría o a la etiqueta de la variable original, y garantiza que no haya duplicados cuando se combinen los resultados de múltiples preguntas a la hora de exportar el proyecto completo. Seleccione entre las siguientes alternativas:
  - Ninguna. Tal como la opción implica, no se añade ningún prefijo.
  - Nombres de pregunta. Los prefijos generan nombres de variables de categoría (ya sea el nombre de categoría o la etiqueta de categoría dependiendo de lo que utilice en el proyecto) con el nombre (pregunta) de variable de texto con final abierto. El nombre de variable de pregunta procede del origen de datos original. Si el nombre de variable de categoría que se genera no se adapta a las convenciones de nombres de variables o sobrepasa los 40 caracteres, se crean nombres por defecto (como con la opción Generar automáticamente).
  - Generar nombres automáticamente. Coloca automáticamente un prefijo *Q1*, *Q2*, *Q3*, etc., a los nombres de categoría. *Q1* representa la primera pregunta de la exportación, y así sucesivamente.
  - Etiquetas de pregunta. Los prefijos generan nombres de variables de categoría (ya sea el nombre de categoría o la etiqueta de categoría dependiendo de lo que utilice en el proyecto) con la etiqueta (pregunta) de variable de texto con final abierto. El nombre de variable de pregunta procede del origen de datos original. Si el nombre de variable de categoría que se genera no se adapta a las convenciones de nombres de variables o sobrepasa los 40 caracteres, se crean nombres por defecto (como con la opción Generar automáticamente).
- ▶ Si tiene marcas de respuesta en los datos, puede optar por exportarlas. Para exportar marcas de respuesta, seleccione la opción correspondiente. [Si desea obtener más información, consulte el tema Marcación de las respuestas el p. 74.](#)
- ▶ En el cuadro de texto Nombre de archivo, seleccione el nombre de proyecto por defecto que aparece, o escriba otro nombre para el archivo.
- ▶ Pulse en Guardar para exportar los resultados.

### **Exportando a Microsoft Excel**

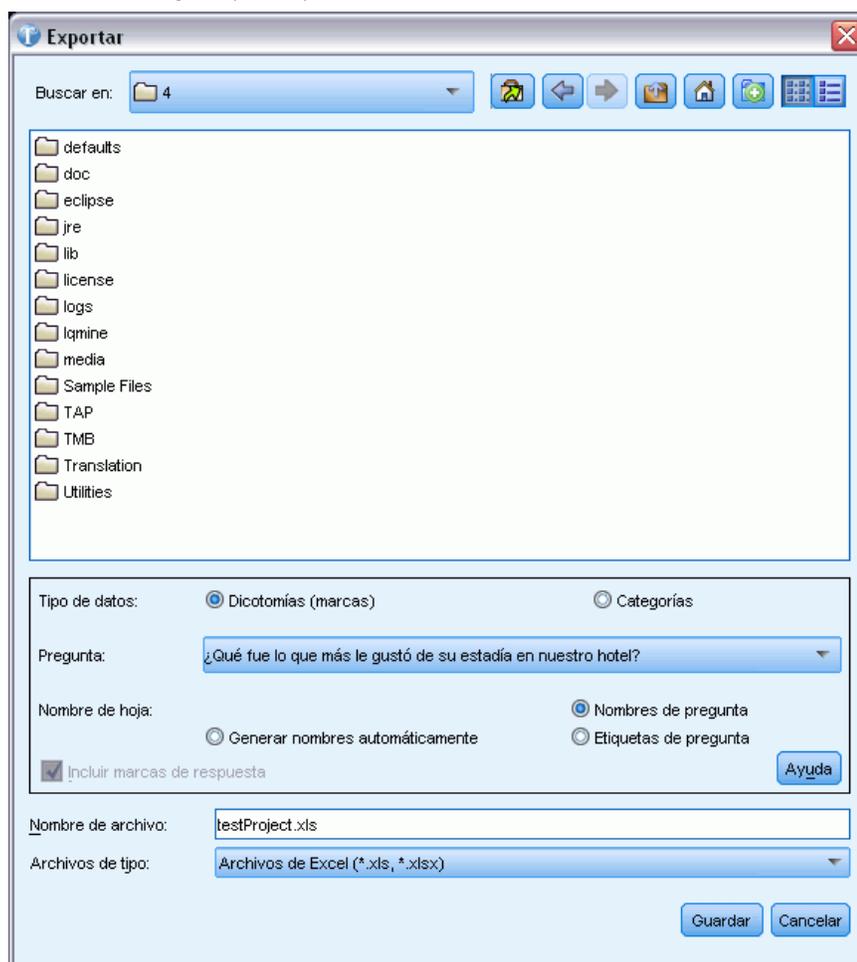
Una vez categorizadas las respuestas, probablemente querrá analizar los resultados mediante procesos estadísticos. IBM® SPSS® Text Analytics for Surveys permite crear un archivo de datos con un formato que puede utilizarse en diferentes productos. Las instrucciones siguientes explican cómo exportar en formato Microsoft Excel. SPSS Text Analytics for Surveys creará automáticamente la variable de múltiples respuestas en el archivo exportado. El formato exacto del archivo depende del tipo de datos que seleccione—dicotomías o categorías. El archivo

resultante contiene los ID de los encuestados y los nombres y etiquetas de las categorías, pero no contiene los valores de las variables de referencia ni de las respuestas con final abierto.

### Para exportar datos

- ▶ En los menús elija Archivo > Exportar resultados > Archivo de Microsoft Excel. Aparecerá el cuadro de diálogo Exportar.

Figura 4-8  
Cuadro de diálogo Exportar para archivos Microsoft Excel



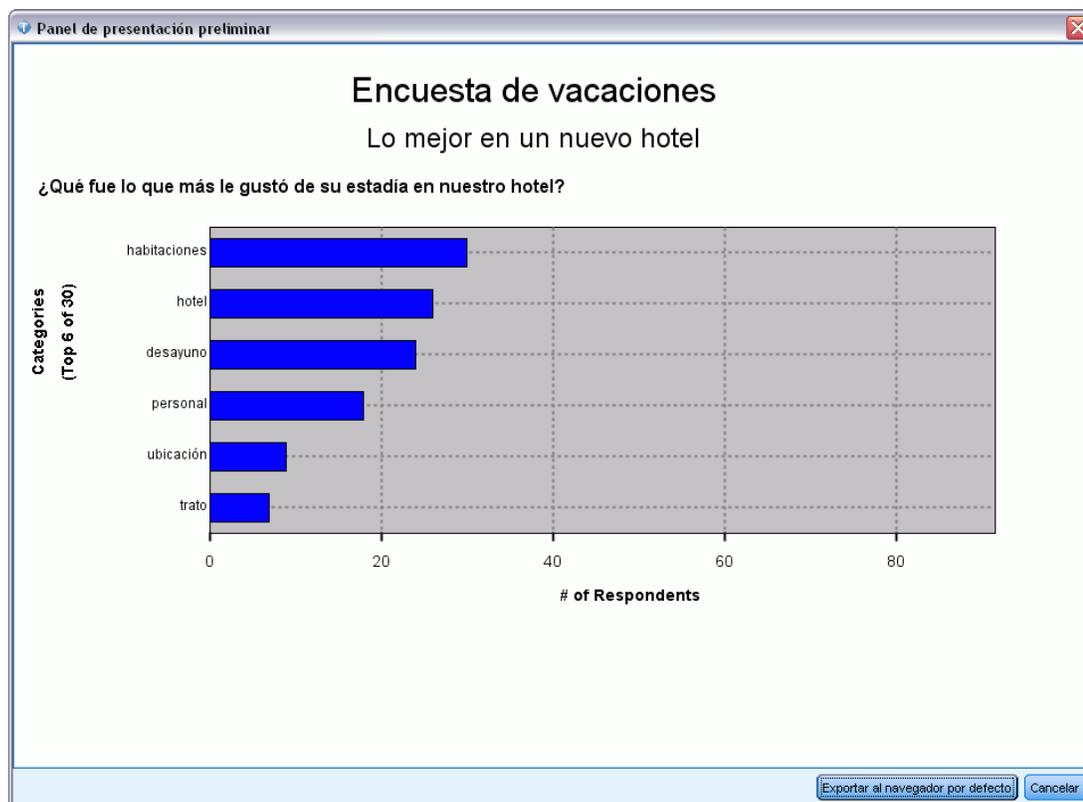
- ▶ En la lista desplegable Guardar en, seleccione la unidad y la carpeta donde desea que se guarde el archivo.
- ▶ Seleccione una opción de Tipo de datos. Si desea obtener más información, consulte el tema [Exportación de los resultados de la categorización](#) el p. 53.
  - Dicotomías.
  - Categorías. Esta opción no está disponible para el archivo de datos IBM® SPSS® Data Collection, y se seleccionará Dicotomías por defecto.

- ▶ En la lista desplegable Pregunta, seleccione la pregunta que desea exportar. Puede elegir si exportar los resultados de la categorización para una sola pregunta o para todo el proyecto. Si desea exportar cada pregunta por separado, deberá seleccionar y exportar cada pregunta individualmente. O puede seleccionar Proyecto completo para exportar los resultados de todas las preguntas con final abierto.
- ▶ Seleccione una opción de nombres de hoja de cálculo para designar cómo deben nombrarse las hojas de cálculo generadas en el archivo Microsoft Excel. Seleccione entre las siguientes alternativas:
  - Nombres de pregunta. Utiliza el nombre (pregunta) de variable de texto con final abierto como nombre de hoja de cálculo. El nombre de pregunta procede del origen de datos original. Si el nombre de variable de categoría que se genera no se adapta a las convenciones de nombres de variables o sobrepasa los 40 caracteres, se crean nombres por defecto (como con la opción Generar automáticamente).
  - Generar nombres automáticamente. Asigna automáticamente *Q1*, *Q2*, *Q3*, etc., a las hojas de cálculo. *Q1* representa la primera pregunta de la exportación, y así sucesivamente.
  - Etiquetas de pregunta. Utiliza la etiqueta (pregunta) de variable de texto con final abierto, si existe alguna, como nombre de hoja de cálculo. Si el nombre de variable de categoría que se genera no se adapta a las convenciones de nombres de variables o sobrepasa los 40 caracteres, se crean nombres por defecto (como con la opción Generar automáticamente).
- ▶ Si tiene marcas de respuesta en los datos, puede optar por exportarlas. Para exportar marcas de respuesta, seleccione la opción correspondiente. [Si desea obtener más información, consulte el tema Marcación de las respuestas el p. 74.](#)
- ▶ En el cuadro de texto Nombre de archivo, seleccione el nombre de proyecto por defecto que aparece, o escriba otro nombre para el archivo.

## ***Exportación de gráficos de resumen***

Cuando haya terminado de trabajar con las categorías y los datos, puede exportar informes gráficos de resumen para compartir los resultados del análisis con otros usuarios. El resultado genera un gráfico de barras por cada pregunta. Puede elegir el número de categorías principales que se utilizarán en cada gráfico para que pueda presentar visualmente las cinco o las diez categorías principales de una pregunta determinada. El gráfico puede exportarse al navegador por defecto, desde el que podrá guardar la imagen para utilizarla en otros productos o presentaciones.

Figura 4-9  
Ejemplos de gráficos de resumen en una ventana de navegación



#### **Para exportar gráficos de resumen**

- ▶ En los menús elija Categorías > Exportar gráficos de resumen. Aparecerá el cuadro de diálogo Exportar gráficos de resumen.
- ▶ Configure el gráfico de resumen utilizando las opciones que se describen en este tema.
- ▶ Pulse en Generar para generar el gráfico y mostrarlo en un panel de Presentación preliminar.
- ▶ Pulse en Exportar al navegador por defecto para ver el gráfico en una ventana del navegador.

Figura 4-10  
Cuadro de diálogo *Exportar gráficos de resumen*

### **Opciones de informe**

**Exportar gráfico(s) de.** Elija si desea generar un gráfico de resumen para todas las preguntas del proyecto o solo para la pregunta actualmente seleccionada.

**Número de categorías principales para incluir.** Seleccione el número máximo de categorías que se han de mostrar en el gráfico. Las categorías que tengan el mayor número de registros se usarán en primer lugar.

**Volver a extraer si se encuentran resultados de extracción desactualizados.** Seleccione esta opción para forzar una nueva extracción antes de generar el gráfico si los resultados de la extracción no están actualizados.

**Título principal.** Escriba un título principal para los gráficos. Por ejemplo, este podría ser el nombre de su encuesta.

**Subtítulo.** Escriba un subtítulo para los gráficos. Por ejemplo, este podría ser el nombre de la empresa o el año de la encuesta.

**Título de la pregunta.** Para ayudarle a identificar cada gráfico, el título procede de la pregunta. Puede elegir utilizar el nombre variable de la pregunta, su etiqueta o ningún nombre.

**Etiqueta del eje X.** Defina una etiqueta para el eje X de los gráficos. Por defecto se propone una etiqueta.

**Etiqueta del eje Y.** Defina una etiqueta para el eje Y de los gráficos. Por defecto se propone una etiqueta.

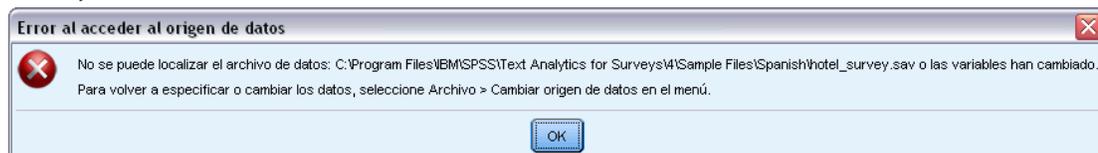
**Color de barra.** Elija un color para las barras del gráfico de resumen. Este color se aplica a todas las preguntas.

**Fuente.** Elija una fuente para los títulos y las etiquetas del gráfico.

## ***Cambio de los orígenes de datos***

Siempre que abra un proyecto, se abrirá también el conjunto de datos correspondiente. Si no se encuentran los datos, aparecerá un mensaje de error. En ocasiones no pueden encontrarse los datos porque se movieron a otra ubicación, alguien los borró accidentalmente o se les cambió el nombre. Como alternativa, puede cambiar a otros orígenes de datos.

**Figura 4-11**  
*Mensaje de error de datos no hallados*

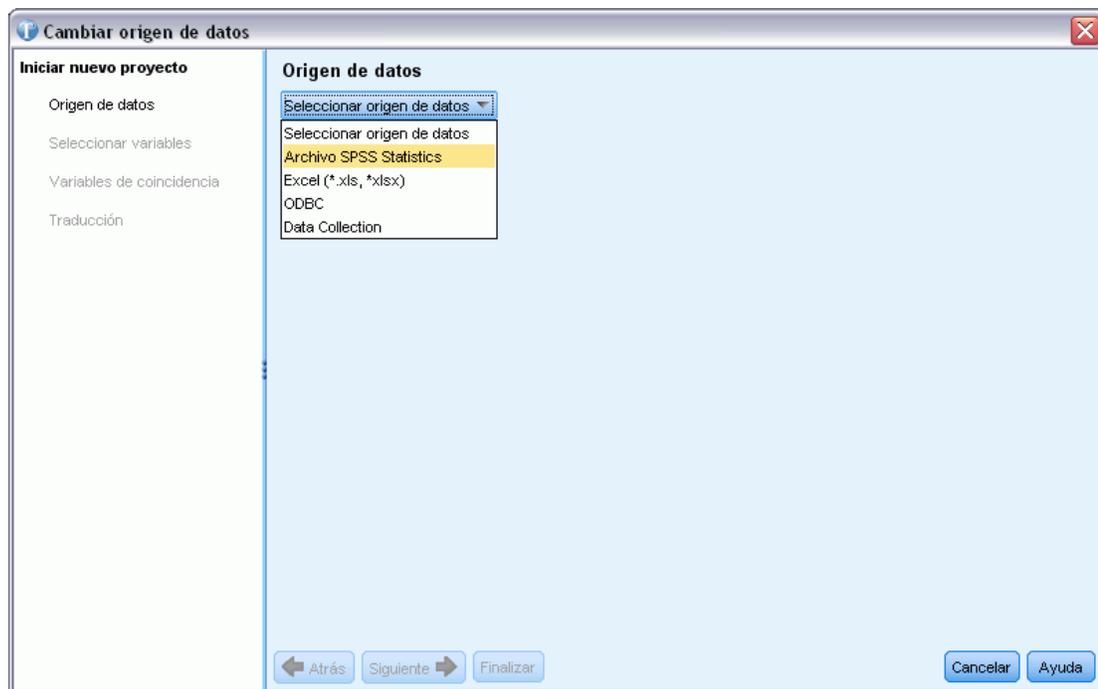


Para seguir trabajando con los datos, debe cambiar la ubicación al origen de datos adecuado. Si se encuentran cambios de variables en los datos, como variables nuevas, con otro nombre o variables ausentes, se le pedirá que compare las variables importadas anteriormente con las nuevas.

### ***Para cambiar el origen de los datos:***

- ▶ Si recibe este tipo de mensaje de error, pulse en Aceptar.
- ▶ En los menús elija Archivo > Cambiar origen de datos. Aparece el diálogo del Asistente para cambiar el origen de datos.

Figura 4-12  
Asistente para cambiar el origen de datos



### ***Selección de los orígenes de datos***

Cuando se abre el asistente, debe empezar seleccionando un origen de datos. IBM® SPSS® Text Analytics for Surveys se ha optimizado para procesar conjuntos de datos de hasta 10.000 registros, aunque el rendimiento variará en función del volumen de texto contenido en estos registros. Consulte las instrucciones de instalación para conocer las estadísticas y recomendaciones sobre rendimiento.

**Importante:** Debe haber una variable de ID con un valor exclusivo para cada registro con el fin de importar los datos.

Puede elegir uno de los siguientes orígenes de datos:

- Archivos **SPSS Statistics** (\*.sav).
- Archivos **Microsoft Excel** (\*.xls / \*.xlsx).
- **Base de datos ODBC** (protocolo Open Database Connectivity de Microsoft).
- Modelo de datos **Data Collection**. Esta opción solo está disponible si tiene instalado el modelo de datos.

### Uso de los archivos IBM SPSS Statistics

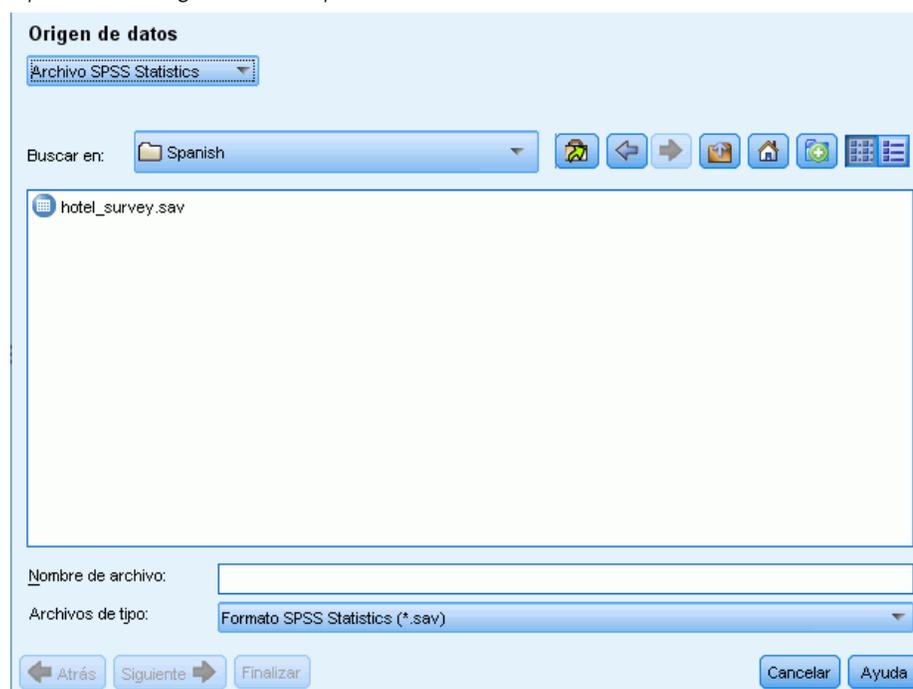
Puede importar un archivo IBM® SPSS® Statistics (.sav) en IBM® SPSS® Text Analytics for Surveys. Debe haber una variable de ID con un valor exclusivo para cada registro con el fin de importar los datos.

**Importante:** No puede importar el archivo SPSS Statistics (.sav) con registros que sobrepasen los 4000 caracteres.

*Nota:* SPSS Text Analytics for Surveys se ha optimizado para procesar conjuntos de datos de hasta 10.000 registros, aunque el rendimiento variará en función del volumen de texto contenido en estos registros. Consulte las instrucciones de instalación para conocer las estadísticas y recomendaciones sobre rendimiento.

Figura 4-13

Opciones de origen de datos para archivos IBM SPSS Statistics



#### Para obtener datos de IBM SPSS Statistics

- ▶ En la primera pantalla del asistente, seleccione el archivo SPSS Statistics de la lista desplegable. El asistente muestra las opciones para los archivos SPSS Statistics.
- ▶ En la lista desplegable Buscar en, seleccione la unidad y la carpeta donde se encuentra el archivo.
- ▶ Seleccione el archivo de la lista. Aparecerá el cuadro de texto Nombre de archivo.
- ▶ Pulse en Siguiente para seleccionar variables. [Si desea obtener más información, consulte el tema Selección de variables el p. 34.](#)

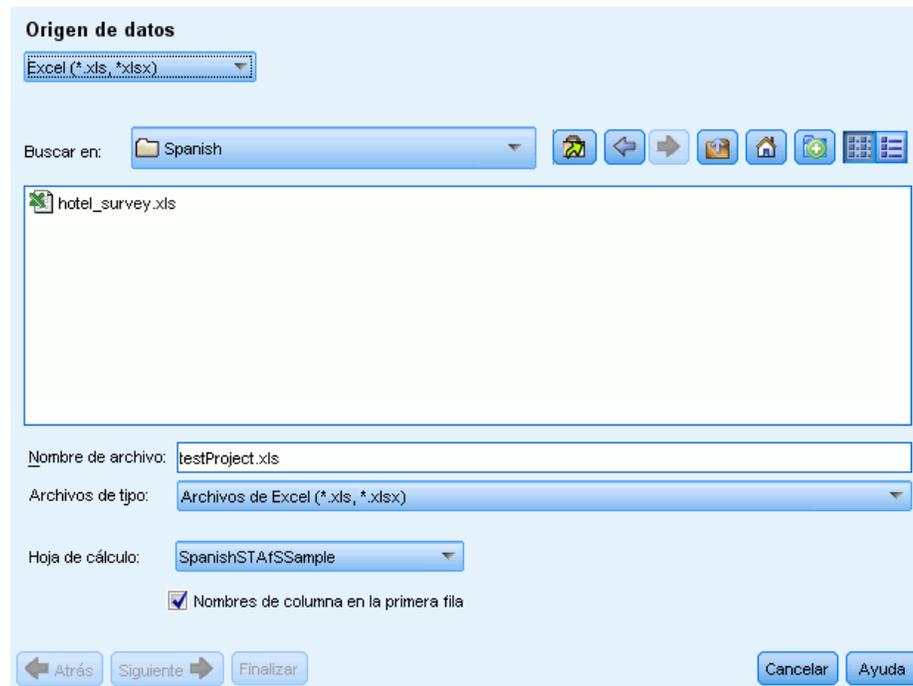
### Uso de los archivos Microsoft Excel

Puede importar un archivo Microsoft Excel (.xls / .xlsx) en IBM® SPSS® Text Analytics for Surveys. Debe haber una variable de ID con un valor exclusivo para cada registro con el fin de importar los datos.

**Importante:** Durante la importación del archivo Microsoft Excel, puede seleccionar una opción para Nombres de columna en la primera fila. Para utilizar esta opción, la primera línea de la hoja de cálculo debe contener nombres de columna—no la fila justo encima donde empiezan los datos. Por ejemplo, si los datos y los nombres de columna empiezan en la línea 7, antes de importar debe eliminar las filas 1–6 para que la importación del archivo sea correcta.

*Nota:* SPSS Text Analytics for Surveys se ha optimizado para procesar conjuntos de datos de hasta 10.000 registros, aunque el rendimiento variará en función del volumen de texto contenido en estos registros. Consulte las instrucciones de instalación para conocer las estadísticas y recomendaciones sobre rendimiento.

Figura 4-14  
Opciones de origen de datos para archivos Microsoft Excel



#### Para obtener datos de Microsoft Excel

- ▶ En la primera pantalla del asistente, seleccione Excel de la lista desplegable. El asistente muestra las opciones para los archivos Microsoft Excel.
- ▶ En la lista desplegable Buscar en, seleccione la unidad y la carpeta donde se encuentra el archivo.
- ▶ Seleccione el archivo de la lista. Aparecerá el cuadro de texto Nombre de archivo.

- ▶ Seleccione la hoja de cálculo en la lista desplegable. Solo puede importar datos de una sola hoja de cálculo. Para trabajar con los datos de varias hojas de cálculo, debe crear varios proyectos.
- ▶ Si la primera fila de esta hoja de cálculo contiene las cabeceras de columna, seleccione Nombres de columna en la primera fila. Para utilizar esta opción, la primera línea de la hoja de cálculo debe contener nombres de columna—no la fila justo encima donde empiezan los datos. Por ejemplo, si los datos y los nombres de columna empiezan en la línea 7, antes de importar debe eliminar las filas 1–6 para que la importación del archivo sea correcta. La aplicación puede utilizar estos (o bien una versión convertida si las cabeceras de columna no cumplen las convenciones de nombres de variables de IBM® SPSS® Statistics) como nombres de variable. En caso contrario, la aplicación utilizará las letras de las columnas de la hoja de cálculo como identificadores.
- ▶ Pulse en Siguiente para seleccionar variables. [Si desea obtener más información, consulte el tema Selección de variables el p. 34.](#)

### ***Uso de los datos a través de ODBC***

Los datos procedentes de orígenes de bases de datos, comúnmente de las bases de datos, se importan fácilmente en IBM® SPSS® Text Analytics for Surveys. El producto puede leer directamente las bases de datos que utilizan controladores ODBC (Open Database Connectivity) después de que los controladores apropiados se hayan instalado en la máquina en la que SPSS Text Analytics for Surveys está instalado. Debe haber una variable de ID con un valor exclusivo para cada registro con el fin de importar los datos.

*Nota:* SPSS Text Analytics for Surveys se ha optimizado para procesar conjuntos de datos de hasta 10.000 registros, aunque el rendimiento variará en función del volumen de texto contenido en estos registros. Consulte las instrucciones de instalación para conocer las estadísticas y recomendaciones sobre rendimiento.

Figura 4-15  
Opciones de origen de datos para ODBC

**Origen de datos**

ODBC

Origen ( DSN )

Nombre	Descripción
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

Usuario:

Contraseña:

Tabla

SQL

Atrás Siguiente Finalizar Cancelar Ayuda

#### **Para utilizar ODBC como medio**

- ▶ En la primera pantalla del asistente, seleccione ODBC de la lista desplegable. El asistente muestra las opciones para ODBC.
- ▶ Especifique el origen de datos seleccionándolo de la lista de orígenes de ODBC registrados o tecleando el nombre en el cuadro de texto Origen (DSN). Si necesita registrar nuevos orígenes de datos que no aparecen en la lista, pulse en ODBC. Así se abrirá el Administrador de orígenes de datos ODBC, que se encuentra en la mayoría de los ordenadores Microsoft Windows. Si no lo encuentra, no podrá utilizar la importación ODBC. Consulte el sistema de ayuda de Microsoft Windows para obtener más información.
- ▶ Si el origen de datos está protegido por contraseña, deberá especificar un nombre de usuario y una contraseña. Se le pedirá que lo haga cada vez que abra el proyecto, puesto que, por razones de seguridad, el nombre de usuario y la contraseña no se guardan en el proyecto.
- ▶ Puede seleccionar los datos de dos maneras distintas: Tabla o SQL. Puede seleccionar una tabla directamente, o bien utilizar comandos SQL para seleccionar los datos.
- ▶ Pulse en Siguiente para seleccionar variables. [Si desea obtener más información, consulte el tema Selección de variables el p. 34.](#)

## **Uso de datos IBM SPSS Data Collection**

### **Para importar mediante IBM SPSS Data Collection**

- ▶ En la primera pantalla del asistente, seleccione Data Collection de la lista desplegable. La opción del modelo de datos IBM® SPSS® Data Collection solo está disponible si tiene el modelo de datos instalado con otro producto.

## **Selección de variables**

Después de seleccionar el origen de datos, el paso siguiente es especificar las variables que se van a importar. Pueden importarse tres tipos de variables en un proyecto.

### **Variable de ID exclusiva (obligatoria)**

La variable de ID es una clave numérica exclusiva o una cadena que identifica a cada encuestado. No es necesario ordenar el archivo de datos por variable de ID exclusiva para que se lea correctamente. Después de que el programa lo haya leído, los registros pueden ordenarse según determinados criterios. [Si desea obtener más información, consulte el tema Ordenación de las variables el p. 50.](#) Esta variable de ID es obligatoria para importar datos. Cada registro (o caso) importado debe tener un valor de ID exclusivo.

Pueden darse dos circunstancias por la que la importación falla:

- Si se detectan valores de ID duplicados
- Si hay registros con valores de ID en blanco

*Nota:* Si se detecta un ID duplicado y IBM® SPSS® Statistics está instalado en el ordenador, puede utilizar el procedimiento Identify Duplicate Cases (Identificar casos duplicados) en ese producto para identificar los duplicados y luego utilizar las opciones para indicar qué registros deben conservarse (casos principales).

### **Variables de texto con final abierto (obligatorio)**

Las variables de texto con final abierto representan las respuestas de texto a las preguntas de la encuesta. Se requiere al menos una de estas variables para importar datos. Estas variables pueden ser variables de cadena o cadena larga en SPSS Statistics, columnas que contienen casillas de texto o generales en Microsoft Excel, o campos de notas o texto procedentes de las bases de datos. Cada variable de texto con final abierto se analizará por separado. Existe un límite de 4.000 caracteres de tamaño (extensión) para cada variable de texto importado de un archivo .SAV.

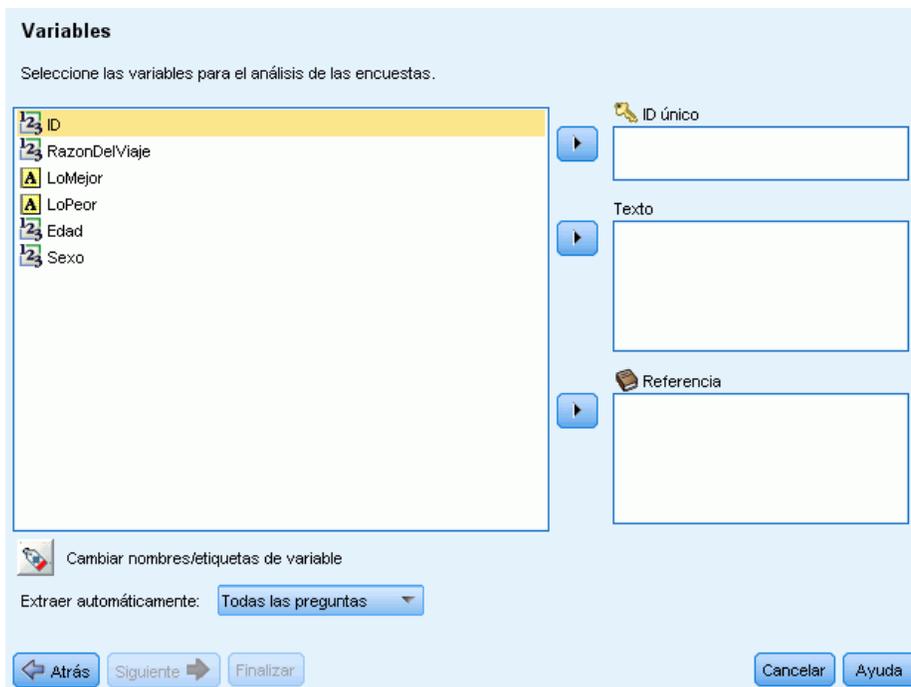
### **Variables de referencia (opcional)**

Las variables de referencia son variables opcionales y adicionales, generalmente de categorías, que pueden importarse con fines de referencia. Las variables de referencia no se utilizan en el análisis de texto, pero proporcionan información adicional que describen al encuestado, lo que puede ayudar a entender e interpretar sus respuestas. Las variables demográficas suelen incluirse como variables de referencia, puesto que pueden contribuir a comprender qué términos o categorías se

están utilizando en cada grupo de personas. Algunos ejemplos son sexo, departamento, trabajo y nivel de estudios (para evaluaciones de formación y estudiantes). Puede ver todas las variables de referencia después de la importación en la vista Proyecto completo. También puede mostrar las variables de referencia en el panel Datos de la vista Pregunta. Además, puede seleccionar variables de referencia en el gráfico de barras del panel de visualización para poder profundizar a un subconjunto de encuestados.

*Nota:* Las variables de referencia que se leen de un archivo de datos SPSS Statistics tendrán etiquetas de variable (si se suministran) que aparecen como cabeceras de columnas, y sus etiquetas de valor correspondientes (si se suministran) se muestran en las casillas del panel Datos.

Figura 4-16  
Selección de variables



### **Para seleccionar variables y opciones de extracción**

- ▶ En la lista de variables disponibles, seleccione la variable que se corresponde con la variable de ID del conjunto de datos, y pulse en el botón de flecha para moverla al cuadro ID único. El ID debe ser un número o una cadena alfanumérica exclusivos para distinguir los registros entre sí. Si el conjunto de datos contiene ID duplicadas, aparecerá un mensaje de error. En tal caso, deberá limpiar los datos antes de repetir la operación.
- ▶ Desde la lista de variables disponibles, seleccione una o más variables que se correspondan con las variables de respuesta con final abierto, y pulse el botón de flecha para mover dichas variables a la lista Variables de texto con final abierto. Las variables se importarán como preguntas independientes cuyas respuestas serán analizadas y categorizadas por usted.
- ▶ Desde la lista de variables disponibles, seleccione una o más variables que se correspondan con las variables de referencia, y pulse el botón de flecha para mover dichas variables a la lista de

referencia. Las variables de referencia no se utilizan en técnicas automáticas de generación de categorías. Sin embargo, puede ver su contenido y utilizarlas para que le ayuden a tomar decisiones informadas a la hora de categorizar las respuestas.

- ▶ Para ver las etiquetas de las variables en lugar de sus nombres, pulse en el botón situado debajo de la lista de variables a la izquierda.
- ▶ Para cambiar el valor de extracción, seleccione una opción en la lista desplegable. La opción Solo la primera pregunta está seleccionada por defecto; significa que si ha seleccionado más de una variable de texto con final abierto, el proceso de extracción empezará automáticamente para la primera pregunta después de que el asistente finalice. La extracción puede tardar un tiempo en el caso de conjuntos de datos extensos. Por lo tanto, puede optar por extraer Ninguna o Todas las preguntas en función del tiempo del que disponga.
- ▶ Pulse en Siguiente > cuando haya seleccionado todas las variables.

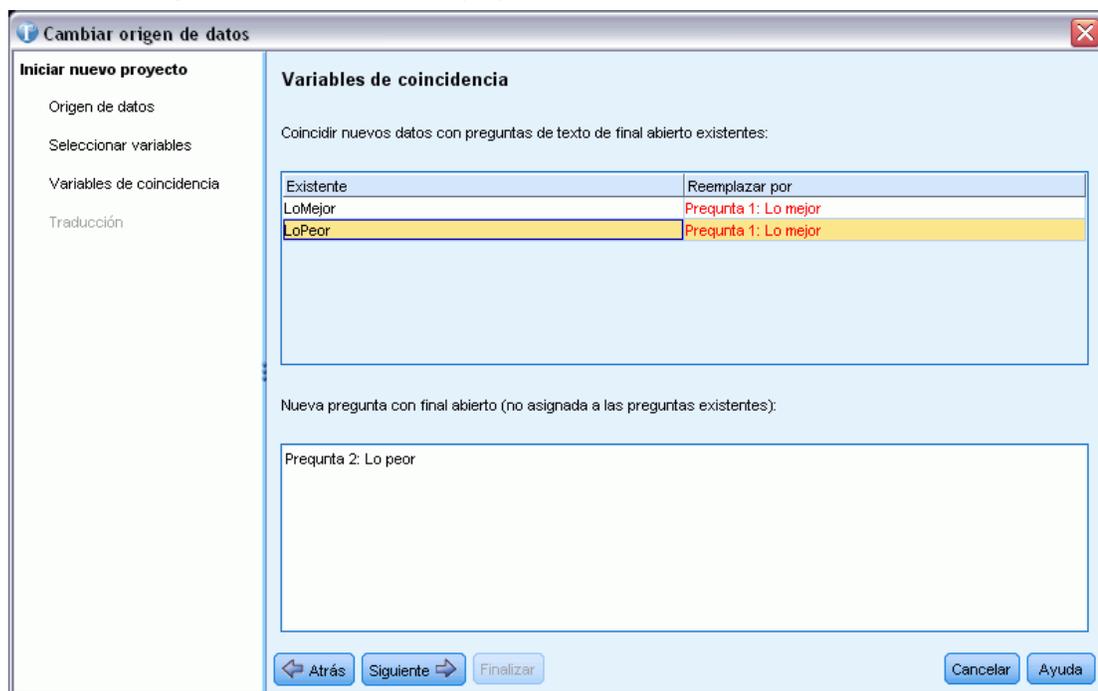
### ***Variables coincidentes***

Después de seleccionar variables en el paso anterior, IBM® SPSS® Text Analytics for Surveys intentará correlacionar las variables importadas con anterioridad con las que acaba de seleccionar. Las coincidencias se proponen automáticamente, pero puede relacionar las variables de otras formas pulsando con el ratón en la columna Reemplazar por y eligiendo otra variable.

Si el nuevo archivo de datos no contiene una variable de texto con final abierto que existiera en el proyecto anterior, puede seleccionar NINGUNO en la lista, y los datos asociados con la pregunta antigua se descartarán del proyecto.

Las variables del nuevo archivo de datos que no puedan asignarse al proyecto existente aparecen en la lista Nueva pregunta con final abierto en la parte inferior del cuadro de diálogo. Después de cambiar los conjuntos de datos, las variables restantes aparecerán como nuevas preguntas en el proyecto.

Figura 4-17  
Cuadro de diálogo *Coincidir contenido con proyecto existente*



#### Para hacer coincidir las variables con otras existentes:

- ▶ Cuando el producto intenta relacionar las nuevas variables con las que ya se encontraban en el archivo de datos, puede cambiar la forma en que las variables relacionan las variables pulsando en la columna Reemplazar por y seleccionando la coincidencia de la variable. Si ha seleccionado más variables en el paso anterior que las especificadas en la columna Reemplazar por, aparecerán en Nueva pregunta con final abierto (no asignada a las preguntas existentes). Después de cambiar los conjuntos de datos, las variables restantes aparecerán como nuevas preguntas en el proyecto.
- ▶ Si la variable existente no tiene ninguna coincidencia en el nuevo conjunto de datos, elija NINGUNO y los datos de esta pregunta existente se descartarán.
- ▶ Si el proyecto contiene respuestas que se hayan forzado dentro o fuera de las categorías o de las marcas, se le indicará que conserve o descarte estos resultados específicos de ID de respuesta. Por lo general, si importa datos diferentes (preguntas nuevas o encuestados distintos, por ejemplo), seguramente querrá descartar esta información para que no se generen resultados falsos. Si está importando el mismo archivo de datos, seguramente querrá conservar esta información puesto que los ID de respuesta coincidirán con los datos antiguos.

### Traducción al inglés

Si trabaja con un texto de origen que no esté en inglés, puede conectarse a Language Weaver para traducirlo a ese idioma. La traducción sólo está disponible hacia el inglés. Debe haber configurado Language Weaver adecuadamente y con conexiones definidas para traducir. [Si desea obtener más información, consulte el tema Opciones: pestaña Traducción en el capítulo 2 el p. 22.](#)

Figura 4-18  
Opciones de traducción

**Traducción mediante Language Weaver**

Traducir texto no inglés a inglés automáticamente con Language Weaver.

Traducir al inglés

Configuración

Conexión de par de idiomas:

Precisión de traducción (1 = más rápida, 3 = máxima calidad):

Utilizar diccionario personalizado

Nombre de diccionario:

### Para traducir al inglés

- ▶ Para traducir los datos de texto de un idioma con licencia al inglés, seleccione la casilla de verificación Traducir al inglés.
- ▶ En la lista Conexión par de idiomas, seleccione la conexión para el par de idiomas Language Weaver que desea utilizar. Si ha configurado Language Weaver en su equipo local, esos pares de idiomas aparecerán automáticamente en esta lista. Puede añadir, cambiar o comprobar la conexión de servicios en línea en la pestaña Traducción del cuadro de diálogo Opciones. [Si desea obtener más información, consulte el tema Opciones: pestaña Traducción en el capítulo 2 el p. 22.](#)
- ▶ Especifique la Precisión de la traducción que desee. Elija un valor entre 1 y 3 indicando el nivel de velocidad contra el nivel de precisión que desea. Un valor bajo hará que los resultados de traducción sean más rápidos pero que la precisión sea menor. Un valor alto producirá resultados con gran precisión pero llevará más tiempo de proceso. Para optimizar el tiempo, se recomienda comenzar con un nivel más bajo e ir aumentándolo sólo si necesita más precisión después de revisar los resultados.
- ▶ Si anteriormente había creado diccionarios personalizados en Language Weaver, puede utilizarlos en relación con la traducción. Para seleccionar un diccionario personalizado, seleccione la casilla de verificación Utilizar diccionario personalizado e introduzca el nombre del diccionario. Para utilizar más de un diccionario, separe los nombres con comas.

- ▶ En el Asistente de nuevo proyecto, pulse en Siguiente> para empezar a seleccionar las categorías y los recursos. [Si desea obtener más información, consulte el tema Selección de categorías y recursos en el capítulo 3 el p. 37.](#)
- ▶ En el asistente Cambiar conjuntos de datos, pulse en Finalizar para completar el cambio de los conjuntos de datos y para iniciar el proceso de traducción.

**Para saltarse la traducción:**

- ▶ Elimine la selección de la opción Traducir al inglés.
- ▶ En el Asistente de nuevo proyecto, pulse en Siguiente> para empezar a seleccionar las categorías y los recursos. [Si desea obtener más información, consulte el tema Selección de categorías y recursos en el capítulo 3 el p. 37.](#)
- ▶ En el asistente Cambiar conjuntos de datos, pulse en Finalizar para completar el cambio de los conjuntos de datos.

## **Actualización de los datos**

A medida que trabaja con los datos del proyecto puede cambiar el origen de datos original. Por ejemplo, puede añadir o eliminar registros. Puede actualizar los datos utilizando la función Actualizar datos. Sin embargo, si ha cambiado los nombres de variable o el nombre de archivo, por ejemplo, tendrá que volver a importar los datos completamente. [Si desea obtener más información, consulte el tema Cambio de los orígenes de datos el p. 61.](#)

**Para actualizar y renovar los datos:**

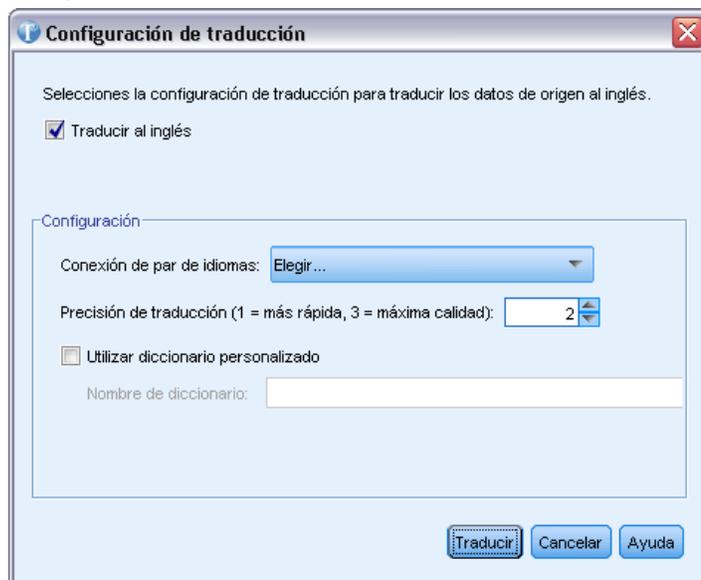
- ▶ En los menús elija Archivo > Actualizar datos. Los datos se vuelven a leer para tener en cuenta los nuevos cambios.
- ▶ Si se ha realizado anteriormente una traducción en inglés, aparecerá el diálogo Configuración de traducción para que pueda seleccionar el par de idiomas y volver a traducir. [Si desea obtener más información, consulte el tema Traducción al inglés el p. 72.](#)

## **Traducción al inglés**

Puede actualizar una traducción cuando lo desee. Tras la traducción, deberá extraer de nuevo dado que los resultados de su traducción no se podrán sincronizar con su nueva traducción.

*Nota:* Si quiere traducir nuevos datos, podrá hacerlo directamente en el asistente Nuevo proyecto cuando cree un proyecto nuevo. [Si desea obtener más información, consulte el tema Traducción al inglés el p. 36.](#)

Figura 4-19  
Configuración de traducción



### **Para traducir al inglés**

- ▶ En los menús seleccione Herramientas > Configuración de traducción. Aparecerá el diálogo Configuración de traducción.
- ▶ Para traducir los datos de texto de un idioma con licencia al inglés, seleccione la casilla de verificación Traducir al inglés.
- ▶ En la lista Conexión par de idiomas, seleccione la conexión para el par de idiomas Language Weaver que desea utilizar. Si ha configurado Language Weaver en su equipo local, esos pares de idiomas aparecerán automáticamente en esta lista. Puede añadir o comprobar conexiones de servicios de red (WAN) o en línea (HTTP) en la pestaña Traducción del diálogo Opciones. [Si desea obtener más información, consulte el tema Opciones: pestaña Traducción en el capítulo 2 el p. 22.](#)
- ▶ Especifique la Precisión de la traducción que desee. Elija un valor entre 1 y 3 indicando el nivel de velocidad contra el nivel de precisión que desea. Un valor bajo hará que los resultados de traducción sean más rápidos pero que la precisión sea menor. Un valor alto producirá resultados con gran precisión pero llevará más tiempo de proceso. Para optimizar el tiempo, se recomienda comenzar con un nivel más bajo e ir aumentándolo sólo si necesita más precisión después de revisar los resultados.
- ▶ Si anteriormente había creado diccionarios personalizados en Language Weaver, puede utilizarlos en relación con la traducción. Para seleccionar un diccionario personalizado, seleccione la casilla de verificación Utilizar diccionario personalizado e introduzca el nombre del diccionario. Para utilizar más de un diccionario, separe los nombres con comas.
- ▶ Pulse en Traducir para empezar el proceso de traducción. Aparecerá el diálogo de progreso de traducción.

## Cómo compartir proyectos

Puede compartir proyectos con otros usuarios o en caso de que quiera trabajar con un proyecto en otra máquina.

### Para compartir un proyecto:

- ▶ En los menús elija Archivo > Guardar proyecto. El proyecto se guarda.
- ▶ Envíe el archivo de proyecto a otra máquina o persona. Este archivo de proyecto contiene una referencia al archivo de datos que importó originalmente. Si desea que el otro usuario utilice los mismos datos de origen para este proyecto, también deberá proporcionarle el archivo de datos original e informarle de la ruta donde deberá copiar dicho archivo de datos para que IBM® SPSS® Text Analytics for Surveys pueda encontrar el archivo de datos cuando abra el archivo del proyecto.
- ▶ Cuando el otro usuario abre el archivo del proyecto en SPSS Text Analytics for Surveys, podrá optar entre utilizar las bibliotecas locales que se encuentran en el archivo del proyecto o bien utilizar las versiones públicas de las bibliotecas que ya tiene. Por lo general, y para garantizar los mismos resultados, conviene utilizar las versiones locales.
- ▶ Si SPSS Text Analytics for Surveys no puede localizar el archivo de datos, aparece un mensaje que avisa al usuario de que debe volver a importar los datos. [Si desea obtener más información, consulte el tema Cambio de los orígenes de datos el p. 61.](#)

## Marcación de las respuestas

Para ayudarle a supervisar el progreso del análisis de la encuesta, puede marcar las respuestas mediante marcas específicas en el panel Datos. Existen varias razones por las que puede marcar una respuesta, por ejemplo:

- Marcar respuestas que ha revisado manualmente y así saber desde dónde deberá continuar más adelante
- Marcar una respuesta que no está seguro sobre cómo manejar
- Señalar y exportar las marcas a otro programa

Una vez que señala una respuesta con una marca, puede continuar trabajando con dicha respuesta. Las marcas solo tienen fines de registro. Puede seleccionar entre las marcas siguientes:

Tabla 4-1  
Descripciones de las marcas

Marca	Descripción
	La marca de Completo denota respuestas que considera terminadas.
	La marca de Importante denota respuestas que considera importantes.

Figura 4-20  
Marcas de respuesta en el panel Datos

	Id.	Respuesta	Categorías
1	1	El <b>hotel cumple</b> con todo lo que se le pide: <b>bien ubicado</b> y <b>limpio</b>	Tipo: Descripción Tipo: Sitio
2	6	Lo que <b>más me gustó</b> fue el <b>trato de los empleados</b> desde el <b>personal de Recepción</b> hasta los <b>empleados de limpieza</b> .	
3	8	El <b>excelente trato del personal</b> de todo el Hotel. La <b>amplitud</b> de las <b>habitaciones</b> y el desayuno	Tipo: Descripción Tipo: Sitio
4	9	La <b>limpieza</b> de las <b>habitaciones</b>	Tipo: Descripción Tipo: Sitio
5	10	Durante mi <b>ESTANCIA</b> , lo que <b>más me gustó</b> fue la <b>Buena comunicación con los medios de transporte públicos</b>	Tipo: Descripción
6	11	<b>Bien comunicado</b> con el <b>centro de la ciudad</b> y <b>próximo</b> al <b>aeropuerto</b> y a las <b>estaciones de tren y autobús</b>	
7	12	Lo <b>mejor</b> es la <b>ubicación</b> . Un <b>paisaje</b> extraordinariamente	
8	16	La <b>localización</b> y <b>tranquilidad interior</b>	
9	19	<b>limpieza</b> de las <b>habitaciones</b>	Tipo: Descripción Tipo: Sitio
10	20	<b>acceso a internet gratis</b>	

#### Para señalar una respuesta con una marca:

- ▶ En el panel Datos seleccione la respuesta que desea marcar.
- ▶ En los menús elija Editar > Marcar respuestas con y luego seleccione el tipo de marca que desea utilizar (Marca de importante o Marca de completo). Se asignará la marca seleccionada. Si la columna Marca del panel Datos no estaba visible, aparecerá. La barra de estado se actualiza con el número de respuestas marcadas.

#### Para borrar las marcas:

- ▶ En el panel Datos pulse el botón derecho del ratón sobre las respuestas de las que desea eliminar la marca.
- ▶ En el menú contextual elija Marcar respuestas con > Borrar marcas. Las marcas seleccionadas se eliminan.

## Barra de estado del proyecto

En función de la ventana o vista en la que esté trabajando, pueden existir diferentes barras de estado. Por defecto, las barras de estado se mostrarán siempre que tenga un proyecto abierto. Estas barras de estado proporcionan información sobre el proyecto y los elementos que contiene. También puede activar o desactivar las barras de estado cuando lo desee.

#### Para desactivar o activar la barra de estado de cualquier ventana:

- ▶ En los menús elija Ver > Barra de estado.

### Ventana Análisis de texto

Esta barra de estado proporciona información sobre las preguntas y las respuestas del proyecto. En función del lugar donde se encuentre en la ventana de análisis de texto, la información de la barra de estado cambiará. También puede ver el número de respuestas que se han marcado como importantes o completas.

Figura 4-21

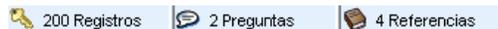
Barra de estado de la ventana de análisis de texto: Vista Pregunta



Cuando está en la vista Pregunta, puede ver el número de categorías para esa pregunta y la categorización por porcentaje de respuesta. Cuando está en la vista Proyecto completo, puede ver la información de todo el proyecto.

Figura 4-22

Barra de estado de la ventana de análisis de texto: vista Proyecto completo



En la tabla siguiente se describe cada elemento de la barra de estado.

Tabla 4-2

Ventana Análisis de texto: Descripción de la barra de estado

Elemento	Descripción
Registros	Número de registros de los datos.
Preguntas	Número de preguntas del proyecto.
Referencia	Número de variables de referencia importadas. Las variables de referencia son variables adicionales que se importan con fines de referencia, pero no se analizan.
Categorías	Número de categorías de una pregunta determinada. Si existen categorías vacías, el número de categorías vacías se indica entre paréntesis.
Categorizadas	Número de respuestas de la pregunta seguido del porcentaje de categorización entre paréntesis.

### Ventana Editor de recursos

Esta barra de estado proporciona información sobre los recursos lingüísticos del proyecto. Puede accionar el área de los términos forzados en la barra, lo que significa que puede pulsar sobre ella para realizar la acción correspondiente. Cuando se trabaja con bibliotecas, puede desactivar elementos de las bibliotecas para excluirlos del proceso. [Si desea obtener más información, consulte el tema Desactivación de bibliotecas locales en el capítulo 9 el p. 207.](#) Si el proyecto contiene elementos desactivados, aparecerán números de dos dígitos en la barra de estado; el primer dígito es el número de elementos presente, y el segundo el número de elementos activados. Por ejemplo, si en la barra de estado se muestra 5(2) bibliotecas, significa que hay cinco bibliotecas en el proyecto pero que solo hay dos activadas.

Figura 4-23

Barra de estado de la ventana Editor de diccionario



En la tabla siguiente se describe cada elemento de la barra de estado.

**Tabla 4-3**  
*Ventana Editor de recursos: Descripción de la barra de estado*

<b>Elemento</b>	<b>Descripción</b>
Biblioteca	Número de bibliotecas en el proyecto.
Tipo	Número de tipos en todo el proyecto.
Término	Número de términos en todas las bibliotecas. Si un término está en la lista Excluido, sigue apareciendo en el recuento. Tenga en cuenta que si un tipo está desactivado, también lo estarán todos los términos de ese tipo.
Excluir	Número de elementos excluidos en todas las bibliotecas del proyecto.
Sinónimo	Número de objetivos de sinónimos en todas las bibliotecas del proyecto.
Opcional	Número de elementos opcionales definidos en todas las bibliotecas del proyecto. Tenga en cuenta que los números incluyen todas las entradas delimitadas en una casilla individualmente.
Términos forzados	Botón que se activa siempre que existen términos forzados en las bibliotecas del proyecto. Si pulsa este botón se muestra el cuadro de diálogo Editar términos forzados. <a href="#">Si desea obtener más información, consulte el tema Forzado de términos en el capítulo 10 el p. 222.</a>

## ***Extracción de datos***

Cuando crea un proyecto a través del asistente Nuevo proyecto, la opción por defecto es realizar una extracción automática para la primera pregunta. Si desea actualizar una extracción, o realizar una extracción de una nueva pregunta, puede hacerlo manualmente (Herramientas > Extraer) o realizar la extracción cuando empiece a generar categorías. El resultado final de esta extracción es un conjunto de conceptos, tipos y patrones. Puede ver y trabajar con estos resultados en el panel Resultados extraídos.

Si no se realizó la extracción cuando creó el proyecto o si eligió no guardar los resultados de la extracción, puede navegar hasta la pregunta con la que desee empezar a trabajar (Ver > Pregunta > “Pregunta”) y realizar la extracción.

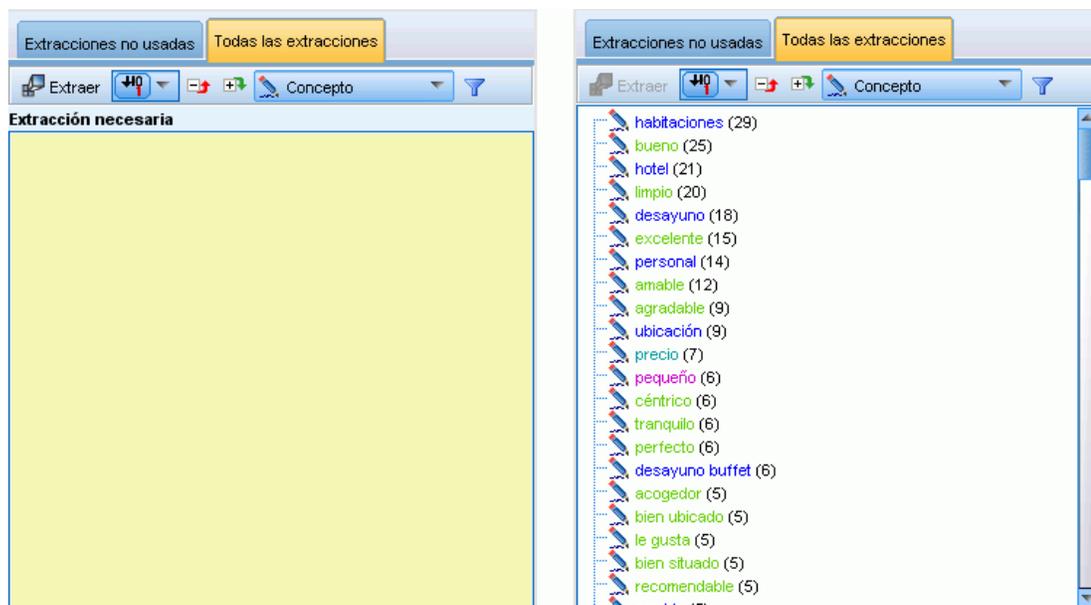
Después de la extracción debe revisar los resultados y realizar los cambios que considere necesarios. [Si desea obtener más información, consulte el tema Refinamiento de los resultados de la extracción el p. 86.](#) Luego puede repetir la extracción para ver los nuevos resultados. Si las respuestas las codifican manualmente dos personas, éstas pueden agrupar las respuestas de manera ligeramente diferente. Sin embargo, la precisión y la continuidad son extremadamente importantes a la hora de categorizar las respuestas de una encuesta.

La potencia de IBM® SPSS® Text Analytics for Surveys reside en su capacidad para proporcionar una reaplicación coherente de las definiciones de categoría. El ajuste preciso desde el principio de los resultados de la extracción permite garantizar que la próxima vez que vuelva a realizar la extracción, obtendrá resultados idénticos en las definiciones de su categoría, perfectamente adaptados al contexto de los datos. De esta forma, las respuestas se asignarán a las definiciones de categoría de una manera más precisa y repetible.

### ***Resultados extraídos: Conceptos, tipos y patrones***

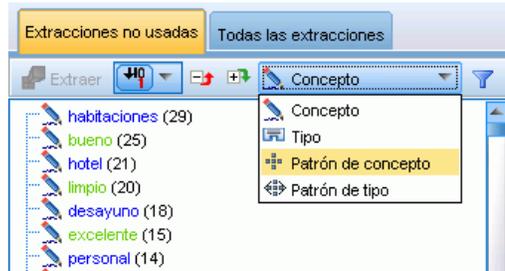
Después de crear un proyecto, la ventana muestra automáticamente la primera pregunta con final abierto que ha importado. El panel Resultados extraídos se encuentra en el ángulo inferior izquierdo de la vista Pregunta. A esta vista se accede desde el menú Ver (Ver > Pregunta > “Pregunta”). Si no dispone de resultados de extracción, deberá realizar una extracción para empezar a trabajar. [Si desea obtener más información, consulte el tema Extracción de datos el p. 82.](#)

Figura 5-1  
Panel Resultados extraídos antes y después de la extracción



Si el panel Resultados extraídos está vacío o desactualizado, aparece de color amarillo. Pulse en el botón Extraer para iniciar el proceso de extracción. Después de la extracción, puede ver los resultados seleccionando lo que desee ver en la lista desplegable.

Figura 5-2  
Lista desplegable del panel Resultados extraídos



Los conceptos, los tipos y los patrones TLA que se han extraído se conocen colectivamente con el nombre de **resultados de extracción**, y actúan como los descriptores, o los cimientos, de las categorías. También puede utilizar conceptos, tipos y patrones en las reglas de categoría. Además, las técnicas automáticas utilizan conceptos y tipos para generar categorías.

El análisis de texto es un proceso repetitivo en el que los resultados de la extracción se revisan de acuerdo con el contexto de los datos de texto, se ajustan para generar resultados nuevos y luego se reevalúan. Después de la extracción debe revisar los resultados y realizar los cambios que considere necesarios modificando los recursos lingüísticos. Puede ajustar los recursos, en parte, directamente desde el panel Resultados extraídos o el panel de Datos a través de los menús contextuales. [Si desea obtener más información, consulte el tema Refinamiento de los resultados de la extracción el p. 86.](#) También puede hacerlo directamente en la vista Editor de

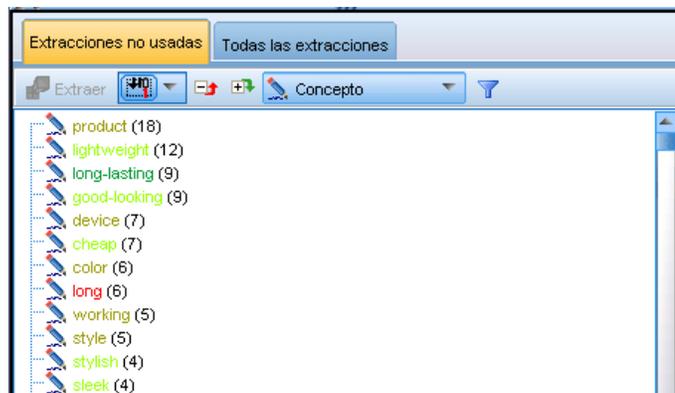
recursos. [Si desea obtener más información, consulte el tema La ventana Editor de recursos en el capítulo 2 el p. 15.](#)

Después del ajuste, puede repetir la extracción para ver los nuevos resultados. El ajuste preciso desde el principio de los resultados de la extracción permite garantizar que la próxima vez que vuelva a realizar la extracción, obtendrá resultados idénticos en las definiciones de su categoría, perfectamente adaptados al contexto de los datos. De esta forma, registros se asignarán a las definiciones de categoría de una manera más precisa y repetible.

### Conceptos

Durante el proceso de extracción, los datos de texto se exploran y se analizan para poder identificar las palabras simples relevantes o interesantes (como *elección* o *paz*) y frases (como *elección presidencial*, *elección del presidente* o *tratados de paz*) en el texto. Estas palabras y frases se conocen colectivamente con el nombre de *términos*. Mediante los recursos lingüísticos, los términos relevantes se extraen y, a continuación, los términos similares se agrupan bajo un término principal llamado **concepto**.

Figura 5-3  
Panel Resultados extraídos: Vista Concepto



Por defecto, los conceptos se muestran en minúscula y en orden descendente según el número de registros donde se encuentra el concepto. Cuando se extraen conceptos, se les asigna un tipo para contribuir a agrupar conceptos similares. Están codificados por colores según este tipo. Los colores están definidos en las propiedades de tipo en el Editor de recursos. [Si desea obtener más información, consulte el tema Diccionarios de tipo en el capítulo 10 el p. 215.](#)

Cuando se utiliza un concepto, tipo o patrón en una definición de categoría, aparece en cursiva en la tabla.

### Tipos

**Tipos** son agrupaciones semánticas de conceptos. Cuando se extraen conceptos, se les asigna un tipo para contribuir a agrupar conceptos similares. Se entregan varios tipos incorporados con IBM® SPSS® Text Analytics for Surveys, como <Location>, <Organization>, <Person>, <Positive>, <Negative>, etc. Por ejemplo, el tipo <Location> agrupa lugares y palabras clave geográficas. Este tipo se asigna a conceptos como *chicago*, *parís* y *tokio*. conceptos que no se encuentran en ningún diccionario de tipo pero que se extraen del texto se tipifican

automáticamente como <Unknown>. Si desea obtener más información, consulte el tema Tipos incorporados en el capítulo 10 el p. 216.

Cuando selecciona la vista Tipo, los tipos extraídos aparecen por defecto en orden descendente por frecuencia. Cuando se expande el árbol, se ven los conceptos que se extrajeron para dicho tipo. Comprobará también que estos tipos están codificados por colores para que sea más fácil distinguirlos. Los colores forman parte de las propiedades de tipo. Si desea obtener más información, consulte el tema Creación de tipos en el capítulo 10 el p. 217. También puede crear sus propios tipos.

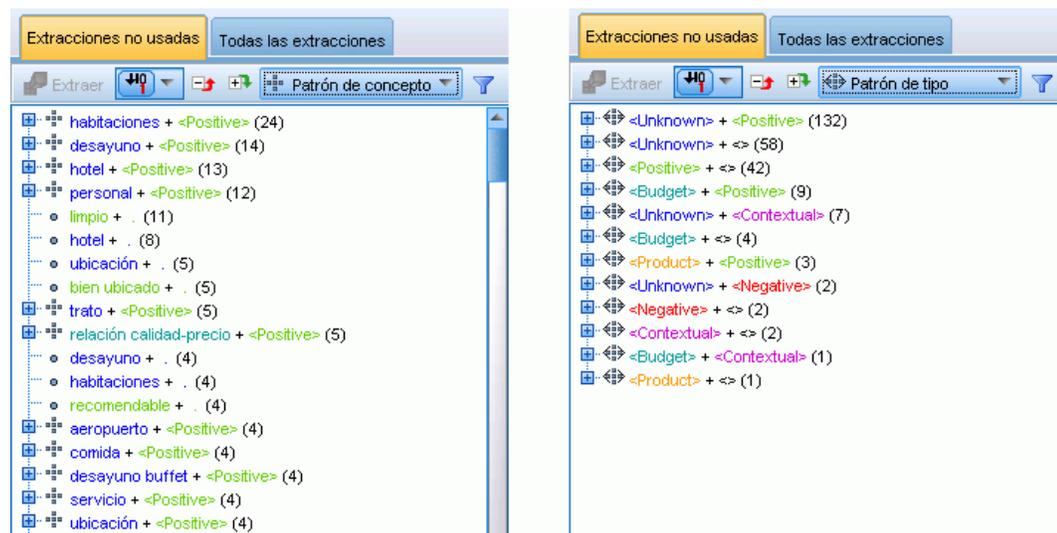
Figura 5-4  
Panel Resultados extraídos: Vista Tipo



## Patrones

Los patrones están compuestos de dos partes: una combinación de conceptos y tipos incorporados que representan cualificadores y adjetivos. Los patrones son muy útiles cuando se intenta detectar opiniones sobre un asunto en particular. Conseguir el nombre del producto de la competencia puede que no sea lo suficientemente interesante. En ese caso, puede comprobar los patrones extraídos por si pudiera encontrar ejemplos donde los encuestados consideraron el producto bueno, malo o caro. Existen dos vistas de patrones diferentes: Patrón de concepto y Patrón de tipo.

Figura 5-5  
Panel Resultados extraídos: La vista Patrón de concepto y la vista Patrón de tipo



**Patrón de concepto.** En esta vista, el nivel superior del árbol del panel Resultados extraídos muestra los patrones con la estructura siguiente: `concepto1 + <Tipo1>` para patrones de conceptos, como análisis de texto + `<Positive>` o coste + `<Negative>`. Cuando se expande el árbol, puede ver los patrones exactos, como análisis de texto + potente o coste + caro.

Los patrones también pueden resultar significativos cuando existen sin la segunda parte del patrón. Por ejemplo, puede querer encontrar situaciones en las que el encuestado no expresó una opinión positiva ni negativa sobre el asunto. En ese caso, esto se representa como `concepto1 + .`, donde `.` designa un cualificador nulo. Por ejemplo, si un encuestado contesta “Coste, ubicación de la tienda” al responder a la pregunta “¿Qué factores tiene en cuenta a la hora de optar por un reproductor de música?”, la extracción podría generar `coste + .` y `ubicación_de_la_tienda + .` como patrones nulos. Cuando se muestran los patrones, se asignan colores a cada elemento del patrón en función de su tipo.

**Patrón de tipo.** En esta vista, el nivel superior del árbol del panel Resultados extraídos muestra los patrones con la estructura siguiente: `<Tipo> + <Tipo>`, como `<Budget> + <Positive>`. Si expande aún más el árbol, verá las relaciones tal como se describen y presentan en la vista Patrón de concepto. Cuando se muestran los patrones, se asignan colores a cada elemento del patrón en función de su tipo.

### ***Pestañas Extracciones no usadas y Todas las extracciones***

El panel Resultados extraídos muestra los resultados del proceso de extracción. Cuando empieza a crear categorías, algunos de sus conceptos extraídos (conceptos, tipos y patrones) se convierten en parte de los descriptores de categoría. Por esta razón, SPSS Text Analytics for Surveys presenta esta información de dos maneras mediante pestañas. Puede pasar de un elemento a otro alternativamente, los elementos que ya se han utilizado en las definiciones de categoría y el conjunto completo de conceptos extraídos. Puede hacerlo pulsando en las pestañas Extracciones no usadas y Todas las extracciones. La pestaña Extracciones no usadas muestra todos los elementos que actualmente no forman parte de un descriptor de categoría. La pestaña Todas las extracciones muestra todos los elementos extraídos con los elementos utilizados en cursiva.

## ***Extracción de datos***

Cuando se necesita una extracción, el panel Resultados extraídos aparece de color amarillo y se muestra el mensaje Pulse el botón Extraer para extraer conceptos debajo de la barra de herramientas de este panel.

Puede que necesite realizar una extracción si todavía no tiene resultados de extracción, si no ha realizado cambios en los recursos lingüísticos y necesita actualizar los resultados de la extracción, o si ha reabierto una sesión proyecto donde no ha guardado los resultados de la extracción (Herramientas > Opciones).

*Nota:* Si cambia el nodo de origen de la ruta después de haber almacenado en caché los resultados de la extracción con la opción Utilizar trabajo de sesión..., para obtener resultados de extracción actualizados necesitará ejecutar una nueva extracción una vez que se inicie la sesión de trabajo interactiva.

Cuando ejecuta una extracción, aparece un indicador de progreso que ofrece información sobre el estado de la extracción. Durante este tiempo, el motor de extracción lee todos los datos de texto e identifica los términos y patrones relevantes, los extrae y les asigna un tipo. A continuación, el motor intenta agrupar los términos sinónimos bajo un término principal, llamado concepto. Cuando finaliza el proceso, los conceptos, tipos y patrones resultantes aparecen en el panel Resultados extraídos. Puede empezar a trabajar con los resultados, y revisarlos.

*Nota:* Existe una relación entre el tamaño del conjunto de datos y el tiempo que se tarda en completar el proceso de extracción. Consulte las instrucciones de instalación para conocer las recomendaciones y las estadísticas de rendimiento.

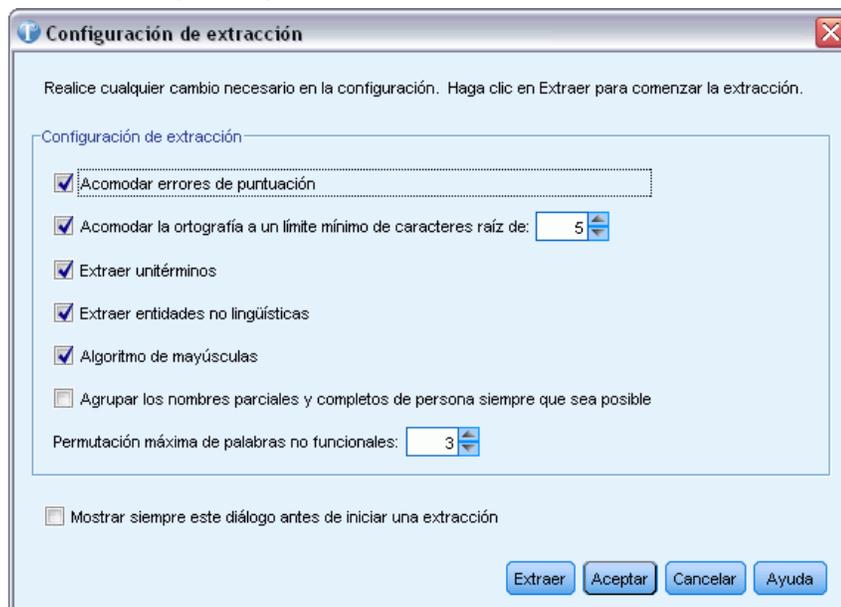
### **Para extraer datos**

- ▶ En los menús elija Herramientas > Extraer. Como alternativa, pulse en el botón de la barra de herramientas Extraer.
- ▶ Si opta por que siempre se muestre el diálogo Configuración de extracción, aparecerá para que pueda realizar cambios. Más adelante en este tema encontrará los descriptores de cada valor de configuración.
- ▶ Pulse en Extraer para empezar el proceso de extracción. Una vez se inicia la extracción, se abre el cuadro de diálogo de progreso. Después de la extracción, los resultados aparecen en el panel Resultados extraídos. Por defecto, los conceptos se muestran en minúscula y en orden descendente según el número de registros donde se encuentra el concepto.

Puede repasar los resultados utilizando las opciones de la barra de herramientas y establecer los resultados con un orden diferente, o cambiar de vista (conceptos, patrones o tipos). También puede refinar los resultados de la extracción trabajando con los recursos lingüísticos. [Si desea obtener más información, consulte el tema Refinamiento de los resultados de la extracción el p. 86.](#)

El cuadro de diálogo Configuración de extracción contiene algunas opciones básicas de extracción.

Figura 5-6  
Cuadro de diálogo Configuración de extracción



**Acomodar errores de puntuación.** Esta opción normaliza temporalmente el texto que contiene errores de puntuación (por ejemplo, el uso inapropiado) durante la extracción para mejorar la capacidad de extracción de los conceptos. Esta opción es extremadamente útil cuando el texto es breve y de calidad mediocre (como, por ejemplo, en respuestas de encuestas con final abierto, correo electrónico y datos CRM) o cuando el texto contiene muchas abreviaturas.

**Acomodar la ortografía a un límite mínimo de caracteres raíz de [n].** Esta opción aplica una técnica de agrupación difusa que ayuda a agrupar bajo un concepto las palabras que suelen escribirse mal o que tienen una ortografía parecida. El algoritmo de agrupación difusa elimina temporalmente todas las vocales (excepto la primera) y las consonantes dobles o triples de las palabras extraídas, y luego las compara para comprobar si son las mismas; en este caso *modelado* y *modulado* se agruparían juntas. Sin embargo, si a cada término se le asigna un tipo diferente, excluyendo el tipo <Unknown>, la técnica de agrupación difusa no se aplicará.

También puede definir el número mínimo de caracteres *raíz* necesarios para poder utilizar la agrupación difusa. El número de caracteres raíz de un término se calcula sumando todos los caracteres y restando los que forman los sufijos de las declinaciones, y en el caso de términos de palabras compuestas, también los determinantes y las preposiciones. Por ejemplo, el término *ejercicios* se considera que tiene 9 caracteres raíz con la forma “ejercicio,” puesto que la letra *s* al final de la palabra es una declinación (forma plural). Asimismo, *agua natural* tiene 11 caracteres raíz (“agua natural”) y *fabricación de coches* tiene 16 caracteres raíz (“fabricación coche”). Este método de recuento de caracteres solo se utiliza para comprobar si debe aplicarse la agrupación difusa, pero no influye en la forma de coincidencia de las palabras.

*Nota:* Si considera que algunas palabras se agrupan incorrectamente, puede excluir parejas de palabras de esta técnica indicándolo explícitamente en la sección Agrupación difusa: Excepciones en la pestaña Recursos avanzados. [Si desea obtener más información, consulte el tema Agrupación difusa en el capítulo 11 el p. 236.](#)

**Extraer unitérminos.** Esta opción extrae palabras simples (unitérminos) siempre que la palabra no forme parte de una palabra compuesta, y si es un sustantivo o una categoría léxica no reconocida.

**Extraer entidades no lingüísticas.** Esta opción extrae entidades no lingüísticas, como números de teléfono, números de la seguridad social, horas, fechas, monedas, dígitos, porcentajes, direcciones de correo electrónico y direcciones de HTTP. Puede incluir o excluir determinados tipos de entidades no lingüísticas en la sección Entidades no lingüísticas: Configuración en la pestaña Recursos avanzados. Si se desactivan las entidades innecesarias, el motor de extracción no malgastará tiempo de proceso. [Si desea obtener más información, consulte el tema Configuración en el capítulo 11 el p. 241.](#)

**Algoritmo de mayúsculas.** Esta opción extrae términos simples y compuestos que no están en los diccionarios incorporados, siempre que la primera letra del término esté en mayúscula. Esta opción supone un buen método para extraer la mayoría de los nombres propios.

**Agrupar los nombres parciales y completos de persona siempre que sea posible.** Esta opción agrupa nombres que aparecen de diferente manera juntos en el texto. Esta función es útil porque a menudo se hace referencia a los nombres completos al principio del texto, y más adelante se utiliza la versión abreviada. Esta opción intenta hacer coincidir cualquier unitérmino que tenga el tipo <Unknown> con la última palabra de cualquier término compuesto que se haya tipificado como <Person>. Por ejemplo, si se encuentra *garcía*, que inicialmente se tipificó como <Unknown>, el motor de extracción comprobará si hay algún término compuesto en el tipo <Person> con el término *garcía* como la última palabra, como en *juan garcía*. Esta opción no se aplica a los nombres propios, porque muchos de ellos no se extraen nunca como unitérminos.

**Permutación máxima de palabras no funcionales.** Esta opción especifica el número máximo de palabras no funcionales que debe haber para poder aplicar la técnica de permutación. Esta técnica de permutación agrupa frases similares que difieren entre sí solo en las palabras no funcionales (por ejemplo, de y el), independientemente de la flexión. Por ejemplo, supongamos que define este valor con al menos dos palabras, y se ha extraído tanto *conductor de autobús* como *el conductor del autobús*. En este caso, los dos términos extraídos se agruparían juntos en la lista de conceptos finales, puesto que ambos términos se consideran el mismo si se pasan por alto las palabras *el del*.

**Mostrar siempre este diálogo antes de iniciar una extracción.** Especifique si desea ver el diálogo Configuración de extracción cada vez que realice una extracción, si no desea verlo nunca a menos que vaya al menú Herramientas, o si desea que cada vez que realice una extracción se le pregunte si desea editar algún valor de la configuración de extracción.

## ***Cómo guardar los resultados de la extracción***

Siempre que realice una extracción, los resultados aparecen en el panel Resultados extraídos y pueden utilizarse para categorizar las respuestas. Durante una sesión de IBM® SPSS® Text Analytics for Surveys, estos resultados de la extracción se guardan en la memoria para que pueda trabajar con ellos. Por defecto, los resultados de la extracción se guardan en los proyectos. Si lo desea, puede guardar estos resultados al cerrar el proyecto mediante un valor de la configuración global que puede cambiar en cualquier momento en el cuadro de diálogo Opciones (Herramientas > Opciones). [Si desea obtener más información, consulte el tema Opciones: pestaña Sistema en el capítulo 2 el p. 18.](#)

Como medida de seguridad, se realiza un proceso de cifrado de dichos resultados de extracción durante el proceso de guardado, y se colocan en la base de datos. Este procedimiento dificulta la posibilidad de que otro usuario vea los datos de la base de datos. Además, los resultados de la extracción nunca se presentarán en SPSS Text Analytics for Surveys hasta que la aplicación haya localizado el origen de datos. De esta manera, si los datos están protegidos por contraseña, el usuario debe especificar el nombre de usuario y la contraseña correspondiente al origen de los datos para que los resultados de la extracción aparezcan en pantalla.

## ***Refinamiento de los resultados de la extracción***

La extracción constituye un proceso repetitivo por el que puede realizar una extracción, revisar los resultados, realizar cambios y volver a realizar la extracción para actualizar los resultados. Puesto que la precisión y la continuidad son esenciales para una minería y categorización de texto exitosos, refinar los resultados de la extracción desde el principio garantiza que cada vez que realice la extracción, obtendrá con exactitud los mismos resultados en las definiciones de categoría. De esta forma, los registros se asignarán a las categorías de una manera más precisa y repetible.

Los resultados de la extracción constituyen los cimientos de las categorías. Cuando crea categorías utilizando estos resultados de extracción, los registros se asignan automáticamente a categorías si contienen texto que coincida con uno o más descriptores de categoría. Aunque empiece a categorizar antes de realizar ajustes en los recursos lingüísticos, conviene revisar los resultados de la extracción al menos una vez antes de comenzar.

Mientras revisa los resultados puede encontrar elementos que desearía que el motor de extracción gestionara de otra forma. Observe los ejemplos siguientes:

- **Sinónimos no reconocidos.** Supongamos que encuentra varios conceptos que considera sinónimos, como *listo*, *inteligente*, *brillante* y *erudito*, y que todos ellos aparecen como conceptos individuales en los resultados extraídos. Puede crear una definición de sinónimo donde se agrupen los términos *inteligente*, *brillante* y *erudito* bajo el concepto objetivo *listo*. Al hacerlo, se agruparían todos estos conceptos con *listo*, y la frecuencia global también sería mayor. [Si desea obtener más información, consulte el tema Adición de sinónimos el p. 87.](#)
- **Conceptos con tipos incorrectos.** Supongamos que los conceptos de los resultados extraídos aparecen en un tipo y desea que se asignen a otro distinto. Imagine, en otro ejemplo, que encuentra 15 conceptos de verduras en los resultados extraídos y desea que todos ellos se añadan a un nuevo tipo llamado `<Verduras>`. Conceptos que no se encuentran en ningún diccionario de tipo pero que se extraen del texto se tipifican automáticamente como `<Unknown>`. Puede añadir los conceptos a los tipos. [Si desea obtener más información, consulte el tema Adición de conceptos a tipos el p. 89.](#)
- **Conceptos insignificantes.** Supongamos que encuentra un concepto extraído y que tiene una frecuencia muy alta, es decir, que aparece en muchos registros. Sin embargo, considera que es un concepto sin trascendencia para su análisis. Puede excluirlo de la extracción. [Si desea obtener más información, consulte el tema Exclusión de conceptos de la extracción el p. 91.](#)
- **Coincidencias incorrectas.** Supongamos que, al revisar los registros que contienen un concepto determinado, descubre que ha habido dos palabras que se han agrupado juntas indebidamente, como *facultad* y *facilidad*. Esta coincidencia puede deberse a un algoritmo interno, al que se conoce como agrupación difusa, que pasa por alto provisionalmente las vocales y las consonantes duplicadas o triplicadas con el objeto de agrupar errores ortográficos comunes.

Puede añadir estas palabras a una lista de parejas de palabras que no deben agruparse. [Si desea obtener más información, consulte el tema Agrupación difusa en el capítulo 11 el p. 236.](#)

- **Conceptos no extraídos.** Supongamos que espera encontrar determinados conceptos extraídos, pero al revisar el texto del registro detecta que algunas palabras o frases no se han extraído. A menudo, este tipo de palabras son verbos o adjetivos en los que no se suele estar interesado. Sin embargo, puede que quiera utilizar una palabra o frase que no se haya extraído como parte de una definición de categoría. Para extraer estos conceptos, puede forzar un término en un diccionario de tipo. [Si desea obtener más información, consulte el tema Forzado de palabras en la extracción el p. 92.](#)

Muchos de estos cambios pueden realizarse directamente desde el panel Resultados extraídos o el panel Datos seleccionando uno o más elementos y pulsando con el botón derecho del ratón para acceder a los menús contextuales.

Una vez realizados los cambios pertinentes, el color de fondo del panel cambia para indicar que debe repetir la extracción para ver los cambios. [Si desea obtener más información, consulte el tema Extracción de datos el p. 82.](#) Si está trabajando con conjuntos de datos voluminosos, puede resultar más eficaz repetir la extracción después de realizar varios cambios en lugar de hacerlo después de cada cambio.

*Nota:* Puede ver el conjunto completo de recursos lingüísticos editables que se utilizan para generar los resultados de la extracción en la vista Editor de recursos (Ver > Editor de recursos). Estos recursos aparecen en esta vista en forma de bibliotecas y diccionarios. Puede personalizar los conceptos y los tipos directamente en las bibliotecas y en los diccionarios. [Si desea obtener más información, consulte el tema Trabajo con bibliotecas en el capítulo 9 el p. 202.](#)

## **Adición de sinónimos**

Los **Sinónimos** asocian dos o más palabras con el mismo significado. Los sinónimos se utilizan a menudo para agrupar términos con sus abreviaturas, o para agrupar palabras que suelen escribirse mal con la ortografía correcta. Con el uso de los sinónimos, la frecuencia del concepto objetivo es mayor, lo cual facilita mucho la detección de información similar que se presenta de distintas formas en los datos de texto.

Las bibliotecas y las plantillas de los recursos lingüísticos que se proporcionan con el producto contienen muchos sinónimos predefinidos. Sin embargo, si detecta sinónimos que no se reconocen, puede definirlos para que la próxima vez que realice una extracción puedan detectarse.

El primer paso es decidir cuál será el concepto objetivo. El **concepto objetivo** es aquel bajo el que desea agrupar todos los términos sinónimos en los resultados finales. Durante la extracción, los sinónimos se agrupan bajo este concepto objetivo. El segundo paso es identificar todos los sinónimos de este concepto. El concepto objetivo se sustituye por todos sus sinónimos en la extracción final. Para que un término sea considerado sinónimo, debe estar extraído. Sin embargo, no es necesario extraer el concepto objetivo para que se produzca la sustitución. Por ejemplo, si desea que *inteligente* se sustituya por *listo*, el término *inteligente* se considera sinónimo y *listo* se considera concepto objetivo.

Si crea una nueva definición de sinónimo, se añade un nuevo concepto objetivo al diccionario. A continuación debe añadir sinónimos al concepto objetivo. Siempre que cree o edite sinónimos, estos cambios quedarán registrados en los diccionarios de sinónimos del Editor de recursos. Si desea ver el contenido completo de estos diccionarios de sinónimos o si desea realizar un número

importante de cambios, puede trabajar directamente en Editor de recursos. [Si desea obtener más información, consulte el tema Diccionarios de sustitución/sinónimos en el capítulo 10 el p. 225.](#)

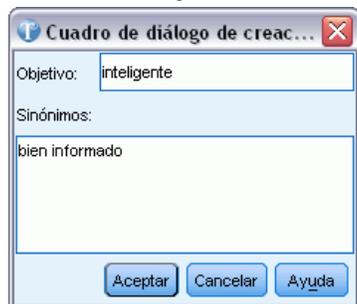
Los sinónimos nuevos se guardarán automáticamente en la primera biblioteca del árbol de bibliotecas de la Editor de recursos vista—por defecto, es la *Biblioteca local*.

*Nota:* Si busca una definición de sinónimo y no puede encontrarla en los menús contextuales o directamente en Editor de recursos, es posible que se produzca una coincidencia con la técnica de agrupación difusa interna. [Si desea obtener más información, consulte el tema Agrupación difusa en el capítulo 11 el p. 236.](#)

### **Para crear un sinónimo nuevo**

- ▶ En el panel Resultados extraídos o en el panel Datos, seleccione el concepto o conceptos para los que desea crear un sinónimo nuevo.
- ▶ Desde los menús, seleccione Edición > Añadir a sinónimo > Nuevo. Aparecerá el cuadro de diálogo Crear sinónimo.

Figura 5-7  
Cuadro de diálogo Crear sinónimo



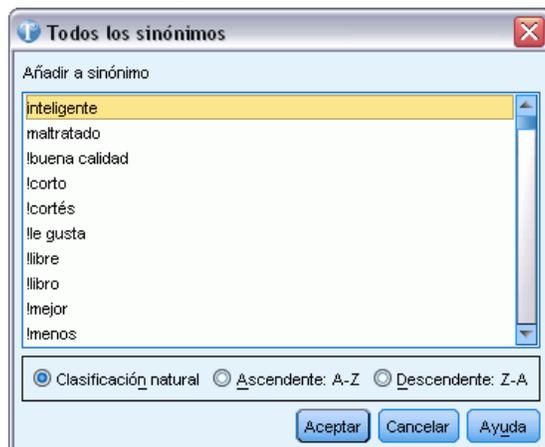
- ▶ Escriba un concepto objetivo en el cuadro de texto Objetivo. Se trata del objetivo bajo el que se agruparán todos los sinónimos.
- ▶ Si desea añadir más sinónimos, escríbalos en el cuadro de lista Sinónimos. Utilice el separador global para separar cada término de sinónimo. [Si desea obtener más información, consulte el tema Opciones: pestaña Sistema en el capítulo 2 el p. 18.](#)
- ▶ Pulse en Aceptar para aplicar los cambios. El cuadro de diálogo se cierra y el color de fondo del panel Resultados extraídos cambia para indicar que debe repetir la extracción para ver los cambios. Si tiene previstos varios cambios, realícelos antes de repetir la extracción.

### **Para añadir un sinónimo**

- ▶ En el panel Resultados extraídos o en el panel Datos, seleccione el concepto o conceptos para los que desea añadir una definición de sinónimo existente.
- ▶ Desde los menús, seleccione Edición > Añadir a sinónimo > . El menú muestra un conjunto de sinónimos, y el creado más recientemente figura al principio de la lista. Seleccione el nombre del sinónimo al que desea añadir los conceptos seleccionados. Si encuentra el sinónimo que está buscando, selecciónelo; acto seguido los conceptos seleccionados se añaden a dicha definición

de sinónimo. Si no lo encuentra, seleccione Más para mostrar el cuadro de diálogo Todos los sinónimos.

Figura 5-8  
Cuadro de diálogo Todos los sinónimos



- En el cuadro de diálogo Todos los sinónimos, puede ordenar la lista por orden de clasificación natural (orden de creación) o en sentido ascendente o descendente. Seleccione el nombre del sinónimo al que desea añadir los conceptos seleccionados y pulse en Aceptar. El cuadro de diálogo se cierra y los conceptos se añaden a la definición de sinónimos.

## Adición de conceptos a tipos

Cuando se ejecuta una extracción, los conceptos extraídos se asignan a tipos determinados con la pretensión de agrupar los términos que tienen algo en común. IBM® SPSS® Text Analytics for Surveys se proporciona con muchos tipos incorporados. [Si desea obtener más información, consulte el tema Tipos incorporados en el capítulo 10 el p. 216.](#) conceptos que no se encuentran en ningún diccionario de tipo pero que se extraen del texto se tipifican automáticamente como <Unknown>.

Cuando revise los resultados, es posible que encuentre algunos conceptos en un tipo que en realidad desea que se asigne a otro concepto, o un grupo de palabras que en realidad pertenece a un nuevo tipo por sí mismo. En estos casos, puede reasignar estos conceptos a otro tipo o crear un tipo nuevo.

Por ejemplo, supongamos que está trabajando con los datos de una encuesta relacionada con el mundo automovilístico, y le interesa que la categorización se centre en diferentes ámbitos de los vehículos. Puede crear un tipo llamado <Salpicadero> para agrupar todos los conceptos relacionados con los contadores y los botones que suelen encontrarse en el tablero de instrumentos de los vehículos. A continuación puede asignar conceptos como medidor de combustible, radiador, radio y cuentakilómetros a este nuevo tipo.

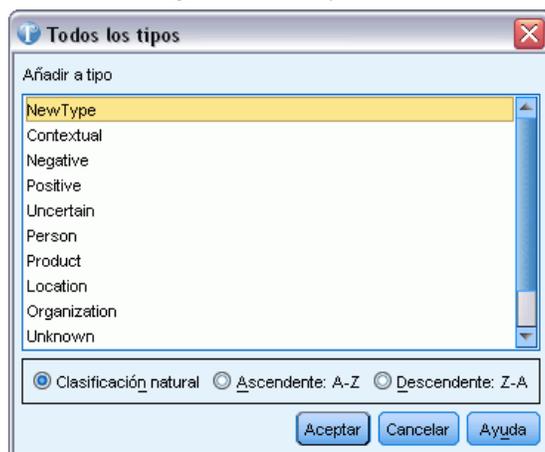
En otro ejemplo, supongamos que está trabajando con los datos de una encuesta relacionada con el ámbito universitario, y en la extracción se ha categorizado el término Ramon Llull (la universidad) como un tipo de <Person> en lugar de hacerlo como un tipo <Organization>. En este caso, puede añadir este concepto al tipo <Organization>.

Siempre que se crea un tipo o se añaden conceptos a un tipo en una lista de términos del tipo, estos cambios quedan registrados en los diccionarios de tipo en las bibliotecas de recursos lingüísticos de Editor de recursos. Si desea ver el contenido de estas bibliotecas o si desea realizar un número importante de cambios, puede trabajar directamente en Editor de recursos. [Si desea obtener más información, consulte el tema Adición de términos en el capítulo 10 el p. 219.](#)

### **Para añadir un concepto a un tipo**

- ▶ En el panel Resultados extraídos o en el panel Datos, seleccione los conceptos que desea añadir a un tipo existente.
- ▶ Pulse el botón derecho del ratón para abrir el menú contextual.
- ▶ Desde los menús, seleccione Edición > Añadir a tipo >. El menú muestra un conjunto de tipos, y el creado más recientemente figura al principio de la lista. Seleccione el nombre del tipo al que desea añadir los conceptos seleccionados. Si encuentra el nombre del tipo que está buscando, selecciónelo; acto seguido los conceptos seleccionados se añaden a dicho tipo. Si no lo encuentra, seleccione Más para mostrar el cuadro de diálogo Todos los tipos.

Figura 5-9  
Cuadro de diálogo Todos los tipos

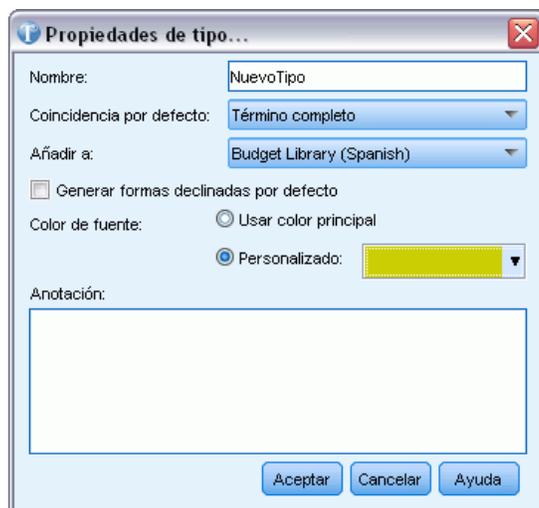


- ▶ En el cuadro de diálogo Todos los tipos, puede ordenar la lista por orden de clasificación natural (orden de creación) o en sentido ascendente o descendente. Seleccione el nombre del tipo al que desea añadir los conceptos seleccionados y pulse en Aceptar. El cuadro de diálogo se cierra y se añadirán como términos al tipo.

### **Para crear un tipo nuevo**

- ▶ En el panel Resultados extraídos o en el panel Datos, seleccione los conceptos para los que desea crear un tipo nuevo.
- ▶ Desde los menús, seleccione Edición > Añadir a tipo > Nuevo. Aparecerá el cuadro de diálogo Propiedades de tipo.

Figura 5-10  
Cuadro de diálogo *Propiedades de tipo*



- ▶ Escriba un nombre nuevo para este tipo en el cuadro de texto Nombre y realice los cambios necesarios en el resto de los campos. [Si desea obtener más información, consulte el tema Creación de tipos en el capítulo 10 el p. 217.](#)
- ▶ Pulse en Aceptar para aplicar los cambios. El cuadro de diálogo se cierra y el color de fondo del panel Resultados extraídos cambia para indicar que debe repetir la extracción para ver los cambios. Si tiene previstos varios cambios, realícelos antes de repetir la extracción.

### ***Exclusión de conceptos de la extracción***

Al revisar los resultados puede encontrar, ocasionalmente, conceptos que no deseaba que se extrajeran ni que se utilizaran en ninguna técnica automática de generación de categorías. En algunos casos, estos conceptos tienen una frecuencia muy alta y son totalmente intrascendentes en su análisis. En ese caso, puede marcar un concepto para que se excluya de la extracción final. Por lo general, los conceptos que añade a esta lista serán palabras o frases de relleno que se utilizan en el texto para conferir continuidad, pero que no aportan información relevante y que pueden cargar innecesariamente los resultados de la extracción. Si añade conceptos al diccionario de exclusión, tendrá la seguridad de que no se extraerán nunca.

Al excluir estos conceptos, la próxima vez que realice la extracción desaparecerán de los resultados de la misma todas las variaciones de los conceptos excluidos. Si el concepto sigue apareciendo como descriptor en una categoría, permanecerá en ella con el indicador cero después de una reextracción.

Cuando realiza la exclusión, estos cambios quedan registrados en un diccionario de exclusión en el Editor de recursos. Si desea ver todas las definiciones excluidas y editarlas, puede trabajar directamente en Editor de recursos. [Si desea obtener más información, consulte el tema Diccionarios de exclusión en el capítulo 10 el p. 231.](#)

**Para excluir conceptos**

- ▶ En el panel Resultados extraídos o en el panel Datos, seleccione los conceptos que desea excluir de la extracción.
- ▶ Pulse el botón derecho del ratón para abrir el menú contextual.
- ▶ Seleccione Excluir de extracción. El concepto se añade al diccionario de exclusión de Editor de recursos y el color de fondo del panel Resultados extraídos cambia para indicar que debe repetir la extracción para ver los cambios. Si tiene previstos varios cambios, realícelos antes de repetir la extracción.

*Nota:* Las palabras que excluya se guardarán automáticamente en la primera biblioteca del árbol de bibliotecas en el Editor de recursos—por defecto, es la *Biblioteca local*.

**Forzado de palabras en la extracción**

Al revisar los datos de texto en el panel Datos después de la extracción, puede encontrar que algunas palabras o frases no se han extraído. A menudo, este tipo de palabras son verbos o adjetivos en los que no se suele estar interesado. Sin embargo, puede que quiera utilizar una palabra o frase que no se haya extraído como parte de una definición de categoría.

Si desea que se extraigan estas palabras o frases, tiene dos opciones:

- Forzar un término dentro de una biblioteca de tipos. [Si desea obtener más información, consulte el tema Forzado de términos en el capítulo 10 el p. 222.](#)
- Añadir las palabras directamente a una definición de categoría existente. Esta alternativa suele utilizarse si la primera opción no proporciona los resultados esperados. [Si desea obtener más información, consulte el tema Coincidencia de texto en las categorías en el capítulo 6 el p. 162.](#)

**Importante:** Marcar un término en el diccionario como forzado no es infalible. Esto significa que aunque haya añadido un término explícitamente en un diccionario, es posible que no siempre esté presente en el panel Resultados extraídos después de haber repetido la extracción, o puede que aparezca pero no exactamente como lo especificó. Aunque esta circunstancia es rara, puede ocurrir cuando una palabra o frase ya se había extraído como parte de una frase más larga. Para evitarlo, aplique la opción Completo (sin compuestos) a este término en el diccionario de tipo. [Si desea obtener más información, consulte el tema Adición de términos en el capítulo 10 el p. 219.](#)

# ***Categorización de los datos de texto***

En IBM® SPSS® Text Analytics for Surveys puede crear **categorías** que representan, en esencia, temas o conceptos de máximo nivel que captan las ideas, conocimientos y actitudes clave expresados en el texto.

Las categorías también pueden tener una estructura jerárquica, lo que significa que pueden contener subcategorías y estas subcategorías pueden también tener sus propias subcategorías y así sucesivamente. Puede importar estructuras de categorías predefinidas, antes denominadas marcos de código, con categorías jerárquicas así como crear estas categorías jerárquicas en el producto.

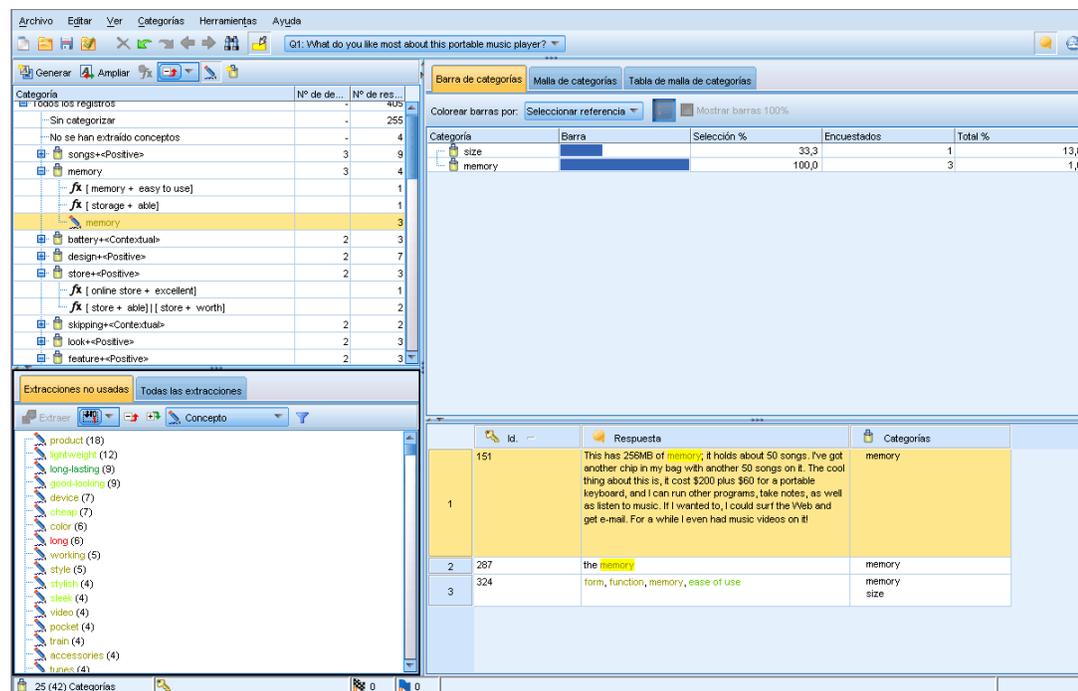
De hecho, las categorías jerárquicas le permiten crear una estructura de árbol con una o más subcategorías para agrupar elementos como áreas de temas o conceptos diferentes de manera más precisa. Un ejemplo sencillo puede estar relacionado con actividades de ocio; al responder a una pregunta como *¿Qué actividad le gustaría hacer si tuviera más tiempo?*, puede tener categorías principales como *deportes*, *actividades artísticas*, *pesca*, etc.; bajando un nivel, debajo de *deportes*, puede tener subcategorías para determinar si son *juegos con pelota*, *deportes acuáticos*, etc.

Las Categorías están formadas por un grupo de descriptores, como *conceptos*, *tipos*, *patrones* y *reglas de categorías*. Todos estos descriptores en conjunción se utilizan para identificar si un registro pertenece a una categoría determinada. El texto de un registro puede ser explorado para comprobar si existe texto que coincida con un descriptor. Si se halla una coincidencia, el registro se asigna a dicha categoría. Este proceso se denomina **categorización**.

Puede trabajar con las categorías, generarlas y explorarlas visualmente mediante los datos que se presentan en los cuatro paneles, cada una de los cuales puede ocultarse o mostrarse seleccionando su nombre en el menú Ver.

- **Panel Categorías.** Genere y administre las categorías en este panel. [Si desea obtener más información, consulte el tema El panel de categorías el p. 94.](#)
- **Panel Resultados extraídos.** Explore y trabaje con los conceptos y tipos extraídos en este panel. [Si desea obtener más información, consulte el tema Resultados extraídos: Conceptos, tipos y patrones en el capítulo 5 el p. 78.](#)
- **Panel Visualización.** Explore visualmente las categorías y compruebe su interacción en este panel. [Si desea obtener más información, consulte el tema Visualización de gráficos en el capítulo 7 el p. 167.](#)
- **Panel Datos.** Explore y revise el texto que contienen los registros que corresponden a las selecciones en este panel. [Si desea obtener más información, consulte el tema El panel de datos el p. 97.](#)

Figura 6-1  
Vista Pregunta



Puede comenzar con un conjunto de categorías a partir de un paquete de análisis de texto (TAP) o importarlo desde un archivo de categoría predefinida, aunque también puede crear el suyo propio. Las categorías pueden crearse automáticamente utilizando el sólido conjunto de técnicas automáticas del producto, que utiliza los resultados de la extracción (conceptos, tipos y patrones) para generar categorías y sus descriptores. Las categorías pueden también crearse manualmente utilizando conocimientos adicionales relacionados con los datos. Puede crear manualmente definiciones de categorías arrastrando los resultados de extracción y soltándolos en las categorías. Puede enriquecer estas categorías o las categorías vacías añadiendo reglas de categoría a una categoría, utilizando sus propias categorías predefinidas, añadiendo una palabra o frase que nunca se haya extraído antes (denominado **coincidencia de texto**), forzando las respuestas directamente en una categoría, o utilizando una combinación de todas estas técnicas.

Cada una de las técnicas y métodos resulta idónea para determinados tipos de datos y situaciones, pero a menudo conviene combinar técnicas en el mismo análisis para capturar el rango completo de registros. Y en el proceso de categorización, puede ver otros cambios que deban realizarse en los recursos lingüísticos.

## El panel de categorías

El panel Categorías es el área donde puede generar y administrar las categorías. Este panel se encuentra en el ángulo superior izquierdo de la vista Pregunta y puede accederse a él desde el menú Ver (Ver > Pregunta > "su\_pregunta"). Después de extraer los conceptos y los tipos de los datos de texto, puede empezar a generar categorías automáticamente utilizando técnicas como

inclusión de conceptos, co-ocurrencia, etc., o bien puede hacerlo manualmente. Si desea obtener más información, consulte el tema [Generación de categorías](#) el p. 110.

Figura 6-2  
Panel Categorías sin categorías y con categorías

Categoría	Nº de desc...	Nº de resp...
Todos los registros	-	200
Sin categorizar	-	200
No se han extraído conceptos	-	200

Categoría	Nº de de...	Nº de res...
Todos los registros	-	405
Sin categorizar	-	282
No se han extraído conceptos	-	4
songs+<Positive>	3	9
fx [ listening to songs + like]		1
fx [ songs + like][ songs + meets needs][ song:		7
drag & dropped songs from their site+<Positive>	1	1
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
fx [ feature + excellent]		2

Cada vez que se crea o actualiza una categoría, los registros se exploran automáticamente para ver si algún texto coincide con un descriptor en una categoría dada. Si se halla una coincidencia, el registro se asigna a dicha categoría. El resultado final es que la mayoría de los registros, si no todos, se asignan a categorías en función de los descriptores de las categorías.

### Tabla de árbol de categorías

La tabla de árbol de este panel presenta el conjunto de categorías, subcategorías y descriptores. El árbol dispone también de algunas columnas que presentan información para cada elemento de árbol. La siguientes columnas pueden estar disponibles para su visualización:

- Código.** Muestra el valor del código para cada categoría. Esta columna está oculta por defecto. Puede mostrar esta columna pulsando con el botón derecho del ratón en la tabla de árbol y seleccionando **Mostrar > Código de categoría**.
- Categoría.** Contiene el árbol de categorías con el nombre de la categoría y subcategorías. Además, si se pulsa el icono de la barra de herramientas de los descriptores, se mostrará el conjunto de descriptores.
- Descriptores.** Proporciona el número de descriptores que componen su definición. Este recuento no incluye el número de descriptores en las subcategorías. No se indica ningún recuento cuando se muestra un nombre de descriptor en la columna **Categorías**. Puede mostrar esta columna pulsando con el botón derecho del ratón en la tabla de árbol y seleccionando **Mostrar > Todos los descriptores**.
- Documentos.** Después de la puntuación, esta columna proporciona el número de registros categorizados en una categoría y todas sus subcategorías. Así, si 5 registros coinciden con su categoría principal en función de sus descriptores, y 7 registros diferentes coinciden con una subcategoría en función de sus descriptores, el recuento total de documentos para la categoría principal es una suma de los dos; en este caso podrían ser 12. Sin embargo, si el mismo registro coincide con la categoría principal y sus subcategorías, el recuento puede ser 11.

Aunque no exista ninguna categoría, la tabla sigue conteniendo dos filas. La fila superior, llamada Todos los registros, es el número total de registros. Una segunda fila, llamada Sin categorizar, muestra el número de documentos/registros que todavía deben categorizarse.

Para cada categoría del panel, hay un pequeño icono amarillo cuadrado delante del nombre de categoría. Si pulsa dos veces en una categoría o pulsa con el botón derecho del ratón en la tabla de árbol y selecciona Definiciones de categoría, se abrirá el cuadro de diálogo Definiciones de categoría y se presentarán todos los elementos, denominados **descriptores**, que conforman su definición, como conceptos, tipos, patrones y reglas de categoría. [Si desea obtener más información, consulte el tema Acerca de las categorías el p. 107](#). Por defecto, la tabla de árbol de categorías no muestra los descriptores en las categorías. Si desea ver los descriptores directamente en el árbol en lugar de en el cuadro de diálogo Definiciones de categorías, pulse en el botón de conmutación con el icono lápiz en la barra de herramientas. Cuando se selecciona el botón de conmutación, puede expandir su árbol para ver los descriptores.

### ***Recuento de categorías***

La columna Documentos en la tabla de árbol de categorías muestra el número de registros categorizados en esa categoría específica. Si los números están fuera de la fecha de caducidad o no están calculados, aparece un icono en esa columna. Tenga en cuenta que el proceso de recuento puede tardar si trabaja con conjuntos de datos grandes.

### ***Selección de categorías en el árbol***

Cuando realice selecciones en el árbol, sólo puede seleccionar categorías hermanas; o sea, si selecciona las categorías de nivel superior, no puede seleccionar una subcategoría. O si selecciona 2 subcategorías de una categoría dada, no puede seleccionar simultáneamente una subcategoría de otra categoría. Seleccionar una categoría que no sea contigua generará la pérdida de la selección anterior.

### ***Mostrar en los paneles de datos y visualización***

Cuando selecciona una fila de la tabla, los paneles Visualización y Datos se actualizan automáticamente con información correspondiente a su selección.

### ***Refinamiento de las categorías***

Es posible que la categorización no arroje un resultado perfecto para sus datos en el primer intento, y bien puede haber categorías que desee eliminar o combinar con otras categorías. También es posible, examinando los resultados de la extracción, que algunas categorías que no haya creado le resulten útiles. En tal caso, puede realizar cambios manuales en los resultados para ajustarlos a su contexto particular. [Si desea obtener más información, consulte el tema Edición y refinamiento de categorías el p. 156](#).

- Editar o añadir elementos a las definiciones de categoría y mover, fusionar o eliminar categorías. Si desea obtener más información, consulte [Edición y refinamiento de categorías a continuación](#).
- Forzar ID de respuesta específicos dentro o fuera de las categorías. [Si desea obtener más información, consulte el tema Forzado de respuestas dentro de categorías el p. 161](#).

- Añada coincidencias de texto a las categorías para capturar respuestas que contengan el mismo texto en la categoría. [Si desea obtener más información, consulte el tema Coincidencia de texto en las categorías el p. 162.](#)
- Añadir reglas de categoría a una categoría para clasificar automáticamente las respuestas en una categoría basada en una expresión lógica. [Si desea obtener más información, consulte el tema Uso de reglas de categoría el p. 145.](#)
- Visualizar de qué forma colaboran las categorías entre sí. [Si desea obtener más información, consulte el tema Visualización de gráficos en el capítulo 7 el p. 167.](#)
- Exportar los resultados de la categorización. [Si desea obtener más información, consulte el tema Exportación de los resultados de la categorización en el capítulo 4 el p. 53.](#)

## ***El panel de datos***

A medida que crea las categorías, en ocasiones puede necesitar revisar parte de los datos del texto con los que está trabajando. Por ejemplo, si crea una categoría donde se categorizan 640 registros, puede ver parte de los registros o todos y comprobar el texto que se ha escrito en realidad. Puede revisar los registros en el panel Datos, que se encuentra abajo a la derecha. Si no está visible por defecto, elija **Ver > Paneles > Datos** en los menús.

Este panel presenta, en formato de tabla, los registros de respuesta de los datos con final abierto. En función de lo que haya seleccionado en los otros paneles de esta vista, en el panel solo aparecerán los registros correspondientes. Por ejemplo, si selecciona un concepto en el panel Resultados extraídos, solo los registros que contengan dicho concepto (y los términos asociados) aparecerán en el panel Datos.

Siempre que seleccione un concepto o una categoría en otro panel y se muestren los datos, los conceptos (y los términos asociados) que se encuentran en dichos registros aparecen resaltados en color para que pueda identificarlos fácilmente en el texto. La codificación por colores corresponde a los tipos a los que pertenecen los conceptos. También puede pasar el ratón por encima de los elementos con codificación por color para mostrar el concepto bajo el que se extrajo y el tipo al que se asignó. El texto que no se ha extraído aparece en negro. Generalmente, estas palabras no extraídas suelen ser conectores (*y* o *con*), pronombres (*mi* o *ellos*), y verbos (*es*, *tiene* o *tomar*).

Figura 6-3  
Panel Datos

	Id	Respuesta	Categorías
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

*Nota:* Para mostrar todos los registros de una pregunta determinada en el panel datos, pulse en el nodo Todos los registros en la parte superior del panel Categorías.

Por defecto, el panel Datos muestra tres columnas (ID, Respuesta y Categorías). Sin embargo, puede añadir columnas adicionales en este panel. Las columnas posibles son las siguientes:

- **ID.** Enumera el registro o el identificador de documento (ID) si se importó alguno.
- **Respuesta.** Enumera los datos de texto de los que se extrajeron conceptos y tipos.
- **Categorías.** Enumera cada una de las categorías a las que pertenece un registro. Cuando se muestra esta columna, la actualización del panel Datos puede tardar un poco más hasta mostrar la información más actualizada. Las categorías se enumeran en esta columna según su relevancia en el registro. [Si desea obtener más información, consulte el tema Relevancia de categoría el p. 99.](#)
- **Forzar dentro.** Enumera las categorías en las que ha forzado una respuesta. Las respuestas pueden forzarse dentro de la categoría a través de la opción de menú Editar > Forzar dentro. [Si desea obtener más información, consulte el tema Forzado de respuestas dentro de categorías el p. 161.](#)
- **Forzar fuera.** Enumera las categorías de las que ha eliminado una respuesta. Las respuestas pueden forzarse fuera de una categoría a través de la opción de menú Editar > Forzar fuera. Generalmente, se utiliza cuando el sarcasmo de un encuestado hace que una respuesta se categorice mal. [Si desea obtener más información, consulte el tema Forzado de respuestas dentro de categorías el p. 161.](#)

- **Coincidencia de texto.** Enumera las coincidencias de texto que se encuentran para cada respuesta. Pueden definirse cadenas de texto para forzar que texto específico forme parte de una definición de categoría independientemente de si dicha cadena se ha extraído o no. [Si desea obtener más información, consulte el tema Coincidencia de texto en las categorías el p. 162.](#)
- **Recuento de categoría.** Proporciona el número total de categorías a la que pertenece la respuesta correspondiente a esta pregunta.
- **Ordenación de relevancia.** Proporciona un orden para cada registro de una categoría. Este orden muestra el grado de adecuación del registro en la categoría si se compara con otros registros de la misma categoría. Seleccione una categoría en el panel Categorías (arriba a la izquierda) para ver el orden en esta columna. [Si desea obtener más información, consulte el tema Relevancia de categoría el p. 99.](#)
- **Marcas de respuesta.** Añade una columna que muestra cualquier marca de respuesta que pueda estar utilizando. Si pulsa en el interior de esta columna, puede cambiar el tipo de marca que asigne a cada respuesta.
- **<cualquier nombres de variable de referencia>.** Añade una columna para la variable de referencia seleccionada en el panel Datos. Si no ha importado ninguna variable de referencia, aquí no se propondrá ninguna. Hay una columna individual disponible para cada variable de referencia. [Si desea obtener más información, consulte el tema Selección de variables en el capítulo 3 el p. 34.](#)

#### **Para mostrar otras columnas del panel Datos**

- ▶ Desde el interior del panel Datos, pulse con el botón derecho del ratón en una cabecera de columna para abrir un menú contextual.
- ▶ En el menú elija Mostrar columnas, y luego seleccione la columna que desea mostrar en el panel Datos. La nueva columna aparece en el panel.

*Nota:* Forzar respuestas dentro y fuera de las categorías permite sustituir las definiciones de categoría creadas por las técnicas automáticas de generación de categorías sin cambiar la definición real de la categoría. [Si desea obtener más información, consulte el tema Forzado de respuestas dentro de categorías el p. 161.](#)

## **Relevancia de categoría**

Para mejorar la generación de categorías, puede revisar la relevancia de los registros de cada categoría, así como la relevancia de todas las categorías a las que pertenece un registro.

#### **Relevancia de una categoría en un registro**

Cuando aparece un registro en el panel Datos, todas las categorías a las que pertenece se muestran en una lista en la columna Categorías. Cuando un registro pertenece a varias categorías, las categorías de esta columna aparecen en orden de mayor a menor relevancia. Se considera que la categoría que se muestra en primer lugar es la que mejor se corresponde con este registro. [Si desea obtener más información, consulte el tema El panel de datos el p. 97.](#)

### Relevancia de un registro en un categoría

Cuando selecciona una categoría, puede revisar la relevancia de cada uno de sus registros en la columna Ordenación de relevancia en el panel Datos. Esta ordenación de relevancia indica el grado de idoneidad con el que el registro se adapta a la categoría seleccionada si se compara con el resto de los registros de dicha categoría. Para ver la ordenación de los registros de una sola categoría, seleccione la categoría en el panel Categorías (situado arriba a la izquierda) y la ordenación del registro aparece en la columna. Esta columna no puede verse de manera predeterminada, pero puede seleccionar la opción para mostrarla. [Si desea obtener más información, consulte el tema El panel de datos el p. 97.](#)

Cuanto menor sea el número del orden del registro, mejor será el ajuste o mayor será la relevancia de este registro para la categoría seleccionada, de tal forma que 1 es el mejor ajuste. Si hay más de un registro con la misma relevancia, cada uno aparece con el mismo grado de ordenación seguido de un signo igual (=) para indicar que tienen la misma relevancia. Por ejemplo, puede tener los valores de ordenación siguientes 1=, 1=, 3, 4, etc.; significa que hay dos registros que se consideran como mejor opción de relación para esta categoría.

**Sugerencia:** Puede añadir el texto del registro más relevante a la anotación de categoría como ayuda para proporcionar una mejor descripción de la categoría. Añada el texto directamente desde el panel Datos seleccionando el texto y eligiendo Categorías > Añadir a anotación en los menús.

Figura 6-4

Panel Datos que muestra la Ordenación de relevancia y de categorías

	Id.	Respuesta	Categorías
1	1	El <b>hotel</b> cumple con todo lo que se le pide: bien ubicado y limpio	hotel
2	7	Servicio de Restaurante ofrecido por el <b>Hotel</b> .	hotel
3	8	El <b>excelente trato del personal</b> de todo el <b>Hotel</b> . La <b>amplitud</b> de las <b>habitaciones</b> y el <b>desayuno</b>	habitaciones desayuno hotel personal
4	22	Fantástico <b>hotel</b> ubicado en un <b>elegante edificio</b>	hotel
5	35	El <b>hotel</b> no está mal, pero en muchos aspectos podría <b>mejorarse</b> .	hotel
6	49	Es un <b>hotel</b> muy agradable, con un <b>jardín</b> muy <b>bonito</b> .	hotel
7	55	<b>comida del hotel</b> <b>excelente</b>	hotel
8	58	la <b>situación</b> ya que el <b>hotel</b> está en el <b>centro</b> y el <b>parking</b>	hotel situación
9	59	El <b>hotel</b> es limpio, silencioso y acogedor.	hotel
10	62	El <b>personal del hotel</b> es muy <b>amable</b>	hotel
11	63	el <b>hotel</b> está fenomenal, muy bien situado	hotel
12	73	El <b>hotel</b> es agradable	hotel
13	75	El <b>hotel</b> está bien comunicado por <b>metro</b> y en una <b>zona</b> tranquila	hotel zona
14	79	La <b>situación del hotel</b> es <b>centrica</b> , <b>cercana</b> a una <b>parada de metro</b> .	hotel parada de metro
15	83	<b>Hotel</b> <b>pequeñito</b> pero muy acogedor.	hotel
16	89	<b>Recomiendo</b> este <b>hotel</b> a <b>parejas</b> que quieran un <b>hotel</b> acogedor y céntrico.	hotel
17	95	muy bueno es el <b>servicio de traslado</b> <b>sin costo</b> <b>aeropuerto-hotel</b>	hotel servicio
18	98	La <b>oferta de restauración</b> y <b>ocio</b> que había dentro del <b>hotel</b>	hotel
	102	Muy <b>próximo</b> al <b>Aeropuerto</b> y con <b>comunicación frecuente</b>	aeropuerto

## ***Métodos y estrategias para crear categorías***

Si todavía no ha realizado la extracción o sus resultados están desactualizados, el uso de una de estas técnicas de ampliación o generación de categorías le indicará la realización de una extracción automática. Una vez aplicada la técnica, los conceptos y tipos que estaban agrupados en una categoría siguen estando disponibles para la generación de categorías mediante otras técnicas. Esto significa que puede ver un concepto en varias categorías a menos que decida no reutilizarlos.

Para ayudarle a crear las mejores categorías, revise lo siguiente:

- **Métodos para crear categorías**
- **Estrategias para crear categorías**
- **Consejos para crear categorías**

### ***Métodos para crear categorías***

Puesto que cada conjunto de datos es exclusivo, el número de métodos para crear categorías y el orden en el que los aplique puede cambiar con el tiempo. Además, puesto que los objetivos de la minería de texto puede diferir de un conjunto de datos a otro, puede que deba experimentar con diferentes métodos para comprobar con cuál de ellos se obtiene el mejor resultado para los datos de texto determinados. Ninguna de las técnicas automáticas categorizará los datos con exactitud, por lo que se recomienda que busque y aplique una o más técnicas automáticas que se adapten bien a sus datos.

Aparte de utilizar paquetes de análisis de texto (TAP, \*.tap) con conjuntos de categorías pregeneradas, también puede categorizar sus respuestas utilizando una combinación de los métodos siguientes:

- **Técnicas de generación automáticas.** Hay disponibles varias opciones de categorías basadas en frecuencia y basadas en lingüística para generar categorías automáticamente. [Si desea obtener más información, consulte el tema Generación de categorías el p. 110.](#)
- **Técnicas de ampliación automáticas.** Hay disponibles varias técnicas lingüísticas para ampliar las categorías existentes añadiendo y mejorando los descriptores para que capturen más registros. [Si desea obtener más información, consulte el tema Ampliación de categorías el p. 125.](#)
- **Técnicas manuales.** Hay varios métodos manuales, como la función de arrastrar y soltar. [Si desea obtener más información, consulte el tema Creación manual de categorías el p. 129.](#)
- **Marcos de código.** Importe sus propios marcos de código, o copie y pegue códigos en el administrador de marcos de códigos. [Si desea obtener más información, consulte el tema Importación de categorías predefinidas el p. 132.](#)

### ***Estrategias para crear categorías***

La siguiente lista de estrategias no es exhaustiva en absoluto, pero puede proporcionarle algunas ideas sobre cómo afrontar la generación de las categorías.

- Cuando inicie un proyecto, seleccione un conjunto de categorías a partir de un paquete de análisis de texto (TAP) para que pueda iniciarlo con algunas categorías pregeneradas. Estas categorías pueden categorizar suficientemente el texto desde el primer momento. Sin

embargo, si desea añadir más categorías, puede editar los valores de Generar categorías (Categorías > Configuración de generación). Abra el cuadro de diálogo Configuración avanzada: lingüística, seleccione la opción de entrada de Categoría Resultados de extracción no utilizados y cree las categorías adicionales.

- Cuando inicie un proyecto, seleccione un conjunto de categorías de un TAP. A continuación, arrastre los conceptos o patrones no usados y suéltelos en las categorías que considere apropiadas. A continuación, amplíe las categorías existentes que acaba de editar (Categorías > Ampliar categorías) para obtener más descriptores relacionados con los descriptores de la categoría existente.
- Genere categorías automáticamente utilizando la configuración avanzada de lingüística (Categorías > Generar categorías). A continuación, refine las categorías manualmente eliminando descriptores, eliminando categorías o fusionando categorías similares hasta que esté conforme con las categorías resultantes. Además, si generó categorías originalmente **sin** utilizar la opción Generalizar con comodines cuando sea posible, también puede intentar simplificar las categorías automáticamente utilizando Ampliar categorías con la opción Generalizar.
- Importe un archivo de categoría predefinida con anotaciones y/o nombres de categorías muy descriptivos. Además, si originalmente realizó una importación **sin** seleccionar la opción de importar o generar descriptores a partir de nombres de categoría, más adelante puede utilizar el cuadro de diálogo Ampliar categorías y seleccionar la opción Extender categorías vacías con descriptores generados desde el nombre de categorías. A continuación, amplíe dichas categorías una segunda vez pero esta vez utilice las técnicas de agrupación.
- Cree manualmente un primer conjunto de categorías clasificando conceptos o patrones de conceptos por frecuencia, y luego arrastrando y soltando los más interesantes en el panel Categorías. Una vez que tenga ese conjunto inicial de categorías, utilice la función Ampliar (Categorías > Ampliar categorías) para ampliar y refinar todas las categorías seleccionadas de manera que incluyan otros descriptores relacionados y pueda haber por lo tanto más registros coincidentes.

Después de aplicar estas técnicas, recomendamos que revise las categorías resultantes y emplee técnicas manuales para realizar pequeños ajustes, eliminar clasificaciones incorrectas o añadir registros o palabras que pueden haberse dejado de lado. Además, puesto que el uso de diferentes técnicas también puede generar categorías repetidas, puede fusionar o eliminar categorías si es necesario. [Si desea obtener más información, consulte el tema Edición y refinamiento de categorías el p. 156.](#)

### ***Consejos sobre la creación de categorías***

Para facilitar la creación de categorías mejores, puede revisar algunos conceptos que le ayudarán a tomar decisiones.

### **Consejos sobre la relación de “categoría a respuesta”**

Cuando se crean códigos para una pregunta con final cerrado, como “¿Cuándo fue la última vez que visitó la tienda de oportunidades?”, las categorías en las que irían las respuestas deben ser mutuamente excluyentes y exhaustivas. Este principio no se aplica necesariamente al análisis de texto cualitativo por al menos dos razones:

- En primer lugar, una regla empírica dice que cuanto más largo sea el registro de texto, más diversas son las ideas y las opiniones que se expresan. Así, las oportunidades de que puedan asignarse numerosas categorías a un registro aumentan considerablemente.
- En segundo lugar, suele haber varias formas de agrupar e interpretar los registros de texto que no están separados de manera lógica. En el caso de una encuesta con preguntas con final abierto sobre las tendencias políticas del encuestado, se pueden crear categorías, como *liberal y conservador* o *republicano y demócrata*, así como otras categorías más específicas, como *socio-liberal*, *económicamente conservador*, etc. Estas categorías no tienen por qué ser mutuamente excluyentes y exhaustivas.

### **Consejos sobre el número de categorías que conviene crear**

Excepto en el caso de una pregunta con final abierto extremadamente simple, nunca puede intuirse con rotundidad cuántas categorías pueden crearse. El número de categorías no es un aspecto del que deba preocuparse. La creación de categorías debería derivar directamente de los propios datos—en cuanto detecte algo interesante en relación a los objetivos de esta encuesta, puede crear una categoría para que represente esas actitudes e ideas.

- **Frecuencia de la categoría.** Para que una categoría sea útil, debe contener un número mínimo de registros. Es posible que encuentre opiniones un tanto intrigantes en unos pocos registros, pero si son solo uno o dos de entre mil registros, puede que esas opiniones no sean tan frecuentes como para que resulten útiles.
- **Complejidad.** Cuantas más categorías cree, más información deberá leer y resumir después de completar el análisis. Sin embargo, el hecho de que haya demasiadas categorías, además de añadir complejidad, no aporta necesariamente detalles útiles.

Desgraciadamente, no existe ninguna regla que determine cuántas categorías se consideran demasiadas, ni para determinar el número mínimo de registros por categoría. Deberá tomar ese tipo de decisiones según las demandas de su situación particular.

Sin embargo, podemos ofrecerle consejos acerca de dónde comenzar. Aunque el número de categorías no debería ser excesiva, en los primeros estadios del análisis se recomienda pecar por exceso que por defecto. Agrupar categorías que sean relativamente similares resulta más fácil que dividir los casos en nuevas categorías, de manera que una estrategia de trabajo de pasar de más a menos categorías suele ser una buena técnica. Teniendo en cuenta la naturaleza repetitiva de la minería de texto y la facilidad con la que puede realizarse con un programa de software, se considera aceptable generar más categorías al comienzo.

## **Selección de los mejores descriptores**

La siguiente información contiene varias directrices sobre cómo seleccionar o generar los mejores descriptores (conceptos, tipos, patrones TLA y reglas de categoría) para sus categorías. Los descriptores son los cimientos de las categorías. Cuando todo el texto de un registro o parte del mismo coincide con un descriptor, el registro se vincula con la categoría.

A menos que un descriptor contenga o se corresponda con un concepto o patrón extraído, no se vinculará con ningún registro. Por lo tanto, utilice conceptos, tipos, patrones y reglas de categoría como se describe en las siguientes secciones.

Como los conceptos no sólo se representan a sí mismos sino que también representan un conjunto de términos subyacentes que pueden ir desde formas en singular/plural hasta sinónimos, pasando por variaciones ortográficas, sólo se debe utilizar el propio concepto como descriptor o como parte de un descriptor. Para obtener más información sobre los términos subyacentes de cualquier concepto específico, pulse en el nombre del concepto en el panel Resultados extraídos. Cuando pase el ratón por encima del nombre del concepto, aparecerá una etiqueta con información y se mostrarán los términos subyacentes encontrados en su texto durante la última extracción. No todos los conceptos tienen términos subyacentes. Por ejemplo, si *coche* y *vehículo* fueran sinónimos pero *coche* se extrajera como concepto con *vehículo* como término subyacente, entonces sólo debe utilizar *coche* en un descriptor, ya que vinculará automáticamente un registro con *vehículo*.

### **Conceptos y tipos como descriptores**

Utilice un concepto como descriptor cuando quiera encontrar todos los registros que contengan dicho concepto (o cualquiera de sus términos subyacentes). En este caso, el uso de una regla de categoría más compleja no es necesario debido a que el nombre de concepto exacto es suficiente. Recuerde que cuando utiliza recursos que extraen opiniones, a veces los conceptos pueden variar durante la extracción de patrón TLA para capturar el verdadero sentido de la frase (consulte el ejemplo de la sección siguiente sobre TLA).

Por ejemplo, la respuesta a una encuesta que indique que las frutas preferidas de cada persona como “*La manzana y la piña son las mejores*” podría dar como resultado la extracción de *manzana* y *piña*. Al añadir el concepto *manzana* como descriptor a su categoría, todas las respuestas que contengan el concepto *manzana* (o cualquiera de sus términos subyacentes) se vinculan a dicha categoría.

Sin embargo, si sólo le interesa saber qué respuestas mencionan *manzana* del modo que sea, puede crear una regla de categoría como *\* manzana \** y también capturaré respuestas que contengan conceptos como *manzana*, *compota de manzana* o *manzana caramelizada*.

También puede capturar todos los registros que contengan conceptos con el mismo tipo mediante el uso de un tipo como descriptor directamente como *<Fruta>*. Recuerde que no puede utilizar *\** con tipos.

Si desea obtener más información, consulte el tema [Resultados extraídos: Conceptos, tipos y patrones en el capítulo 5 el p. 78](#).

### **Descriptores y patrones TLA (Análisis de enlace de texto)**

Utilice un resultado de patrón TLA como descriptor cuando quiera capturar ideas más finas con más matices. Cuando el texto se analiza durante la extracción de TLA, el texto se procesa frase a frase o cláusula a cláusula en lugar de tomar el texto completo (el registro). Al considerar todas las partes de una frase en conjunto, TLA puede identificar opiniones, relaciones entre dos elementos o, por ejemplo, una negación y comprender el sentido verdadero. Puede utilizar patrones de conceptos o patrones de tipo como descriptores.

Por ejemplo, si tuviéramos el texto “*la habitación no estaba tan limpia*”, se podrían extraer los siguientes conceptos: `habitación` y `limpia`. Sin embargo, si se hubiera activado la extracción de TLA en la configuración de extracción, TLA podría detectar que `limpia` se utilizó con un sentido negativo y realmente se corresponde con `no limpia`, lo que es un sinónimo del concepto `sucia`. Aquí, puede ver que el uso del concepto `limpia` como descriptor por sí solo coincidiría con este texto pero también capturaría otros registros que mencionaran la limpieza. Por lo tanto, sería mejor utilizar el patrón de concepto TLA con `sucia` como concepto de salida ya que coincidiría con este texto y sería un descriptor más adecuado.

### **Reglas empresariales de categoría como descriptores**

Las reglas de categoría son sentencias que clasifican automáticamente registros en una categoría que se basa en una expresión lógica utilizando los conceptos, tipos y patrones extraídos, así como operadores booleanos. Por ejemplo, puede escribir una expresión que signifique *incluir todos los registros que contienen el concepto extraído* `embajada pero no argentina en esta categoría`.

Puede escribir y utilizar reglas de categoría como descriptores en sus categorías para expresar diferentes ideas mediante los operadores booleanos `&`, `|` y `!` (`()`). Para obtener información detallada sobre la sintaxis de estas reglas y sobre cómo crearlas y editarlas, consulte [Uso de reglas de categoría el p. 145](#)

- Utilice una regla de categoría con el operador booleano `&` (AND) para buscar registros en los que se produzcan 2 o más conceptos. Los 2 o más conceptos conectados por operadores `&` no tienen que aparecer en la misma frase u oración; pueden hacerlo en cualquier parte del mismo registro para que se les considere una coincidencia con la categoría. Por ejemplo, si crea la regla de categoría `comida & barata` como descriptor, coincidiría con un registro que contenga el texto “*la comida era muy cara pero la habitación era barata*” a pesar del hecho de que `comida` no fuera el sustantivo que va con `barata`, ya que el texto contiene tanto `comida` como `barata`.
- Use una regla de categoría con el operador booleano `!` (`()`) (NOT) como descriptor para que le ayude a encontrar registros en los que ocurren algunas cosas y otras no. Esto puede evitar la agrupación de información que puede parecer que está relacionada basándose en las palabras pero no en el contexto. Por ejemplo, si crea la regla de categoría `<Organización> & !(ibm)` como descriptor, coincidirá con el siguiente texto *SPSS Inc. es una empresa fundada en 1967* y no coincidirá con el siguiente texto *la empresa de software fue adquirida por IBM*.

- Utilice una regla de categoría con el operador booleano | (OR) como descriptor para buscar registros que contengan uno de varios conceptos o tipos. Por ejemplo, si crea la regla de categoría `(personal|equipo) & malo` como descriptor, coincidirá con cualquier registro en el que se encuentre cualquiera de estos sustantivos junto con el concepto `malo`.
- Utilice tipos en reglas de categoría para hacer que sean más genéricas y puedan implementarse en más casos. Por ejemplo, si está trabajando con datos de hotel, puede que esté muy interesado en saber lo que piensan los clientes sobre el personal del hotel. Los términos relacionados podrían incluir palabras como `repcionista`, `camarero`, `camarera`, `recepción`, `mostrador de recepción`, etc. En este caso podría crear un nuevo tipo denominado `<PersonalHotel>` y añadir todos los términos anteriores a dicho tipo. Aunque es posible crear una regla de categoría para cada clase de personal como `[* camarera * & agradable]`, `[* mostrador * & simpático]`, `[* recepcionista * & servicial]`, podría crear una única regla de categoría más genérica utilizando el tipo `<PersonalHotel>` para capturar todas las respuestas que obtuvieron opiniones favorables del personal del hotel con la forma `[<PersonalHotel> & <Positivo>]`.

*Nota:* Puede utilizar tanto + como & en reglas de categoría al incluir patrones TLA en dichas reglas. [Si desea obtener más información, consulte el tema Uso de patrones TLA en las reglas de categoría el p. 148.](#)

### ***Ejemplo de cómo los conceptos, TLA o reglas de categoría como descriptores coinciden de manera diferente***

El siguiente ejemplo muestra el modo en que el uso de un concepto como descriptor, regla de categoría como descriptor o patrón TLA como descriptor afecta a cómo se categorizan los registros. Supongamos que tiene los 5 registros siguientes.

- A: *“Magnífico personal del restaurante, comida excelente y habitación cómoda y limpia.”*
- B: *“El personal del restaurante era pésimo, pero la habitación estaba limpia.”*
- C: *“Habitación cómoda y limpia.”*
- D: *“Mi habitación no estaba demasiado limpia.”*
- E: *“Limpia.”*

Como los registros incluyen la palabra *limpia* y desea capturar esta información, podría crear uno de los descriptores que aparecen en la tabla siguiente. Basándose en lo esencial de lo que esté intentando capturar, podrá ver cómo el uso de una clase de descriptor antes que otra puede producir resultados diferentes.

Tabla 6-1  
Modo de coincidencia entre descriptores y registros del ejemplo

Descriptor	A	B	C	D	E	Explicación
limpia	coincidencia	coincidencia	coincidencia	coincidencia	coincidencia	El descriptor es un concepto extraído. Todos los registros contenían el concepto <i>limpia</i> , incluso el registro D, ya que sin TLA no se sabe automáticamente que “ <i>no limpia</i> ” significa <i>sucia</i> según las reglas TLA.
limpia + .	-	-	-	-	coincidencia	El descriptor es un patrón TLA que representan <i>limpia</i> por sí mismo. Sólo ha coincidido con el registro en el que <i>limpia</i> se extrajo sin ningún concepto asociado durante la extracción de TLA.
[limpia]	coincidencia	coincidencia	coincidencia	-	coincidencia	El descriptor es una regla de categoría que busca una regla de TLA que contenga <i>limpia</i> en solitario o con algo más. Se vincularon todos los registros donde se encontró un resultado de TLA con <i>limpia</i> independientemente de si <i>limpia</i> estaba vinculado a otro concepto como <i>habitación</i> y en cualquier espacio.

## Acerca de las categorías

**Categorías** hace referencia a un grupo de conceptos, opiniones o actitudes estrechamente relacionados. Para que resulte útil, una categoría debe describirse fácilmente mediante una breve frase o etiqueta que capte su significado esencial.

Por ejemplo, si está analizando las respuestas de los usuarios en una encuesta sobre un nuevo detergente, puede crear una categoría etiquetada como *olor* que contenga todas las respuestas que describan el aroma del producto. Sin embargo, una categoría así no diferenciaría entre aquellos que consideraron que el detergente tenía un aroma agradable de los que lo encontraron molesto. Puesto que IBM® SPSS® Text Analytics for Surveys es capaz de extraer opiniones cuando se utilizan los recursos adecuados, se podrían crear otras dos categorías que identificaran a los encuestados a los que *les gustó el olor* y a los encuestados a los que *no les gustó el olor*.

Puede crear y trabajar con categorías en el panel Categorías del panel superior izquierdo de la ventana de análisis de texto. Cada categoría está definida por uno o más descriptores. Los **Descriptores** son conceptos, tipos y patrones, así como reglas de categoría, que se han utilizado para definir una categoría.

Si quiere ver los descriptores que conforman una categoría específica, puede expandir la categoría en el árbol. Se muestran unos iconos en el árbol para que pueda identificar fácilmente cada elemento. Solo el primer nivel define la categoría. Si sigue ampliando las definiciones, podrá ver ejemplos de lo que se ha encontrado en los datos.

Cuando se generan categorías automáticamente mediante las técnicas de generación de categorías, como la inclusión de conceptos, dichas técnicas utilizarán conceptos y tipos como descriptores para crear las categorías. Puede también añadir patrones o partes de dichos patrones

como descriptores de categorías. Por último, puede crear reglas de categoría manualmente para utilizarlas como descriptores en las categorías. [Si desea obtener más información, consulte el tema Uso de reglas de categoría el p. 145.](#)

Por ejemplo, si añade un tipo a una definición de categoría, los conceptos asignados a dicho tipo se incluirán automáticamente, aunque solo haya unos cuantos presentes en los datos en este momento. De esta forma es más fácil reutilizar las definiciones de categoría con los datos nuevos. [Si desea obtener más información, consulte el tema Copia de categorías el p. 164.](#) También puede crear reglas de categoría manualmente para utilizarlas como descriptores en las categorías. [Si desea obtener más información, consulte el tema Uso de reglas de categoría el p. 145.](#)

Tabla 6-2

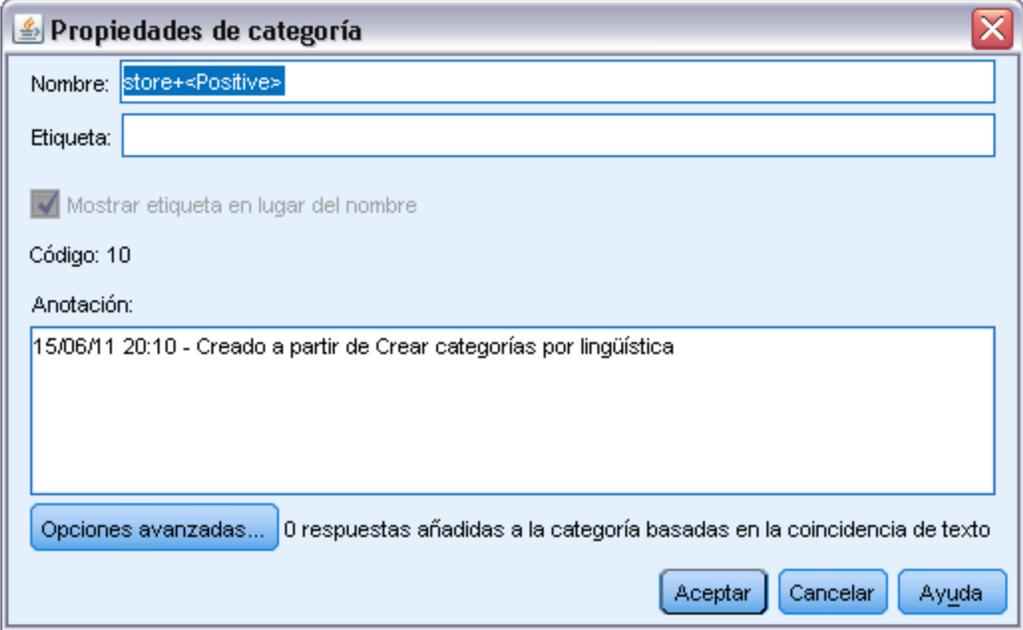
*Iconos para identificar elementos en las definiciones*

Icono	Descripción
	Concepto.
	Tipo, que puede expandirse para ver los conceptos que contiene.
	Patrón de conceptos, que puede expandirse para ver los conceptos específicos de los patrones.
	Patrón de tipo, que puede expandirse al nivel de patrón de conceptos.
	Reglas de categoría de la categoría. Pulse el botón derecho del ratón en el nombre de la regla para editarla.

### ***Propiedades de categoría***

Además de los descriptores, las categorías también tienen propiedades que pueden editarse para cambiar el nombre de las categorías, añadir una etiqueta o anotación o acceder al cuadro de diálogo de coincidencia de texto.

Figura 6-5  
Diálogo Propiedades de categoría



Propiedades de categoría

Nombre: store+<Positive>

Etiqueta:

Mostrar etiqueta en lugar del nombre

Código: 10

Anotación:  
15/06/11 20:10 - Creado a partir de Crear categorías por lingüística

Opciones avanzadas... 0 respuestas añadidas a la categoría basadas en la coincidencia de texto

Aceptar Cancelar Ayuda

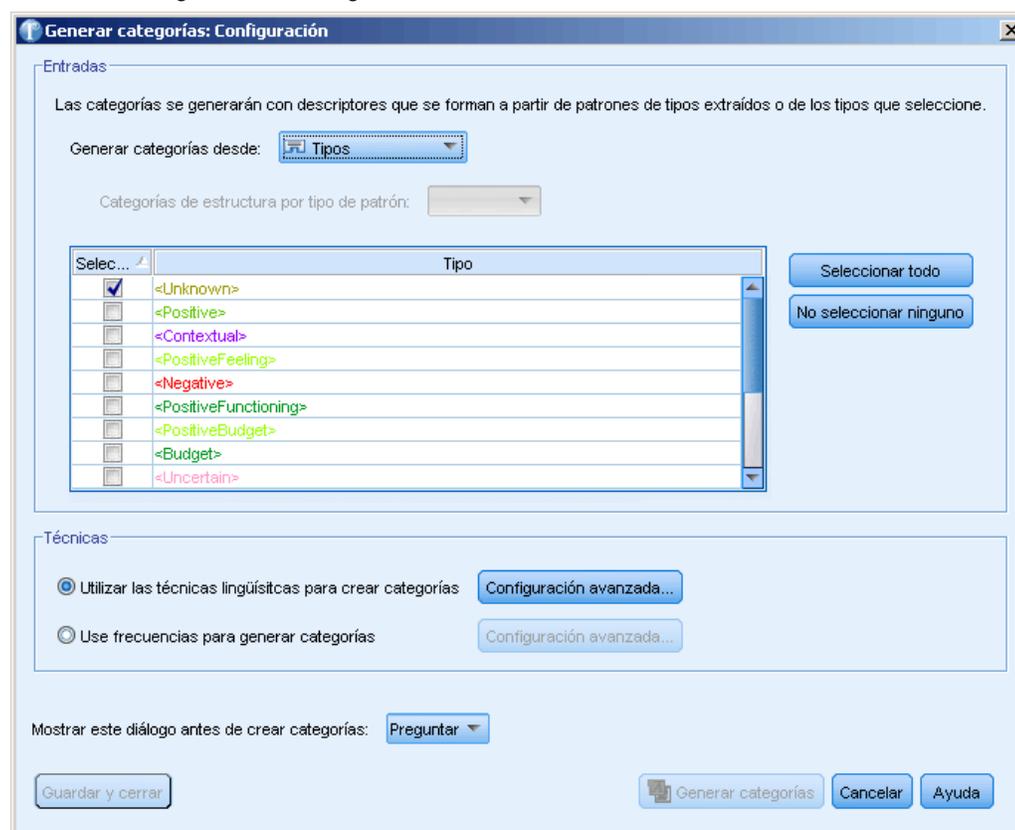
Existen las propiedades siguientes:

- **Nombre.** Este nombre aparece por defecto en el árbol. Cuando se crea una categoría utilizando una técnica automática, se le asigna un nombre automáticamente.
- **Etiqueta.** El uso de etiquetas resulta útil para crear descripciones de categoría con un significado más detallado que se utilizan en otros productos o en otras tablas o gráficos. Si elige la opción de visualizar la etiqueta, ésta se utilizará en la interfaz para identificar la categoría. En IBM® SPSS® Statistics, se exportan como etiquetas variables. En Microsoft Excel, se exportan como filas individuales.
- **Código.** El número de código se corresponde con el valor de código de esta categoría. Puede editar este código en el Administrador de marcos de código. El Administrador de marcos de código permite editar el nombre, la etiqueta y el código de cada categoría, así como copiar y pegar marcos de código enteros. .
- **Anotación.** Puede añadir una descripción breve a cada categoría de este campo. Cuando el diálogo Generar categorías genera una categoría, automáticamente se añade una nota a esta anotación. También puede añadir texto de muestra a una anotación directamente desde el panel Datos seleccionando el texto y eligiendo Categorías > Añadir a anotación en los menús.
- **Opciones avanzadas de Coincidencia de texto** Con el botón Opciones avanzadas, puede añadir palabras o frases a una definición de categoría. A menudo se utiliza para sustituir una extracción extraviada. [Si desea obtener más información, consulte el tema Coincidencia de texto en las categorías el p. 162.](#)

## Generación de categorías

Por un lado puede tener categorías procedentes de un paquete de análisis de texto, y también puede generar categorías automáticamente utilizando una serie de técnicas lingüísticas y de frecuencia. A través del cuadro de diálogo de configuración Generar categorías, puede aplicar las técnicas automáticas lingüísticas y de frecuencia para generar categorías a partir de conceptos o de patrones de conceptos.

Figura 6-6  
Cuadro de diálogo Generar categorías



En general, las categorías pueden estar formadas por diferentes tipos de descriptores (tipos, conceptos, patrones TLA, reglas de categoría). Cuando se generan categorías utilizando las técnicas automáticas de generación de categorías, las categorías resultantes se indican después de un concepto o de un patrón de conceptos (según la entrada que haya seleccionado) y contiene un conjunto de descriptores. Estos descriptores pueden presentarse en forma de reglas de categoría o conceptos e incluyen todos los conceptos relacionados descubiertos por las técnicas.

Después de generar las categorías, puede aprender mucho sobre ellas revisándolas en el panel Categorías o explorándolas en gráficos y diagramas. Luego puede utilizar técnicas manuales para realizar pequeños ajustes, eliminar clasificaciones incorrectas o añadir registros o palabras que pueden haberse dejado de lado. Una vez aplicada la técnica, los conceptos, tipos y patrones que estaban agrupados en una categoría siguen estando disponibles para otras técnicas. Y puesto que el uso de diferentes técnicas también puede generar categorías repetidas o inapropiadas, puede

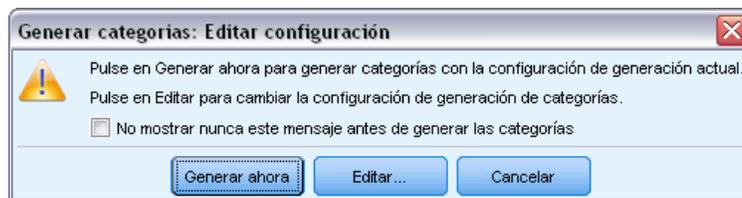
fusionar o eliminar categorías. Si desea obtener más información, consulte el tema [Edición y refinamiento de categorías](#) el p. 156.

**Importante:** En versiones anteriores, las reglas de co-ocurrencia y de sinónimos se colocaban entre corchetes. En esta versión, los corchetes ahora indican un resultado de patrón de . En su lugar, las reglas de co-ocurrencia y de sinónimos estarán entre paréntesis, por ejemplo (sistemas de sonido|altavoces).

### Para generar categorías

- ▶ En los menús elija **Categorías > Generar categorías**. A menos que haya elegido que no se le avise nunca, aparecerá un cuadro de mensaje.

Figura 6-7  
Aviso antes de la generación



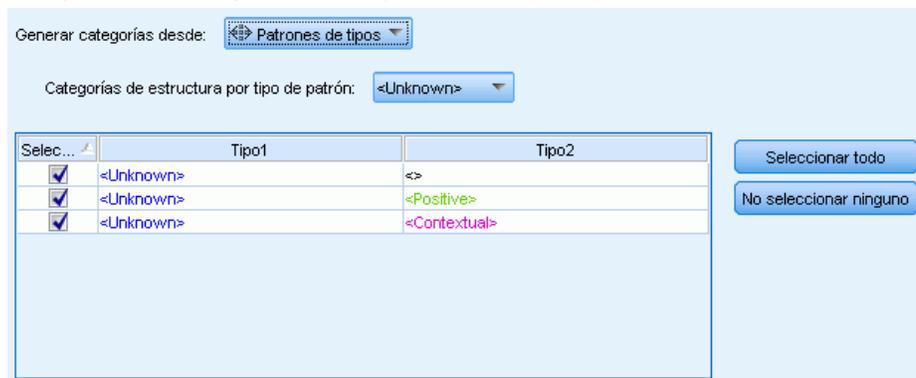
- ▶ Elija si desea generar ahora o editar primero la configuración.
  - Pulse en **Generar ahora** para empezar a generar categorías utilizando la configuración actual. La configuración seleccionada de manera predeterminada suele ser suficiente para comenzar el proceso de categorización. El proceso de generación de categorías comienza y aparece un diálogo de progreso.
  - Pulse en **Editar** para revisar y modificar la configuración de la generación.

### Entradas

Las categorías se generan a partir de descriptores derivados de cada tipo o patrón de tipos. Por defecto, en la lista desplegable aparecen seleccionados los patrones de tipo. En esta tabla, puede seleccionar los tipos individuales o los patrones que se incluirán en el proceso de generación de categorías.

**Patrones de tipo.** Si selecciona patrones de tipo, las categorías se generarán a partir de patrones en lugar de tipos y conceptos individuales. De esta forma se categorizarán los registros que contengan un patrón de conceptos que pertenezca al patrón de tipo seleccionado. Así pues, si selecciona el patrón de tipo <Presupuesto> y <Positivo> en la tabla, pueden generarse categorías como `coste & <Positivo>` o `tarifas & excelentes`. La tabla muestra únicamente una fila por cada combinación de tipo como <Ubicación> + <Positiva> y <Positiva> + <Ubicación>, y el orden no es relevante para la forma en que se generan las categorías.

**Figura 6-8**  
Diálogo *Generar categorías con los patrones de tipo disponibles*



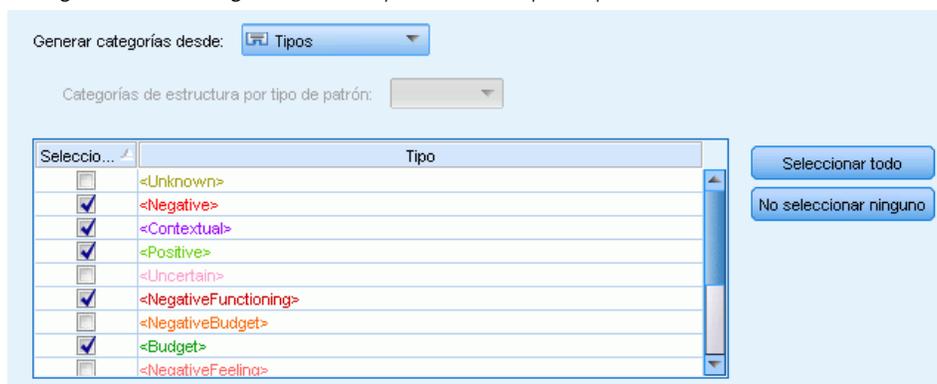
Al utilizar patrones de tipo como entrada para la generación de categorías automática, a veces las técnicas identifican varias maneras de formar la estructura de la categoría. Técnicamente, no hay una única manera correcta de producir las categorías; sin embargo, puede que encuentre una estructura más adecuada para su análisis que otra. Para ayudar a personalizar el resultado en este caso, puede designar un tipo como el foco preferido. Elija este tipo en el campo Estructurar categorías por tipo de patrón: y la tabla se actualizará para mostrar únicamente los patrones aplicables que contienen el tipo seleccionado. La mayoría de las veces, <Desconocido> estará preseleccionado. Esto dará como resultado la selección de todos los patrones que contengan el tipo <Desconocido>. La tabla muestra los tipos en orden descendente empezando por el que tiene el número más alto de registros.

**Tipos.** Si selecciona tipos, las categorías se generarán a partir de los conceptos que pertenecen a los tipos seleccionados. Por lo tanto, si selecciona el tipo <Presupuesto> en la tabla, podrían producirse categorías como `coste` o `precio` debido a que `coste` y `precio` son conceptos asignados al tipo <Presupuesto>.

De manera predeterminada, sólo se seleccionan los tipos que capturan el número máximo de registros. Esta preselección permite centrarse rápidamente en los tipos más interesantes y evitar la generación de categorías irrelevantes. La tabla muestra los tipos en orden descendente empezando por el que tiene el número más alto de registros. Los tipos de la biblioteca `Opinions` no están seleccionados de manera predeterminada en la tabla de tipos.

La entrada que ha seleccionado afecta a las categorías que obtiene. Cuando decida utilizar Tipos como entrada, podrá ver los conceptos relacionados claramente con más facilidad. Por ejemplo, si genera categorías utilizando Tipos como entrada, podría obtener una categoría `Fruta` con conceptos como `manzana`, `pera`, `cítricos`, `naranja`, etc. Si por el contrario selecciona Patrones de tipo como entrada y, a continuación, selecciona el patrón <Desconocido> + <Positive>, por ejemplo, entonces podría obtener una categoría `fruta + <Positiva>` con uno o dos tipos de fruta como `fruta + sabrosa` y `manzana + buena`. Este segundo resultado sólo muestra 2 patrones de conceptos porque las otras apariciones de frutas no tienen por qué haberse calificado positivamente. Aunque esto puede ser suficiente para sus datos de texto actuales, en estudios longitudinales donde utiliza diferentes conjuntos de documentos, puede que desee añadir manualmente otros descriptores como `cítrico + positivo` o utilizar tipos. Si sólo utiliza tipos como entrada, esto le ayudará a encontrar toda la fruta posible.

Figura 6-9  
Diálogo Generar categorías con los patrones de tipo disponibles



### Técnicas

Puesto que cada conjunto de datos es exclusivo, el número de métodos y el orden en el que los aplique puede cambiar con el tiempo. Puesto que los objetivos de la minería de texto puede diferir de un conjunto de datos a otro, puede que deba experimentar con diferentes técnicas para comprobar con cuál de ellas se obtiene el mejor resultado para los datos de texto determinados.

No es necesario que sea un experto en este tipo de configuración para poder utilizarla. Por defecto, los valores más comunes ya aparecen seleccionados. Por lo tanto, puede pasar por alto los diálogos de configuración avanzada y empezar a generar categorías directamente. Igualmente, si realiza cambios aquí, no es necesario que vuelva al diálogo de configuración cada vez, puesto que siempre se conservan los últimos valores.

Seleccione las técnicas lingüísticas o de frecuencia y pulse en el botón Configuración avanzada para que se muestren los valores de las técnicas seleccionadas. Ninguna de las técnicas automáticas categorizará los datos con exactitud, por lo que se recomienda que busque y aplique una o más técnicas automáticas que se adapten bien a sus datos. No se pueden utilizar técnicas lingüísticas y de frecuencia simultáneamente para el proceso de generación.

- **Técnicas lingüísticas avanzadas.** Si desea obtener más información, consulte el p. 113.
- **Técnicas de frecuencia avanzadas.** Si desea obtener más información, consulte el p. 123.

### Configuración avanzada: Lingüística

Cuando se generan categorías, puede seleccionar entre una serie de técnicas lingüísticas avanzadas para la generación de categorías, como por ejemplo la *derivación de raíz de conceptos*, *inclusión de conceptos*, *redes semánticas* (solo en inglés) y *reglas de co-ocurrencia*. Estas técnicas pueden utilizarse solas o combinadas entre sí para crear categorías.

Tenga en cuenta que, puesto que cada conjunto de datos es exclusivo, el número de métodos y el orden en el que los aplique puede cambiar con el tiempo. Puesto que los objetivos de la minería de texto puede diferir de un conjunto de datos a otro, puede que deba experimentar con diferentes técnicas para comprobar con cuál de ellas se obtiene el mejor resultado para los datos de texto determinados. Ninguna de las técnicas automáticas categorizará los datos con exactitud, por lo que se recomienda que busque y aplique una o más técnicas automáticas que se adapten bien a sus datos.

Figura 6-10  
Cuadro de diálogo Configuración avanzada: Lingüística para generar categorías

**Configuración avanzada: Lingüística**

**Entrada y salida**

Entrada de categorías:  Resultados de extracción no usados  Todos los resultados de la extracción

Salida de categorías:

Jerárquico con subcategorías Niveles máximos creados: 5

Categorías planas (nivel único)

**Técnicas de agrupación**

Derivación de raíz de concepto  Inclusión de conceptos

Red semántica  Co-ocurrencia

Número mínimo de registros: 2

Distancia máxima de búsqueda: 3

Evitar el emparejado de conceptos específicos [Gestionar pares...](#)

Generalizar con comodines cuando sea posible (por ejemplo, apple \*)

**Otras opciones**

Se ha creado el número máximo de categorías de nivel superior: 30

Número mínimo de descriptores y/o subcategorías por categoría: 2

Permitir que los descriptores aparezcan en más de una categoría

Para resolver los nombres de categorías duplicados: Fusionar los nuevos con los existentes

[Restablecer a valor por defecto](#) [Guardar y cerrar](#) [Cancelar](#) [Ayuda](#)

### Entrada y salida

**Entrada de categorías.** Seleccione a partir de qué se generarán las categorías:

- **Resultados de extracción no usados.** Esta opción permite que se generen categorías a partir de los resultados de extracción que no se utilizan en ninguna categoría existente. De esta manera se minimiza la tendencia que tienen los registros de hacer coincidir varias categorías y limita el número de categorías que se generan.
- **Todos los resultados de la extracción.** Esta opción permite generar categorías utilizando los resultados de extracción. Esto resulta especialmente útil cuando no existen categorías, o existen muy pocas.

**Salida de categorías.** Seleccione la estructura general de las categorías que se generarán:

- Jerárquico con subcategorías. Esta opción permite la creación de subcategorías y sub-subcategorías. Puede establecer la profundidad de sus categorías seleccionando el número máximo de niveles (campo Niveles máximos creados) que se puede crear. Si selecciona 3, las categorías podrían contener subcategorías y dichas subcategorías también podrían tener subcategorías.
- Categorías planas (nivel único). Esta opción sólo permite la creación de un nivel de categorías, lo que significa que no se generará ninguna subcategoría.

### **Agrupación de técnicas**

Cada una de las técnicas disponibles resulta idónea para determinados tipos de datos y situaciones, pero a menudo conviene combinar técnicas en el mismo análisis para capturar el rango completo de registros. Puede ver un concepto en diversas categorías o detectar categorías redundantes.

**Derivación de raíz de conceptos.** Esta técnica crea categorías tomando un concepto y buscando otros conceptos que estén relacionados con el primero analizando si alguno de los componentes de los conceptos están morfológicamente relacionados o comparten raíces. Esta técnica es muy útil para identificar conceptos de palabras compuestas sinónimas, puesto que los conceptos de cada categoría generada son sinónimos o tienen un significado muy similar. Funciona con datos de extensión diversa y genera un número más reducido de categorías compactas. Por ejemplo, el concepto `ocasiones de progreso` se agruparía con los conceptos `ocasión de progresar` y `ocasión de progresión`. [Si desea obtener más información, consulte el tema Derivación de raíz de conceptos el p. 118.](#)

**Red semántica.** Esta técnica comienza identificando los posibles sentidos de cada concepto a partir de un amplio índice de relaciones de palabras, y luego crea categorías agrupando los conceptos relacionados. Esta técnica resulta idónea cuando los conceptos son conocidos en la red semántica y no son muy ambiguos. Es menos idónea si el texto contiene terminología específica o jerga desconocida en la red. Por ejemplo, el concepto `manzana golden` se podría agrupar con `manzana reineta` y `manzana fuji` puesto que son familia de la `golden`. En otro ejemplo, el concepto `animal` se agruparía con `gato` y `canguro` puesto que ambos son hipónimos de `animal`. En esta versión esta técnica está disponible solo para texto en inglés. [Si desea obtener más información, consulte el tema Redes semánticas el p. 121.](#)

**Inclusión de conceptos.** Esta técnica genera categorías agrupando los conceptos multitérmino (palabras compuestas) basándose en si contienen palabras que son subconjuntos o superconjuntos de una palabra en la otra. Por ejemplo, el concepto `seguridad` estaría agrupado en `asiento de seguridad`, `cinturón de seguridad` y `silla infantil de seguridad`. [Si desea obtener más información, consulte el tema Inclusión de conceptos el p. 120.](#)

**Co-ocurrencia.** Esta técnica crea categorías a partir de las co-ocurrencias que se encuentran en el texto. La idea radica en que cuando en los documentos y registros a menudo se encuentran conceptos o patrones de conceptos que aparecen juntos, esa co-ocurrencia refleja una relación subyacente que probablemente sea valiosa para las definiciones de categorías. Cuando la co-ocurrencia de algunas palabras es significativa, se crea una regla de co-ocurrencia que puede utilizarse como un descriptor de categoría para una nueva subcategoría. Por ejemplo, si muchos registros contienen las palabras `precio` y `disponibilidad` (pero hay pocos registros que

contengan sólo una de las dos), estos conceptos se podrían agrupar en una regla de co-ocurrencia, (`precio & disponible`) y asignarse a una subcategoría de la categoría `precio` por ejemplo. [Si desea obtener más información, consulte el tema Reglas de co-ocurrencia el p. 122.](#)

- **Número mínimo de registros** Para ayudar a determinar la relevancia de las co-ocurrencias, defina el número mínimo de registros que deben contener una co-ocurrencia determinada para que se utilice como descriptor en una categoría.

**Distancia máxima de búsqueda.** Seleccione el alcance de búsqueda de las técnicas antes de generar las categorías. Cuanto más bajo sea el valor, menos resultados obtendrá; sin embargo, dichos resultados serán más precisos y será más probable que pueda asociarlos o vincularlos significativamente entre sí. Cuanto más alto sea el valor, más resultados se obtienen; sin embargo, estos resultados pueden ser también menos fiables o relevantes. Esta opción se aplica globalmente en todas las técnicas, pero sus resultados son mejores en las co-ocurrencias y en las redes semánticas.

**Evitar el emparejamiento de conceptos específicos.** Seleccione esta casilla de verificación para detener el proceso de agrupación o emparejamiento de dos conceptos en el resultado. Para crear o administrar pares de conceptos, pulse en Administrar parejas. [Si desea obtener más información, consulte el tema Administración de pares de excepciones de enlace el p. 117.](#)

**Generalizar con comodines cuando sea posible.** Seleccione esta opción para que el producto genere reglas genéricas en las categorías utilizando el comodín de asterisco. Por ejemplo, en lugar de generar varios descriptores como `[naranja de mesa + .]` y `[naranja de zumo + .]`, el uso de comodines generaría `[naranja * + .]`. Si generaliza con comodines, obtendrá a menudo exactamente el mismo número de registros que antes. Sin embargo, esta opción tiene la ventaja de reducir el número y simplificar los descriptores de categoría. Además, esta opción aumenta la posibilidad de categorizar más registros utilizando estas categorías en datos de texto nuevos (por ejemplo, en estudios por fases o longitudinales).

### ***Otras opciones para generar categorías***

Además de seleccionar las técnicas de agrupación que se aplicarán, puede editar otras muchas opciones de generación, como se indica a continuación:

**Número máximo de categorías de creadas.** Utilice esta opción para limitar el número de categorías que pueden generarse cuando pulsa en el botón Generar categorías. En algunos casos, puede obtener mejores resultados si establece un valor alto y luego suprime cualquiera de las categorías irrelevantes.

**Número mínimo de descriptores y/o subcategorías por categoría.** Utilice esta opción para definir el número mínimo de descriptores y subcategorías que debe contener una categoría para que pueda crearse. Esta opción ayuda a limitar la creación de categorías que no capturen un número significativo de registros.

**Permitir que los descriptores aparezcan en más de una categoría.** Cuando está seleccionada, esta opción permite el uso de descriptores en más de una de las categorías que se generarán a continuación. Esta opción suele generarse porque los elementos se clasifican de forma natural o habitual en dos o más categorías, y dejar que esto ocurra conlleva la creación de categorías de mayor calidad. Si no selecciona esta opción, se reduce el solapamiento de registros en varias categorías, y en función del tipo de datos que tenga, puede ser una situación deseable. Sin

embargo, con la mayoría de los tipos de datos, el hecho de restringir los descriptores a una sola categoría suele resultar en una pérdida de calidad o de cobertura de la categoría. Por ejemplo, supongamos que tiene el concepto `fabricante` de `asientos de automóviles`. Con esta opción, este concepto podría aparecer en una categoría basada en el texto `asientos de automóviles` y en otra basada en `fabricante`. Pero si no se selecciona esta opción, aunque igualmente puede obtener las dos categorías, el concepto `fabricante de asientos de automóviles` solo aparecerá como descriptor en la categoría que mejor se adapte en función de diversos factores, como el número de registros en los que aparezcan `asientos de automóviles` y `fabricante`.

**Solucionar nombres de categoría duplicados por.** Seleccione cómo tratar las categorías o subcategorías nuevas cuyos nombres sean iguales que los de categorías existentes. Puede fusionar las nuevas (y sus descriptores) con las categorías existentes que tengan el mismo nombre. Asimismo, puede omitir la creación de categorías si se encuentra un nombre duplicado en las categorías existentes.

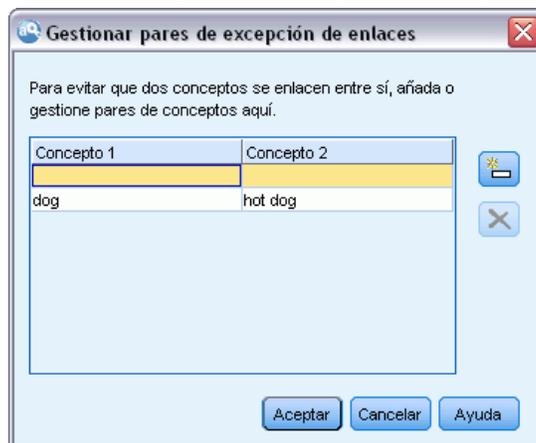
### ***Administración de pares de excepciones de enlace***

Durante el proceso de generación de categorías y asignación de conceptos, los algoritmos internos agrupan las palabras en asociaciones conocidas. Para evitar que dos conceptos se emparejen o vinculen entre sí, puede activar esta función en el cuadro de diálogo `Generar categorías Configuración avanzada` y el cuadro de diálogo `Configuración de índice de mapa de conceptos` y pulsar en el botón `Administrar parejas`.

En el cuadro de diálogo resultante `Administrar excepciones de enlace`, puede añadir, editar o eliminar parejas de conceptos. Introduzca un par por cada línea. La introducción de pares aquí evitará que se produzca el emparejamiento al generar o ampliar categorías y asignación de conceptos. Introduzca las palabras exactas que quiera, por ejemplo, una palabra con tilde no es igual que una palabra sin tilde.

Por ejemplo, si quiere asegurarse de que `perrito caliente` y `perrito` no estén agrupadas, puede añadir el par en una línea separada de la tabla.

Figura 6-11  
Cuadro de diálogo `Administrar pares de excepciones de enlace`



## ***Acerca de las técnicas lingüísticas***

Cuando se generan o amplían categorías, puede seleccionar entre una serie de técnicas lingüísticas avanzadas para la generación de categorías, como por ejemplo la *derivación de raíz de conceptos*, *inclusión de conceptos*, *redes semánticas* (solo en inglés) y *reglas de co-ocurrencia*. Estas técnicas pueden utilizarse solas o combinadas entre sí para crear categorías.

No es necesario que sea un experto en este tipo de configuración para poder utilizarla. Por defecto, los valores más comunes ya aparecen seleccionados. Puede pasar por alto los diálogos de configuración avanzada y empezar a generar o ampliar categorías directamente. Igualmente, si realiza cambios aquí, no es necesario que vuelva al diálogo de configuración cada vez, puesto que siempre se conservan los últimos valores.

Tenga en cuenta que, puesto que cada conjunto de datos es exclusivo, el número de métodos y el orden en el que los aplique puede cambiar con el tiempo. Puesto que los objetivos de la minería de texto puede diferir de un conjunto de datos a otro, puede que deba experimentar con diferentes técnicas para comprobar con cuál de ellas se obtiene el mejor resultado para los datos de texto determinados. Ninguna de las técnicas automáticas categorizará los datos con exactitud, por lo que se recomienda que busque y aplique una o más técnicas automáticas que se adapten bien a sus datos.

Las principales técnicas lingüísticas automáticas para la generación de categorías son:

- **Derivación de raíz de conceptos.** Esta técnica crea categorías tomando un concepto y buscando otros conceptos que estén relacionados con el primero analizando si alguno de los componentes de los conceptos están morfológicamente relacionados. [Si desea obtener más información, consulte el tema Derivación de raíz de conceptos el p. 118.](#)
- **Inclusión de conceptos.** Esta técnica crea categorías tomando un concepto y buscando otros conceptos que lo incluyan. [Si desea obtener más información, consulte el tema Inclusión de conceptos el p. 120.](#)
- **Red semántica.** Esta técnica comienza identificando los posibles sentidos de cada concepto a partir de un amplio índice de relaciones de palabras, y luego crea categorías agrupando los conceptos relacionados. [Si desea obtener más información, consulte el tema Redes semánticas el p. 121.](#) Esta opción sólo se encuentra disponible para texto en inglés.
- **Co-ocurrencia.** Esta técnica crea reglas de co-ocurrencia que pueden utilizarse para crear una categoría nueva, para ampliar una categoría o como entrada a otra técnica de categoría. [Si desea obtener más información, consulte el tema Reglas de co-ocurrencia el p. 122.](#)

### ***Derivación de raíz de conceptos***

La técnica de derivación de raíz de conceptos crea categorías tomando un concepto y buscando otros conceptos que estén relacionados con el primero analizando si alguno de los componentes de los conceptos está morfológicamente relacionado. Un componente es una palabra. Esta técnica intenta agrupar conceptos observando la terminación (el sufijo) de cada componente de un concepto y buscando otros conceptos que puedan derivar de los primeros. La idea radica en que cuando las palabras derivan unas de otras, probablemente tienen el mismo significado o parecido. Para poder identificar las terminaciones, se utilizan reglas específicas de idioma. Por ejemplo, el concepto `ocasiones de progreso` se agruparía con los conceptos `ocasión de progresar` y `ocasión de progresión`.

Puede utilizar la derivación de raíz de conceptos en cualquier tipo de texto. Por sí mismo genera pocas categorías, y cada una de ellas suele contener pocos conceptos. Los conceptos de cada categoría son sinónimos o están relacionados por posición. Puede resultarle útil emplear este algoritmo aunque esté generando categorías manualmente; los sinónimos que encuentre pueden ser sinónimos de los conceptos en los que está particularmente interesado.

*Nota:* Puede evitar que los conceptos se agrupen entre sí especificándolos explícitamente. Si desea obtener más información, consulte el tema [Administración de pares de excepciones de enlace](#) el p. 117.

### ***Estructuración en componentes y desarticulación***

Cuando se aplican las técnicas de derivación de raíz o inclusión de conceptos, en primer lugar los términos se desglosan en componentes (palabras) y luego los componentes se desarticulan. Cuando se aplica una técnica, los conceptos y sus términos asociados se cargan y se dividen en componentes basándose en separadores, como espacios, guiones y apóstrofes. Por ejemplo, el término `jefe administrador` se divide en los componentes `{administrador, jefe}`.

Sin embargo, es posible que algunas partes del término original no se utilicen, las cuales se consideran palabras vacías. En inglés y en otros idiomas, algunos de estos componentes ignorables pueden incluir palabras como `a`, `y`, `en`, `por`, `para`, `desde`, `un`, `de`, `sin`, `o`, `el`, `hasta` y `con`.

Por ejemplo, el término `examen de los datos` tiene el conjunto de componentes `{datos, examen}`, y los términos `de` y `los` se consideran ignorables. Además, el orden de los componentes no se refleja en un conjunto de componentes. De esta forma, los tres términos siguientes podrían ser equivalentes: `un coche seguro para viajar`, `viajar en un coche seguro` y `seguro para viajar en coche` puesto que todos ellos tienen el mismo conjunto de componentes `{coche, seguro, viajar}`. Cada vez que una pareja de términos se identifica como equivalente, los conceptos correspondientes se fusionan para formar un concepto nuevo que haga referencia a todos los términos.

Además, puesto que los componentes de un término pueden estar declinados, internamente se aplican reglas específicas del idioma para identificar los términos equivalentes independientemente de la variación de la declinación, por ejemplo, las formas plurales. De esta forma, los términos `nivel de soporte` y `soporte de niveles` pueden identificarse como equivalentes porque la forma singular que se deriva sería `nivel`.

### ***Funcionamiento de la derivación de raíz de conceptos***

Cuando se ha aplicado la estructuración en componentes y la desarticulación de los términos (consulte la sección anterior), el algoritmo de derivación de raíz de conceptos analiza las terminaciones o sufijos de los componentes con el fin de encontrar la raíz del componente y luego agrupar los conceptos con otros que tengan raíces iguales o similares. Las terminaciones se identifican a partir de un conjunto de reglas de derivación lingüística específicas del idioma del texto. Por ejemplo, en inglés existe una regla de derivación por la que la terminación del componente de un término con el sufijo `ical` puede derivar de un término que tenga la misma raíz y terminación con el sufijo `ic`. Si se utiliza esta regla (y la desarticulación), el algoritmo debería agrupar los conceptos ingleses `epidemiologic study` y `epidemiological studies`.

Puesto que los términos ya están estructurados en componentes y se han identificado los que son ignorables (por ejemplo, *in* y *of*), el algoritmo de derivación de raíz de conceptos también debería agrupar el concepto inglés *studies in epidemiology* con *epidemiological studies*.

Se ha elegido el conjunto de reglas de derivación de componentes de manera que la mayoría de los conceptos agrupados mediante este algoritmo son sinónimos: los conceptos ingleses *epidemiologic studies*, *epidemiological studies*, *studies in epidemiology* son términos equivalentes. Para aumentar la precisión, existen algunas reglas de derivación que permiten al algoritmo agrupar conceptos que están relacionados por posición. Por ejemplo, el algoritmo puede agrupar conceptos ingleses como *empire builder* y *empire building*.

### ***Inclusión de conceptos***

La técnica de inclusión de conceptos genera categorías tomando un concepto y, mediante los algoritmos de series léxicas, identifica los conceptos que están incluidos en otros conceptos. La idea radica en que cuando las palabras de un concepto forman un subconjunto de otro concepto, refleja una relación semántica subyacente. La inclusión es una potente técnica que puede utilizarse con cualquier tipo de texto.

Esta técnica funciona bien en combinación con las redes semánticas, pero puede utilizarse por separado. La inclusión de conceptos puede arrojar mejores resultados cuando los registros contienen una gran cantidad de jerga o terminología específica del dominio. Esto es especialmente cierto si ha ajustado los diccionarios de antemano para permitir la extracción y agrupación apropiada de términos especiales (con sinónimos).

### ***Funcionamiento de la inclusión de conceptos***

Antes de aplicar el algoritmo de inclusión de conceptos, los términos se estructuran en componentes y se desarticulan. [Si desea obtener más información, consulte el tema Derivación de raíz de conceptos el p. 118.](#) A continuación, el algoritmo de inclusión de conceptos analiza los conjuntos de componentes. Para cada conjunto de componentes, el algoritmo busca otro conjunto de componentes que sea un subconjunto del primer conjunto de componentes.

Por ejemplo, si tiene el concepto *desayuno continental*, que tiene el conjunto de componentes {*continental*, *desayuno*}, y tiene el concepto *desayuno*, que tiene el conjunto de componentes {*desayuno*}, el algoritmo llegará a la conclusión de que *desayuno continental* es un tipo de *desayuno* y los agrupará juntos.

En un ejemplo más extenso, si tenemos el término *seguridad* en el panel Resultados extraídos y aplica este algoritmo, en dicha categoría también se agruparán los conceptos *asiento de seguridad*, *seguridad adicional*, *cinturón de seguridad*, *hebilla del cinturón de seguridad*, *silla infantil de seguridad* y *normativa de seguridad en el automóvil*.

Puesto que los términos ya se han estructurado en componentes y se han identificado los que son ignorables (por ejemplo, *de* y *en*), el algoritmo de inclusión de conceptos reconocerá que el concepto *curso avanzado de español* incluye el concepto *curso de español*.

*Nota:* Puede evitar que los conceptos se agrupen entre sí especificándolos explícitamente. [Si desea obtener más información, consulte el tema Administración de pares de excepciones de enlace el p. 117.](#)

## **Redes semánticas**

En esta versión, la técnica de redes semánticas solo está disponible para texto en inglés.

Esta técnica genera categorías utilizando una red incorporada de relaciones de palabras. Por esta razón, esta técnica puede generar resultados muy buenos cuando los términos son concretos y no contienen demasiadas ambigüedades. Sin embargo, no confíe en que esta técnica encuentre muchos vínculos entre conceptos especializados y muy técnicos. Cuando trabaje con este tipo de conceptos, el empleo de las técnicas de inclusión y derivación de raíz de conceptos le resultará más útil.

### **Funcionamiento de la red semántica**

La idea que encierra la técnica de la red semántica es aprovechar las relaciones de las palabras comunes para crear categorías de sinónimos o hipónimos. Un **hipónimo** es un concepto que constituye una especie de concepto secundario en una relación jerárquica, también conocida como relación “ISA”. Por ejemplo, si `animal` es un concepto, `gato` y `canguro` serían hipónimos de `animal`, puesto que son especies de animales.

Además de las relaciones de sinónimos e hipónimos, la técnica de red semántica también examina enlaces parciales y completos entre los conceptos del tipo <Ubicación>. Por ejemplo, la técnica agruparía los conceptos `normandía`, `provenza` y `francia` en una categoría, porque Normandía y Provenza forman parte de Francia.

Las redes semánticas empiezan identificando los sentidos posibles de cada concepto de la red semántica. Cuando los conceptos se identifican como sinónimos o hipónimos, se agrupan en una sola categoría. Por ejemplo, la técnica crearía una única categoría que contenga estos tres conceptos: `manzana comestible`, `manzana de postre` y `clase reineta`, puesto que la red semántica contiene la información de que: 1) `manzana de postre` es un sinónimo de `manzana comestible`, y 2) `clase reineta` es un tipo de `manzana comestible` (significa que es un hipónimo de `manzana comestible`).

Si se consideran individualmente, muchos conceptos, sobre todo los unitérminos, son ambiguos. Por ejemplo, el concepto `buffet` puede significar un tipo de comida o un mueble. Si el conjunto de conceptos incluye `comida`, `mueble` y `buffet`, el algoritmo se verá forzado a elegir entre agrupar `buffet` con `comida` o con `mueble`. Tenga en cuenta que en algunos casos, las opciones que elige el algoritmo pueden no ser apropiadas en el contexto de un conjunto particular de registros.

La técnica de la red semántica puede generar un mejor rendimiento de la inclusión de conceptos con determinados tipos de datos. Mientras que tanto la red semántica como la inclusión de conceptos reconoce que `pastel de manzana` es un tipo de `pastel`, solo la red semántica reconoce que `tarta` también es un tipo de `pastel`.

Las redes semánticas funcionarán en conjunción con el resto de las técnicas. Por ejemplo, supongamos que ha seleccionado las técnicas de red semántica y de inclusión, y que la red semántica ha agrupado el concepto `profesor` con el concepto `tutor` (porque un tutor es un tipo de profesor). El algoritmo de inclusión puede agrupar el concepto `tutor de doctorado` con `tutor` y, como resultado, los dos algoritmos colaboran para generar una categoría resultante que contiene los tres conceptos: `tutor`, `tutor de doctorado` y `profesor`.

### **Opciones de la red semántica**

Existe una serie de valores adicionales que pueden ser de interés para esta técnica.

- Cambie la Distancia máxima de búsqueda. Seleccione el alcance de búsqueda de las técnicas antes de generar las categorías. Cuanto más bajo sea el valor, menos resultados obtendrá—sin embargo, dichos resultados serán más precisos y será más probable que pueda asociarlos o vincularlos significativamente entre sí. Cuanto más alto sea el valor, más resultados se obtienen; sin embargo, estos resultados pueden ser también menos fiables o relevantes.

Por ejemplo, dependiendo de la distancia, el algoritmo busca desde `melindro` hasta `bollo dulce` (su elemento de grado superior), luego `pastelito` (elemento de grado más superior) y así hacia arriba hasta `pan`.

Al reducir la distancia de búsqueda, esta técnica produce categorías más pequeñas con las que debería ser más fácil trabajar si cree que las categorías que se están produciendo son demasiado grandes o agrupan demasiados elementos.

**Importante:** Además, se recomienda no aplicar la opción Acomodar la ortografía a un límite mínimo de caracteres raíz de (definido en el cuadro de diálogo Extraer) para la agrupación difusa cuando se utiliza esta técnica, puesto que algunas agrupaciones falsas pueden tener un impacto muy negativo en los resultados.

### **Reglas de co-ocurrencia**

Las reglas de co-ocurrencia permiten detectar y agrupar conceptos que están estrechamente relacionados dentro del conjunto de registros. La idea radica en que cuando en los registros a menudo se encuentran conceptos que aparecen juntos, esa co-ocurrencia refleja una relación subyacente que probablemente sea valiosa para las definiciones de categorías. Esta técnica crea reglas de co-ocurrencia que pueden utilizarse para crear una categoría nueva, para ampliar una categoría o como entrada a otra técnica de categoría. Se considera que la co-ocurrencia de dos conceptos es muy alta si estos aparecen con frecuencia juntos en un conjunto de registros y lo hacen raramente separados en el resto de los registros. Esta técnica puede generar buenos resultados con conjuntos de datos más extensos que tengan al menos varios centenares de registros.

Por ejemplo, si muchos registros contienen las palabras `manzanas` y `naranjas`, estos conceptos se podrían agrupar en una regla de co-ocurrencia, `(manzanas & naranjas)`. En otro ejemplo, si los conceptos `mantequilla de cacahuete`, `mermelada` y `bocadillo` aparecen más a menudo juntos que separados, se agruparán en una regla de co-ocurrencia de conceptos, `(mantequilla de cacahuete&mermelada & bocadillo)`.

**Importante:** En versiones anteriores, las reglas de co-ocurrencia y de sinónimos se colocaban entre corchetes. En esta versión, los corchetes ahora indican un resultado de patrón de . En su lugar, las reglas de co-ocurrencia y de sinónimos estarán entre paréntesis, por ejemplo `(sistemas de sonido|altavoces)`.

### **Funcionamiento de las reglas de co-ocurrencia**

Esta técnica explora los registros en busca de dos o más conceptos que tiendan a aparecer juntos. Se considera que dos o más conceptos son co-ocurrentes cuando aparecen con frecuencia juntos en un conjunto de registros y si raramente aparecen separados en cualquiera de los otros registros.

Cuando se encuentran conceptos co-ocurrentes, se genera una regla de categoría. Estas reglas constan de dos o más conceptos conectados entre sí mediante un operador booleano &. Estas reglas son sentencias lógicas que clasifican automáticamente a un registro en una categoría siempre que el conjunto de conceptos de la regla co-ocurrán en ese registro.

### ***Opciones de las reglas de co-ocurrencia***

Si utiliza la técnica de reglas de co-ocurrencia, puede ajustar varios de los valores de configuración que influyen en las reglas resultantes:

- Cambie la Distancia máxima de búsqueda. Seleccione el alcance de búsqueda de las técnicas antes de generar las categorías. Cuanto más bajo sea el valor, menos resultados obtendrá; sin embargo, dichos resultados serán más precisos y será más probable que pueda asociarlos o vincularlos significativamente entre sí. Cuanto más alto sea el valor, más resultados se obtienen; sin embargo, estos resultados pueden ser también menos fiables o relevantes. Al trabajar en co-ocurrencias, el valor predeterminado de la distancia de búsqueda da como resultado muchas co-ocurrencias, muchas de las cuales tienen un vínculo muy débil y por lo tanto no son interesantes. Si reduce la distancia de búsqueda, filtrará las co-ocurrencias más débiles y obtendrá resultados más significativos.
- Número mínimo de registros. Para ayudar a determinar la relevancia de las co-ocurrencias, defina el número mínimo de registros que deben contener una co-ocurrencia determinada para que se utilice como descriptor en una categoría. Con conjuntos de datos más pequeños, cuanto menor sea el valor de esta opción, más fácil será encontrar co-ocurrencias.

*Nota:* Puede evitar que los conceptos se agrupen entre sí especificándolos explícitamente. [Si desea obtener más información, consulte el tema Administración de pares de excepciones de enlace el p. 117.](#)

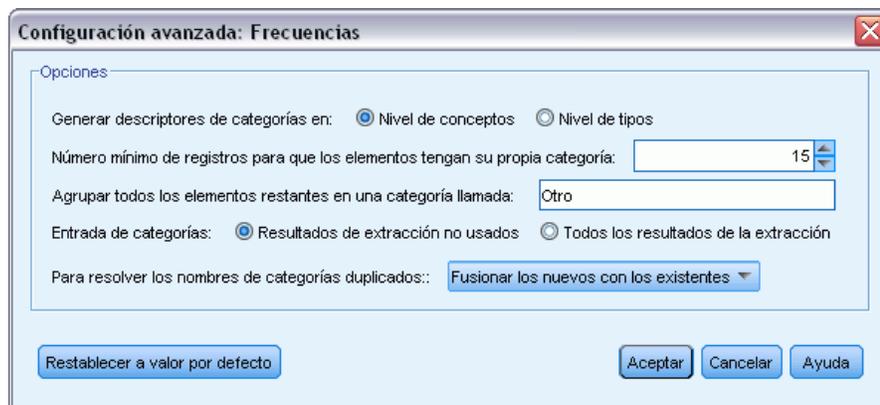
### ***Configuración avanzada de frecuencia***

Puede generar categorías basándose en una técnica de frecuencia mecánica o directa. Con esta técnica, puede generar una categoría para cada elemento (tipo, concepto o patrón) que se haya encontrado por encima del recuento de un registro determinado. También puede generar una sola categoría para todos los elementos que se produzcan con menos frecuencia. Por recuento se entiende el número de registros que contienen el concepto extraído (y cualquiera de sus sinónimos) o el tipo o patrón en cuestión, en contraposición al número total de apariciones en todo el texto.

La agrupación de elementos que aparecen con frecuencia puede arrojar resultados interesantes, ya que puede indicar una respuesta común o significativa. Esta técnica es muy útil si se ejecuta sobre los resultados de extracción sin utilizar después de haber aplicado otras técnicas. Otra aplicación es ejecutar esta técnica inmediatamente después de la extracción si no existe ninguna otra categoría, editar los resultados para suprimir las categorías que no interesen, y luego ampliar esas categorías para que coincidan con más registros. [Si desea obtener más información, consulte el tema Ampliación de categorías el p. 125.](#)

En lugar de utilizar esta técnica, puede clasificar los conceptos o los patrones de conceptos disminuyendo el número de registros en el panel Resultados extraídos y luego arrastrar los principales y soltarlos en el panel Categorías para crear las categorías correspondientes.

Figura 6-12  
Cuadro de diálogo Configuración avanzada: Cuadro de diálogo Frecuencias



**Generar descriptores de categorías en.** Seleccione este tipo de entrada para los descriptores. Si desea obtener más información, consulte el tema [Generación de categorías](#) el p. 110.

- **Nivel de conceptos.** Si selecciona esta opción significa que se utilizarán las frecuencias de los conceptos o de los patrones de conceptos. Se utilizarán conceptos si se han seleccionado tipos como entrada para la generación de categorías, y patrones de conceptos si se han seleccionado patrones de tipo. En general, la aplicación de esta técnica al nivel de conceptos genera resultados más específicos, ya que los conceptos y los patrones de conceptos representan un nivel inferior de medición.
- **Nivel de tipos.** Si selecciona esta opción significa que se utilizarán las frecuencias de tipos o de patrones de tipo. Se utilizarán tipos si se han seleccionado tipos como entrada para la generación de categorías, y patrones de tipos si se han seleccionado patrones de tipo. La aplicación de esta técnica al nivel de tipo permite obtener una perspectiva rápida del amplio rango de respuestas que se da.

**Número mínimo de registros para elementos que tengan su propia categoría.** Esta opción permite generar categorías a partir de elementos que aparecen con frecuencia. Esta opción restringe el resultado sólo a las categorías que contengan un descriptor que haya aparecido como mínimo en X registros, donde X es el valor que hay que introducir para esta opción.

**Agrupar todos los elementos restantes en una categoría llamada.** Esta opción permite agrupar en una única categoría con el nombre de su elección todos los conceptos o tipos que aparecen rara vez. Por defecto, esta categoría se llama *Otros*.

**Entrada de categorías.** Seleccione el grupo al que aplicar las técnicas:

- **Resultados de extracción no usados.** Esta opción permite que se generen categorías a partir de los resultados de extracción que no se utilizan en ninguna categoría existente. De esta manera se minimiza la tendencia que tienen los registros de hacer coincidir varias categorías y limita el número de categorías que se generan.
- **Todos los resultados de la extracción.** Esta opción permite generar categorías utilizando los resultados de extracción. Esto resulta especialmente útil cuando no existen categorías, o existen muy pocas.

**Solucionar nombres de categoría duplicados por.** Seleccione cómo tratar las categorías nuevas cuyos nombres sean iguales que los de categorías existentes. Puede fusionar las nuevas (y sus descriptores) con las categorías existentes que tengan el mismo nombre. Asimismo, puede omitir la creación de categorías si se encuentra un nombre duplicado en las categorías existentes.

## ***Ampliación de categorías***

La ampliación es un proceso a través del cual se añaden descriptores o se mejoran automáticamente para “aumentar” las categorías existentes. El objetivo es generar una categoría mejor que capture los registros relacionados que no se asignaron originalmente a dicha categoría.

Las técnicas automáticas de agrupación que seleccione intentarán identificar conceptos, patrones TLA y reglas de categoría relacionadas con los descriptores de categorías existentes. Estos nuevos conceptos, patrones y reglas de categoría se añadirán como nuevos descriptores, o lo harán a los descriptores existentes. Las técnicas de agrupación para ampliación incluyen la *derivación de raíz de conceptos*, *inclusión de conceptos*, *redes semánticas* (solo para el idioma inglés) y *reglas de co-ocurrencia*. El método Ampliar categorías vacías con descriptores generados desde el nombre de categorías genera descriptores utilizando las palabras de los nombres de categoría, por lo tanto, cuanto más descriptivos sean los nombres de las categorías, mejores serán los resultados.

*Nota:* Las técnicas de frecuencia no están disponibles cuando se amplían categorías.

La ampliación es una excelente manera de mejorar interactivamente las categorías. He aquí algunos ejemplos de cuándo ampliar una categoría:

- Después de arrastrar y soltar patrones de conceptos para crear categorías en el panel Categorías
- Después de crear categorías manualmente y añadir reglas de categoría y descriptores simples
- Después de importar un marco de códigos en el que las categorías tienen nombres muy descriptivos
- Después de refinar las categorías que procedían del TAP que eligió durante la creación del proyecto

Puede ampliar una categoría varias veces. Por ejemplo, si ha importado un archivo de categoría predefinida con nombres muy descriptivos, puede realizar la ampliación utilizando la opción Ampliar categorías vacías con descriptores generados desde el nombre de categorías para obtener un primer conjunto de descriptores, y luego volver a ampliar estas categorías. Sin embargo, en otros casos, realizar la ampliación en distintas ocasiones puede dar como resultado una categoría demasiado genérica si los descriptores se amplían cada vez más. Puesto que las técnicas de agrupación de generación y de ampliación utilizan algoritmos subyacentes similares, es improbable que la ampliación directa después de generar las categorías genere resultados más interesantes.

### **Sugerencias:**

- Si realiza una ampliación y no desea utilizar los resultados, siempre puede deshacer la operación (Editar > Deshacer) inmediatamente después de haber realizado la ampliación.
- La ampliación puede generar dos o más reglas de categoría en una categoría que coincidan exactamente con el mismo conjunto de documentos, puesto que las reglas se generan de manera independiente durante el proceso. Si lo desea, puede revisar las categorías y eliminar

redundancias editando manualmente la descripción de la categoría. [Si desea obtener más información, consulte el tema Edición de descriptores de categoría el p. 158.](#)

### ***Para ampliar categorías***

- ▶ En el panel Categorías, seleccione las categorías que desea ampliar.
- ▶ En los menús elija Categorías > Ampliar categorías. A menos que haya elegido que no se le avise nunca, aparecerá un cuadro de mensaje.
- ▶ Elija si desea generar ahora o editar primero la configuración.
  - Pulse en Ampliar ahora para empezar a ampliar categorías utilizando la configuración actual. El proceso comienza y aparece un diálogo de progreso.
  - Pulse en Editar para revisar y modificar la configuración.

Después de intentar la ampliación, las categorías para las que se encuentren nuevos descriptores se marcan mediante la palabra Ampliado en el panel Categorías, para que pueda identificarlas rápidamente. El texto Ampliado permanecerá hasta que amplíe de nuevo, edite la categoría de otra forma o lo borre mediante el menú contextual.

Figura 6-13  
Cuadro de diálogo Ampliar categorías

Cada una de las técnicas disponibles al generar o ampliar categorías resulta idónea para determinados tipos de datos y situaciones, pero a menudo conviene combinar técnicas en el mismo análisis para capturar el rango completo de registros. Los conceptos y los tipos que se agruparon en una categoría seguirán estando disponibles la próxima vez que genere categorías. Esto significa que puede ver un concepto en diversas categorías o detectar categorías redundantes.

**Entrada de categorías.** Seleccione qué entrada se utilizará para ampliar las categorías:

- **Resultados de extracción no usados.** Esta opción permite que se generen categorías a partir de los resultados de extracción que no se utilizan en ninguna categoría existente. De esta manera se minimiza la tendencia que tienen los registros de hacer coincidir varias categorías y limita el número de categorías que se generan.
- **Todos los resultados de la extracción.** Esta opción permite generar categorías utilizando los resultados de extracción. Esto resulta especialmente útil cuando no existen categorías, o existen muy pocas.

### **Agrupación de técnicas**

Para obtener una breve descripción de cada técnica, consulte [Configuración avanzada: Lingüística](#) el p. 113. Estas técnicas incluyen:

- Derivación de raíz de conceptos (*no disponible para el japonés*)
- Red semántica (*sólo en inglés*)
- Inclusión de conceptos
- Co-ocurrencia y subopción Número mínimo de documentos.

Hay un número de tipos que están permanentemente excluidos de la técnica de redes semánticas, porque dichos tipos no generan resultados relevantes. Incluyen <Positivo>, <Negativo>, <IP>, otros tipos no lingüísticos, etc.

**Distancia máxima de búsqueda.** Seleccione el alcance de búsqueda de las técnicas antes de generar las categorías. Cuanto más bajo sea el valor, menos resultados obtendrá; sin embargo, dichos resultados serán más precisos y será más probable que pueda asociarlos o vincularlos significativamente entre sí. Cuanto más alto sea el valor, más resultados se obtienen; sin embargo, estos resultados pueden ser también menos fiables o relevantes. Esta opción se aplica globalmente en todas las técnicas, pero sus resultados son mejores en las co-ocurrencias y en las redes semánticas.

**Evitar el emparejamiento de conceptos específicos.** Seleccione esta casilla de verificación para detener el proceso de agrupación o emparejamiento de dos conceptos en el resultado. Para crear o administrar pares de conceptos, pulse en Administrar parejas. [Si desea obtener más información, consulte el tema Administración de pares de excepciones de enlace el p. 117.](#)

**Cuando sea posible:** Seleccione si desea simplemente ampliar, generalizar los descriptores mediante comodines o ambas opciones.

- **Ampliar y generalizar.** Esta opción ampliará las categorías seleccionadas y, a continuación, generalizará los descriptores. Si selecciona generalizar, el producto creará reglas de categoría genéricas en categorías mediante el comodín de asterisco. Por ejemplo, en lugar de generar varios descriptores como [naranja de mesa + .] y [naranja de zumo + .], el uso de comodines generaría [naranja \* + .]. Si generaliza con comodines, obtendrá a menudo exactamente el mismo número de registros que antes. Sin embargo, esta opción tiene la ventaja de reducir el número y simplificar los descriptores de categoría. Además, esta opción aumenta la posibilidad de categorizar más registros utilizando estas categorías en datos de texto nuevos (por ejemplo, en estudios por fases o longitudinales).
- **Sólo ampliar.** Esta opción ampliará sus categorías sin generalizar. Puede ser de utilidad seleccionar primero la opción Sólo ampliar para las categorías creadas manualmente y, a continuación, ampliar las mismas categorías de nuevo mediante la opción Ampliar y generalizar.
- **Sólo generalizar.** Esta opción generalizará los descriptores sin ampliar sus categorías de ningún otro modo.

### **Otras opciones para ampliar categorías**

Además de seleccionar las técnicas que se aplicarán, puede editar cualquiera de las opciones siguientes:

**Número máximo de elementos por los que ampliar un descriptor.** Cuando se amplía un descriptor con elementos (conceptos, tipos y otras expresiones), define el número máximo de elementos que pueden añadirse a un solo descriptor. Si establece este límite en 10, no podrá añadir más de 10 elementos adicionales a un descriptor existente. Si hay más de 10 elementos para añadir, las técnicas dejan de añadir elementos nuevos cuando se alcanza el número diez. Con ello se puede reducir la lista de un descriptor, pero no se garantiza que se utilicen en primer lugar los elementos más interesantes. Puede resultar preferible reducir el tamaño de la ampliación sin comprometer la calidad; para ello utilice la opción Generalizar con comodines cuando sea posible. Esta opción sólo se aplica a descriptores que contengan los operadores booleanos & (AND) o ! (NOT).

**Ampliar también subcategorías.** Esta opción también ampliará cualquier subcategoría por debajo de las categorías seleccionadas.

**Ampliar las categorías con descriptores basados en el nombre de las categorías.** Esta opción intenta generar descriptores automáticamente para cada categoría basándose en las palabras que conforman el nombre de la categoría. El nombre de la categoría se explora para comprobar si las palabras que conforman el nombre coinciden con alguno de los conceptos extraídos. Si se reconoce un concepto, se utilizará para buscar patrones de conceptos coincidentes, y ambos se utilizarán para generar descriptores para la categoría. Esta opción genera los mejores resultados cuando los nombres de categoría son largos y descriptivos. Se trata de un método rápido para generar descriptores de categorías, que a su vez permiten a la categoría capturar registros que contienen dichos descriptores. Esta opción es muy útil cuando se importan categorías desde otro punto o cuando crea categorías manualmente con nombres descriptivos largos. Este método solo se aplica a categorías vacías, que tienen 0 descriptores. Si una categoría ya contiene descriptores, no se ampliará de esta forma.

**Generar descriptores como.** Esta opción sólo se aplica si la opción anterior está seleccionada.

- **Conceptos.** Seleccione esta opción para producir los descriptores resultantes en forma de conceptos, independientemente de si se han extraído del texto de origen.
- **Patrones.** Seleccione esta opción para producir los descriptores resultantes en forma de patrones, independientemente de si se han extraído los patrones resultantes o cualquier patrón.

## **Creación manual de categorías**

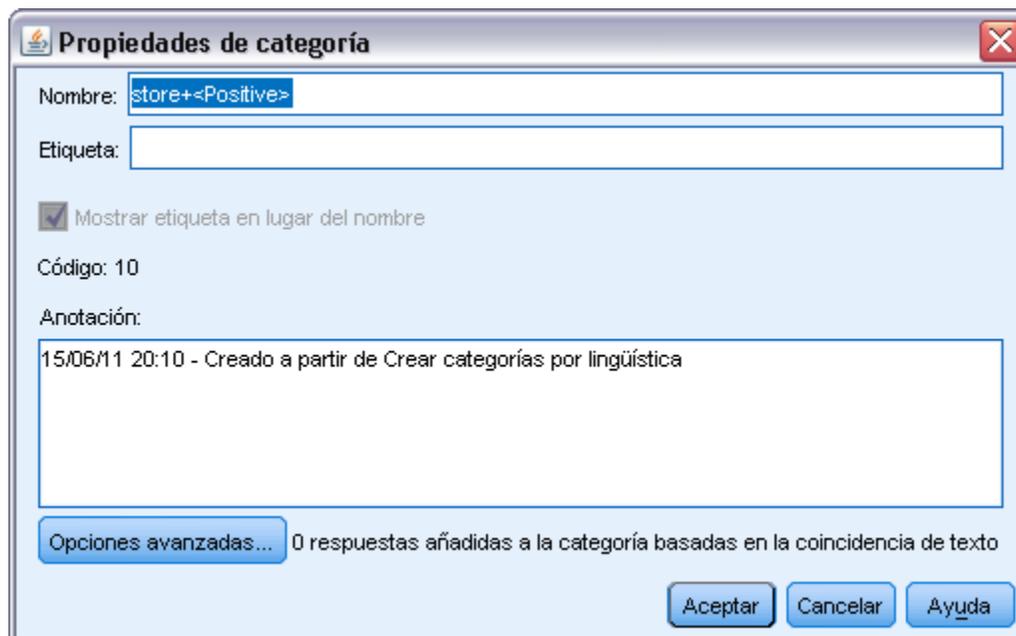
Además de crear categorías utilizando las técnicas de generación automática de categorías, el Administrador de marcos de código y el editor de reglas, también puede crear categorías manualmente. Existen los métodos manuales siguientes:

- Crear una categoría vacía en la que se añadirán elementos uno a uno. [Si desea obtener más información, consulte el tema Creación de categorías nuevas o cambio de nombre de categorías el p. 130.](#)
- Arrastrar términos, tipos y patrones al panel de categorías. [Si desea obtener más información, consulte el tema Creación de categorías mediante el método de arrastrar y soltar el p. 131.](#)

## Creación de categorías nuevas o cambio de nombre de categorías

Puede crear categorías vacías en las que añadir conceptos y tipos. También puede cambiar el nombre de las categorías.

Figura 6-14  
Diálogo Propiedades de categoría



### Para crear una categoría vacía nueva

- ▶ Vaya al panel Categorías.
- ▶ En los menús elija Categorías > Crear categoría vacía. Se abrirá el cuadro de diálogo.
- ▶ Escriba un nombre para esta categoría en el campo Nombre.
- ▶ Pulse en Aceptar para aceptar el nombre y cerrar el cuadro de diálogo. El cuadro de diálogo se cierra y el nombre de la nueva categoría aparece en el panel.

Ahora ya puede empezar a añadir elementos a esta categoría. [Si desea obtener más información, consulte el tema Añadir descriptores a las categorías el p. 157.](#)

### Para cambiar el nombre de una categoría

- ▶ Seleccione una categoría y elija Categorías > Cambiar nombre de categoría. Se abrirá el cuadro de diálogo.
- ▶ Escriba un nuevo nombre para esta categoría en el campo Nombre.
- ▶ Pulse en Aceptar para aceptar el nombre y cerrar el cuadro de diálogo. El cuadro de diálogo se cierra y el nombre de la nueva categoría aparece en el panel.

## Creación de categorías mediante el método de arrastrar y soltar

La técnica de arrastrar y soltar es manual y no se basa en algoritmos. Puede crear categorías en el panel Categorías arrastrando a éste los elementos siguientes:

- Conceptos, tipos o patrones extraídos desde el panel Resultados extraídos al panel Categorías.
- Conceptos extraídos desde el panel Datos del panel Categorías.
- Filas completas desde el panel Datos del panel Categorías. De esta forma se creará una categoría formada por todos los conceptos y patrones extraídos contenidos en dicha fila.

*Nota:* El panel Resultados extraídos da soporte a la selección múltiple para facilitar la función de arrastrar y soltar múltiples elementos.

**Importante:** No pueden arrastrarse y soltarse conceptos desde el panel Datos que no se hayan extraído del texto. Si desea forzar la extracción de un concepto que ha encontrado en los datos, debe añadir este concepto a un tipo. Luego vuelva a ejecutar la extracción. Los nuevos resultados de la extracción contendrán el concepto que acaba de añadir. Luego puede utilizarlo en la categoría. [Si desea obtener más información, consulte el tema Adición de conceptos a tipos en el capítulo 5 el p. 89.](#)

### Para crear categorías utilizando la función de arrastrar y soltar:

- ▶ En el panel Resultados extraídos o en el panel Datos, seleccione uno o más conceptos, patrones, tipos, registros o registros parciales.
- ▶ Mientras mantiene pulsado el botón del ratón, arrastre el elemento hasta una categoría existente o a un área del panel para crear una categoría nueva.
- ▶ Cuando haya alcanzado el área donde desea soltar el elemento, suelte el botón del ratón. El elemento se añade al panel Categorías. Las categorías que se han modificado aparecen con un color de fondo especial. Este color se llama fondo de comentario de categoría. [Si desea obtener más información, consulte el tema Opciones de configuración en el capítulo 2 el p. 17.](#)

*Nota:* A la categoría resultante se le asigna un nombre automáticamente. Si desea cambiar el nombre, puede hacerlo. [Si desea obtener más información, consulte el tema Edición de las propiedades de las categorías el p. 156.](#)

Si desea ver qué registros están asignados a una categoría, seleccione la categoría en el panel Categorías. El panel de datos se actualiza automáticamente y muestra todos los registros de dicha categoría. Para ver el conjunto completo de respuestas de una pregunta, seleccione el nodo Todos los registros en la parte superior del árbol de categorías.

## Importación y exportación de categorías predefinidas

Si tiene sus propias categorías almacenadas en un archivo Microsoft Excel (\*.xls, \*.xlsx), puede importarlas a IBM® SPSS® Text Analytics for Surveys.

También puede exportar las categorías que tenga en un proyecto a un archivo Microsoft Excel (\*.xls, \*.xlsx). Al exportar sus categorías, puede decidir incluir o excluir información adicional como descriptores y puntuaciones. [Si desea obtener más información, consulte el tema Exportación de categorías el p. 142.](#)

*Importante:* En SPSS Text Analytics for Surveys versión 4.0.1, las categorías predefinidas han sustituido en gran medida el uso de marcos de código. Por ejemplo, el Asistente para Importar marcos de código ha sido sustituido por el Asistente de importación de categorías predefinidas. Sin embargo, este nuevo asistente sigue permitiéndole importar los marcos de código existentes que tenga. Además, el Administrador de marcos de código ya no es compatible; para editar valores de código, seleccione Mostrar > Códigos de categoría en los menús para mostrar la columna Código en el panel Categorías y edite los códigos según sea necesario.

Si sus categorías predefinidas no tienen códigos o desea nuevos códigos, puede generar automáticamente un nuevo conjunto de códigos para el conjunto de categorías en el panel Categorías seleccionando Categorías > Administrar categorías > Generar códigos automáticamente en los menús. Esto eliminará los códigos existentes y volverá a numerarlos a todos automáticamente.

### ***Importación de categorías predefinidas***

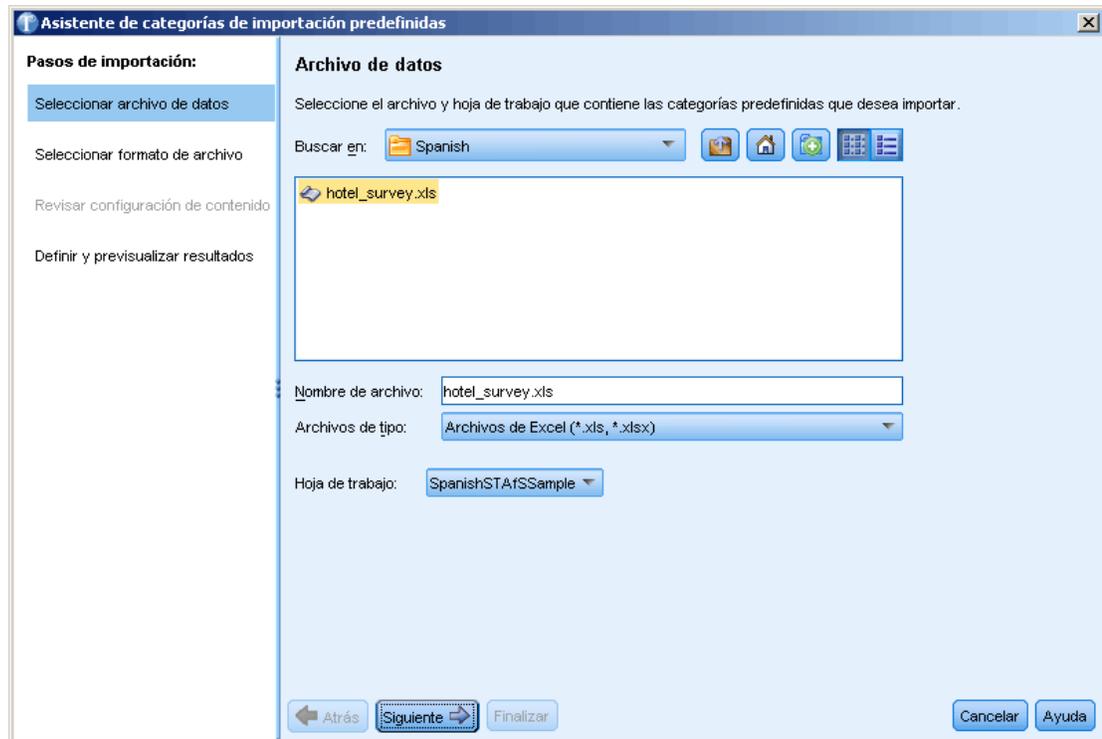
Puede importar sus categorías predefinidas a IBM® SPSS® Text Analytics for Surveys. Antes de la importación, asegúrese de que el archivo de categoría predefinida se encuentre en un archivo Microsoft Excel (\*.xls, \*.xlsx) y esté estructurado en uno de los formatos compatibles. También puede hacer que el producto detecte automáticamente el formato en su lugar. Los siguientes formatos son compatibles:

- Formato de lista plana: [Si desea obtener más información, consulte el tema Formato de lista plana el p. 137.](#)
- Formato compacto: [Si desea obtener más información, consulte el tema Formato compacto el p. 138.](#)
- Formato con sangrado: [Si desea obtener más información, consulte el tema Formato con sangrado el p. 140.](#)

#### ***Para importar categorías predefinidas***

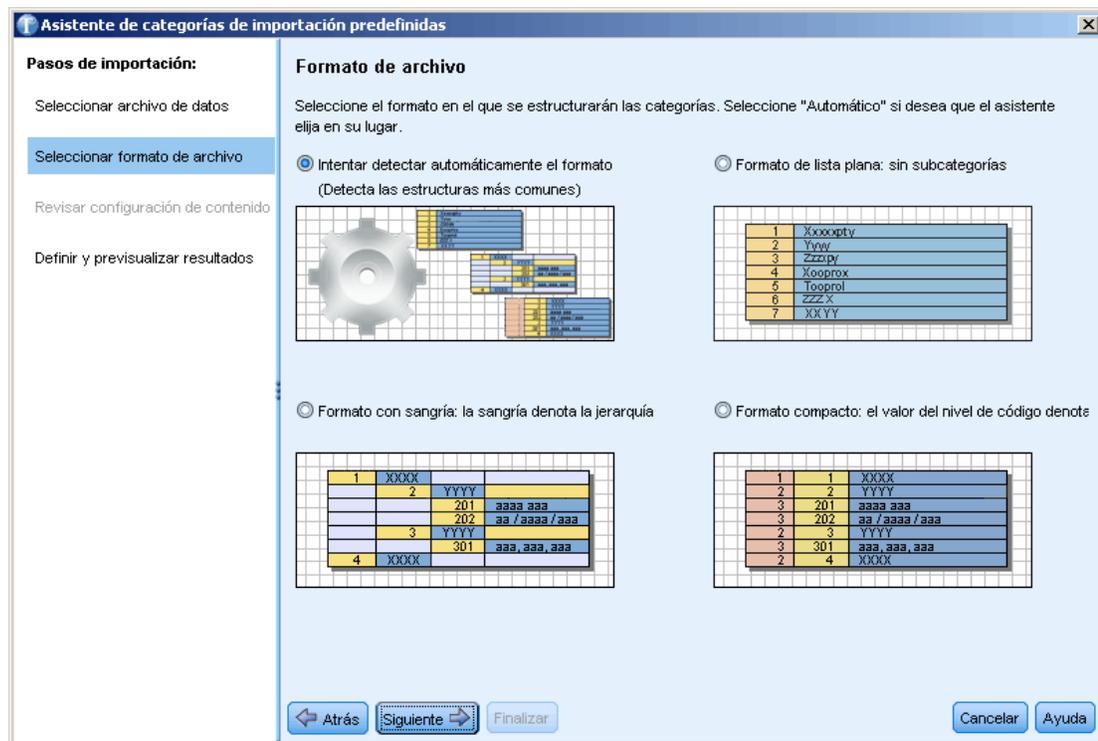
- ▶ En los menús, seleccione Categorías > Administrar categorías > Importar categorías predefinidas. Aparecerá un asistente de importación de categorías predefinidas.

Figura 6-15  
Asistente de importación de categorías predefinidas



- ▶ En la lista desplegable Buscar en, seleccione la unidad y la carpeta donde se encuentra el archivo.
- ▶ Seleccione el archivo de la lista. El nombre del archivo aparece en el cuadro de texto Nombre de archivo.
- ▶ Seleccione la hoja de cálculo que contenga las categorías predefinidas de la lista. El nombre de la hoja de cálculo aparece en el campo Hoja de cálculo.
- ▶ Pulse en Siguiente para empezar a elegir el formato de los datos.

Figura 6-16  
Cuadro de diálogo *Importar categorías predefinidas*, paso *Seleccionar formato de datos*



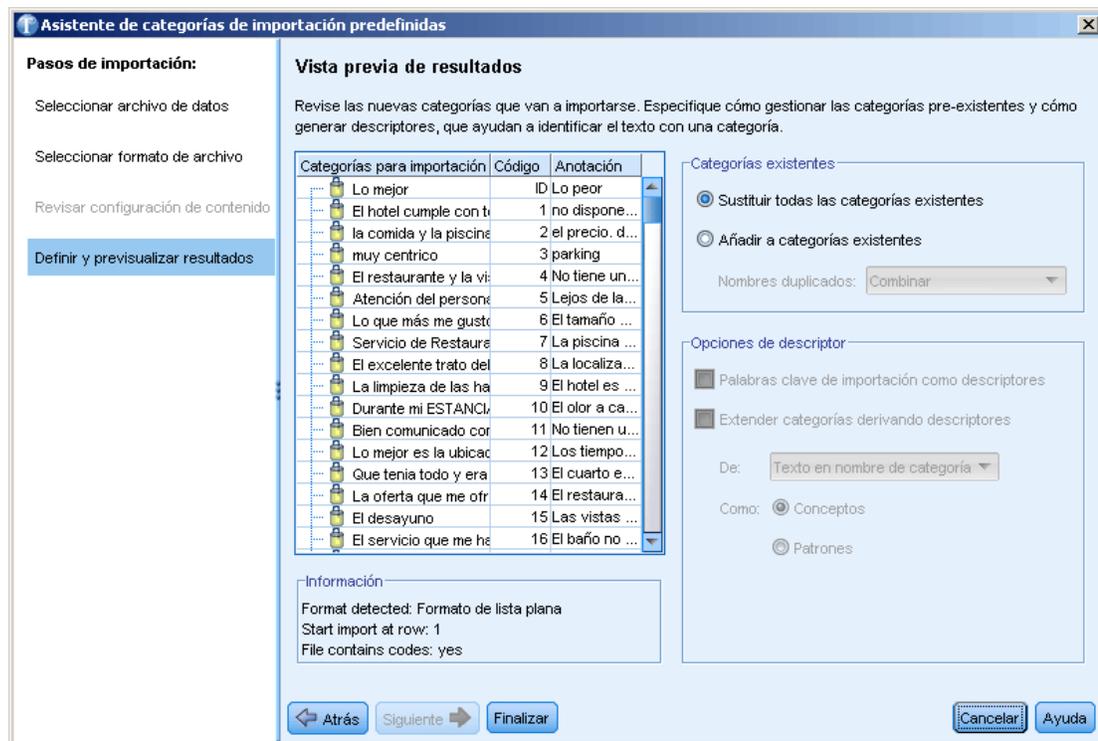
- ▶ Seleccione el formato de su archivo o la opción para permitir que el producto intente detectar el formato automáticamente. La detección automática funciona mejor con los formatos más comunes.
  - Formato de lista plana: [Si desea obtener más información, consulte el tema Formato de lista plana el p. 137.](#)
  - Formato compacto: [Si desea obtener más información, consulte el tema Formato compacto el p. 138.](#)
  - Formato con sangrado: [Si desea obtener más información, consulte el tema Formato con sangrado el p. 140.](#)
- ▶ Pulse en Siguiente para definir las opciones de importación adicionales. Si decide que se detecte el formato automáticamente, se le llevará al paso final.

Figura 6-17  
 Importar categorías predefinidas, paso Opciones de importación



- ▶ Si una o más filas contienen cabeceras de columna u otra información externa, seleccione el número de fila desde el que desee comenzar la importación en la opción Comenzar importación en la fila. Por ejemplo, si sus nombres de categoría empiezan en la fila 7, debe introducir el número 7 en esta opción para que el archivo se importe correctamente.
- ▶ Si su archivo contiene códigos de categorías, seleccione la opción Contiene códigos de categorías. De este modo ayuda a que el asistente reconozca sus datos correctamente.
- ▶ Revise las casillas codificadas por colores y la leyenda para asegurarse de que los datos se han identificado correctamente. Los errores detectados en el archivo se muestran de color rojo y se hace referencia a ellos debajo de la tabla de presentación preliminar del formato. Si se ha seleccionado un formato incorrecto, retroceda y seleccione otro formato. Si necesita realizar correcciones en su archivo, haga dichos cambios y reinicie el asistente volviendo a seleccionar el archivo. Debe corregir todos los errores antes de poder finalizar el asistente.
- ▶ Pulse en Siguiente para revisar el conjunto de categorías y subcategorías que se importarán y definir el modo de crear descriptores para estas categorías.

Figura 6-18  
Cuadro de diálogo *Importar categorías predefinidas*, paso *Presentación preliminar*



- ▶ Revise el conjunto de categorías que se importará a la tabla. Si no ve las palabras clave que esperaba ver como descriptores, puede que no se hayan reconocido durante la importación. Asegúrese de que tienen los prefijos adecuados y de que aparecen en la casilla correcta.
- ▶ Elija de qué forma desea manejar las categorías preexistentes en el proyecto.
  - Sustituir todas las categorías existentes. Esta opción depura todas las categorías existentes y, a continuación, las categorías recién importadas se utilizan en solitario en su lugar.
  - Agregar a las categorías existentes. Esta opción importará las categorías y fusionará las categorías comunes con las categorías existentes. Cuando esté añadiendo categorías a categorías existentes, debe determinar cómo desea tratar los duplicados. Una opción (opción: Fusionar) es fusionar las categorías que se estén importando con las categorías existentes si comparten el nombre de categoría. Otra opción (opción: Excluir de la importación) es prohibir la importación de categorías si existe una con el mismo nombre.
- ▶ Importar palabras clave como descriptores es una opción para importar las palabras clave identificadas en sus datos como descriptores para la categoría asociada.
- ▶ Ampliar categorías mediante la derivación de descriptores es una opción que generará descriptores a partir de las palabras que representan el nombre de la categoría o subcategoría y/o las palabras que componen la anotación. Si las palabras coinciden con resultados extraídos, entonces se añaden como descriptores a la categoría. Esta opción genera los mejores resultados cuando los nombres de categoría o anotaciones son largos y descriptivos. Se trata de un método rápido

para generar los descriptores de categorías que permiten que la categoría capture registros que contengan dichos descriptores.

- El campo De le permite seleccionar de qué texto se derivarán los descriptores, los nombres o categorías y subcategorías, las palabras de las anotaciones, o ambas.
  - El campo Como le permite crear estos descriptores en forma de conceptos o patrones TLA. Si la extracción de TLA no ha tenido lugar, las opciones de patrones están desactivadas en este asistente.
- Pulse en Finalizar para importar las categorías predefinidas en el panel Categorías.

### Formato de lista plana

En este formato de lista plana, sólo hay un nivel superior de categorías sin ninguna jerarquía, lo que significa que no hay subcategorías ni subredes. Los nombres de categorías están en una única columna.

Figura 6-19

Ejemplo de formato de lista plana

	A	B	C	D	E	F	G	H	I	J	K
1		1 Dormitorios									
2		10 Dormitorios/Limpieza									
3		11 Dormitorios/Tranquilidad									
4		silencioso									
5		12 Dormitorios/Comodidad									
6		_confortable									
7		2 Ubicación		medios de transporte públicos							
8		_céntrico									
9		_bien ubicado									
10		_bien situado									
11		3 Personal		trato de los empleados, hospitalidad							
12		_amable									
13		_simpático									
14		_cordial									
15		4 Restaurante		se trata de la comida, del desayuno							
16		5 Precio									
17		_relación calidad-precio									
18		_barato									
19		6 Servicios									
20		61 Servicios/Conexión internet		se trata del acceso a internet							
21		62 Servicios/Piscina		piscina cubierta							
22											
23											
24											

La siguiente información puede incluirse en un archivo con este formato:

- La columna opcional **códigos** contiene valores numéricos que identifican de manera exclusiva cada categoría. Si especifica que el archivo de datos contiene códigos (opción Contiene códigos de categorías en el paso Configuración de contenido), entonces debe existir una columna que contenga códigos únicos para cada categoría en la casilla justo a la izquierda del nombre de categoría. Si sus datos no contienen códigos, pero desea crear códigos posteriormente, siempre podrá hacerlo (Categorías > Administrar categorías > Generar códigos automáticamente). Puede editar los códigos más adelante seleccionando Mostrar > Código de categoría; los códigos se muestran en la columna Código en el panel Categorías donde puede modificarlos manualmente.

- Una columna *necesaria* de **nombres de categoría** contiene todos los nombres de las categorías. Esta columna es obligatoria para importar mediante este formato.
- **Anotaciones** opcionales en la casilla justo a la derecha del nombre de categoría. Esta anotación consiste en texto que describe sus categorías/subcategorías.
- Se pueden importar **palabras clave** opcionales como descriptores para categorías. Para que se las reconozca, estas palabras clave deben existir en la casilla justo debajo del nombre de categoría/subcategoría asociada y la lista de palabras clave debe tener un guión bajo ( \_ ) como prefijo, como en *\_armamento, armas / pistolas*. La casilla de palabras clave puede contener una o más palabras utilizadas para describir cada categoría. Estas palabras se importarán como descriptores o se ignorarán dependiendo de lo que especifique en el último paso del asistente. Posteriormente, los descriptores se comparan con los resultados extraídos del texto. Si se encuentra una coincidencia, entonces ese registro o documento se puntúa en la categoría que contiene este descriptor.

Tabla 6-3

*Formato de lista plana con códigos, palabras clave y anotaciones*

Columna A	Columna B	Columna C
Código de categoría ( <i>opcional</i> )	Nombre de categoría	Anotación
	Lista _descriptor/palabra clave ( <i>opcional</i> )	

### **Formato compacto**

El formato compacto se estructura de forma parecida al formato de lista plana excepto en que el formato compacto se utiliza con categorías jerárquicas. Por lo tanto, se necesita una columna de nivel de código para definir el nivel jerárquico de cada categoría y subcategoría.

Figura 6-20  
Ejemplo de un archivo de categoría predefinida compacto en Microsoft Excel

	A	B	C	D	E	F	G	H	I	J	K	L
1		1	1	Dormitorios								
2		2	10	Limpieza								
3		2	11	Tranquilidad								
4				_silencioso								
5		2	12	Comodidad								
6				_confortable								
7		1	2	Ubicación	medios de transporte públicos							
8				_céntrico								
9				_bien ubicado								
10				_bien situado								
11		1	3	Personal	trato de los empleados, hospitalidad							
12				_amable								
13				_simpático								
14				_cordial								
15		1	4	Restaurante	se trata de la comida, del desayuno							
16		1	5	Precio								
17				_relación calidad-precio								
18				_barato								
19		1	6	Servicios								
20		2	60	Conexión internet	se trata del acceso a internet							
21		2	61	Piscina	piscina cubierta							
22												
23												
24												

La siguiente información puede incluirse en un archivo con este formato:

- La columna **necesaria nivel de código** contiene números que indican la posición jerárquica de la información siguiente de dicha línea. Por ejemplo, si se especifican los valores 1, 2 ó 3 y tiene tanto categorías como subcategorías, entonces 1 es para categorías, 2 para subcategorías y 3 para sub-subcategorías. Si sólo tiene categorías y subcategorías, entonces 1 es para categorías y 2 para subcategorías. Así sucesivamente, hasta la profundidad de categoría deseada.
- La columna opcional **códigos** contiene valores que identifican de manera exclusiva cada categoría. Si especifica que el archivo de datos contiene códigos (opción Contiene códigos de categorías en el paso Configuración de contenido), entonces debe existir una columna que contenga códigos únicos para cada categoría en la casilla justo a la izquierda del nombre de categoría. Si sus datos no contienen códigos, pero desea crear códigos posteriormente, siempre podrá hacerlo (Categorías > Administrar categorías > Generar códigos automáticamente). Puede editar los códigos más adelante seleccionando Mostrar > Código de categoría; los códigos se muestran en la columna Código en el panel Categorías donde puede modificarlos manualmente.
- Una columna **necesaria de nombres de categoría** contiene todos los nombres de las categorías y subcategorías. Esta columna es obligatoria para importar mediante este formato.
- **Anotaciones** opcionales en la casilla justo a la derecha del nombre de categoría. Esta anotación consiste en texto que describe sus categorías/subcategorías.
- Se pueden importar **palabras clave** opcionales como descriptores para categorías. Para que se las reconozca, estas palabras clave deben existir en la casilla justo debajo del nombre de categoría/subcategoría asociada y la lista de palabras clave debe tener un guión bajo ( \_ ) como prefijo, como en *\_armamento*, *armas / pistolas*. La casilla de palabras clave puede contener una o más palabras utilizadas para describir cada categoría. Estas palabras se



- **Los códigos de categoría de nivel superior y nombres de categoría** ocupan las columnas A y B, respectivamente. No obstante, si no hay ningún código, el nombre de categoría está en la columna A.
- **Los códigos de subcategoría y los nombres de subcategoría** ocupan las columnas B y C, respectivamente. No obstante, si no hay ningún código, el nombre de subcategoría está en la columna B. La subcategoría es miembro de una categoría. No puede tener subcategorías si no tiene categorías de nivel superior.

Tabla 6-6

Estructura sangrada con códigos

Columna A	Columna B	Columna C	Columna D
Código de categoría (opcional)	Nombre de categoría		
	Código de subcategoría (opcional)	Nombre de subcategoría	
		Código de sub-subcategoría (opcional)	Nombre de sub-subcategoría

Tabla 6-7

Estructura sangrada sin códigos

Columna A	Columna B	Columna C
Nombre de categoría		
	Nombre de subcategoría	
		Nombre de sub-subcategoría

La siguiente información puede incluirse en un archivo con este formato:

- Los **códigos** opcionales deben ser valores que identifiquen de manera exclusiva cada categoría o subcategoría. Si especifica que el archivo de datos contiene códigos (opción Contiene códigos de categorías en el paso Configuración de contenido), entonces debe existir un código único para cada categoría o subcategoría en la casilla justo a la izquierda del nombre de categoría/subcategoría. Si sus datos no contienen códigos, pero desea crear códigos posteriormente, siempre podrá hacerlo (Categorías > Administrar categorías > Generar códigos automáticamente). Puede editar los códigos más adelante seleccionando Mostrar > Código de categoría; los códigos se muestran en la columna Código en el panel Categorías donde puede modificarlos manualmente.
- Un **nombre obligatorio** para cada categoría y subcategoría. Las subcategorías deben tener un sangrado con respecto a las categorías de una casilla a la derecha en una fila separada.
- **Anotaciones** opcionales en la casilla justo a la derecha del nombre de categoría. Esta anotación consiste en texto que describe sus categorías/subcategorías.
- Se pueden importar **palabras clave** opcionales como descriptores para categorías. Para que se las reconozca, estas palabras clave deben existir en la casilla justo debajo del nombre de categoría/subcategoría asociada y la lista de palabras clave debe tener un guión bajo ( \_ ) como prefijo, como en `_armamento`, `armas / pistolas`. La casilla de palabras clave puede contener una o más palabras utilizadas para describir cada categoría. Estas palabras se importarán como descriptores o se ignorarán dependiendo de lo que especifique en el último paso del asistente. Posteriormente, los descriptores se comparan con los resultados extraídos

del texto. Si se encuentra una coincidencia, entonces ese registro o documento se puntúa en la categoría que contiene este descriptor.

**Importante:** Si utiliza un código a un nivel, debe incluir un código para cada categoría y subcategoría. De lo contrario, el proceso de importación fallará.

## **Exportación de categorías**

También puede exportar las categorías que tenga en un proyecto a un archivo Microsoft Excel (\*.xls, \*.xlsx). Los datos que se exportarán provienen en gran medida del contenido actual del panel Categorías o de las propiedades de categoría. Por lo tanto, le recomendamos que vuelva a obtener la puntuación si tiene la intención de exportar también el valor de puntuación de Documentos.

### **Siempre se exporta...**

- Códigos de categorías, si los hay
- Nombres de categoría (y subcategoría)
- Niveles de código, si los hay (formato *plano/compacto*)
- Cabeceras de columna (formato *plano/compacto*)

### **Se exporta opcionalmente...**

- Puntuaciones de Documentos
- Anotaciones de categoría
- Nombres de descriptor
- Recuentos de descriptores

**Importante:** Cuando exporta descriptores, se convierten en cadenas de texto y se les añade un guión bajo como prefijo. Si vuelve a realizar una importación en este producto, se perderá la capacidad de distinguir entre descriptores que sean patrones, los que sean reglas de categoría y los que sean conceptos planos. Si tiene la intención de volver a utilizar estas categorías en este producto, recomendamos encarecidamente que cree un archivo de paquete de análisis de texto (TAP) en su lugar, ya que el formato TAP conservará todos los descriptores del modo en que estén definidos actualmente, así como todas sus categorías, códigos y los recursos lingüísticos utilizados. Los archivos TAP pueden utilizarse tanto en IBM® SPSS® Text Analytics como en IBM® SPSS® Text Analytics for Surveys. [Si desea obtener más información, consulte el tema Uso de los paquetes de análisis de texto en el capítulo 3 el p. 41.](#)

### **Para exportar categorías predefinidas**

- ▶ En los menús, seleccione Categorías > Administrar categorías > Exportar categorías. Aparecerá un asistente de exportación de categorías.

Figura 6-22  
Asistente de exportación de categorías, paso 1

**Exportar asistente de categorías**

**Pasos de exportación:**

- Seleccionar archivo de datos
- Seleccionar formato de archivo
- Revisar resultados

**Archivo de datos**

Defina un nombre de archivo y ubicación para la exportación de los datos de categoría.

Buscar en: Spanish    

Nombre de archivo:

Archivos de tipo: Archivos de Excel (\*.xls)

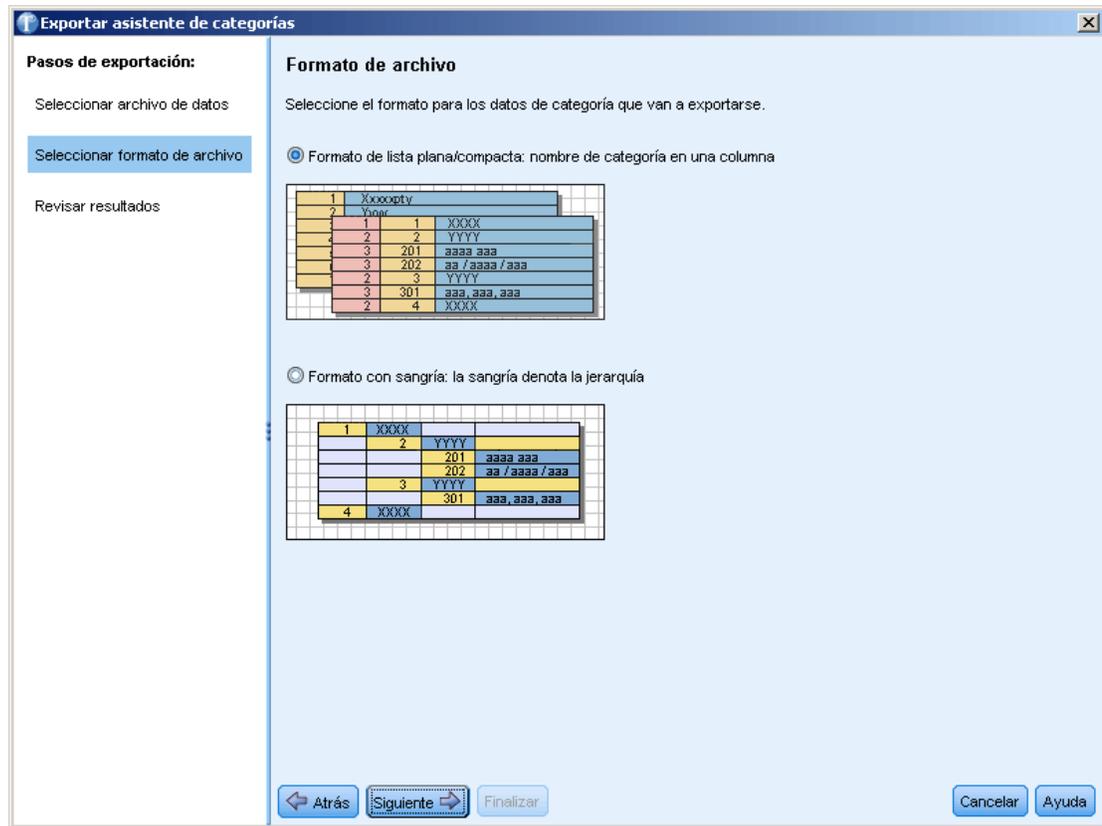
Los descriptores se exportan en forma de cadenas de texto. Por ello, si intenta importar posteriormente, el asistente no puede distinguir entre los distintos descriptores (patrones, reglas, etc). En su lugar, se tratan como conceptos.

Para volver a usar estas categorías en este producto, cree un paquete de análisis de texto (menú Archivo), que conserve los datos de categoría, descriptores y recursos lingüísticos asociados.

← Atrás Siguiete → Finalizar Cancelar Ayuda

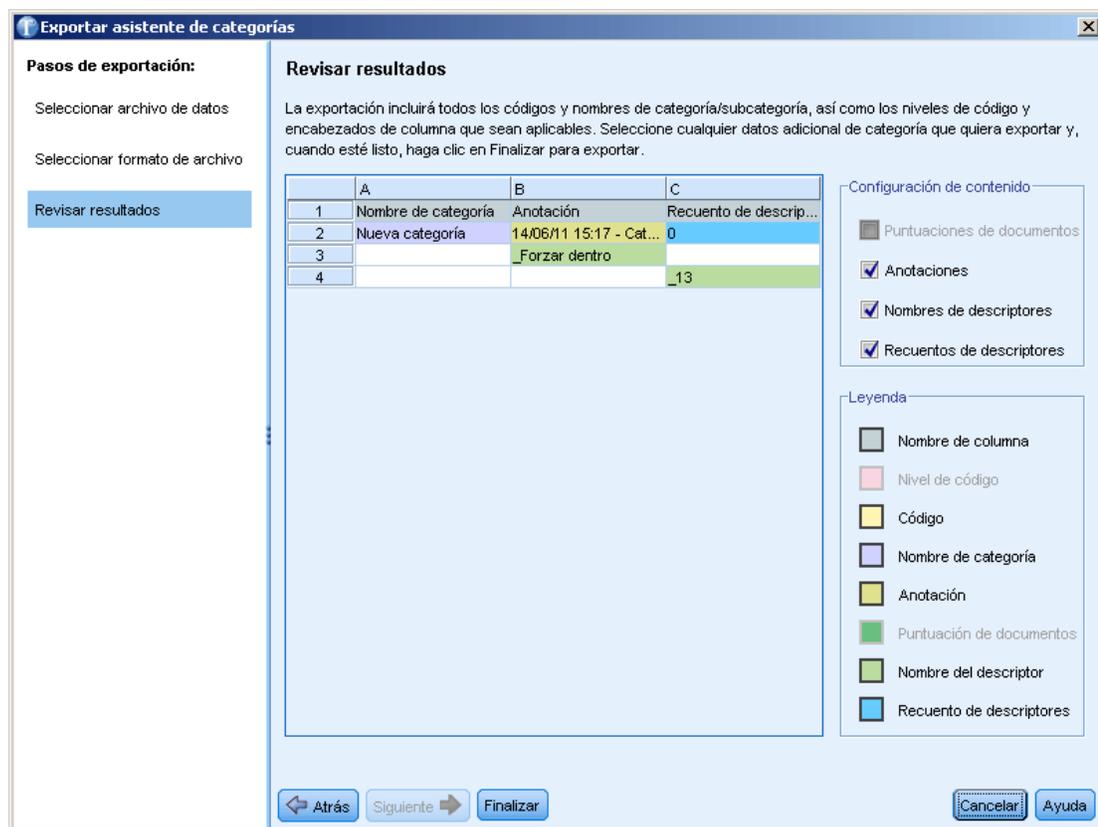
- ▶ Seleccione la ubicación e introduzca el nombre del archivo que se exportará.
- ▶ Introduzca un nombre para el archivo de salida en el cuadro de texto Nombre de archivo.
- ▶ Pulse en Siguiete para seleccionar el formato al que exportará sus datos de categoría.

Figura 6-23  
Asistente de exportación de categorías, paso 2



- ▶ Seleccione el formato entre las siguientes alternativas:
  - Formato de lista plana o compacta: [Si desea obtener más información, consulte el tema Formato de lista plana el p. 137.](#) La lista plana no contiene ninguna subcategoría. [Si desea obtener más información, consulte el tema Formato compacto el p. 138.](#) El formato de lista compacta contiene categorías jerárquicas.
  - Formato con sangrado: [Si desea obtener más información, consulte el tema Formato con sangrado el p. 140.](#)
- ▶ Pulse en Siguiente para empezar a seleccionar el contenido que se exportará y revisar los datos propuestos.

Figura 6-24  
Asistente de exportación de categorías, paso 3



- ▶ Revise el contenido del archivo exportado.
- ▶ Seleccione o elimine la selección de la configuración de contenido adicional que se exportará como Anotaciones o Nombres de descriptor.
- ▶ Pulse en Finalizar para exportar las categorías.

## Uso de reglas de categoría

Puede crear categorías de muchas formas. Una de estas formas es definir reglas de categoría para expresar ideas. Las reglas de categoría son sentencias que clasifican automáticamente registros en una categoría que se basa en una expresión lógica utilizando los conceptos, tipos y patrones extraídos, así como operadores booleanos. Por ejemplo, puede escribir una expresión que signifique *incluir todos los registros que contienen el concepto extraído* embajada *pero no argentina en esta categoría*.

Aunque algunas reglas de categoría se generan de forma automática cuando se crean categorías utilizando técnicas de agrupación como *co-ocurrencia* y *derivación de concepto de raíz* (Categorías > Configuración de generación > Configuración avanzada: Lingüística), también puede crear reglas de categoría manualmente en el editor de reglas utilizando sus conocimientos sobre

categorías en los datos y en el contexto. Cada regla se adjunta a una única categoría, de manera que cada registro que coincida con la regla se registre en dicha categoría.

Las reglas de categoría ayudan a mejorar la calidad y la productividad de los resultados de la minería de datos de texto y del análisis cuantitativo al permitirle categorizar las respuestas con mayor grado de especificidad. Tanto su experiencia como los conocimientos empresariales pueden proporcionarle mayor profundización en los datos y en el contexto. Puede aprovechar esta profundización para trasladar los conocimientos a reglas de categoría y categorizar así los registros con mayor eficacia y precisión, ya que podrá combinar los elementos extraídos con la lógica booleana.

La capacidad de crear estas reglas mejora la precisión de la codificación, la eficacia y la productividad, ya que le permite desglosar en capas sus conocimientos empresariales en la tecnología de extracción del producto.

*Nota:* Entre los ejemplos de cómo coincide la regla con el texto, véase [Ejemplos de reglas de categoría el p. 151](#)

## **Sintaxis de regla de categoría**

Aunque algunas reglas de categoría se generan de forma automática cuando se crean categorías utilizando técnicas de agrupación como *co-ocurrencia* y *derivación de concepto de raíz* (Categorías > Configuración de generación > Configuración avanzada: lingüístico), también pueden crearse reglas de categoría manualmente en el editor de reglas. Cada regla se adjunta a una única categoría, de manera que cada registro que coincida con la regla es anotado automáticamente en dicha categoría.

*Nota:* Entre los ejemplos de cómo coincide la regla con el texto, véase [Ejemplos de reglas de categoría el p. 151](#)

Cuando cree o edite una regla, deberá tenerla abierta en un editor de reglas. Puede añadir conceptos, tipos o patrones y utilizar comodines para ampliar las coincidencias. Cuando se utilizan conceptos, tipos y patrones extraídos, puede beneficiarse de encontrar todos los conceptos relacionados.

**Importante:** para evitar errores comunes, se recomienda arrastrar los conceptos directamente desde el panel Resultados extraídos, desde el panel de análisis de enlace de texto o desde el panel Datos y soltarlos en el editor de reglas o añadirlos mediante los menús contextuales siempre que sea posible.

Cuando se reconocen los conceptos, tipos y patrones, aparece un icono junto al texto.



Concepto extraído



Tipo extraído



Patrón extraído

## **Sintaxis y operadores de reglas**

En la tabla siguiente encontrará los caracteres con los que puede definir la sintaxis de las reglas. Utilice estos caracteres junto con los conceptos, tipos y patrones para crear la regla.

Tabla 6-8  
Sintaxis soportada

Carácter	Descripción
&	El booleano “and”. Por ejemplo, a & b contiene tanto a <i>como</i> b como: - invasión & estados unidos - 2016 & olimpiadas - bueno & manzana
	El booleano “or” es inclusivo, lo que significa que si se encuentran algunos o todos los elementos, existe coincidencia. Por ejemplo, a   b contiene tanto a <i>como</i> b como: - ataque   francia - condominio   apartamento
!()	El booleano “not”. Por ejemplo, !(a) no contiene a. como, !(buen & hotel)
*	Comodín que representa cualquier cosa, desde un solo carácter a una palabra completa dependiendo de cómo se utilice. <a href="#">Si desea obtener más información, consulte el tema Uso de comodines en reglas de categoría el p. 150.</a>
()	Un delimitador de expresión. Las expresiones que están entre paréntesis se evalúan en primer lugar.
+	Conector de patrones que se utiliza para formar un patrón específico de orden. Cuando está presente, deben utilizarse corchetes. <a href="#">Si desea obtener más información, consulte el tema Uso de patrones TLA en las reglas de categoría el p. 148.</a>
[]	Se requiere el delimitador de patrón si está buscando coincidir basándose en un patrón TLA extraído de una regla de categoría. El contenido de los corchetes representa los patrones TLA y no coincidirá nunca con los conceptos o tipos basados en una simple co-ocurrencia. Si no ha extraído este patrón TLA, ninguna coincidencia será posible. <a href="#">Si desea obtener más información, consulte el tema Uso de patrones TLA en las reglas de categoría el p. 148.</a> No utilice corchetes si está buscando coincidir conceptos y tipos en lugar de patrones. <i>Nota:</i> En versiones anteriores, las reglas de co-ocurrencia y de sinónimos generadas por las técnicas de generación de categorías solían estar entre corchetes. En todas las versiones nuevas, los corchetes ahora indican la presencia de un patrón TLA. En su lugar, las reglas generadas por la técnica de co-ocurrencia y de sinónimos estarán entre paréntesis, por ejemplo (sistemas de sonido   altavoces).

El & y | operadores son conmutativos como  $a \& b = b \& a$  y  $a | b = b | a$ .

### ***Salto de caracteres con la barra invertida***

Si tiene un concepto que contiene cualquier carácter que sea también un carácter de sintaxis, deberá colocar una barra inclinada invertida delante de dicho carácter para que la regla se interprete correctamente. El carácter de barra inclinada invertida se utiliza para caracteres de escape que pueden tener un significado especial. Cuando realiza la acción de arrastrar y soltar en el editor, las barras inclinadas invertidas se colocan automáticamente.

Los caracteres de sintaxis de reglas deben ir precedidos por una barra inclinada invertida si quiere que se consideren como tal en vez de como sintaxis de regla:

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

Por ejemplo, puesto que el concepto `r&d` contiene el operador “and” (&), deberá incluir la barra inclinada invertida cuando la escriba en el editor de reglas, así: `r\&d`.

## Uso de patrones TLA en las reglas de categoría

Los patrones de análisis de enlaces de texto se pueden especificar explícitamente en reglas de categoría para permitirle obtener resultados aún más específicos y contextuales. Cuando define un patrón en una regla de categoría, está omitiendo los resultados más sencillos de la extracción de conceptos, y sólo compara los documentos y registros basados en los resultados de patrones de análisis de enlaces de texto extraídos.

**Delimitación con corchetes.** Un patrón TLA debe colocarse entre corchetes [ ] si lo está utilizando dentro de una regla de categoría. Se requiere el delimitador de patrón si está buscando coincidir basándose en un patrón TLA extraído. Puesto que las reglas de categoría pueden contener tipos, conceptos o patrones; los corchetes sirven para aclarar a la regla que su contenido representa el patrón TLA extraído. Si no ha extraído este patrón TLA, ninguna coincidencia será posible. Si ve un patrón sin corchetes como `pastel + bueno` en el panel de Categorías, probablemente signifique que el patrón fue añadido directamente a la categoría fuera del editor de reglas de categoría. Por ejemplo, si añade un patrón de concepto directamente a la categoría desde panel de resultados extraídos, no aparecerá con corchetes. Sin embargo, cuando se utiliza un patrón en una regla de categoría, deberá incluir el patrón entre corchetes dentro de la regla de categoría como `[plátano + !(bueno)]`.

**Uso del signo + en patrones.** En IBM® SPSS® Text Analytics for Surveys, puede tener patrones de dos partes. Cuando desee indicar que el orden es importante, utilice el signo + para conectar los elementos. Si el orden no es importante, puede utilizar el booleano &. En los dos textos siguientes, la posición de la palabra *mejor* es importante: “*el bróquil es mejor*” y “*es mejor que el bróquil*”.

Por ejemplo, supongamos que tiene estas dos expresiones en los textos de muestra siguientes: “*Me gusta la piña*” y “*Detesto la piña. Sin embargo, me gustan las fresas*”. La expresión `gusta & piña` puede hacer coincidir ambos textos ya que se trata de una expresión de conceptos, no una regla de enlace de texto (no está entre corchetes). La expresión `piña+ gusta` sólo coincide con “*Me gusta la piña*” ya que en el segundo texto, la palabra *gusta* está asociada a *fresas* en su lugar.

**Agrupación con patrones.** Puede simplificar las reglas con sus propios patrones. Supongamos que desea capturar las tres expresiones siguientes, `pimienta de cayena + gusta`, `pimienta de chile + gusta`, y `pimientas + gusta`. Puede agruparlos en una sola regla de categoría, como `[* pimientas & gusta]`. Si tiene otra expresión `pimientas picantes + buenas`, puede agrupar las cuatro en una regla como `[* pimientas + <Positivo>]`.

**Orden de patrones.** Para ordenar mejor los resultados, las reglas de análisis de enlace de texto proporcionadas en las plantillas que ha instalado con su producto intentan sacar patrones básicos en el mismo orden independientemente del orden de las palabras en la frase. Por ejemplo, si tiene un registro que contiene el texto “*Buenas presentaciones.*” y otro registro que contiene “*las presentaciones fueron buenas*”, ambos textos coinciden en la misma regla y aparecen en el mismo orden que `presentación + buena` en los resultados de patrón de concepto en vez de `presentación + buena` así como `buena+ presentación`. Y en patrones de dos espacios como en el ejemplo, los conceptos asignados en la biblioteca *Opinions* se presentarán por último en los resultados predeterminados como `pastel + malo`.

Tabla 6-9  
Sintaxis de patrones y uso de booleanos

Expresión	Coincide con un registro que
[ ]	Contiene cualquier patrón TLA. Se requiere el delimitador de patrón <i>en las reglas de categoría</i> si está buscando coincidir basándose en un patrón TLA extraído. El contenido entre los corchetes se refiere a los patrones TLA y no a conceptos y tipos sencillos. Si no ha extraído este patrón TLA, ninguna coincidencia será posible. Si quisiera crear una regla que no incluyera ningún patrón, podría utilizar ! ( [ ] ).
[a]	Contiene un patrón del que al menos un elemento es a independientemente de su posición en el patrón. Por ejemplo, [negocio] puede coincidir con [negocio + bueno] o solo con [negocio + .]
[a + b]	Contiene un patrón de concepto. Por ejemplo, [negocio + bueno]. <i>Nota:</i> Si solo desea capturar este patrón sin añadir ningún otro elemento, se recomienda añadir el patrón directamente a su categoría en lugar de crear una regla con él.
[<A> + <B>]	Contiene cualquier patrón con el tipo <A> en el primer espacio y el tipo <B> en el segundo espacio, y hay exactamente dos espacios. El signo + denota que el orden de los elementos coincidentes es importante. Por ejemplo, [<Budget> + <Negativo>]. <i>Nota:</i> Si solo desea capturar este patrón sin añadir ningún otro elemento, se recomienda añadir el patrón directamente a su categoría en lugar de crear una regla con él.
[<A> & <B>]	Contiene cualquier tipo de patrón con el tipo <A> y el tipo <B>. Por ejemplo, [<Budget> & <Negativo>]. Este patrón TLA no se extraerá nunca; sin embargo, cuando se escribe como tal, es igual a [<Presupuesto> + <Negativo>]   [<Negativo> + <Presupuesto>]. El orden de los elementos coincidentes no es importante.
[a + .]	Contiene un patrón donde a es el único concepto y no existe nada en ningún otro espacio de este patrón. Por ejemplo: [negocio + .] coincide con el patrón de concepto en el que el único resultado es el concepto negocio. Si ha añadido el concepto negocio como descriptor de categoría, obtendrá todos los registros con negocio como concepto incluyendo frases positivas sobre un negocio. Sin embargo, el uso de [negocio + .] coincidirá sólo los resultados de patrón de registros que representen negocio y no otras relaciones u opiniones, y no coincidirá negocio+ fantástico. <i>Nota:</i> Si solo desea capturar este patrón sin añadir ningún otro elemento, se recomienda añadir el patrón directamente a su categoría en lugar de crear una regla con él.
[<A> + <>]	Contiene un patrón donde <A> es el único tipo. Por ejemplo, [<Budget> + <>] coincide con el patrón en el que el único resultado es un concepto del tipo <Budget>. <i>Nota:</i> Puede utilizar <> para denotar un tipo vacío solo cuando se coloca detrás del símbolo + del patrón en el patrón de tipo, como [<Budget> + <>] pero no en [precio + <>]. <i>Nota:</i> Si solo desea capturar este patrón sin añadir ningún otro elemento, se recomienda añadir el patrón directamente a su categoría en lugar de crear una regla con él.
[a + !(b)]	Contiene al menos un patrón que incluye el concepto a pero no incluye el concepto b. Debe incluir al menos un patrón. Por ejemplo, [precio + !(alto)] o para tipos, [!(<Fruta> <Verduras>) + <Positivo>]
! ([<A> & <B>])	No contiene un patrón específico. Por ejemplo, ! ([<Budget> & <Negativo>]).

*Nota:* Entre los ejemplos de cómo coincide la regla con el texto, véase [Ejemplos de reglas de categoría](#) el p. 151

## Uso de comodines en reglas de categoría

Pueden añadirse comodines a los conceptos de las reglas para ampliar las posibilidades de coincidencia. El comodín asterisco (\*) puede colocarse delante y/o detrás de una palabra para indicar cómo deben coincidir los conceptos. Existen dos tipos de usos de los comodines:

- **Comodines afijos.** Estos comodines se colocan inmediatamente antes o después de una cadena sin dejar espacios en blanco entre ésta y el asterisco. Por ejemplo, *opera\** puede coincidir con *operad*, *operar*, *operado*, *operaciones*, *operativo*, etc.
- **Comodines de palabras.** Estos comodines se colocan delante o detrás de un concepto con un espacio entre éste y el asterisco. Por ejemplo, *operación \** puede coincidir con *operación*, *operación quirúrgica*, *postoperación*, etc. Un comodín de palabra puede utilizarse también junto con un comodín afijo, así: *\* opera\* \**, que coincidiría con *operación*, *operación quirúrgica*, *operadora telefónica*, *área de ópera*, etc. Como puede comprobar en este último ejemplo, se recomienda utilizar los comodines con precaución para que el rango no sea excesivamente amplio y no se capturen coincidencias no deseadas.

### ¡Excepciones!

- Un comodín nunca puede ser un elemento individual. Por ejemplo, (*manzana| \**) no se aceptaría.
- Un comodín no se puede utilizar nunca para hacer coincidir nombres de tipo. <Negativo\*> no podrá hacer coincidir ningún nombre de tipo.
- No puede filtrar algunos tipos para evitar que coincidan con conceptos encontrados a través de los comodines. El tipo al que está asignado el concepto se utiliza automáticamente.
- Un comodín no puede nunca estar situado en el centro de una secuencia de palabras, ni al final o el comienzo de una palabra (*abrir\* cuenta*) ni como componente independiente (*abrir \* cuenta*). Tampoco puede utilizar comodines en nombres de tipo. Por ejemplo, *palabra\* palabra*, como *pastel\* carne*, no coincidirá en absoluto con *pastelillo de carne* ni ninguna otra opción. Sin embargo, *pastel\* \** coincidirá con *pastelillo de manzana*, *pastel de chocolate*, *pastel* etc. En otro ejemplo, *palabra \* palabra*, como *pastel \* manzana*, no coincidirá ni con *pastel de manzana ácida* ni con nada ya que el asterisco aparece entre otras dos palabras. Sin embargo, *pastel \* palabra* coincidirá con *pastel de manzana ácida*, *pastel*, *pastel hojaldrado* etc.

Tabla 6-10

Uso de los comodines

Expresión	Coincide con un registro que
*tina	Contiene un concepto que termina con las letras especificadas pero puede tener un número indefinido de letras como prefijo. Por ejemplo: *tina termina con las letras <i>tina</i> pero puede tener unos prefijos como éstos: - mar - cantina - pegatina
mar*	Contiene un concepto que empieza con las letras especificadas pero puede tener un número indefinido de letras como sufijo. Por ejemplo: mar* empieza con las letras <i>mar</i> pero puede tener unos sufijos como éstos: - mar - marinero - maraca Por ejemplo, <i>mar* &amp; !(sal*   barco)</i> , que contiene un concepto que empieza con las letras <i>mar</i> pero no un concepto que empieza con las letras <i>sal</i> o el concepto <i>barco</i> , NO coincidirían: <i>mar&amp; barco</i>

Expresión	Coincide con un registro que
	pero podría coincidir con: - marinero - mar& montaña
*producto*	Contiene un concepto que contiene las letras especificadas producto, pero puede tener un número indefinido de letras como prefijo y/o sufijo. Por ejemplo: *producto* podría coincidir con: - producto - subproducto - productor
* ficción	Contiene un concepto que contiene la palabra ficción pero puede ser un compuesto de otra palabra colocada delante de ella. Por ejemplo, * ficción puede coincidir con: - ficción - ciencia ficción - cine ciencia ficción Por ejemplo, [* telefónico + <Negative>] contiene un concepto que termina con la palabra telefónico en la primera posición y contiene un tipo <Negative> en la segunda posición, lo que podría coincidir con los patrones de conceptos siguientes: - operador telefónico + ineficaz - acceso telefónico + lento
evento *	Contiene un concepto que contiene la palabra evento pero puede ser un compuesto seguido de otra palabra. Por ejemplo, evento * puede coincidir con: - evento - evento deportivo - evento musical multitudinario
* pastel *	Contiene un concepto que puede comenzar con cualquier palabra seguida de la palabra pastel probablemente seguida por otra palabra. * significa 0 o n; por ello coincide también con pastel. Por ejemplo, * pastel * puede coincidir con: - cielo estrellado - primera gran estrella fugaz - famosa estrella polar - mar Por ejemplo, [* reserva* * + <Positivo>], que contiene un concepto con la palabra reserva (independientemente del lugar en que se encuentre el concepto) en la primera posición y contiene un tipo <Positivo> en la segunda posición, puede coincidir con los patrones de concepto siguientes: - reserva de hoteles + bien - reserva en línea + bien

*Nota:* Entre los ejemplos de cómo coincide la regla con el texto, véase [Ejemplos de reglas de categoría](#) el p. 151

## **Ejemplos de reglas de categoría**

Para ayudar a demostrar cómo coinciden las reglas con los registros basados de manera diferente en la sintaxis utilizada para expresarlos, observe el siguiente ejemplo.

### **Ejemplo de registros**

Imagine que tiene dos registros:

- **Registro A:** “Cuando he comprobado mi cartera, he visto que he perdido 5 dólares.”
- **Registro B:** “Se han encontrado 5 dólares en el área de picnic, pero se ha perdido la manta.”

Las dos tablas siguientes muestran qué se puede extraer de los conceptos y tipos así como los patrones de concepto y patrones de tipo.

### **Conceptos y tipos extraídos de un ejemplo**

Tabla 6-11

*Ejemplo de conceptos y tipos extraídos*

Concepto extraído	Conceptos escritos como
cartera	<Desconocido>
perdido	<Negativo>
5 dólares	<Moneda>
manta	<Desconocido>
área de picnic	<Desconocido>

### **Patrones TLA extraídos de un ejemplo**

Tabla 6-12

*Ejemplo de resultados extraídos de patrones TLA*

Patrones de concepto extraídos	Patrones de tipo extraídos	En el registro
área de picnic + .	<Desconocido> + <>	Registro B
cartera + .	<Desconocido> + <>	Registro A
manta + perdido	<Desconocido> + <Negativo>	Registro B
5 dólares + .	<Moneda> + <>	Registro B
5 dólares + perdido	<Moneda> + <Negativo>	Registro A

### **Qué reglas de categoría pueden coincidir**

La siguiente tabla contiene algunas sintaxis que se pueden introducir en el editor de reglas de categoría. No todas las reglas funcionan aquí y no todas coinciden con los mismos registros. Vea cómo afectan las diferentes sintaxis a los registros coincidentes.

Tabla 6-13

*Reglas de muestra*

Sintaxis de la regla	Resultado
5 dólares & perdido	Coincidencias de los registros A y B desde que ambos contienen el concepto extraído perdido y el concepto extraído 5 dólares. Esto equivale a: (5 dólares & perdido)
perdido & 5 dólares	Coincidencias de los registros A y B desde que ambos contienen el concepto extraído perdido y el concepto extraído 5 dólares. Esto equivale a: (perdido & 5 dólares)
perdido & <Moneda>	Coincide con los registros A y B puesto que ambos contienen el concepto extraído perdido y un concepto que coincide con el tipo <Moneda>. Esto equivale a: (perdido & <Moneda>)
<Moneda> & perdido	Coincide con los registros A y B puesto que ambos contienen el concepto extraído perdido y un concepto que coincide con el tipo <Moneda>. Esto equivale a: (<Moneda> & perdido)

Sintaxis de la regla	Resultado
[5 dólares + perdido]	Coincide con A pero <b>no</b> con B puesto que el registro B no genera ningún resultado de patrón TLA que contiene 5 dólares + perdido (consulte la tabla anterior). Esto es equivalente al resultado de patrón TLA: 5 dólares + perdido
[perdido + 5 dólares]	No coincide ni con el registro A ni con B puesto que ningún patrón TLA extraído (consulte la tabla anterior) coincide con el orden expresado aquí usando <i>perdido</i> en la primera posición. Esto es equivalente al resultado de patrón TLA: 5 dólares + perdido
[perdido & 5 dólares]	Coincide con A pero <b>no</b> con B puesto que no se extrajo dicho patrón TLA del registro B. El uso del carácter & indica que el orden no es importante al realizar la coincidencia; por lo que esta regla busca un patrón que coincida indistintamente con [perdido + 5 dólares] o [5 dólares + perdido]. Sólo [5 dólares + perdido] del registro A tiene una coincidencia.
[perdido+ <Moneda>]	No coincide ni con el registro A ni con B ya que ningún patrón TLA extraído coincide con este orden. Esto no tiene equivalente, puesto que el resultado de TLA sólo se basa en términos (5 dólares + perdido) o en tipos (<Moneda> + <Negativo>), pero no mezcla conceptos y tipos.
[<Moneda> + <Negativo>]	Coincide con el registro A pero no con B puesto que no se ha extraído ningún patrón TLA del registro B. Esto equivale al resultado de TLA: <Moneda> + <Negativo>
[<Negativo> + <Moneda>]	No coincide ni con el registro A ni con B ya que ningún patrón TLA extraído coincide con este orden. En la plantilla <i>Opinions</i> , por defecto, cuando un <i>tema</i> se halla con una <i>opinión</i> , el <i>tema</i> (<Moneda>) ocupa el primer espacio y <i>opinión</i> (<Negativo>) ocupa el segundo espacio.

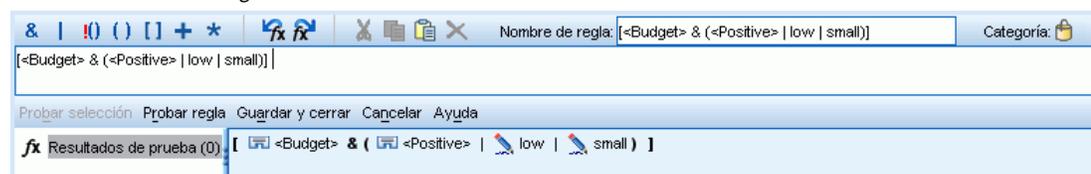
## Creación de reglas de categoría

Cuando cree o edite una regla, deberá tenerla abierta en un editor de reglas. Puede añadir conceptos, tipos o patrones y utilizar comodines para ampliar las coincidencias. Cuando se utilizan conceptos, tipos y patrones reconocidos, tiene la ventaja de que encontrará todos los conceptos relacionados. Por ejemplo, cuando utiliza un concepto, todos sus términos asociados, formas plurales y sinónimos también se asocian con la regla. De la misma forma, cuando utiliza un tipo, la regla también captura todos sus conceptos.

Puede abrir el editor de reglas editando una regla existente o pulsando el botón derecho del ratón en el nombre de la categoría y eligiendo Crear regla.

Figura 6-25

Panel del editor de reglas



Puede utilizar menús contextuales, la acción de arrastrar y soltar o especificar manualmente conceptos, tipos y patrones en el editor. Luego puede combinarlos con operadores booleanos (&, !(), |) y corchetes para crear las expresiones de las reglas. Para evitar errores comunes, se recomienda arrastrar los conceptos directamente desde el panel Resultados extraídos o desde el panel Datos y soltarlos en el editor de reglas. Preste especial atención en la sintaxis de las reglas para evitar errores. [Si desea obtener más información, consulte el tema Sintaxis de regla de categoría el p. 146.](#)

*Nota:* Entre los ejemplos de cómo coincide la regla con el texto, véase [Ejemplos de reglas de categoría el p. 151](#)

### **Para crear una regla**

- ▶ Si no ha extraído aún ningún dato o su extracción está fuera de la fecha, hágalo ahora. [Si desea obtener más información, consulte el tema Extracción de datos en el capítulo 5 el p. 82.](#)
- ▶ En el panel Categorías, seleccione la categoría en la que desea añadir una regla.
- ▶ En los menús elija Categorías > Crear regla. El panel del editor de reglas de categoría se abre en la ventana.
- ▶ En la columna Nombre de regla, escriba un nombre para la regla. Si no proporciona un nombre, se utilizará automáticamente la expresión como nombre. Puede cambiar el nombre de la regla en otro momento.
- ▶ En el campo de texto de expresión más largo puede:
  - Escribir texto directamente en el campo o arrastrarlo desde otro panel y soltarlo aquí. Utilice solo conceptos extraídos, tipos y patrones. Por ejemplo, si especifica la palabra `gatos` pero en el panel Resultados extraídos solo aparece en singular, `gato`, el editor no podrá reconocer `gatos`. En este último caso, el singular puede incluir automáticamente el plural, de lo contrario deberá utilizar un comodín. [Si desea obtener más información, consulte el tema Sintaxis de regla de categoría el p. 146.](#)
  - Seleccione los conceptos, tipos o patrones que desee añadir a las reglas y utilice los menús.
  - Añada operadores booleanos para enlazar los elementos de la regla. Utilice los botones de la barra de herramientas para añadir el booleano “and” `&`, el booleano “or” `|`, el booleano “not” `!`, paréntesis `()` y corchetes para patrones `[]` en la regla.
- ▶ Pulse en el botón Probar regla para verificar que la regla tiene un formato correcto. [Si desea obtener más información, consulte el tema Sintaxis de regla de categoría el p. 146.](#) El número de registros encontrados aparece entre paréntesis junto al texto Resultado de la prueba. A la derecha de este texto puede ver los elementos de la regla que se han reconocido o posibles mensajes de error. Si el gráfico junto al tipo, patrón o concepto aparece con un signo de interrogación rojo, significa que el elemento no coincide con ninguna extracción conocida. Si no coincide, la regla no encontrará ningún registro.
- ▶ Para probar una parte de la regla, seleccione dicha parte y pulse en Probar selección.
- ▶ Realice los cambios necesarios y vuelva a probar la regla si encontró problemas.

- ▶ Cuando termine, pulse en Guardar & Cerrar para guardar otra vez la regla y cerrar el editor. El nuevo nombre de la regla aparece en la categoría.

Figura 6-26  
Regla en el panel Categorías

Categoría	Nº de de...	Nº de res...
Todos los registros	-	405
Sin categorizar	-	255
No se han extraído conceptos	-	4
songs+<Positive>	3	9
memory	3	4
fx [ <Budget> & ( <Positive>   low   small)]		1
fx [ storage + able]		1
memory		3
battery+<Contextual>	2	3
design+<Positive>	2	7
store+<Positive>	2	3
fx [ online store + excellent]		1
fx [ store + able ] [ store + worth]		2
skipping+<Contextual>	2	2
look+<Positive>	2	3

## Edición y eliminación de reglas

Después de crear y guardar una regla, puede editarla en cualquier momento. [Si desea obtener más información, consulte el tema Sintaxis de regla de categoría el p. 146.](#)

Si ya no desea una regla, puede eliminarla.

### Para editar reglas

- ▶ En el panel de categorías, seleccione la regla que desea editar.
- ▶ En los menús elija Categorías > Editar regla o pulse dos veces en el nombre de la regla. El editor se abre con la regla seleccionada.
- ▶ Realice cambios en la regla utilizando los resultados extraídos y los botones de la barra de herramientas.
- ▶ Vuelva a probar la regla para asegurarse de que arroja los resultados esperados.
- ▶ Pulse en Guardar & Cerrar para guardar otra vez la regla y cerrar el editor.

### Para eliminar una regla

- ▶ En el panel de categorías, seleccione la regla que desea eliminar.
- ▶ En los menús elija Editar > Eliminar. La regla se elimina de la categoría.

## **Edición y refinamiento de categorías**

Una vez creadas algunas categorías, querrá examinarlas y realizar en ellas algunos ajustes. Además de refinar los recursos lingüísticos, debe revisar las categorías y averiguar maneras de combinar o limpiar las definiciones, así como comprobar algunos de los registros categorizados. También puede revisar los registros de una categoría y realizar ajustes para que las categorías se definan de tal forma que puedan captarse todos los matices y distinciones.

Puede utilizar las técnicas automáticas de generación para crear las categorías; sin embargo, seguramente querrá realizar algunos ajustes en estas categorías. Después de utilizar una técnica o más, en la ventana aparecerán una serie de categorías nuevas. Luego puede revisar los datos de una categoría y realizar ajustes hasta que esté conforme con las definiciones de categoría. [Si desea obtener más información, consulte el tema Acerca de las categorías el p. 107.](#)

A continuación mostramos algunas opciones para refinar sus categorías, la mayoría de las cuales están descritas en las páginas siguientes:

- Editar las propiedades de las categorías (cambiar el nombre, añadir etiquetas, añadir anotaciones)
- Adición de descriptores a las categorías
- Edición de categorías
- Mover categorías
- Aplanar categorías jerárquicas
- Fusionar categorías
- Añadir texto coincidente
- Forzar respuestas dentro de las categorías
- Copiar y reutilizar categorías
- Eliminar categorías
- Realizar cambios en los recursos lingüísticos y repetir la extracción.
- Visualizar el modo de funcionamiento conjunto de las categorías y realizar ajustes. [Si desea obtener más información, consulte el tema Visualización de gráficos en el capítulo 7 el p. 167.](#)

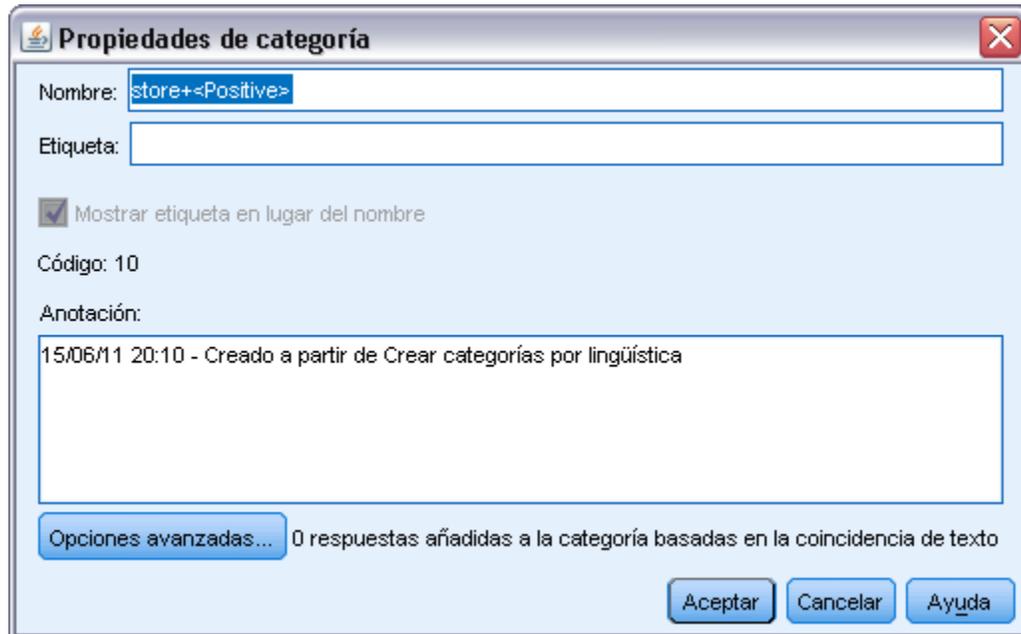
### **Edición de las propiedades de las categorías**

Al igual que ocurre con otros muchos elementos en IBM® SPSS® Text Analytics for Surveys, puede editar las propiedades de las categorías: nombre, etiqueta, anotaciones y entradas avanzadas de coincidencia de texto. [Si desea obtener más información, consulte el tema Propiedades de categoría el p. 108.](#) Además de las propiedades que puede editar, también puede ver el número de elementos que se incluye en la definición de la categoría, es decir, el número de términos, tipos, patrones TLA o reglas de categoría que conforman dicha categoría. También se muestra el número de código y se corresponde con el valor de código encontrado en el Administrador de marcos de código.

**Para editar las propiedades de categorías:**

- ▶ En los menús elija Categorías > Propiedades de categoría. Aparecerá el cuadro de diálogo Propiedades de categoría.

Figura 6-27  
Diálogo Propiedades de categoría



- ▶ Si lo desea puede cambiar el nombre de la categoría especificando un nombre nuevo en el campo Nombre.
- ▶ Cambie el nombre o la etiqueta de la categoría.
- ▶ Para utilizar la etiqueta en la interfaz, como en el panel Categoría, en lugar del nombre de categoría, seleccione Mostrar etiqueta en lugar del nombre.
- ▶ Si lo desea, escriba una anotación en el campo Anotación.
- ▶ Para forzar una palabra o frase en una definición de categoría, pulse en Opciones avanzadas y escriba las coincidencias del texto en la tabla. [Si desea obtener más información, consulte el tema Coincidencia de texto en las categorías el p. 162.](#)
- ▶ Pulse en Aceptar para aplicar los cambios.

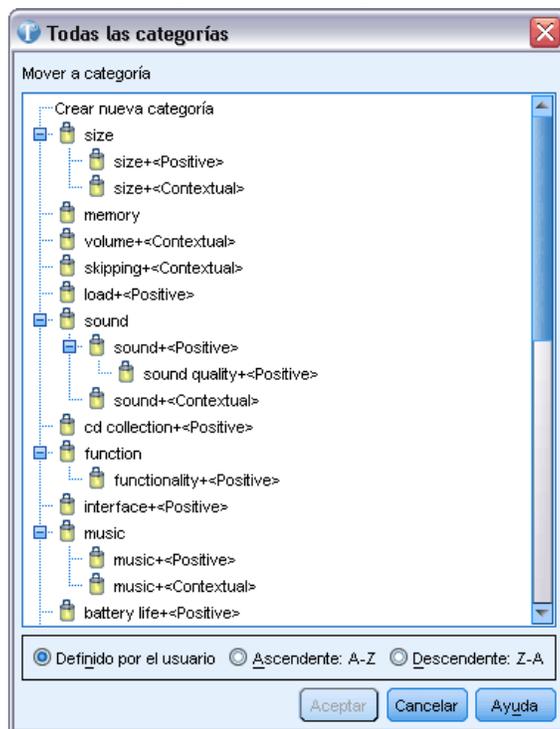
**Añadir descriptores a las categorías**

Después de utilizar las técnicas automáticas, lo más probable es que aún tenga resultados extraídos que no se utilizaron en ninguna definición de categoría. Debe revisar esta lista en el panel Resultados extraídos. Si encuentra elementos que desearía mover a una categoría, puede añadirlos a una categoría nueva o existente.

**Para añadir un concepto o un tipo a una categoría**

- ▶ Desde los paneles Datos y Resultados extraídos, seleccione los elementos que desea añadir a una categoría nueva o existente.
- ▶ En los menús elija Categorías > Añadir a categoría. El cuadro de diálogo Todas las categorías presenta el conjunto de categorías. Seleccione la categoría a la que desee añadir los elementos seleccionados. Si desea añadir los elementos a una categoría nueva, seleccione Nueva categoría. Aparece una categoría nueva en el panel Categorías con el nombre del primer elemento seleccionado.

Figura 6-28

*Cuadro de diálogo Todas las categorías***Edición de descriptores de categoría**

Una vez que haya creado algunas categorías, puede abrir cada una de ellas para ver todos los descriptores que conforman su definición. En el cuadro de diálogo Definiciones de categoría, puede realizar una serie de ediciones en los descriptores de categoría. Además, si se muestran las categorías en el árbol de categorías, puede también trabajar con ellos allí mismo.

**Para editar una categoría**

- ▶ Seleccione la categoría que desea editar en el panel Categorías.
- ▶ En los menús elija Ver > Definiciones de categoría. Aparecerá el cuadro de diálogo Definiciones de categoría.

Figura 6-29  
Cuadro de diálogo Definiciones de categoría



- Seleccione el descriptor que desea editar y pulse en el botón correspondiente de la barra de herramientas.

En la tabla siguiente se describe cada botón de la barra de herramientas con los que podrá editar las definiciones de categoría.

Tabla 6-14  
Descripciones y botones de la barra de herramientas

Iconos	Descripción
	Elimina los descriptores seleccionados de la categoría.
	Mueve los descriptores seleccionados a una categoría nueva o existente.
	Mueve los descriptores seleccionados en forma de una regla de categoría & a una categoría. <a href="#">Si desea obtener más información, consulte el tema Uso de reglas de categoría el p. 145.</a>
	Mueve cada uno de los descriptores seleccionados como su propia categoría nueva
Representación	Actualiza lo que se muestra en el panel Datos y en el panel Visualización en función de los descriptores seleccionados

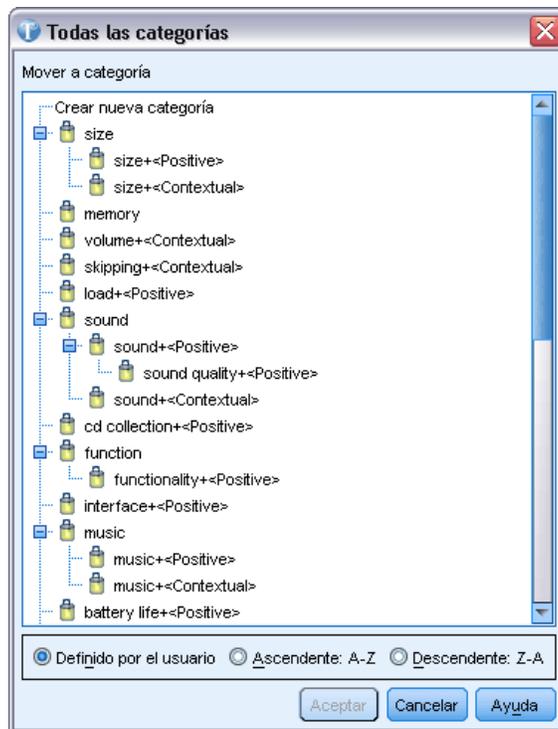
## Cómo mover categorías

Si desea colocar una categoría en otra categoría, o mover descriptores a otra categoría, puede moverla.

**Para mover una categoría**

- ▶ En el panel Categorías, seleccione las categorías o descriptores que desea mover a otra categoría.
- ▶ En los menús elija Categorías > Mover a categoría. El menú muestra un conjunto de categorías, y la creada más recientemente figura al principio de la lista. Seleccione el nombre de la categoría a la que desea mover los conceptos seleccionados.
  - Si puede ver el nombre que está buscando, selecciónelo; acto seguido los elementos seleccionados se añaden a dicha categoría.
  - Si no lo encuentra, seleccione Más para mostrar el cuadro de diálogo Todas las categorías, y seleccione la categoría de la lista que aparece.

Figura 6-30  
Cuadro de diálogo Todas las categorías

**Fusión o combinación de categorías**

Si desea combinar dos o más categorías en una categoría nueva, puede fusionarlas. Cuando se fusionan categorías, se crea una categoría nueva con un nombre genérico. Todos los conceptos, tipos y patrones que se utilizan en los descriptores de categorías se mueven a la nueva categoría. Más tarde podrá cambiar el nombre de la categoría editando sus propiedades. [Si desea obtener más información, consulte el tema Edición de las propiedades de las categorías el p. 156.](#)

**Para fusionar una categoría o parte de una categoría**

- ▶ En el panel Categorías, seleccione los elementos que desea fusionar.

- ▶ En los menús elija Categorías > Fusionar categorías. Aparecerá el cuadro de diálogo Propiedades de categoría en el que debe introducir un nombre para la categoría recién creada. Las categorías seleccionadas se combinan en la nueva categoría como subcategorías.

## **Forzado de respuestas dentro de categorías**

Forzar respuestas dentro y fuera de las categorías permite sustituir las definiciones de categoría creadas por las técnicas automáticas de generación de categorías sin cambiar la definición real de la categoría. Es posible que, aunque la respuesta contenga términos que se utilizan para definir una categoría determinada, la respuesta propiamente dicha no debería estar en dicha categoría. En tal caso, puede forzar la respuesta fuera de dicha categoría sin necesidad de eliminar los términos de la definición de categoría.

La acción de Forzar se utiliza en casos especiales en que una respuesta se adapta, o no, a una categoría, pero que por alguna razón, por ejemplo, porque contiene un término determinado, se asigna o no a dicha categoría. Esto suele ocurrir cuando el encuestado utiliza el sarcasmo para responder, por ejemplo “*La pizza estaba riquísima. Estoy segura que a todo el mundo le gusta la pizza fría y requemada.*” Supongamos que tiene una categoría llamada Pos: [`<Comida>` + `<Positive>`] para capturar las opiniones positivas sobre la comida que se sirve en un restaurante; esta respuesta podría asignarse a dicha categoría. En este caso, probablemente quiera forzar esta respuesta fuera de la categoría.

### **Para forzar respuestas dentro o fuera de categorías**

- ▶ En el panel Datos seleccione la respuesta que desea forzar dentro o fuera de una categoría determinada.
- ▶ En los menús elija Categorías > Forzar respuesta dentro o Categorías> Forzar respuesta fuera. En un submenú aparecerá la lista de categorías para seleccionar.
- ▶ Seleccione la categoría en la que, o fuera de la que, desea forzar la respuesta. Si ha creado muchas categorías, es posible que no todas aparezcan en el submenú.
  - En ese caso, elija Más en la parte inferior del submenú. Se abrirá el cuadro de diálogo Todas las categorías, en el que podrá seleccionar la categoría y pulsar en Aceptar para aplicar el cambio.
  - Si desea forzar la respuesta dentro de una categoría nueva, seleccione Nueva categoría vacía. Aparece una nueva categoría en el árbol de categorías con un nombre genérico.

Cuando una categoría contiene una o más respuestas forzadas, aparece una pseudo-categoría denominada Forzar dentro o Forzar fuera debajo del nombre de categoría en el árbol. También puede indicar qué respuestas se fuerzan dentro o fuera de una categoría mostrando las columnas Forzar dentro y Forzar fuera en el panel Datos. [Si desea obtener más información, consulte el tema El panel de datos el p. 97.](#)

### **Para borrar un estado de respuesta forzada**

- ▶ En el panel Datos seleccione la respuesta que ya no desea forzar dentro o fuera de una categoría.

- En los menús elija Categorías > Forzar respuesta dentro para forzar la respuesta dentro, o elija Categorías> Forzar respuesta fuera para forzarla fuera. Las categorías en las que la respuesta se ha forzado dentro o fuera están precedidas por una marca de verificación.

Figura 6-31  
Forzar respuestas desde el panel Datos

	Id.	Respuesta	Categorías
1	5	Atención del personal	personal
2	8	El excelente trato del personal de todo el Hotel. La amplitud de las habitaciones y el desayuno	habitaciones desayuno hotel personal
3	34	el personal es muy amable	personal
4	42	e	personal
5	44		personal
6	46		personal
7	69		personal
8	87		personal
9	91		personal
10	92		personal
11	100		habitaciones personal
12	113		personal
13	119		personal
14	121	personal exquisito	personal
15	136	el personal y las habitaciones	habitaciones personal
16	182	personal muy amable y atento	personal

Añadir a regla Ctrl+U		Nueva categoría vacía...
<input checked="" type="checkbox"/>	Forzar respuesta dentro	<input type="checkbox"/> hotel
<input checked="" type="checkbox"/>	Forzar respuesta fuera	<input type="checkbox"/> desayuno
<input type="checkbox"/>	Marcar respuestas con	<input type="checkbox"/> comida
<input type="checkbox"/>	Ordenar	<input checked="" type="checkbox"/> empleados
<input type="checkbox"/>	Seleccionar todo Ctrl+A	<input type="checkbox"/> transporte
<input type="checkbox"/>	Copiar Ctrl+C	<input type="checkbox"/> dueños
<input type="checkbox"/>	Buscar Ctrl+F	<input type="checkbox"/> comunicación
<input type="checkbox"/>	Mostrar columnas	<input type="checkbox"/> camas
		<input type="checkbox"/> ubicación
		<input type="checkbox"/> gente
		<input type="checkbox"/> acceso a internet

- Seleccione la categoría en el submenú que está marcada y para la que desea eliminar la acción de forzar. La marca de verificación se borra y la respuesta deja de estar forzada.

#### **Para borrar un estado de respuesta forzada**

#### **Para borrar todos los estados de respuesta forzados:**

- En los menús elija Categorías > Borrar todos > Forzar dentro o Categorías > Borrar todos > Forzar fuera. El estado forzado de las respuestas se borra y éstas dejan de estar forzadas dentro o fuera de las categorías.

### **Coincidencia de texto en las categorías**

Si ha intentado forzar la extracción de una palabra o frase a través de los recursos lingüísticos y todavía no ha conseguido la extracción debido a otras reglas de gestión lingüística, puede crear una entrada de coincidencia de texto para asignar directamente las categorías que contienen dicho texto a una categoría determinada sin utilizar los resultados extraídos.

Cuando se añaden entradas de coincidencia de texto (una palabra o frase) en una categoría, IBM® SPSS® Text Analytics for Surveys asignará automáticamente las respuestas que contienen la palabra o frase a esta categoría. La coincidencia de texto solo debe utilizarse si ya ha intentado añadir esta palabra a los recursos lingüísticos para beneficiarse al máximo de esta definición. [Si desea obtener más información, consulte el tema Forzado de términos en el capítulo 10 el p. 222.](#)

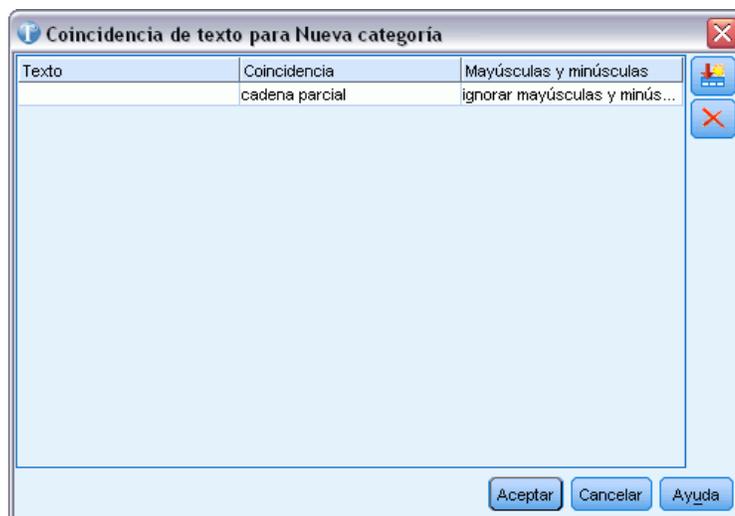
Cuando una coincidencia de texto se añade a una definición de categoría, aparece una pseudo-categoría denominada Coincidencia de texto debajo de la categoría en el árbol de categorías. [Si desea obtener más información, consulte el tema El panel de datos el p. 97.](#)

También puede eliminar entradas de esta tabla seleccionando la fila o filas que desea eliminar y pulsando en el botón Eliminar.

**Para añadir una entrada de coincidencia de texto a una definición de categoría:**

- ▶ En el panel Datos, identifique la palabra o frase que desea forzar dentro de una definición de categoría.
- ▶ En el panel Categorías, seleccione la categoría dentro de la que desea forzar esta palabra o frase.
- ▶ En los menús elija Categorías > Propiedades de categoría. Aparecerá el cuadro de diálogo Propiedades de categoría.
- ▶ Pulse en Opciones avanzadas. Aparecerá el cuadro de diálogo Coincidencia de texto.

Figura 6-32  
Cuadro de diálogo Coincidencia de texto



- ▶ En la tabla, escriba la palabra o frase en la primera casilla de la columna Texto.
- ▶ Seleccione la forma en que esta palabra o frase debe coincidir con el texto que se encuentra en las respuestas. Para que la palabra o frase coincida exactamente con lo que haya escrito, seleccione la palabra o frase completa. Para que la palabra o frase coincida con frases más largas, seleccione cadena parcial.
- ▶ Si la palabra o frase que ha introducido distingue entre mayúsculas y minúsculas, seleccione coincidir mayúsculas y minúsculas en la columna Mayúsculas y minúsculas.
- ▶ Pulse en Aceptar para guardar los cambios y cerrar el cuadro de diálogo. El número de respuestas asignadas a la categoría utilizando entradas de coincidencia de texto se actualiza y se muestra en el cuadro de diálogo.
- ▶ Pulse en Aceptar para aplicar los cambios.

## ***Copia de categorías***

Si utiliza las mismas preguntas u otras similares en una o más encuestas, la reutilización de definiciones de categorías es una opción que le ahorrará mucho tiempo. Puede copiar las categorías de una pregunta a otra en el mismo proyecto. Al reutilizar categorías debe repetir la extracción para poder hacer coincidir las categorías con los datos de respuestas. Antes de volver a extraer, las categorías aparecen en el panel Categorías con un signo de interrogación (?) en lugar del recuento de frecuencia.

*Nota:* Para reutilizar categorías en otro proyecto se recomienda hacer un paquete de análisis de texto (TAP) con las categorías y los recursos del proyecto y utilizar dicho paquete cuando cree el nuevo proyecto con el asistente. [Si desea obtener más información, consulte el tema Uso de los paquetes de análisis de texto en el capítulo 3 el p. 41.](#)

### ***Para copiar categorías en el mismo proyecto***

- ▶ Vaya a la pregunta cuyas categorías desea copiar.
- ▶ En el árbol del panel Categorías seleccione todas las categorías.
- ▶ En los menús elija Editar > Copiar para copiar las categorías.
- ▶ Vaya a la pregunta (Ver > Pregunta) en la que desea pegar estas categorías.
- ▶ En los menús elija Editar > Pegar para pegar las categorías. Las categorías se añaden al panel. Los recuentos de frecuencia no están disponibles porque no ha realizado la reextracción. Por lo tanto, en lugar de los recuentos aparecen signos de interrogación (?).

## ***Impresión de categorías***

Puede imprimir la vista de árbol del panel Categorías.

### ***Para imprimir la vista de árbol de categorías***

- ▶ En el árbol del panel Categorías, expanda, contraiga u ordene los tres elementos según lo que desea que se imprima.
- ▶ En los menús elija Archivo > Imprimir > Imprimir categorías. Aparece el cuadro de diálogo Presentación preliminar.

Figura 6-33  
Cuadro de diálogo Presentación preliminar

The screenshot shows a window titled 'Presentación preliminar' with a toolbar at the top containing icons for home, print, page number (Página 1), file name (de4), zoom (100%), and close. The main area displays a tree view of categorized text data. The root node is 'Todos los registros' with a count of 405. Below it are 'Sin categorizar' (255) and 'No se han extraído conceptos' (4). The tree continues with various categories like 'songs+<Positive>', 'memory', 'battery+<Contextual>', 'design+<Positive>', etc. The 'memory' category is highlighted in yellow. At the bottom, there is a print icon.

Todos los registros	-	405
Sin categorizar	-	255
No se han extraído conceptos	-	4
songs+<Positive>	3	9
memory	3	4
fx [ memory + easy to use]		1
fx [ storage + able]		1
memory		3
battery+<Contextual>	2	3
design+<Positive>	2	7
store+<Positive>	2	3
fx [ online store + excellent]		1
fx [ store + able]   [ store + worth]		2
skipping+<Contextual>	2	2
look+<Positive>	2	3
feature+<Positive>	2	3
fx [ feature + excellent]		2
features of product+<Positive>	1	1
software+<Positive>	1	3
function	2	2
fx [ function + like]		1
functionality+<Positive>	1	1
playlists+<Positive>	1	2
fx [ playlists + like]   [ playlists + able to build]		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
music+<Positive>	9	24
music+<Contextual>	3	5
sound	7	23
sound+<Positive>	5	23

- Pulse en el botón de impresión para imprimir la vista tal como aparece en el cuadro de diálogo.

## ***Eliminación de categorías***

Si ya no desea conservar una categoría, puede eliminarla. Cuando se elimina una categoría, los conceptos que no se utilizan en otra categoría quedan visibles en la pestaña Extracciones no usadas en el panel Resultados extraídos.

### ***Para eliminar una categoría***

- ▶ En el panel Categorías, seleccione la categoría o categorías que desea eliminar.
- ▶ En los menús elija Editar > Eliminar.

# Visualización de gráficos

Cuando se generan categorías, es importante dedicar un tiempo a revisar las definiciones de categorías, las respuestas que contienen y de qué modo se superponen las categorías. El panel de visualización ofrece varias perspectivas sobre sus categorías. El panel de visualización se encuentra en el ángulo superior derecho de la vista Pregunta. Si no está visible, puede acceder a este panel desde el menú Ver (Ver > Paneles > Visualización).

En esta vista, el panel de visualización ofrece tres perspectivas sobre las similitudes en la categorización de respuestas. Todos estos gráficos de este panel pueden utilizarse para analizar los resultados de la categorización y ayudarle a ajustar las categorías o los informes. Cuando se refinan las categorías, puede utilizar este panel para revisar las definiciones de categoría con el objeto de descubrir las que resultan demasiado similares (por ejemplo, si comparten más del 75% de las respuestas) o demasiado distintas. Si hay dos categorías demasiado similares entre sí, puede resultar útil combinar las dos categorías. Como alternativa, puede refinar las definiciones de categoría eliminando determinados descriptores de una categoría o de otra. Puede copiar, pegar e imprimir los resultados de este panel para ayudarle en el análisis o para generar informes.

Figura 7-1  
Panel Categoría y panel Visualización



En función de lo que haya seleccionado en el panel Resultados extraídos o Categorías, puede ver las interacciones correspondientes entre las respuestas y las categorías de cada una de las pestañas del panel. Cada una presenta información similar, pero de una forma distinta o con un nivel diferente de detalle. En caso necesario, puede personalizar los colores utilizados en estos diagramas y gráficos en el cuadro de diálogo Opciones. [Si desea obtener más información, consulte el tema Opciones: Pestaña Mostrar en el capítulo 2 el p. 19.](#)

*Nota:* También puede generar gráficos de resumen, por ejemplo, un gráfico de barras Las 5 categorías principales. Estos gráficos, que se exportan en HTML, pueden utilizarse en presentaciones. [Si desea obtener más información, consulte el tema Exportación de gráficos de resumen en el capítulo 4 el p. 58.](#)

El panel de visualización ofrece los diagramas y gráficos siguientes:

- **Gráfico de barras de categorías** Una tabla y un gráfico de barras que muestra la superposición entre las respuestas que corresponden a su selección y a las categorías asociadas. El gráfico de barras también muestra la proporción de respuestas en las categorías con el número total de respuestas. También puede seleccionar una variable de referencia, si es que ha importado

alguna, para comparar los valores de las variables de referencia de los registros de cada categoría. [Si desea obtener más información, consulte el tema Gráfico de barras de categorías el p. 168.](#)

- **Gráfico de malla de categorías.** Este gráfico presenta la superposición de respuestas correspondientes a las categorías a las que pertenecen las respuestas según la selección en los otros paneles. [Si desea obtener más información, consulte el tema Gráfico de malla de categorías el p. 169.](#)
- **Tabla de malla de categorías.** Esta tabla presenta la misma información que la pestaña Malla de categorías, pero en formato de tabla. La tabla contiene tres columnas que pueden ordenarse pulsando en las cabeceras de columna. [Si desea obtener más información, consulte el tema Tabla de malla de categorías el p. 170.](#)

[Si desea obtener más información, consulte el tema Categorización de los datos de texto en el capítulo 6 el p. 93.](#)

## ***Gráfico de barras de categorías***

Esta pestaña presenta una tabla y un gráfico de barras que muestra la superposición entre las respuestas que corresponden a su selección y a las categorías asociadas. El gráfico de barras también muestra la proporción de respuestas en las categorías con el número total de respuestas. No se puede editar el diseño de este gráfico.

También puede utilizar los menús contextuales de este gráfico de barras para ordenar columnas, cambiar los colores del gráfico, seleccionar el contenido del gráfico, copiar el contenido y mostrar u ocultar la leyenda.

El gráfico contiene las columnas siguientes:

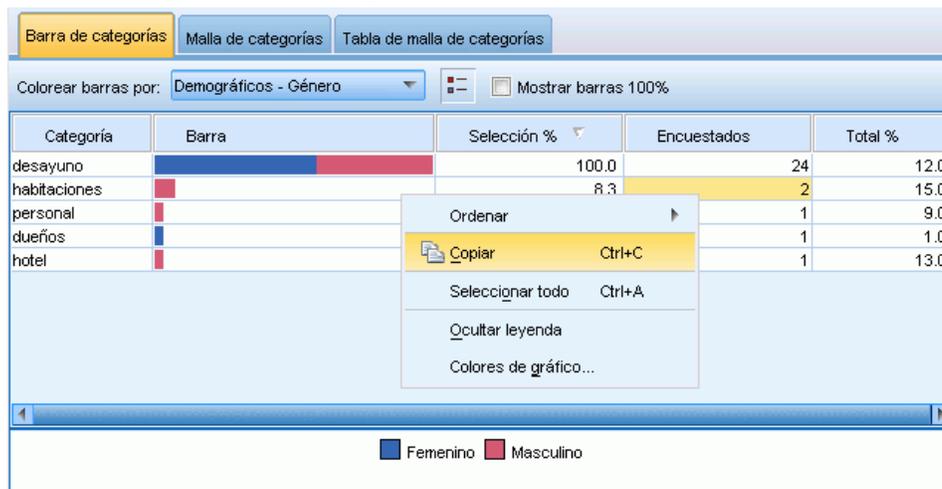
- **Categoría.** Esta columna presenta el nombre de las categorías que ha seleccionado. Por defecto, la categoría más común de su selección aparece en el primer lugar de la lista.
- **Barra.** Esta columna presenta, de una manera visual, la proporción de registros de una categoría determinada con el número total de registros.
- **Selección %.** Esta columna presenta el porcentaje basado en la proporción del número total de registros de una categoría con el número total de registros representados en la selección.
- **Respuestas** Esta columna presenta el número de respuestas de una selección correspondientes a una categoría determinada.
- **Total %.** Esta columna presenta un porcentaje basado en la proporción del número total de registros de una categoría determinada comparado con el número total de registros de esta pregunta (no de la selección).

También puede seleccionar una variable de referencia disponible desde la lista desplegable, para comparar su valores. Cuando selecciona una variable de referencia, las barras de la tabla están divididas y codificadas por colores según los valores de las variables de referencia. Si pulsa en cada valor de referencia coloreado de la barra, el panel Datos se actualiza para mostrar la subselección de respuestas en función del valor de la variable de referencia. Para ver una leyenda de los valores de las variables de referencia, pulse en el botón de la barra de herramientas Leyenda.

Figura 7-2  
Botón de la barra de herramientas Leyenda



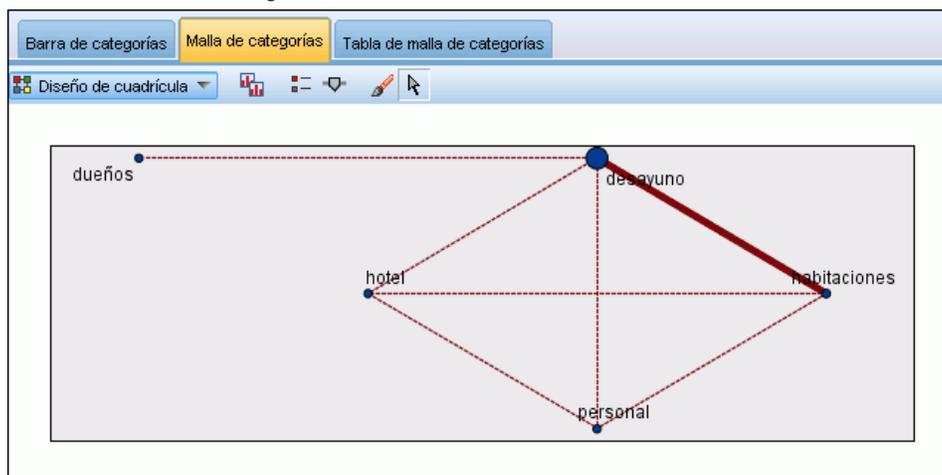
Figura 7-3  
Panel Visualización: Barra de categorías



## Gráfico de malla de categorías

Esta pestaña muestra un gráfico de malla de categorías. Este gráfico presenta la superposición de respuestas correspondientes a las categorías a las que pertenecen las respuestas según la selección en los otros paneles. Si existen etiquetas de categoría, estas aparecen en el gráfico. Puede elegir un diseño de gráfico (de red, circular, direccional o de cuadrícula) mediante los botones de la barra de herramientas de este panel.

Figura 7-4  
Gráfico de malla de categorías, diseño de cuadrícula



En la malla, cada nodo representa una categoría. Con el ratón puede seleccionar y mover los nodos en el panel. El tamaño del nodo representa el tamaño relativo basado en el número de registros de dicha categoría en su selección. El grosor y el color de la línea entre dos categorías denotan el número de registros comunes que tienen. Si pasa el ratón por encima de un nodo en el modo Explorar, la información de la herramienta muestra el nombre (o etiqueta) de la categoría y el número global de registros en la categoría.

*Nota:* Por defecto está activado el modo Explorar para los gráficos, donde puede mover nodos. Sin embargo, puede cambiar al modo Editar para modificar el diseño de los gráficos, incluidos los colores, las fuentes, las leyendas, etc. [Si desea obtener más información, consulte el tema Uso de barras de herramientas y paletas de gráficos el p. 170.](#)

## Tabla de malla de categorías

Esta tabla presenta la misma información que la pestaña Malla de categorías, pero en formato de tabla. La tabla contiene tres columnas que pueden ordenarse pulsando en las cabeceras de columna:

- **Recuento.** Esta columna presenta el número de registros compartidos o comunes entre las dos categorías.
- **Categoría 1.** Esta columna presenta el nombre de la primera categoría seguido por el número total de registros que contiene, entre paréntesis.
- **Categoría 2.** Esta columna presenta el nombre de la segunda categoría seguido por el número total de registros que contiene, entre paréntesis.

Figura 7-5

Panel Visualización: Tabla de malla de categorías

Recuento	Categoría 1	Categoría 2
2 habitaciones(30)	desayuno(24)	dueños(2)
1 desayuno(24)	personal(18)	hotel(26)
1 habitaciones(30)	desayuno(24)	hotel(26)
1 habitaciones(30)	hotel(26)	desayuno(24)
1 hotel(26)	hotel(26)	desayuno(24)
1 personal(18)	desayuno(24)	
1 personal(18)		

## Uso de barras de herramientas y paletas de gráficos

El gráfico de malla de categorías dispone de una barra de herramientas que ofrece acceso rápido a algunas paletas comunes en las que podrá realizar un número de acciones en los gráficos. Puede elegir entre el modo de vista *Exploración* o el modo de vista *Edición*.

Mientras el modo de exploración le permite explorar analíticamente los datos y valores representados por la visualización, el modo de edición le permite cambiar el diseño y aspecto de la visualización. Por ejemplo, puede cambiar las fuentes y colores para que coincidan con el manual

de estilo de su organización. Para seleccionar este modo, seleccione Ver > Panel Visualización > Modo edición de los menús (o pulse en el icono de la barra de herramientas).

En el modo de edición hay varias barras de herramientas que afectan a distintos aspectos del diseño de la visualización. Si no utiliza algunas de ellas, puede ocultarlas para ampliar el espacio del cuadro de diálogo en el que aparece el gráfico. Para seleccionar o anular la selección de las barras de herramientas, pulse en el nombre de la barra de herramientas o paleta correspondiente en el menú Ver.

Tabla 7-1  
Botones de la barra de herramientas de Text Analytics

Botón/Lista	Descripción
	Activa el modo de edición. Pase al modo edición para cambiar el aspecto del gráfico: puede agrandar la fuente, cambiar los colores para que se adapten al estilo corporativo de su empresa o eliminar etiquetas y leyendas. <a href="#">Si desea obtener más información, consulte el tema Edición de visualizaciones el p. 171.</a>
	Activa el modo de exploración. Por defecto, el modo de exploración está activado, lo que significa que puede mover y arrastrar nodos por el gráfico y pasar el ratón sobre los objetos del gráfico para revelar información adicional.
	<p>Seleccione un tipo de malla que mostrar para los gráficos.</p> <ul style="list-style-type: none"> <li>■ Diseño circular. Diseño general que puede aplicarse a cualquier gráfico. Organiza un gráfico teniendo en cuenta que los enlaces no tienen dirección y trata a todos los nodos por igual. Los nodos solo se colocan alrededor del perímetro de un círculo.</li> <li>■ Diseño de red. Diseño general que puede aplicarse a cualquier gráfico. Organiza un gráfico teniendo en cuenta que los enlaces no tienen dirección y trata a todos los nodos por igual. Los nodos se colocan libremente por el diseño.</li> <li>■ Diseño direccional. Un diseño que solo debería utilizarse para gráficos direccionales. Este enlace genera estructuras de árbol a partir de los nodos raíz hasta los nodos ramales, y las organiza por colores.</li> <li>■ Diseño de cuadrícula. Diseño general que puede aplicarse a cualquier gráfico. Organiza un gráfico teniendo en cuenta que los enlaces no tienen dirección y trata a todos los nodos por igual. Los nodos solo se colocan en los puntos de la cuadrícula en el espacio.</li> </ul>
	Botón de conmutación que, cuando se pulsa, muestra la leyenda. Cuando el botón no está pulsado, la leyenda queda oculta.
	Botón de conmutación que, cuando se pulsa, muestra el Control deslizante de enlaces bajo el gráfico. Puede filtrar los resultados deslizando la flecha.

## Edición de visualizaciones

Tiene varias opciones para editar una visualización en el **Modo de edición**. Puede:

- Editar texto y darle formato.
- Cambiar el color de relleno, la transparencia y el patrón de marcos y elementos gráficos.
- Cambiar el color y el estilo de guión de los bordes y las líneas.
- Girar y cambiar la forma y la relación de aspecto de los elementos de punto.
- Cambiar el tamaño de los elementos gráficos (como barras y puntos).
- Ajustar el espacio que rodea a los elementos mediante márgenes y relleno.
- Especificar un formato para los números.
- Cambiar la configuración del eje y la escala.

- Ordenar, excluir y contraer categorías en un eje categórico.
- Definir la orientación de los paneles.
- Aplicar transformaciones a un sistema de coordenadas.
- Cambiar estadísticos, tipos de elementos gráficos y modificadores de colisión.
- Cambiar la posición de la leyenda.
- Aplicar hojas de estilo de visualización.

Los siguientes temas describen cómo se pueden realizar estas distintas tareas. También se recomienda leer las reglas generales de edición de gráficos.

#### ***Para cambiar al modo de edición***

- ▶ Elija en los menús:  
Ver > Modo de edición

## ***Reglas generales de edición de visualizaciones***

### ***Modo de edición***

Todas las ediciones se realizan en modo de edición. Para activarlo, elija en los menús:  
Ver > Modo de edición

### ***Selección***

Las opciones disponibles para edición dependen de la selección. Se activan diferentes opciones de la barra de herramientas y la paleta de propiedades según lo que se seleccione. Sólo se aplican a la selección actual los elementos activados. Por ejemplo, si se selecciona un eje, las pestañas Escalas, Marca mayor y Marcas menores estarán disponibles en la paleta de propiedades.

A continuación mostramos algunos consejos para seleccionar elementos de la visualización:

- Pulse en un elemento para seleccionarlo.
- Seleccione un elemento gráfico (como los puntos de un diagrama de dispersión o las barras de un gráfico de barras) pulsando una vez. Tras la selección inicial, vuelva a pulsar para reducir la selección a grupos de elementos gráficos o a un único elemento gráfico.
- Pulse en Esc para anular la selección de todo.

### ***Paletas***

Cuando se selecciona un elemento de la visualización, las diferentes paletas se actualizan para reflejar la selección. Las paletas contienen controles para realizar las modificaciones en la selección. Las paletas pueden ser barras de herramientas o un panel con múltiples controles y pestañas. Las paletas pueden estar ocultas, por lo que hay que asegurarse de que se muestra la paleta adecuada para realizar las ediciones. Compruebe en el menú Ver las paletas que se muestran actualmente.

Puede cambiar la posición de las paletas pulsando y arrastrando el espacio vacío de una paleta de barra de herramientas o en la parte izquierda de otras paletas. La información visual le permite saber dónde anclar la paleta. En el caso de paletas que no son de la barra de herramientas, también puede hacer en el botón cerrar para ocultar la paleta y en el botón desanclar para mostrar la paleta en una ventana diferente. Pulse el botón ayuda para mostrar la ayuda de la paleta específica.

### **Configuración automática**

Algunos parámetros de configuración incluyen la opción *-automático-*. Esto indica que se aplican valores automáticos. Los ajustes automáticos que se utilizarán dependen de la visualización específica y los valores de los datos. Puede introducir un valor que sustituya a la configuración automática. Si desea restaurar la configuración automática, elimine el valor actual y pulse en Entrar. El parámetro volverá a mostrar la opción *-automático-*.

### **Eliminación y ocultación de elementos**

Puede eliminar u ocultar varios elementos en la visualización. Por ejemplo, puede ocultar la leyenda o la etiqueta del eje. Seleccione un elemento y pulse la tecla Supr para eliminarlo. Si no se permite eliminar ese elemento, no se realizará ninguna acción. Si elimina un elemento por accidente, pulse Ctrl+Z para deshacer la eliminación.

### **Estado**

Algunas barras de herramientas muestran el estado de la selección actual, mientras que otras no. La paleta de propiedades siempre muestra el estado. Si una barra de herramientas *no* muestra el estado, se indicará en el tema que describe dicha barra de herramientas.

## **Edición y formato de texto**

Puede editar texto directamente y cambiar el formato de un bloque de texto entero. Tenga en cuenta que no podrá editar texto directamente enlazado a valores de datos. Por ejemplo, no podrá editar una etiqueta de marca, ya que su contenido se deriva de los datos subyacentes. Sin embargo, puede formatear cualquier texto de la visualización.

### **Cómo editar texto directamente**

- ▶ Pulse dos veces en el bloque de texto. De esta forma se selecciona todo el texto. Todas las barras de texto se desactivarán en este momento, ya que no podrá cambiar ninguna otra parte de la visualización mientras edita el texto.
- ▶ Escriba sobre el texto que desea sustituir. También puede volver a pulsar en el texto para que aparezca un cursor. Coloque el cursor en la posición deseada y escriba el texto adicional.

### **Cómo dar formato al texto**

- ▶ Seleccione el marco que contiene el texto. No pulse dos veces en el texto.

- Use la barra de herramientas de fuentes para dar formato al texto. Si la barra de herramientas no está activada, asegúrese de seleccionar sólo el *marco* que contiene el texto. Si el texto está seleccionado, se desactivará la barra de herramientas.

Figura 7-6  
Barra de herramientas de fuentes



Puede cambiar distintas características de la fuente:

- Color
- Familia (por ejemplo, Arial o Verdana)
- Tamaño (la unidad es pt, a menos que se indique otra distinta, como pc)
- Ponderación
- Alineación relativa al marco del texto

El formato se aplica a todo el texto de un marco. No puede cambiar el formato de letras o palabras individuales de ningún bloque de texto concreto.

### ***Cambio de colores, tramas, estilos de guión y transparencia***

Muchos elementos diferentes de una visualización tienen un relleno y un borde. El ejemplo más obvio es la barra de un gráfico de barras. El color de las barras es el color de relleno. También pueden llevar un borde negro y continuo alrededor.

Hay otros elementos menos obvios en la visualización que tienen colores de relleno. Si el color de relleno es transparente, es posible que no se vea que existe dicho relleno. Tomemos como ejemplo el texto de una etiqueta de eje. Da la sensación de que el texto es “flotante”, pero en realidad aparece en un marco con color de relleno transparente. Puede ver el marco si selecciona la etiqueta del eje.

Cualquier marco de la visualización puede tener un estilo de relleno y borde, incluyendo el marco alrededor de toda la visualización. Además, cualquier relleno tiene un nivel de opacidad/transparencia asociado que puede ajustarse.

#### ***Para cambiar de colores, tramas, estilos de guión y transparencia***

- Seleccione el elemento cuyo formato desea cambiar. Por ejemplo, seleccione las barras de un gráfico de barras o un marco que contenga texto. Si la visualización está dividida por un campo o variable categórica, puede seleccionar también el grupo que corresponda a una categoría individual con el fin de cambiar el aspecto estético asignado por defecto a dicho grupo. Por ejemplo, puede cambiar el color de uno de los grupos de apilado de un gráfico de barras apiladas.
- Use la barra de herramientas de colores para cambiar el color de relleno o del borde, o bien la trama de relleno.

Figura 7-7  
Barra de herramientas de colores

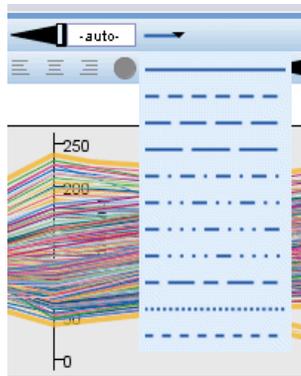


*Nota:* esta barra de herramientas no muestra el estado de la selección actual.

Para cambiar un color o relleno, puede pulsar el botón para seleccionar la opción mostrada o en la flecha desplegable para seleccionar otra opción. En el caso de los colores, observe que hay uno representado como blanco y atravesado por una línea diagonal roja. Se trata del color transparente. Puede usarlo, por ejemplo, para ocultar los bordes de las barras de un histograma.

- El primer botón controla el color de relleno.
  - El segundo botón controla el color del borde.
  - El tercer botón controla la trama de relleno. La trama de relleno utiliza el color del borde. Por tanto, la trama de relleno sólo se ve si el color del borde es visible.
  - El cuarto control es una barra deslizante y un cuadro de texto que controlan la opacidad del color y el patrón de relleno. Si el porcentaje es menor, habrá menor opacidad y más transparencia. 100% significa que es totalmente opaco (sin transparencia).
- Para cambiar el estilo de guión de un borde o una línea, use la barra de herramientas de líneas.

Figura 7-8  
Barra de herramientas de líneas



*Nota:* esta barra de herramientas no muestra el estado de la selección actual.

Al igual que con la otra barra de herramientas, puede pulsar en el botón para seleccionar la opción mostrada o en la flecha desplegable para seleccionar otra opción.

## ***Rotación y cambio de la forma y la relación de aspecto de los elementos de punto***

Puede girar elementos de punto, asignar una forma predefinida diferente o cambiar la relación de aspecto (la relación entre anchura y altura).

### ***Cómo modificar elementos de punto***

- Seleccione los elementos de punto. No puede girar ni cambiar la forma y relación de aspecto de elementos de punto individuales.
- Use la barra de herramientas de símbolos para modificar los puntos.

Figura 7-9  
Barra de herramientas de símbolos



- El primer botón permite cambiar la forma de los puntos. Pulse en la flecha desplegable y seleccione una forma predefinida.
- Con el segundo botón puede girar los puntos hasta una posición específica como si lo hiciera con un compás. Pulse en la flecha desplegable y, a continuación, arrastre la punta hasta la posición deseada.
- El tercero permite cambiar la relación de aspecto. Pulse en la flecha desplegable y, a continuación, arrastre el rectángulo que aparece. La forma del rectángulo representa la relación de aspecto.

### **Cambio de tamaño de los elementos gráficos**

Puede cambiar el tamaño de los elementos gráficos de la visualización. Estos incluyen barras, líneas y puntos entre otros. Si el tamaño del elemento gráfico depende de una variable o un campo, el tamaño especificado es el *mínimo*.

#### **Cómo cambiar el tamaño de los elementos gráficos**

- ▶ Seleccione los elementos gráficos cuyo tamaño desea cambiar.
- ▶ Use el control deslizante o introduzca un tamaño específico para la opción disponible en la barra de herramientas de símbolos. La unidad es píxeles, a menos que se indique otra distinta (consulte a continuación la lista completa de abreviaturas de unidad). También puede especificar un porcentaje (como 30 %), lo que significa que un elemento gráfico usa el porcentaje especificado del espacio disponible. El espacio disponible dependerá del tipo de elemento gráfico y la visualización específica.

Tabla 7-2  
Abreviaturas de unidad válidas

Abreviatura	Unidad
cm	centímetro
en	pulgada
mm	milímetro
pc	pica
pt	punto
px	pixel

Figura 7-10  
Control del tamaño en la barra de herramientas de símbolos



## Especificación de márgenes y relleno

Si hay demasiado espacio o un espacio insuficiente alrededor o dentro de un marco de la visualización, puede cambiar este margen y los ajustes de relleno. El **margen** es la cantidad de espacio existente entre el marco y otros elementos que lo rodean. El **relleno** es la cantidad de espacio existente entre el borde del marco y el *contenido* de dicho marco.

### Cómo especificar los márgenes y el relleno

- ▶ Seleccione el marco cuyos márgenes y relleno desea especificar. Puede tratarse de un marco de texto, el marco que rodea a la leyenda o incluso el marco de datos que muestra los elementos gráficos (como barras y puntos).
- ▶ Use la pestaña Márgenes de la paleta de propiedades para especificar la configuración. Todos los tamaños se expresan en píxeles, a menos que se indique una unidad distinta (como cm o pulg).

Figura 7-11  
Pestaña Márgenes



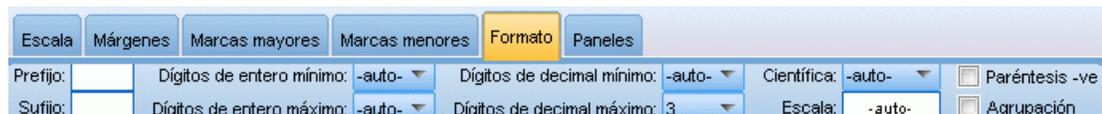
## Formato de numeración

Puede especificar el formato de los números de las etiquetas señalizadoras en el eje continuo o las etiquetas de valor de datos que muestran un número. Por ejemplo, puede especificar que los números que aparecen en las etiquetas de marcas se muestren en miles.

### Para especificar formatos de numeración

- ▶ Seleccione las etiquetas señalizadoras del eje continuo o las etiquetas de valor de datos si contienen números.
- ▶ Pulse en la pestaña Formato de la paleta Propiedades.

Figura 7-12  
Pestaña Formato



- ▶ Seleccione las opciones de formato de numeración que desee:

**Prefijo.** Carácter que se mostrará al comienzo del número. Introduzca, por ejemplo, un signo de dólar (\$) si los números se refieren a salarios en dólares estadounidenses.

**Sufijo.** Carácter que se mostrará al final del número. Introduzca, por ejemplo, un signo de porcentaje (%) si los números equivalen a porcentajes.

**Dígitos de entero mínimo.** Número mínimo de dígitos que mostrar en la parte entera de una representación decimal. Si el valor real no contiene el número mínimo de dígitos, la parte entera del valor se rellenará con ceros.

**Dígitos de entero máximo.** Número máximo de dígitos que mostrar en la parte entera de una representación decimal. Si el valor real supera el número máximo de dígitos, la parte entera del valor se rellenará con asteriscos.

**Dígitos de decimal mínimo.** Número mínimo de dígitos que mostrar en la parte decimal de una representación decimal o científica. Si el valor real no contiene el número mínimo de dígitos, la parte decimal del valor se rellenará con ceros.

**Dígitos de decimal máximo.** Número máximo de dígitos que mostrar en la parte decimal de una representación decimal o científica. Si el valor real supera el número mínimo de dígitos, la parte decimal se redondeará al número adecuado de dígitos.

**Científica.** Determina si se muestran o no los números en notación científica. La notación científica resulta útil para números muy grandes o muy pequeños. -auto- permite a la aplicación determinar cuándo es adecuado utilizar la notación científica.

**Escalamiento.** Un factor de escala, que es un número por el que se divide el valor original. Utilice un factor de escala si los números son grandes, pero no desea que la etiqueta se extienda demasiado para adaptarse al número. Si cambia el formato de numeración de las etiquetas señaladoras, asegúrese de editar el título del eje para indicar cómo se debe interpretar el número. Supongamos, por ejemplo, que el eje de escala muestra salarios y las etiquetas son 30.000, 50.000 y 70.000. Puede introducir un factor de escala de 1000 para mostrar 30, 50 y 70. A continuación, debe editar el título del eje para incluir el texto en miles.

**Paréntesis -ve.** Indica si deben incluirse paréntesis alrededor de los valores negativos.

**Agrupación.** Indica si se debe mostrar un carácter entre grupos de dígitos. La configuración regional actual de su equipo determinará qué carácter se utiliza para la agrupación de dígitos.

## ***Cambio de la configuración del eje y la escala***

Existen varias opciones para modificar ejes y escalas.

### ***Cómo cambiar la configuración del eje y la escala***

- ▶ Seleccione cualquier parte del eje (por ejemplo, la etiqueta del eje o las etiquetas de marca).
- ▶ Use las pestañas Escalas, Marca mayor y Marcas menores de la paleta de propiedades para cambiar la configuración del eje y la escala.

Figura 7-13  
Paleta de propiedades



### **Pestaña Escalas**

**Tipo.** Especifica si la escala es lineal o transformada. Las transformaciones de escala permiten comprender los datos o realizar los supuestos necesarios para la inferencia estadística. En los diagramas de dispersión, puede utilizar una escala transformada si la relación entre las variables o los campos independientes y dependientes es no lineal. Las transformaciones de escala también se pueden utilizar para lograr que un histograma asimétrico sea más simétrico y se parezca a una distribución normal. Tenga en cuenta que sólo se transforma la escala en que se muestran los datos; los datos reales no se transforman.

- **Lineal.** Especifica una escala lineal sin transformar.
- **Log.** Especifica una escala transformada de logaritmo en base 10. Para adaptarse a los valores cero y negativos, esta transformación usa una versión modificada de la función logarítmica. Esta función de “registro seguro” se define como  $\text{sign}(x) * \log(1 + \text{abs}(x))$ . De modo que  $\text{safeLog}(-99)$  es igual a:

$$\text{sign}(-99) * \log(1 + \text{abs}(-99)) = -1 * \log(1 + 99) = -1 * 2 = -2$$

- **Potencia.** Especifica una escala transformada de potencia con un exponente de 0,5. Para adaptarse a los valores negativos, esta transformación usa una versión modificada de la función exponencial. Esta función de “potencia segura” se define como  $\text{sign}(x) * \text{pow}(\text{abs}(x), 0.5)$ . De modo que  $\text{safePower}(-100)$  es igual a:

$$\text{sign}(-100) * \text{pow}(\text{abs}(-100), 0.5) = -1 * \text{pow}(100, 0.5) = -1 * 10 = -10$$

**Mín/Máx/Bajo redondeado/Alto redondeado.** Especifica el rango de la escala. Si selecciona Bajo redondeado y Alto redondeado, la aplicación podrá seleccionar una escala apropiada de acuerdo con los datos. Los valores mínimo y máximo están redondeados porque suelen ser valores enteros mayores o menores que los valores de datos máximos y mínimos. Por ejemplo, si los datos oscilan entre 4 y 92, los valores alto y bajo redondeados para la escala pueden estar comprendidos entre 0 y 100 en lugar de los valores mínimo y máximo reales de los datos. Preste atención para no definir un rango demasiado pequeño que oculte elementos importantes. Tenga en cuenta también que no puede definir un valor mínimo y máximo explícito si está seleccionada la opción Incluir cero.

**Margen bajo/Margen alto.** Crea márgenes en los extremos inferior y/o superior del eje. El margen aparece en posición perpendicular con respecto al eje seleccionado. La unidad es píxeles, a menos que se indique otra distinta (como cm o pulg). Por ejemplo, si define Margen alto en 5 para el eje vertical, se extenderá un margen horizontal de 5 px a lo largo de la parte superior del marco de datos.

**Inversa.** Especifica si la escala está invertida.

**Incluir cero.** Indica que la escala debe incluir el valor 0. Esta opción suele utilizarse en gráficos de barras para asegurarse de que las barras empiezan en el 0, y no en un valor cercano a la altura de la barra más pequeña. Si se selecciona esta opción, Mín y Máx se desactivan ya que no se pueden definir los valores mínimo y máximo del rango de la escala.

### **Pestañas Marca mayor/Marcas menores**

Las **marcas** o **marcas de comprobación** son líneas que aparecen sobre un eje. Indican valores en intervalos o categorías específicos. Las **marcas mayores** son marcas de comprobación con etiquetas. Son también más largas que las otras marcas de comprobación. Las **marcas menores**

son marcas de comprobación que aparecen entre las marcas de comprobación mayores. Algunas opciones son específicas del tipo de marca, pero la mayoría de las opciones están disponibles para marcas mayores y menores.

**Mostrar marcas.** Especifica si aparecen marcas mayores o menores en el gráfico.

**Mostrar cuadrículas.** Especifica si las cuadrículas aparecen en las marcas mayores o menores. Las **cuadrículas** son líneas que atraviesan todo el gráfico de eje a eje.

**Posición.** Especifica la posición de las marcas de comprobación en relación con el eje.

**Longitud.** Especifica la longitud de las marcas de comprobación. La unidad es píxeles, a menos que se indique otra distinta (como cm o pulg).

**Base.** *Sólo se aplica a las marcas mayores.* Especifica el valor en el que aparece la primera marca mayor.

**Delta.** *Sólo se aplica a las marcas mayores.* Especifica la diferencia entre marcas mayores. Esto es, las marcas mayores aparecerán cada  $n$  valores, donde  $n$  es el valor delta.

**Divisiones.** *Sólo se aplica a las marcas menores.* Especifica el número de divisiones de marcas menores existentes entre las marcas mayores. El número de marcas menores es igual al número de divisiones menos uno. Por ejemplo, supongamos que hay marcas mayores en 0 y 100. Si introduce 2 como número de divisiones de marcas menores, habrá *una* marca menor en 50, dividiendo el intervalo de 0 a 100, lo que creará *dos* divisiones.

## **Edición de categorías**

Puede editar las categorías en un eje categórico de varias formas:

- Cambie el orden de clasificación para mostrar las categorías.
- Excluya categorías específicas.
- Añada una categoría que no aparezca en el conjunto de datos.
- Contraiga/combine categorías pequeñas en una categoría.

### **Cómo cambiar el orden de clasificación de categorías**

- ▶ Seleccione un eje categórico. La paleta Categorías muestra las categorías en el eje.

*Nota:* Si la paleta no está visible, asegúrese de haberla activado.

- ▶ En la paleta Categorías, seleccione una opción de ordenación de la lista desplegable.

**Personalizado.** Ordena las categorías basándose en el orden en el que aparecen en la paleta. Utilice los botones de flecha para mover categorías a la parte superior de la lista, arriba, abajo y a la parte inferior de la lista.

**Datos.** Ordena las categorías basándose en el orden en el que aparecen en el conjunto de datos.

**Nombre.** Ordena las categorías alfabéticamente, utilizando los nombres como aparecen en la paleta.

**Valor.** Ordena las categorías según el valor de datos subyacente, utilizando los valores que aparecen entre paréntesis en la paleta. Sólo los orígenes de datos con metadatos (como los archivos de datos de IBM® SPSS® Statistics) admiten esta opción.

**Estadístico.** Ordena las categorías basándose en el estadístico calculado para cada categoría. Algunos ejemplos de estadísticos son recuentos, porcentajes y medias. Esta opción sólo está disponible si se utiliza un estadístico en el gráfico.

### **Para añadir una categoría**

Por defecto, únicamente las categorías que aparecen en el conjunto de datos están disponibles. Puede añadir una categoría a la vista si fuera necesario.

- ▶ Seleccione un eje categórico. La paleta Categorías muestra las categorías en el eje.

*Nota:* Si la paleta no está visible, asegúrese de haberla activado.

- ▶ En la paleta de categorías, pulse en el botón Añadir categoría:

Figura 7-14  
Botón Añadir categoría



- ▶ En el cuadro de diálogo Añadir una nueva categoría, introduzca un nombre para la categoría.
- ▶ Pulse en Aceptar.

### **Cómo excluir categorías específicas**

- ▶ Seleccione un eje categórico. La paleta Categorías muestra las categorías en el eje.

*Nota:* Si la paleta no está visible, asegúrese de haberla activado.

- ▶ En la paleta Categorías, seleccione un nombre de categoría en la lista Incluir y, a continuación, pulse en el botón X. Para mover la categoría hacia atrás, seleccione su nombre en la lista Excluido y, a continuación, pulse en la flecha situada a la derecha de la lista.

### **Cómo contraer/combinar categorías pequeñas**

Puede combinar categorías que sean tan pequeñas que no necesite mostrarlas por separado. Por ejemplo, si tiene un gráfico de sectores con muchas categorías, considere la opción de contraer categorías con un porcentaje inferior a 10. Sólo se pueden contraer las estadísticas aditivas. Por ejemplo, no puede añadir medias porque las medias no son aditivas. Por lo tanto, no se pueden combinar/contrar categorías utilizando una media.

- ▶ Seleccione un eje categórico. La paleta Categorías muestra las categorías en el eje.

*Nota:* Si la paleta no está visible, asegúrese de haberla activado.

- ▶ En la paleta Categorías, seleccione Contraer y especifique un porcentaje. Cualquier categoría cuyo porcentaje del total sea inferior al número especificado se combinará en una categoría. El

porcentaje se basa en el estadístico mostrado en el gráfico. La contracción sólo está disponible para estadísticos basados en recuentos y de suma.

## ***Cambio de la orientación de paneles***

Si utiliza paneles en su visualización, puede cambiar su orientación.

### ***Cómo cambiar la orientación de los paneles***

- ▶ Seleccione cualquier parte de la visualización.
- ▶ Pulse en la pestaña Paneles de la paleta Propiedades.

Figura 7-15  
Pestaña Paneles



- ▶ Seleccione una opción en Diseño:

**Tabla.** Organiza los paneles como una tabla, de manera que hay una fila o columna asignada a cada valor individual.

**Transpuesto.** Organiza los paneles como una tabla, aunque también intercambia las filas y columnas originales. Esta opción no equivale a transponer el gráfico en sí. Observe que los ejes  $x$  e  $y$  no cambian al seleccionar esta opción.

**Lista.** Organiza los paneles como una lista, de manera que cada casilla representa una combinación de valores. Las columnas y filas ya no estarán asignadas a valores individuales. Esta opción permite que los paneles se ajusten si es necesario.

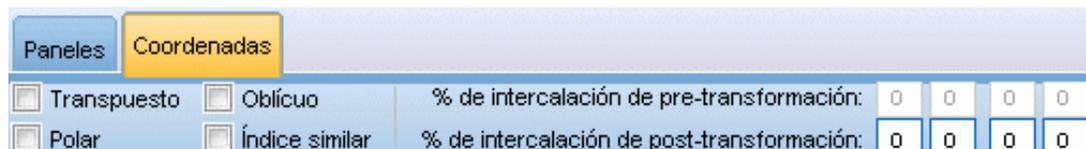
## ***Transformación del sistema de coordenadas***

Muchas visualizaciones se muestran en un sistema de coordenadas plano y rectangular. Puede transformar el sistema de coordenadas según sea necesario. Por ejemplo puede aplicar una transformación polar al sistema de coordenadas, añadir efectos de sombra en sentido oblicuo y transponer los ejes. También puede deshacer cualquiera de estas transformaciones si ya se aplicaron a la visualización actual. Por ejemplo, se dibuja un gráfico de sectores en un sistema de coordenadas polares. Si lo desea, puede deshacer la transformación polar y mostrar el gráfico de sectores como una única barra apilada en un sistema de coordenadas rectangular.

### ***Para transformar el sistema de coordenadas***

- ▶ Seleccione el sistema de coordenadas que desee transformar. El sistema de coordenadas se selecciona eligiendo el marco que rodea el gráfico individual.
- ▶ Pulse en la pestaña Coordenadas de la paleta Propiedades.

Figura 7-16  
Pestaña Coordenadas



- Seleccione las transformaciones que desee aplicar al sistema de coordenadas. También puede cancelar la selección de una transformación para deshacerla.

**Transpuesto.** El cambio de orientación de los ejes se denomina **transposición**. Es como intercambiar los ejes vertical y horizontal en una visualización 2-D.

**Polar.** Una transformación polar dibuja los elementos gráficos con un ángulo y una distancia específicos desde el centro del gráfico. Un gráfico de sectores es una visualización de una dimensión con una transformación polar que dibuja las barras individuales con ángulos específicos. Un gráfico radial es una visualización 2-D con una transformación polar que dibuja elementos gráficos con ángulos y distancias específicos desde el centro del gráfico. Una visualización 3-D también incluye una dimensión de profundidad adicional.

**Oblicuo.** Una transformación oblicua añade un efecto 3-D a los elementos gráficos. Esta transformación añade profundidad a los elementos gráficos, pero su profundidad es meramente decorativa. No se ve influenciado por valores de datos concretos.

**Misma proporción.** Si se aplica la misma proporción se especifica que la misma distancia en cada escala represente la misma diferencia de valores de datos. Por ejemplo, 2 cm en ambas escalas representan una diferencia de 1000.

**% del recuadro antes de transformación.** Si tras la transformación se recortan los ejes, puede que desee añadir recuadros al gráfico antes de aplicar la transformación. Los recuadros reducen las dimensiones en un porcentaje antes de aplicar transformaciones al sistema de coordenadas. Tiene control sobre las dimensiones  $x$  inferior,  $x$  superior,  $y$  inferior e  $y$  superior, en ese orden.

**% del recuadro tras la transformación.** Si desea cambiar la relación de aspecto de un gráfico, puede añadirle recuadros tras aplicar transformación. Los recuadros reducen las dimensiones en un porcentaje después de aplicar transformaciones al sistema de coordenadas. Estos recuadros también pueden aplicarse aunque no se haya realizado ninguna transformación en el gráfico. Tiene control sobre las dimensiones  $x$  inferior,  $x$  superior,  $y$  inferior e  $y$  superior, en ese orden.

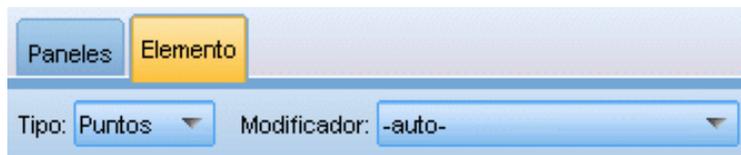
## ***Cambio de estadísticos y elementos gráficos***

Puede convertir un en otro tipo, cambiar el estadístico utilizado para dibujar el elemento gráfico o especificar el modificar de colisión que determina lo que ocurre cuando los elementos gráficos se solapan.

### ***Para convertir un elemento gráfico***

- Seleccione el elemento gráfico que desee convertir.
- Pulse en la pestaña Elemento de la paleta Propiedades.

Figura 7-17  
Pestaña Elemento.



- Seleccione un nuevo tipo de elemento gráfico en la lista Tipo.

Tipo de elemento gráfico	Descripción
Punto	Un marcador que especifica un punto de datos concreto. Un elemento de punto se utiliza en gráficos de dispersión y en otras visualizaciones relacionadas.
Intervalo	Una forma rectangular dibujada en un valor de datos específicos y que rellena el espacio entre un origen y otro valor de datos. Un elemento de intervalo se utiliza en gráficos de barras e histogramas.
Línea	Una línea que conecta valores de datos.
Ruta de acceso	Una línea que conecta valores de datos en el orden en que aparecen en el conjunto de datos.
Área	Una línea que conecta elementos de datos en el área entre la línea y un origen relleno.
Polígono	Una forma de varias caras que encierra una región de datos. Se puede usar un elemento poligonal en un diagrama de dispersión en intervalos o un mapa.
Esquema	Un elemento que consiste en una caja con bigotes y marcadores que indican valores atípicos. Un elemento de esquema se utiliza para diagramas de caja.

### **Para cambiar el estadístico**

- Seleccione el elemento gráfico cuyo estadístico desee cambiar.
- Pulse en la pestaña Elemento de la paleta Propiedades.
- En la lista desplegable Resumen, seleccione un nuevo estadístico. Tenga en cuenta que al seleccionar un estadístico se agregan los datos. Si en su lugar desea que la visualización muestra datos sin agregar, seleccione (sin estadístico) en la lista Resumen.

### **Estadísticos de resumen calculados a partir de un campo continuo**

- **Media.** Una medida de tendencia central. El promedio aritmético, la suma dividida por el número de casos.
- **Mediana.** Es el valor por encima y por debajo del cual se encuentran la mitad de los casos, el percentil 50. Si hay un número par de casos, la mediana es la media de los dos valores centrales, cuando los casos se ordenan en orden ascendente o descendente. La mediana es una medida de tendencia central que no es sensible a los valores atípicos (a diferencia de la media, que puede resultar afectada por unos pocos valores extremadamente altos o bajos).
- **Modo.** El valor que ocurre con mayor frecuencia. Si varios valores comparten la mayor frecuencia de aparición, cada uno de ellos es un modo.
- **Mínimo.** Valor más pequeño de una variable numérica.

- **Máximo.** El mayor valor de una variable numérica.
- **Rango.** Diferencia entre los valores mínimos y máximos.
- **Medio intervalo.** El punto medio del intervalo, es decir, el valor cuya diferencia con el mínimo es el mismo que el de su diferencia con el máximo.
- **Suma.** Suma o total de todos los valores, a lo largo de todos los casos que no tengan valores perdidos.
- **Suma acumulada.** La suma acumulada de los valores. Cada elemento gráfico muestra la suma de un subgrupo más la suma total de todos los grupos anteriores.
- **Suma porcentual.** El porcentaje de cada subgrupo basado en un campo sumado comparado con la suma de todos los grupos.
- **Suma porcentual acumulada.** El porcentaje acumulado de cada subgrupo basado en un campo sumado comparado con la suma de todos los grupos. Cada elemento gráfico muestra el porcentaje de un subgrupo más el porcentaje total de todos los grupos anteriores.
- **Variación.** Es una medida de dispersión en torno a la media, igual a la suma de las desviaciones al cuadrado respecto a la media, dividida por el número de casos menos 1. La varianza se mide en unidades que son el cuadrado de las de la variable en cuestión.
- **Desviación típica.** Es una medida de la dispersión en torno a la media. En una distribución normal, el 68% de los casos se encuentra dentro de una desviación típica de la media y el 95% queda entre dos desviaciones típicas. Por ejemplo, si la edad media es de 45 años, con una desviación típica de 10, el 95% de los casos estaría entre los 25 y 65 en una distribución normal.
- **Error típico.** Medida de cuánto puede variar el valor de un estadístico de contraste de muestra en muestra. Es la desviación típica de la distribución muestral de un estadístico. Por ejemplo, el error típico de la media es la desviación típica de las medias muestrales.
- **Curtosis.** Medida del grado en que las observaciones están agrupadas en torno al punto central. Para una distribución normal, el valor del estadístico de curtosis es 0. Una curtosis positiva indica que, con respecto a una distribución normal, las observaciones se concentran más en el centro de la distribución y presentan colas más estrechas hasta los valores extremos de la distribución, en cuyo punto las colas de la distribución leptocúrtica son más gruesas con respecto a una distribución normal. Una curtosis negativa indica que, con respecto a una distribución normal, las observaciones se concentran menos y presentan colas más gruesas hasta los valores extremos de la distribución, en cuyo punto las colas de la distribución platicúrtica son más estrechas con respecto a una distribución normal.
- **Asimetría.** Medida de la asimetría de una distribución. La distribución normal es simétrica y tiene un valor de asimetría igual a 0. Una distribución que tenga una asimetría positiva significativa tiene una cola derecha larga. Una distribución que tenga una asimetría negativa significativa tiene una cola izquierda larga. Como regla aproximada, un valor de la asimetría mayor que el doble de su error típico se asume que indica una desviación de la simetría.

Las siguientes estadísticas de región pueden dar como resultado más de un elemento gráfico por subgrupo. Cuando se utiliza el elemento gráfico de intervalo, área o borde, las estadísticas de región dan como resultado un elemento gráfico que muestra el rango. El resto de los elementos gráficos dan como resultado dos elementos independientes, uno muestra el principio del rango y el otro el final del rango.

- **Región: Rango.** El rango de valores entre los valores mínimos y máximos.
- **Región: 95% de intervalo de confianza de la media.** Un rango de valores que el 95% de las veces incluye la media prevista.
- **Región: 95% de intervalo de confianza de individual.** Rango de valores que el 95% de las veces incluye el valor previsto según el caso individual.
- **Región: 1 desviación estándar por encima/debajo de la media.** Rango de valores entre 1 desviación típica por encima y por debajo de la **media**.
- **Región: 1 error estándar por encima/debajo de la media.** Rango de valores entre 1 **error típico** por encima y por debajo de la **media**.

#### ***Estadísticos de resumen basados en recuentos***

- **Contar apariciones.** El número de filas/casos.
- **Recuento acumulado.** El número acumulado de filas/casos. Cada elemento gráfico muestra el recuento de un subgrupo más la suma total de todos los grupos anteriores.
- **Porcentaje de recuento.** El porcentaje de filas/casos de cada subgrupo comparado con el número total de filas/casos.
- **Porcentaje acumulado de recuento.** El porcentaje acumulado de filas/casos de cada subgrupo comparado con el número total de filas/casos. Cada elemento gráfico muestra el porcentaje de un subgrupo más el porcentaje total de todos los grupos anteriores.

#### ***Para especificar el modificador de colisión***

El modificador de colisión determina lo que ocurre cuando los elementos gráficos se superponen.

- ▶ Seleccione el elemento gráfico cuyo modificador de colisión desee especificar.
- ▶ Pulse en la pestaña Elemento de la paleta Propiedades.
- ▶ En la lista desplegable Modificador, seleccione un modificador de colisión. -auto- permite a la aplicación determinar qué modificador de colisión es adecuado para el tipo de elemento gráfico y estadístico.

**Superpuesto.** Dibuje elementos gráficos encima de otros si tienen el mismo valor.

**Pila.** Elementos gráficos de pila que normalmente se superpondrían al tener los mismos valores de datos.

**Esquivar.** Mueve los elementos gráficos junto a otros elementos gráficos que aparecen con el mismo valor, en vez de superponerlos. Los elementos gráficos se ordenan simétricamente. Es decir, los elementos gráficos se trasladan a lados opuestos de una posición central. Esquivar es similar a conglomerar.

**Pila.** Mueve los elementos gráficos junto a otros elementos gráficos que aparecen con el mismo valor, en vez de superponerlos. Los elementos gráficos se ordenan asimétricamente. Es decir, los elementos gráficos se apilan unos sobre otros, con el elemento gráfico más bajo colocado en un valor específico de la escala.

**Perturbar (normal).** Recoloca los elementos gráficos en el mismo valor de datos aleatoriamente utilizando la distribución normal.

**Perturbar (uniforme).** Recoloca los elementos gráficos en el mismo valor de datos aleatoriamente utilizando la distribución uniforme.

## ***Cambio de la posición de la leyenda***

Si un gráfico incluye una leyenda, ésta suele mostrarse a la derecha de un gráfico. Si lo desea, puede cambiar esta posición.

### ***Cómo cambiar la posición de la leyenda***

- ▶ Seleccione la leyenda.
- ▶ Pulse en la pestaña Leyenda de la paleta Propiedades.

Figura 7-18  
Pestaña Leyenda



- ▶ Seleccione una posición.

## ***Copia de una visualización y de datos de visualización***

La paleta General incluye botones para copiar la visualización y sus datos.

Figura 7-19  
Botón Copiar visualización



**Copia de visualización** Esta acción copia la visualización en el portapapeles como una imagen. Hay múltiples formatos de imagen disponibles. Cuando pegue la imagen en otra aplicación, puede seleccionar una opción de pegado especial para seleccionar uno de los formatos de imagen disponibles para pegar.

Figura 7-20  
Botón Copiar datos de visualización



**Copia de los datos de visualización.** Esta acción copia los datos subyacentes utilizados para dibujar la visualización. Los datos se copian en el portapapeles como texto normal o como texto con formato HTML. Cuando pegue los datos en otra aplicación, puede seleccionar una opción de pegado especial para seleccionar uno de los formatos disponibles para pegar.

**Métodos abreviados de teclado**

Tabla 7-3  
*Métodos abreviados de teclado*

<b>Tecla de método abreviado</b>	<b>Función</b>
Ctrl + Espacio	Cambiar entre el modo de exploración y edición
Supr	Eliminar un elemento de visualización
Ctrl + Z	Deshacer
Ctrl + Y	Rehacer la última acción
F2	Mostrar esquema para seleccionar los elementos del gráfico

# ***Parte III:***

## ***Editor de recursos***

## Plantillas y recursos

IBM® SPSS® Text Analytics for Surveys captura rápidamente y con precisión y extrae los conceptos clave de los datos de texto. Este proceso de extracción se basa mayoritariamente en los recursos lingüísticos para determinar de qué forma se extrae la información de los datos de texto. [Si desea obtener más información, consulte el tema \*Cómo funciona la extracción en el capítulo 1 el p. 3\*](#). Puede ajustar estos recursos en la vista Editor de recursos.

Cuando se instala el software, también se obtiene un conjunto de recursos especializados. Estos recursos incluidos permiten que se beneficie de muchos años de investigación y ajustes para idiomas específicos y aplicaciones específicas. Puesto que no siempre los recursos incluidos se adaptan a la perfección al contexto de sus datos, puede editar estas plantillas de recursos o incluso crear y utilizar bibliotecas personalizadas exclusivamente adaptadas a los datos de su organización. Estos recursos se encuentran en varios formatos y puede utilizarlos en su proyecto. Los recursos pueden encontrarse en los siguientes lugares:

- **Plantilla de recursos.** Las plantillas se componen de un conjunto de bibliotecas, tipos y algunos recursos avanzados que juntos conforman un conjunto especializado de recursos que se adapta a un dominio o contexto particular, como las opiniones sobre productos.
- **Paquetes de análisis de texto (TAP).** Además de los recursos almacenados en una plantilla, los TAP unen además uno o más conjuntos de categorías especializados que se han generado utilizando esos recursos, de modo que tanto las categorías como los recursos se almacenan y reutilizan conjuntamente. [Si desea obtener más información, consulte el tema \*Uso de los paquetes de análisis de texto en el capítulo 3 el p. 41\*](#).
- **Bibliotecas.** Las bibliotecas se utilizan como los cimientos sobre los que se basan tanto los TAP como las plantillas. También pueden añadirse individualmente a los recursos de su proyecto. Cada biblioteca está compuesta por varios diccionarios que se utilizan para definir y administrar listas de tipos, sinónimos y exclusiones. Aunque las bibliotecas también se entregan individualmente, se empaquetan juntas en plantillas y en los TAP. [Si desea obtener más información, consulte el tema \*Trabajo con bibliotecas en el capítulo 9 el p. 202\*](#).

*Nota:* Durante la extracción, también se utilizan algunos recursos internos compilados. Estos recursos compilados contienen un gran número de definiciones que complementan a los tipos de la biblioteca Core. Estos recursos compilados no se pueden editar.

El Editor de recursos ofrece acceso al conjunto de recursos que se utilizan para generar los resultados de la extracción (conceptos, tipos y patrones). Existe una serie de tareas que puede realizar en el Editor de recursos, e incluyen:

- **Trabajar con bibliotecas.** [Si desea obtener más información, consulte el tema \*Trabajo con bibliotecas en el capítulo 9 el p. 202\*](#).
- **Crear diccionarios de tipo.** [Si desea obtener más información, consulte el tema \*Creación de tipos en el capítulo 10 el p. 217\*](#).
- **Añadir términos a los diccionarios.** [Si desea obtener más información, consulte el tema \*Adición de términos en el capítulo 10 el p. 219\*](#).

- **Crear sinónimos.** Si desea obtener más información, consulte el tema Definición de sinónimos en el capítulo 10 el p. 227.
- **Actualizar los recursos en los TAP.** Si desea obtener más información, consulte el tema Actualización de los paquetes de análisis de texto en el capítulo 3 el p. 43.
- **Crear plantillas.** Si desea obtener más información, consulte el tema Creación y actualización de plantillas el p. 193.
- **Importar y exportar plantillas.** Si desea obtener más información, consulte el tema Importación y exportación de plantillas el p. 196.
- **Publicar bibliotecas.** Si desea obtener más información, consulte el tema Publicación de bibliotecas en el capítulo 9 el p. 211.

## La interfaz del editor

Las operaciones que realiza en el Editor de recursos se concentran en torno a la administración y ajuste de los recursos lingüísticos. Estos recursos se almacenan en forma de plantillas y bibliotecas. Si desea obtener más información, consulte el tema Diccionarios de tipo en el capítulo 10 el p. 215.

### Pestaña Recursos de la biblioteca

Figura 8-1

The screenshot displays the 'Recursos' (Resources) tab in the IBM SPSS Text Analytics for Surveys application. The interface is organized as follows:

- Left Pane (Library Management):** Shows a tree view of libraries under 'Opinions (English)', including 'Biblioteca local', 'Opinions Library (English)', 'Budget Library (English)', 'Core Library (English)', 'Variations Library (English)', 'Slang Library (English)', and 'Emoticon Library (English)'. A red '1' is placed over this pane.
- Central Pane (Resources Table):** A table with columns: 'Término', 'Coincidencia', 'Flexionar', 'Tipo', and 'Biblioteca'. It lists various terms like '0 problem', '10 stars', '100% correct', etc. A red '2' is placed over the 'Flexionar' column.
- Right Pane (Exclusions):** A table with columns: 'Excluidos' and 'Biblioteca'. It lists terms like 'any kind of problem', 'any problems', etc. A red '3' is placed over this pane.
- Bottom Pane (Synonyms):** Shows a table with columns: 'Objetivo', 'Sinónimos', and 'Biblioteca'. It displays the goal 'able to log-on' and its corresponding synonyms like 'able to log-in', 'able to login', etc. A red '4' is placed over this pane.

At the bottom of the window, a status bar shows: 7 Bibliotecas, 35 Tipos, 13425 Términos, 28 Excluidos, 1334 Sinónimos, and 27 Opcional.

La interfaz está organizada en cuatro partes:

**1. Panel del árbol de bibliotecas.** Situada en el ángulo superior izquierdo, este panel muestra un árbol con las bibliotecas. Puede activar y desactivar las bibliotecas de este árbol y filtrar las vistas en los otros paneles seleccionando una biblioteca del árbol. Puede realizar muchas operaciones en este árbol utilizando los menús contextuales. Si expande una biblioteca en el árbol, podrá ver el conjunto de tipos que contiene. También puede filtrar esta lista a través del menú Ver si desea centrarse únicamente en una biblioteca determinada.

**2. Listas de términos del panel Diccionarios de tipo.** Situado a la derecha del árbol de bibliotecas, este panel muestra las listas de términos de los diccionarios de tipo de las bibliotecas seleccionadas en el árbol. Un **diccionario de tipo** es una recopilación de términos que se agrupan bajo una misma etiqueta, tipo o nombre. Cuando el motor de extracción lee los datos de texto, compara las palabras que encuentra en el texto con los términos definidos en los diccionarios de tipo. Si un concepto extraído aparece como término en un diccionario de tipo, se asignará ese nombre de tipo. Puede considerar que el diccionario de tipo es un diccionario específico de términos que tienen algo en común. Por ejemplo, el tipo <Location> de la biblioteca Core contiene conceptos como *nueva orleans*, *gran bretaña* y *nueva york*. Todos estos términos representan ubicaciones geográficas. Una biblioteca puede contener uno o más diccionarios de tipo. [Si desea obtener más información, consulte el tema Diccionarios de tipo en el capítulo 10 el p. 215.](#)

**3. Panel del diccionario de exclusión.** Situado en el lado derecho, este panel muestra la colección de términos que se excluirán de los resultados de extracción finales. Los términos que aparecen en este diccionario de exclusión no aparecen en el panel Resultados extraídos. Los términos excluidos pueden almacenarse en la biblioteca que usted elija. Sin embargo, el panel Diccionario de exclusión muestra todos los términos extraídos de todas las bibliotecas visibles en el árbol de bibliotecas. [Si desea obtener más información, consulte el tema Diccionarios de exclusión en el capítulo 10 el p. 231.](#)

**4. Panel del diccionario de sustitución.** Situado en la parte inferior izquierda, este panel muestra los sinónimos y los elementos opcionales, cada uno en su propia pestaña. Los sinónimos y los elementos opcionales ayudan a agrupar términos similares bajo un concepto principal u objetivo en los resultados de extracción finales. Este diccionario puede contener sinónimos conocidos y sinónimos definidos por el usuario y elementos, así como los errores ortográficos más comunes emparejados con la ortografía correcta. Las definiciones de sinónimos y los elementos opcionales pueden almacenarse en la biblioteca que elija. Sin embargo, el panel del diccionario de sustitución muestra todos los contenidos de todas las bibliotecas visibles en el árbol de bibliotecas. Mientras que este panel muestra todos los sinónimos o elementos opcionales de todas las bibliotecas, las sustituciones para todas las bibliotecas del árbol se muestran conjuntamente en este panel. Una biblioteca puede contener tan solo un diccionario de sustitución. [Si desea obtener más información, consulte el tema Diccionarios de sustitución/sinónimos en el capítulo 10 el p. 225.](#)

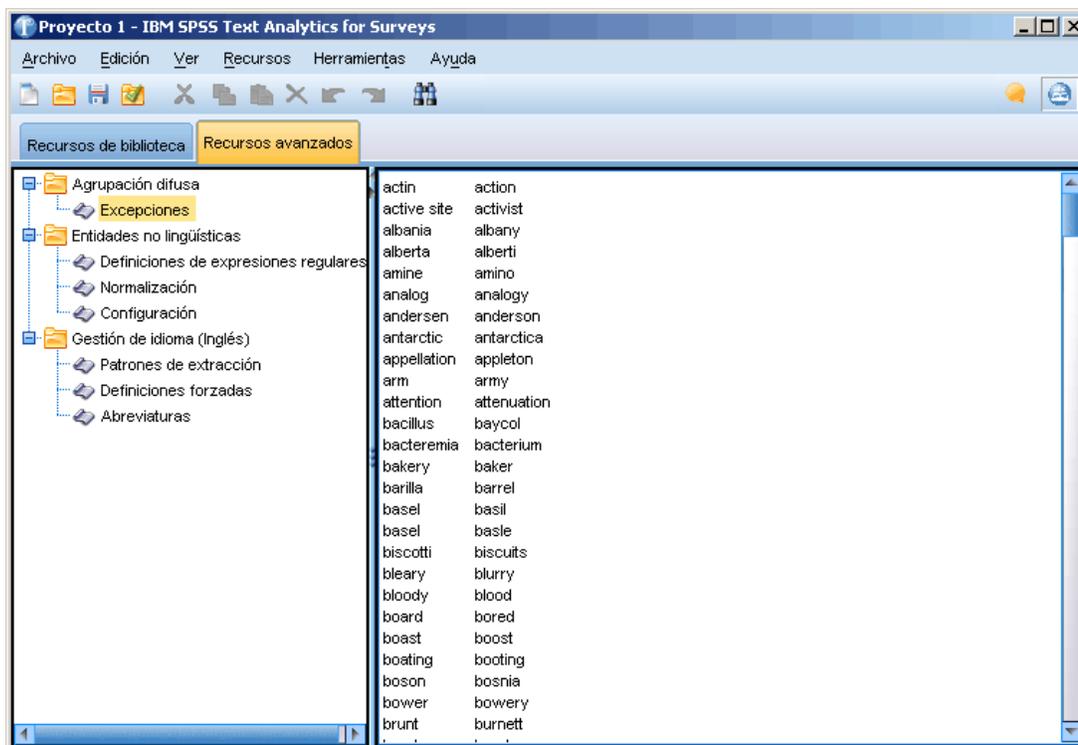
*Notas:*

- Si desea filtrar esta ventana de manera que solo se vea la información que pertenece a una única biblioteca, puede cambiar la vista de la biblioteca mediante la lista desplegable de la barra de herramientas. Contiene una entrada de nivel superior llamada Todas las bibliotecas así como una entrada adicional para cada biblioteca individual. [Si desea obtener más información, consulte el tema Visión de bibliotecas en el capítulo 9 el p. 206.](#)

### ***Pestaña Recursos avanzados***

Los recursos avanzados están disponibles en la segunda pestaña de la vista del editor. Puede revisar y editar los recursos avanzados en esta pestaña. [Si desea obtener más información, consulte el tema Acerca de los recursos avanzados en el capítulo 11 el p. 233.](#)

Figura 8-2  
*Recursos avanzados*



## ***Creación y actualización de plantillas***

Cuando se realizan cambios en los recursos y prevé que los utilizará en otro momento, puede guardar los recursos en forma de plantilla. Al hacerlo, puede optar por guardar utilizando un nombre de plantilla existente o asignando un nombre nuevo. Más adelante, cuando quiera cargar esta plantilla, podrá disponer de los mismos resultados.

*Nota:* también puede publicar y compartir las bibliotecas. [Si desea obtener más información, consulte el tema Compartimiento de bibliotecas en el capítulo 9 el p. 210.](#)

Figura 8-3  
Cuadro de diálogo Crear plantilla de recursos



### ***Para crear (o actualizar) una plantilla***

- ▶ En los menús de la vista del Editor de recursos, elija Recursos > Crear plantilla de recursos. Aparecerá el cuadro de diálogo Crear plantilla de recursos.
- ▶ Escriba un nombre nuevo en el campo Nombre de plantilla si desea crear una plantilla nueva. Seleccione una plantilla de la tabla si desea sobrescribir una plantilla existente con los recursos actualmente cargados.
- ▶ Pulse en Guardar para crear la plantilla.

## ***Cambio de plantillas de recursos***

Si desea sustituir los recursos actualmente cargados por una copia de los recursos de otra plantilla, puede cambiar a esos otros recursos. Al hacerlo se sobrescribirán los recursos actualmente cargados.

Puede seleccionar la plantilla cuyo contenido desea copiar en el Editor de recursos y pulsar en Aceptar. Así se sustituyen los recursos que tiene en este proyecto.

Figura 8-4  
Cuadro de diálogo Cambiar de recursos



### Para cambiar de recursos

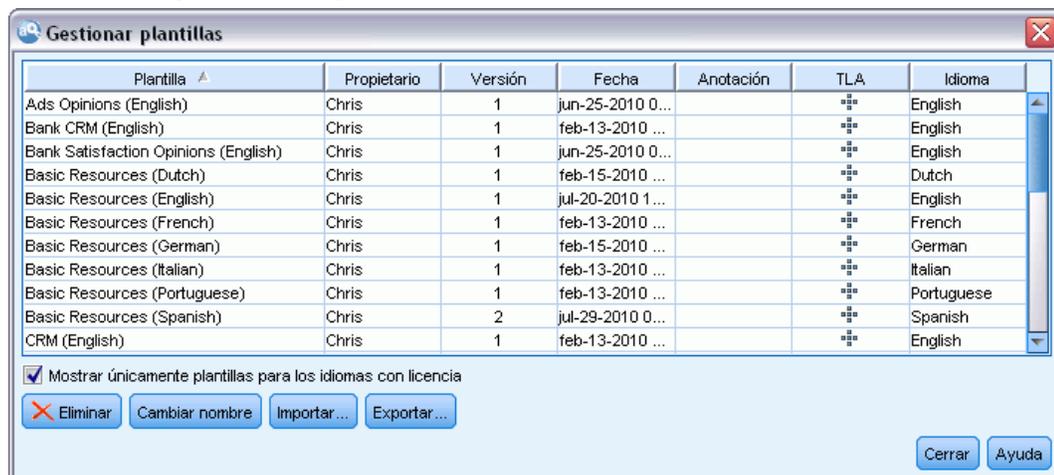
- ▶ En los menús de la vista del Editor de recursos, elija Recursos > Cambiar de plantillas de recursos. Aparecerá el cuadro de diálogo Cambiar de plantillas de recursos.
- ▶ Seleccione la plantilla que desea utilizar entre las que aparecen en la tabla.
- ▶ Pulse en Aceptar para salir de los recursos actualmente cargados y cargar una copia de los recursos de la plantilla seleccionada. Si ha realizado cambios en los recursos y desea guardar las bibliotecas para su uso futuro, puede publicarlas, actualizarlas y compartirlas antes realizar el cambio. [Si desea obtener más información, consulte el tema Compartimiento de bibliotecas en el capítulo 9 el p. 210.](#)

## Administración de plantillas

Existen también algunas tareas básicas de administración que puede realizar de vez en cuando en las plantillas, como cambiarles el nombre, importar y exportar plantillas o eliminar las que hayan quedado obsoletas. Estas tareas se realizan en el cuadro de diálogo Administrar plantillas. La importación y exportación de plantillas permite compartir plantillas con otros usuarios. [Si desea obtener más información, consulte el tema Importación y exportación de plantillas el p. 196.](#)

*Nota:* No puede cambiar el nombre ni eliminar las plantillas que están instaladas (o se envían) con el producto. En su lugar, si desea cambiarles el nombre, abra la plantilla instalada y cree una nueva con el nombre que desee. Puede eliminar las plantillas personalizadas, pero si intenta eliminar una de las plantillas incorporadas, se restaurará a la versión originalmente instalada.

Figura 8-5  
Cuadro de diálogo Administrar plantillas



#### **Para cambiar el nombre de una plantilla**

- ▶ En los menús elija Recursos > Administrar plantillas de recursos. Aparecerá el cuadro de diálogo Administrar plantillas.
- ▶ Seleccione la plantilla a la que desea cambiar el nombre y pulse en Cambiar nombre. El cuadro de nombre pasará a ser un campo editable en la tabla.
- ▶ Teclee un nombre nuevo y pulse la tecla Intro. Se abrirá un cuadro de diálogo de confirmación.
- ▶ Si está conforme con el cambio de nombre, pulse en Sí. De lo contrario, pulse en No.

#### **Para eliminar una plantilla**

- ▶ En los menús elija Recursos > Administrar plantillas de recursos. Aparecerá el cuadro de diálogo Administrar plantillas.
- ▶ En el cuadro de diálogo Administrar plantillas, seleccione la plantilla que desea eliminar.
- ▶ Pulse en Eliminar. Se abrirá un cuadro de diálogo de confirmación.
- ▶ Pulse en Sí para eliminar o en No para cancelar la solicitud. Si pulsa en Sí, la plantilla se elimina.

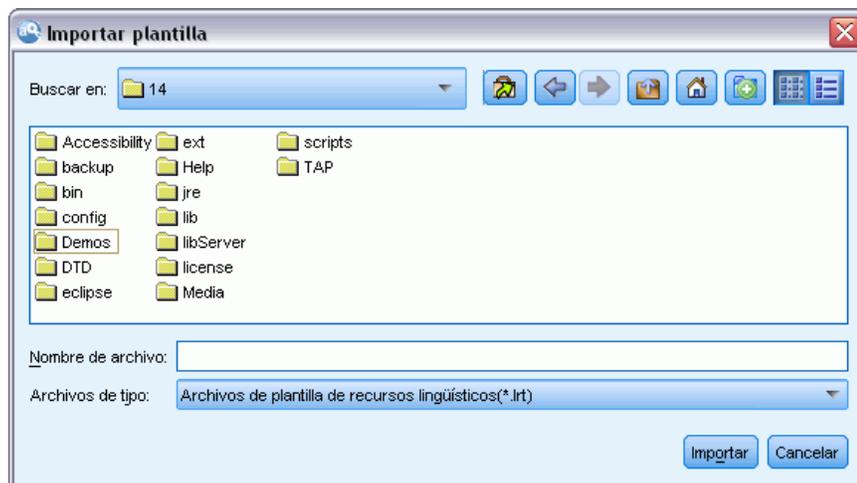
## **Importación y exportación de plantillas**

Puede compartir plantillas con otros usuarios u otras máquinas importándolas y exportándolas. Las plantillas se guardan en una base de datos interna, pero pueden exportarse en forma de archivos \*.lrt en el disco duro. Puede importar y exportar plantillas en el cuadro de diálogo Administrar plantillas en el Editor de recursos.

#### **Para importar una plantilla**

- ▶ En el cuadro de diálogo, pulse en Importar. Aparecerá el cuadro de diálogo Importar plantilla.

Figura 8-6  
Cuadro de diálogo Importar plantilla

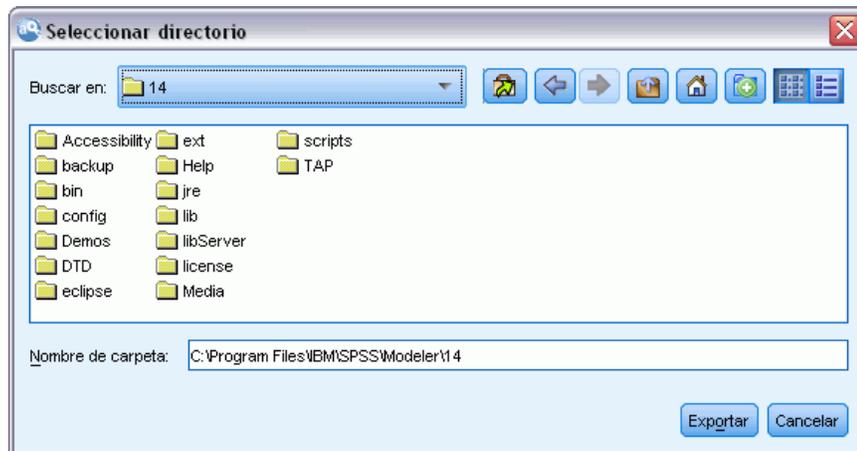


- Seleccione el archivo de la plantilla de recursos (\*.lrt) para importar y pulse en Importar. Puede guardar la plantilla que está importando con otro nombre o sobrescribir la existente. El cuadro de diálogo se cierra y la plantilla aparece ahora en la tabla.

#### **Para exportar una plantilla**

- En el cuadro de diálogo, seleccione la plantilla que desea exportar y pulse en Exportar. Aparecerá el cuadro de diálogo Seleccionar directorio.

Figura 8-7  
Cuadro de diálogo Seleccionar directorio



- Seleccione el directorio al que desea exportar y pulse en Exportar. Este cuadro de diálogo se cierra y la plantilla se exporta y adopta la extensión de archivo (\*.lrt)

## **Copia de seguridad de los recursos**

Puede hacer una copia de seguridad de sus recursos periódicamente como medida de seguridad.

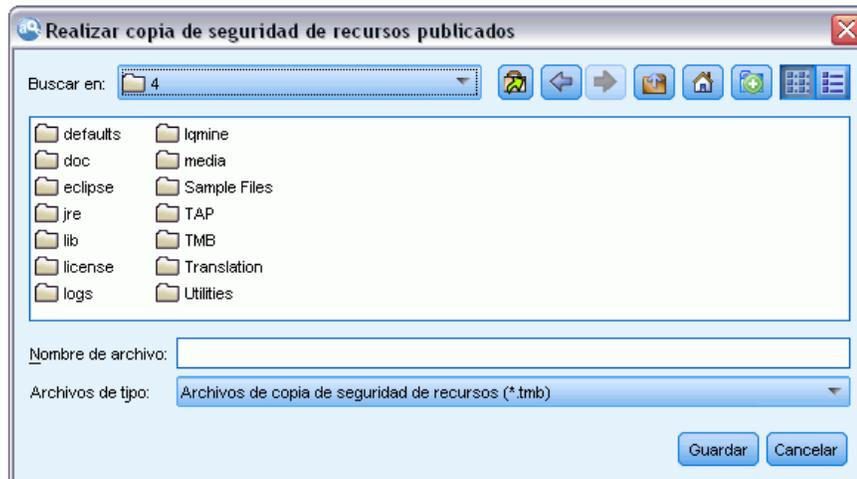
**Importante:** A la hora de restaurar, los contenidos completos de los recursos se limpiarán y sólo podrá accederse al contenido del archivo de seguridad en el producto. Esta acción incluye todos los trabajos abiertos.

### **Para hacer una copia de seguridad de los recursos**

- En los menús elija Recursos > Realizar copia de seguridad de herramientas> Realizar copia de seguridad de recursos. Aparecerá el cuadro de diálogo Copia de seguridad.

Figura 8-8

Cuadro de diálogo Realizar copia de seguridad de recursos



- Introduzca un nombre para el archivo de copia de seguridad y pulse en Guardar. El cuadro de diálogo se cierra y se crea el archivo de copia de seguridad.

### **Para restaurar los recursos**

- En los menús elija Recursos > Realizar copia de seguridad de herramientas> Restaurar recursos. Un mensaje de alerta le advierte que al restaurar se sobrescriben todos los contenidos actuales de la base de datos.

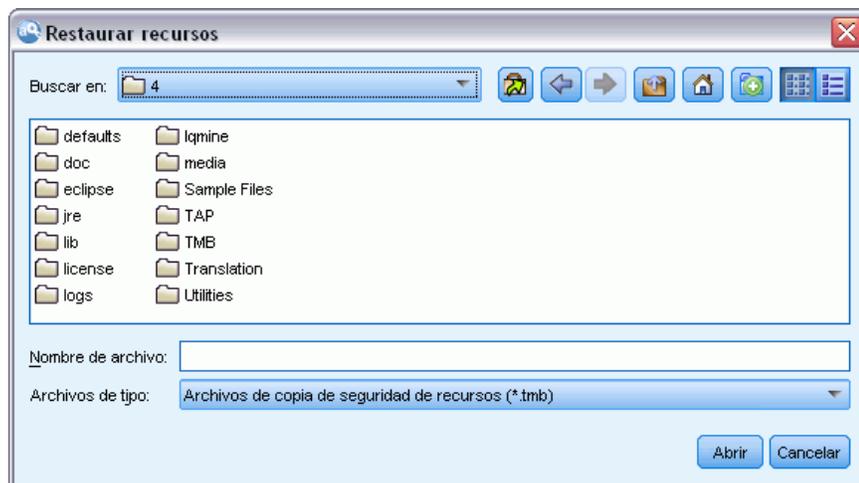
Figura 8-9

Mensaje de alerta de sobrescritura



- Pulse en Sí para continuar. Si hay un proyecto abierto, éste se conservará porque está en la memoria; sin embargo, deberá volver a guardarlo en la base de datos recién restaurada. Se abrirá el cuadro de diálogo.

Figura 8-10  
Cuadro de diálogo Restaurar recursos



- Seleccione el archivo de copia de seguridad que desea restaurar y pulse en Abrir. El cuadro de diálogo se cierra y los recursos se restauran en la aplicación.

**Importante:** A la hora de restaurar, los contenidos completos de los recursos se limpiarán y sólo podrá accederse al contenido del archivo de seguridad en el producto. Esta acción incluye todos los trabajos abiertos.

## Importación de los archivos de recursos

Si ha introducido cambios directamente en los archivos de recursos fuera de este producto, puede importarlos a la biblioteca que quiera seleccionando esta biblioteca y procediendo con la importación. Cuando se importa a un directorio, también puede importar todos los archivos admitidos en una biblioteca abierta específica. Sólo puede importar archivos *\*.txt*.

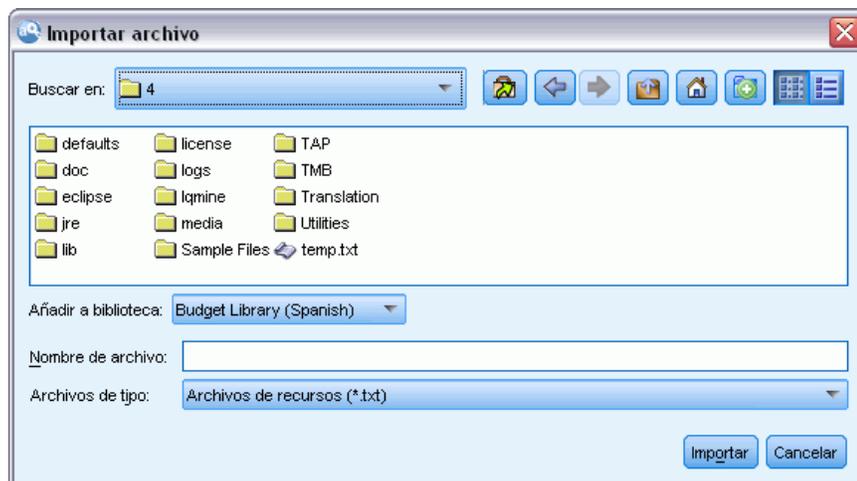
Cada archivo importado debe contener únicamente una entrada por línea, y si el contenido está estructurado así:

- Una lista de palabras o frases (una por línea). El archivo se importa como una lista de términos para un diccionario de tipo, donde el diccionario de tipo toma el nombre del archivo menos la extensión.
- Una lista de entradas como *term1*<PESTAÑA>*term2*, se importa como una lista de sinónimos, donde *term1* es el conjunto del término subyacente y *term2* es el término objetivo.

### Para importar un solo archivo de recursos

- En los menús elija Recursos > Importar archivos > Importar archivo único. Aparecerá el cuadro de diálogo Importar archivo.

Figura 8-11  
Cuadro de diálogo Importar archivo

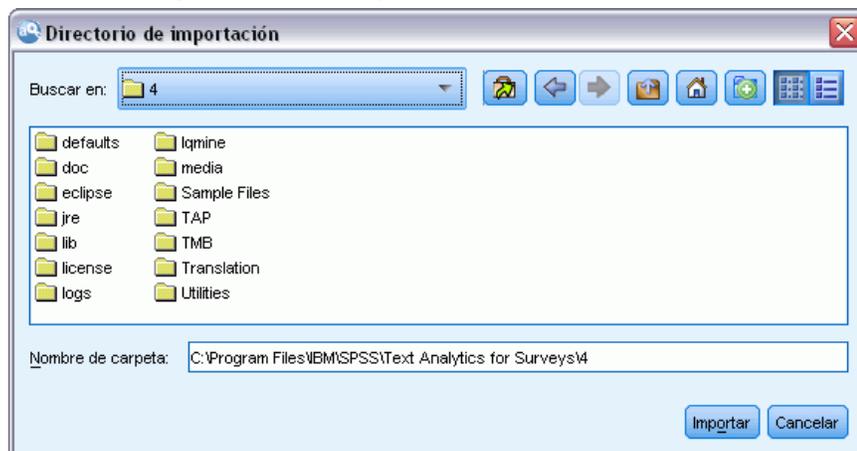


- ▶ Seleccione el archivo que desea importar y pulse en Importar. El contenido del archivo se transforma en un formato interno y se añade a la biblioteca.

#### **Para importar todos los archivos de un directorio**

- ▶ En los menús elija Recursos > Importar archivos > Importar directorio completo. Aparecerá el cuadro de diálogo Directorio de importación.

Figura 8-12  
Cuadro de diálogo Directorio de importación



- ▶ Seleccione la biblioteca en la que desee que se importen todos los archivos de recursos de la lista Importar. Si selecciona la opción Por defecto, se creará una biblioteca nueva con el mismo nombre del directorio.
- ▶ Seleccione el directorio desde el que importar los archivos. Los subdirectorios no se leerán.

- ▶ Pulse en Importar. El cuadro de diálogo se cierra y el contenido de los archivos de recursos importados aparece ahora en el editor en forma de diccionarios y de archivos de recursos avanzados.

## Trabajo con bibliotecas

Los recursos que utiliza el motor de extracción para extraer y agrupar términos de los datos de texto siempre contienen una o más bibliotecas. Puede ver el conjunto de bibliotecas en el árbol de bibliotecas situado en la parte superior izquierda de Editor de recursos. Las bibliotecas están compuestas por tres tipos de diccionarios: De tipo, de sustitución y de exclusión. [Si desea obtener más información, consulte el tema Acerca de los diccionarios de biblioteca en el capítulo 10 el p. 215.](#)

La plantilla de recursos o los recursos del TAP que seleccionó incluye varias bibliotecas que le permiten empezar a extraer inmediatamente conceptos de los datos de texto. Sin embargo, puede crear y publicar sus propias bibliotecas para poder reutilizarlas. [Si desea obtener más información, consulte el tema Publicación de bibliotecas el p. 211.](#)

Por ejemplo, supongamos que suele trabajar con datos de texto relacionados con la industria del automóvil. Después de analizar los datos, decide que desea crear algunos recursos personalizados para gestionar el vocabulario o la jerga específica de dicha industria. Con el Editor de recursos, puede crear una plantilla nueva, y en ella una biblioteca para extraer y agrupar los términos automovilísticos. Puesto que volverá a necesitar la información de esta biblioteca, la biblioteca se publica en un repositorio central que es accesible desde el cuadro de diálogo **Administrar bibliotecas**, para que pueda reutilizarse independientemente en diferentes proyectos.

Supongamos que también desea agrupar términos específicos de diferentes subsectores, como dispositivos electrónicos, motores, sistemas de refrigeración o incluso de un fabricante o mercado en particular. Puede crear una biblioteca para cada grupo y luego publicarlas para que puedan utilizarse en diferentes conjuntos de datos de texto. De esta forma, puede añadir las bibliotecas que mejor se correspondan con el contexto de los datos de texto.

*Nota:* En la pestaña Recursos avanzados pueden configurarse y administrarse recursos adicionales. Algunos se aplican a todas las bibliotecas y administran entidades no lingüísticas, excepciones de agrupación difusa, etc. [Si desea obtener más información, consulte el tema Acerca de los recursos avanzados en el capítulo 11 el p. 233.](#)

### Bibliotecas enviadas

Por defecto se instalan varias bibliotecas con IBM® SPSS® Text Analytics for Surveys. Puede utilizar estas bibliotecas preformateadas para acceder a miles de términos y sinónimos predefinidos, y a muchos tipos diferentes. A estas bibliotecas enviadas se les realizan ajustes según los diferentes dominios y están disponibles en varios idiomas distintos.

Existe un gran número de bibliotecas pero las que se utilizan habitualmente son las siguientes:

- **Biblioteca local.** Se utiliza para almacenar diccionarios definidos por el usuario. Es una biblioteca vacía que se añade por defecto a todos los recursos. También contiene un diccionario vacío de tipo. Resulta más útil cuando se realizan cambios o reajustes en los recursos directamente (como añadir una palabra a un tipo) desde ventana de análisis de texto. En este caso, estos cambios y reajustes se almacenan automáticamente en la primera

biblioteca de la lista del árbol de bibliotecas, en el Editor de recursos; por defecto, es la *Biblioteca local*. Esta biblioteca no puede publicarse porque es específica de los datos del proyecto. Si desea publicar el contenido, primero deberá cambiar el nombre de la biblioteca.

- **Biblioteca Core.** Se utiliza en la mayoría de los casos, pues contiene los cinco tipos básicos incorporados, que representan a personas, ubicaciones, organizaciones, productos y desconocido. Es posible que sólo vea unos cuantos términos en uno de sus diccionarios de tipo, pero los tipos que están representados en la biblioteca Core son en realidad complementos de los tipos más sólidos que se encuentran en los recursos compilados internos que se entregan junto con el producto de minería de datos de texto. Estos recursos compilados internos contienen miles de términos por cada tipo. Por esta razón, aunque un término no pueda verse en la lista de términos del diccionario de tipo, todavía puede extraerse y escribirse con un tipo Core. Esto explica cómo nombres como *Jorge* pueden extraerse y tipificarse como <Person> mientras que en el diccionario de tipo <Person> de la biblioteca Core solo aparece el nombre inglés *John*. Del mismo modo, si no incluye la biblioteca Core, puede que siga viendo estos tipos en los resultados de extracción, puesto que el motor de extracción utilizará los recursos compilados que contienen estos tipos.
- **Biblioteca Opinions.** Se utiliza habitualmente para extraer opiniones e impresiones de los datos de texto. Esta biblioteca incluye miles de palabras que representan actitudes, calificadores y preferencias que—cuando se utilizan en conjunción con otros términos—indican una opinión sobre un tema. Esta biblioteca incluye un número de tipos, sinónimos y exclusiones incorporadas. También incluye un voluminoso conjunto de reglas de patrones que se utilizan para el análisis de enlace de texto.
- **Biblioteca Budget.** Se utiliza para extraer los términos relacionados con el coste de las cosas. Esta biblioteca incluye muchas palabras y frases que representan adjetivos, calificadores y juicios sobre el precio o la calidad de las cosas.
- **Biblioteca Variations.** Se utiliza para incluir casos donde algunas variaciones del idioma requieren definiciones de sinónimos para poder agruparlas adecuadamente. Esta biblioteca solo contiene definiciones de sinónimos.

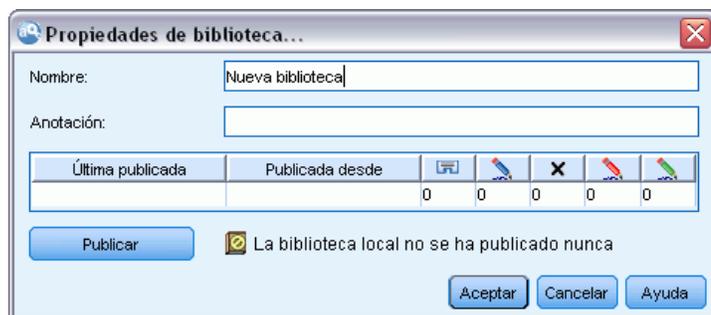
Aunque algunas de las bibliotecas enviadas fuera de las plantillas tienen un contenido similar al de algunas plantillas, éstas se han ajustado específicamente a aplicaciones determinadas y contienen recursos avanzados adicionales. Es recomendable que intente utilizar una plantilla que se haya diseñado para el tipo de datos de texto con el que está trabajando y realice sus cambios en aquellos recursos en lugar de añadir simplemente bibliotecas individuales a una plantilla más general.

Con SPSS Text Analytics for Surveys también se entregan recursos compilados. Siempre se utilizan durante el proceso de extracción y contienen un gran número de definiciones complementarias a los diccionarios de tipo incorporados en las bibliotecas por defecto. Puesto que estos recursos están compilados, no pueden verse ni editarse. Sin embargo, puede forzar que un término especificado en los recursos compilados se coloque en otro diccionario. [Si desea obtener más información, consulte el tema Forzado de términos en el capítulo 10 el p. 222.](#)

## ***Creación de bibliotecas***

Puede crear un número indefinido de bibliotecas. Después de crear una biblioteca nueva, puede empezar a crear diccionarios de tipo dentro de la biblioteca e introducir términos, sinónimos y exclusiones.

Figura 9-1  
Cuadro de diálogo *Propiedades de biblioteca*



### **Para crear una biblioteca**

- ▶ En los menús elija Recursos > Nueva biblioteca. Se abre el cuadro de diálogo Añadir biblioteca al proyecto.
- ▶ Introduzca un nombre para la biblioteca en el cuadro de texto Nombre.
- ▶ Si lo desea, introduzca un comentario en el cuadro de texto Anotación.
- ▶ Pulse en Publicar si desea publicar esta biblioteca ahora antes de introducir nada más en ella. [Si desea obtener más información, consulte el tema Compartimiento de bibliotecas el p. 210.](#) También puede publicarla en cualquier otro momento.
- ▶ Pulse en Aceptar para crear la biblioteca. El cuadro de diálogo se cierra y la biblioteca aparece en la vista de árbol. Si expande las bibliotecas del árbol, comprobará que en la biblioteca se ha incluido automáticamente un diccionario vacío de tipo. Puede empezar a añadir términos en él inmediatamente. [Si desea obtener más información, consulte el tema Adición de términos en el capítulo 10 el p. 219.](#)

## **Adición de bibliotecas públicas**

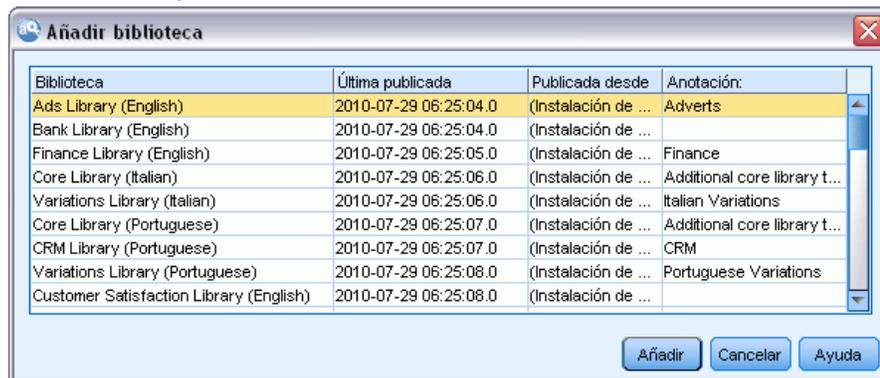
Si desea reutilizar una biblioteca de los datos de otra o proyecto, puede añadirla a sus recursos actuales siempre que se trate de una biblioteca pública. Una **biblioteca pública** es aquella que ha sido publicada. [Si desea obtener más información, consulte el tema Publicación de bibliotecas el p. 211.](#)

Cuando añade una biblioteca pública, se incrusta una copia **local** en los datos de la o proyecto. Puede realizar cambios en esta biblioteca, pero deberá volver a publicar su versión pública para poder compartir los cambios.

Cuando se añade una biblioteca pública, puede aparecer el cuadro de diálogo Resolver conflictos si se detectan conflictos entre los términos y los tipos de una biblioteca y las otras bibliotecas locales. Deberá resolver estos conflictos o aceptar las resoluciones propuestas para poder finalizar la operación. [Si desea obtener más información, consulte el tema Resolución de conflictos el p. 213.](#)

*Nota:* si siempre actualiza las bibliotecas cuando abre un proyecto o las publica cuando cierra un proyecto, será menos probable que tenga bibliotecas desincronizadas. [Si desea obtener más información, consulte el tema Compartimiento de bibliotecas el p. 210.](#)

Figura 9-2  
Cuadro de diálogo Añadir biblioteca



### **Para añadir una biblioteca**

- ▶ En los menús elija Recursos > Añadir biblioteca. Aparecerá el cuadro de diálogo Añadir biblioteca.
- ▶ Seleccione la biblioteca o bibliotecas de la lista.
- ▶ Pulse en Añadir. Si se producen conflictos entre las bibliotecas recién añadidas y las que ya existían, se le pedirá que verifique las resoluciones de conflictos o que las cambie antes de finalizar la operación. [Si desea obtener más información, consulte el tema Resolución de conflictos el p. 213.](#)

## **Búsqueda de términos y tipos**

Puede buscar en los diversos paneles del editor utilizando la función Buscar. En los menús del editor puede elegir Editar > Buscar; aparecerá la barra de herramientas Buscar. Puede utilizar esta barra de herramientas para buscar una aparición cada vez. Si pulsa Buscar de nuevo, puede buscar apariciones subsiguientes del término de búsqueda.

Al realizar la búsqueda, el editor solo busca en la biblioteca o bibliotecas de la lista desplegable de la barra de herramientas Buscar. Si se selecciona Todas las bibliotecas, el programa buscará en todas las bibliotecas del editor.

Cuando se inicia una búsqueda, empieza por el área que tiene el foco. La búsqueda continúa a través de cada sección, y vuelve al principio hasta que llega a la casilla activa. Puede invertir el orden de la búsqueda mediante las flechas direccionales. También puede elegir si su búsqueda distingue o no entre mayúsculas y minúsculas.

### **Para buscar cadenas en la vista**

- ▶ En los menús elija Editar > Buscar. Aparece la barra de herramientas Buscar.
- ▶ Escriba la cadena que desea buscar.

- ▶ Pulse en el botón Buscar para empezar la búsqueda. La siguiente aparición del término o del tipo quedará resaltada.
- ▶ Pulse otra vez en el botón para pasar de una aparición a otra.

## **Visión de bibliotecas**

Puede mostrar los contenidos de una biblioteca particular o de todas las bibliotecas. Esto puede resultar útil cuando trabaja con muchas bibliotecas o cuando desea revisar los contenidos de una biblioteca específica antes de publicarla. El cambio de la vista sólo incide en lo que ve en esta pestaña Recursos de la biblioteca, pero no impide el uso de las bibliotecas durante la extracción. [Si desea obtener más información, consulte el tema Desactivación de bibliotecas locales el p. 207.](#)

La vista por defecto es Todas las bibliotecas, que muestra todas las bibliotecas del árbol y sus contenidos en otros paneles. Puede cambiar esta selección utilizando la lista desplegable en la barra de herramientas o mediante la selección de un menú (Ver > Bibliotecas). Cuando se está viendo una sola biblioteca, todos los elementos del resto de las bibliotecas desaparecen de la vista pero siguen leyéndose durante la extracción.

### **Para cambiar la vista de Biblioteca**

- ▶ En los menús de la pestaña Recursos de la biblioteca, elija Ver > Bibliotecas. Se abre un menú con todas las bibliotecas locales.
- ▶ Seleccione la biblioteca que desea ver o seleccione la opción Todas las bibliotecas para ver los contenidos de todas las bibliotecas. Los contenidos de la vista se filtran de acuerdo con su selección.

## **Administración de las bibliotecas locales**

Las bibliotecas locales son aquellas que se encuentran dentro de del proyecto o dentro de una plantilla, en contraposición a las bibliotecas públicas. [Si desea obtener más información, consulte el tema Administración de bibliotecas públicas el p. 208.](#) Existen también algunas tareas de administración de bibliotecas locales básicas que puede realizar, por ejemplo: cambiar el nombre, desactivar o eliminar una biblioteca local.

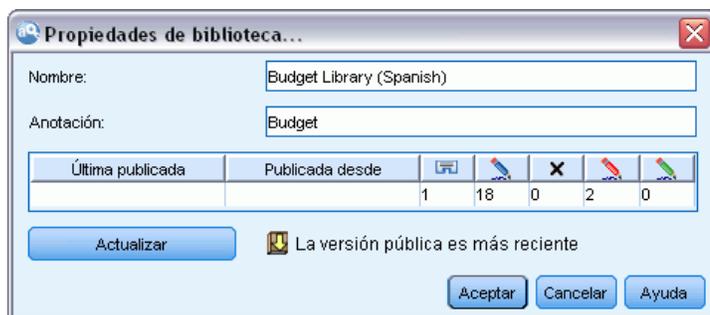
### **Cambio de nombre de las bibliotecas locales**

Puede cambiar el nombre de las bibliotecas locales. Si cambia el nombre de una biblioteca local, elimina la asociación que tiene de la versión pública, si es que existe. Esto significa que si realiza cambios posteriores no podrá compartirlos en la versión pública. Puede volver a publicar esta biblioteca local con un nombre nuevo. Esto significa también que no podrá actualizar la versión pública original con cambios que realice en esta versión local.

*Nota:* No puede cambiar el nombre de una biblioteca pública.

- ▶ En los menús elija Editar > Propiedades de biblioteca. Aparecerá el cuadro de diálogo Propiedades de biblioteca.

Figura 9-3  
Cuadro de diálogo *Propiedades de biblioteca*



#### **Para cambiar el nombre de una biblioteca local**

- ▶ En la vista de árbol, seleccione la biblioteca a la que desea cambiar el nombre.
- ▶ Introduzca un nombre nuevo para la biblioteca en el cuadro de texto Nombre.
- ▶ Pulse en Aceptar para aceptar el nombre nuevo de la biblioteca. El cuadro de diálogo se cierra y el nombre de la biblioteca se actualiza en la vista de árbol.

### **Desactivación de bibliotecas locales**

Si desea excluir temporalmente una biblioteca del proceso de extracción, puede anular la selección del cuadro de verificación a la izquierda del nombre de la biblioteca de la vista de árbol. Esto indica que desea conservar la biblioteca pero que su contenido se pase por alto cuando se realice la comprobación de conflictos y durante el proceso de extracción.

#### **Para desactivar una biblioteca**

- ▶ En el panel del árbol de bibliotecas, seleccione la biblioteca que desea desactivar.
- ▶ Pulse la barra espaciadora. Se borra la marca de la casilla de verificación a la izquierda del nombre.

### **Eliminación de bibliotecas locales**

Puede suprimir una biblioteca sin eliminar la versión pública de la misma y viceversa. La eliminación de una biblioteca local puede eliminar la biblioteca y todo su contenido solamente de del proyecto. La eliminación de una versión local de una biblioteca no supone la eliminación de dicha biblioteca de otras o proyectos ni de la versión pública. [Si desea obtener más información, consulte el tema Administración de bibliotecas públicas el p. 208.](#)

#### **Para eliminar una biblioteca local**

- ▶ En la vista de árbol, seleccione la biblioteca que desea eliminar.
- ▶ En los menús elija Editar > Eliminar para eliminar la biblioteca. La biblioteca se elimina.

- Si nunca había publicado la biblioteca, se abre un mensaje que le pregunta si desea eliminar la biblioteca o conservarla. Pulse en Eliminar para continuar o Conservar si desea conservarla.

*Nota:* siempre debe quedar una biblioteca.

## Administración de bibliotecas públicas

Para poder reutilizar las bibliotecas públicas, puede publicarlas y luego trabajar con ellas y verlas en el cuadro de diálogo Administrar bibliotecas (Recursos > Administrar bibliotecas). [Si desea obtener más información, consulte el tema Compartimiento de bibliotecas el p. 210.](#) Algunas de las tareas básicas de administración de bibliotecas públicas que puede realizar son importar, exportar o eliminar una biblioteca pública. No puede cambiar el nombre de una biblioteca pública.

Figura 9-4

Cuadro de diálogo Administrar bibliotecas

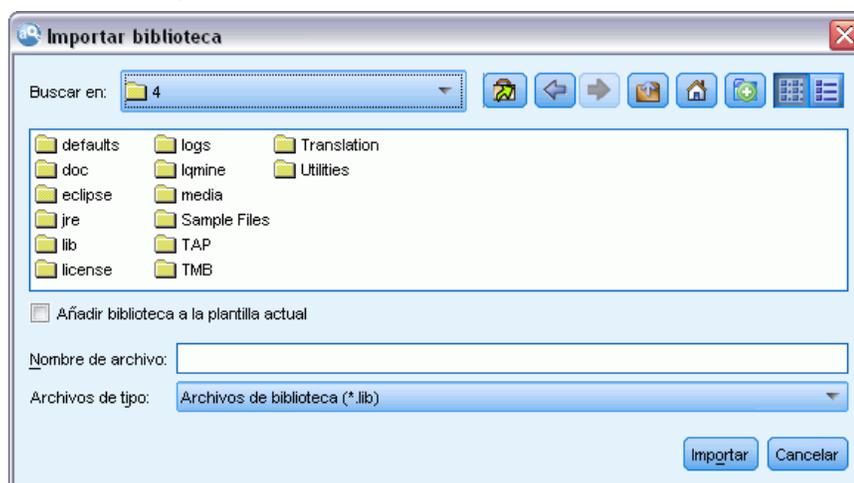


### Importación de bibliotecas públicas

- En el cuadro de diálogo Administrar bibliotecas, pulse en Importar.... Aparecerá el cuadro de diálogo Importar biblioteca.

Figura 9-5

Cuadro de diálogo Importar biblioteca



- ▶ Seleccione el archivo de biblioteca (\*.lib) que desea importar, y si también desea añadirla localmente seleccione Añadir biblioteca al proyecto actual.
- ▶ Pulse en Importar. El cuadro de diálogo se cierra. Si ya existe una biblioteca pública con el mismo nombre, se le pedirá que cambie el nombre de la biblioteca que está importando o que sobrescriba la biblioteca pública actual.

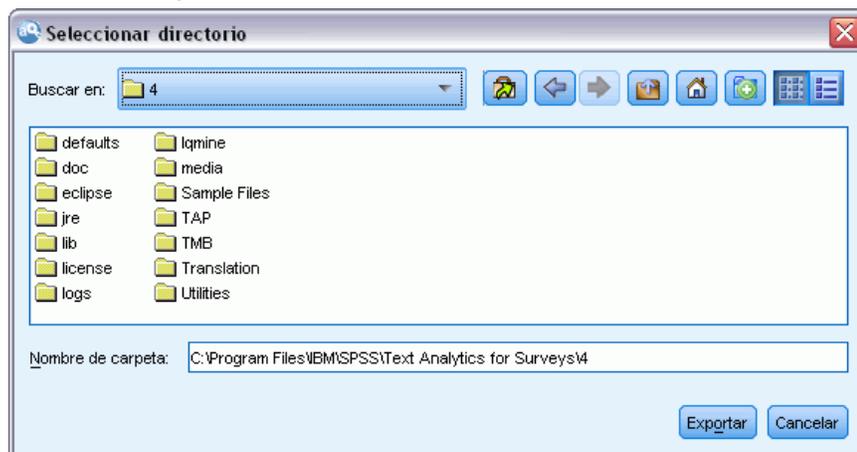
### **Exportación de bibliotecas públicas**

Puede exportar bibliotecas públicas en formato .lib para poder compartirlas.

- ▶ En el cuadro de diálogo Administrar bibliotecas, seleccione la biblioteca que desea exportar a la lista.
- ▶ Pulse en Exportar. Aparecerá el cuadro de diálogo Seleccionar directorio.

Figura 9-6

Cuadro de diálogo Seleccionar directorio



- ▶ Seleccione el directorio al que desea exportar y pulse en Exportar. El cuadro de diálogo se cierra y el archivo de biblioteca (\*.lib) se exporta.

### **Eliminación de bibliotecas públicas**

Puede suprimir una biblioteca local sin eliminar la versión pública de la misma y viceversa. Sin embargo, si se elimina la biblioteca de este cuadro de diálogo, ya no podrá añadirla a ningún proyecto hasta que vuelva a publicar una versión local.

Si elimina una biblioteca que estaba instalada en el producto, se restaurará la versión originalmente instalada.

- ▶ En el cuadro de diálogo Administrar bibliotecas, seleccione la biblioteca que desea eliminar. Puede ordenar la lista pulsando en la cabecera apropiada.
- ▶ Pulse en Eliminar para eliminar la biblioteca. IBM® SPSS® Text Analytics for Surveys verifica si la versión local de la biblioteca coincide con la biblioteca pública. En caso afirmativo, la biblioteca

se elimina sin emitir ningún mensaje de alerta. Si las versiones de las bibliotecas difieren, se abre un mensaje de alerta para preguntarle si desea conservar o eliminar la versión pública.

## Compartimiento de bibliotecas

Las bibliotecas le permiten trabajar con recursos que le facilitan poder compartirlas entre varios proyectos. Las bibliotecas pueden existir en dos estados o versiones posibles. Las bibliotecas que están asociadas con un proyecto determinado se denominan **bibliotecas locales**. Cuando trabaja en un proyecto, puede realizar muchos cambios en la biblioteca *Vegetales*, por ejemplo. Si los cambios realizados fueron útiles en otros datos, puede hacer que estos recursos estén disponibles creando una versión de **biblioteca pública** de la biblioteca *Vegetales*. Una biblioteca pública, como su nombre indica, está disponible para cualquier otro proyecto.

Las bibliotecas públicas pueden verse en el cuadro de diálogo Administrar bibliotecas. Cuando haya creado esta versión de biblioteca pública, podrá añadirla a los recursos de otros contextos para poder compartir estos recursos lingüísticos personalizados.

Las bibliotecas enviadas inicialmente son bibliotecas públicas. Se pueden editar los recursos de estas bibliotecas y luego crear una versión pública nueva. Estas nuevas versiones podrían accederse en otros proyectos nuevos.

A medida que trabaje con las bibliotecas y realice cambios, las versiones de la biblioteca quedarán desincronizadas. En algunos casos, una versión local puede ser más reciente que una versión pública, y en otros casos, la versión pública puede ser más reciente que la versión local. También es posible que tanto la versión pública como la local contengan cambios que la otra biblioteca no contenga si la versión pública se actualizó desde otro proyecto. Si las versiones de las bibliotecas se desincronizan, puede sincronizarlas de nuevo. La sincronización de las versiones de las bibliotecas consiste en volver a publicar y/o actualizar las bibliotecas locales.

Cuando abra o cierre un proyecto, se le pedirá que sincronice todas las bibliotecas que deban actualizarse o volver a publicarse. También puede identificar fácilmente el estado de sincronización de la biblioteca local mediante el icono que aparece junto al nombre de la biblioteca en la vista de árbol, o bien en el cuadro de diálogo Propiedades de biblioteca. También puede optar por hacerlo en cualquier momento mediante las selecciones de menú apropiadas. En la tabla siguiente se describen los cinco estados posibles y sus iconos asociados.

Tabla 9-1  
Estados de sincronización de la biblioteca local

Icono	Descripción del estado de la biblioteca local
	Sin publicar—la biblioteca local no se ha publicado nunca.
	Sincronizada—Las versiones de la biblioteca local y pública son idénticas. También se aplica a la <i>Biblioteca local</i> , que no puede publicarse porque está diseñada para que contenga únicamente recursos específicos de o proyecto.
	Desactualizada—La versión de la biblioteca pública es más reciente que la versión local. Puede actualizar la versión local con los cambios.
	Reciente—La versión de la biblioteca local es más reciente que la versión pública. Puede volver a publicar la versión local a la versión pública.
	Desincronizada—Tanto la biblioteca local como la pública contienen cambios que la otra no contiene. Debe decidir si actualizar o publicar la biblioteca local. Si decide actualizar, perderá los cambios realizados hasta la última vez que la actualizó o publicó. Si decide publicar, se sobrescribirán los cambios de la versión pública.

*Nota:* si siempre actualiza las bibliotecas cuando abre un proyecto o las publica cuando cierra un proyecto, será menos probable que tenga bibliotecas desincronizadas.

Puede volver a publicar una biblioteca en cualquier momento que considere que los cambios realizados en la biblioteca pueden beneficiar a los proyectos que también pueden contener esta biblioteca. A continuación, si los cambios pueden beneficiar a otros proyectos, puede actualizar las versiones locales de dichos proyectos. De esta forma, puede crear proyectos para cada contexto o dominio que se aplique a los datos mediante la creación de nuevas bibliotecas y/o la adición de un número indefinido de bibliotecas públicas a los recursos.

Si se comparte una versión pública de una biblioteca, habrá mayores probabilidades de que las diferencias entre las versiones local y pública aumenten. Cuando abra o publique al cerrar un proyecto, aparecerá un mensaje que le permitirá publicar y/o actualizar las bibliotecas cuyas versiones no estén sincronizadas con las del cuadro de diálogo Administrar bibliotecas. Si la versión de la biblioteca pública es más reciente que la versión local, un cuadro de diálogo le preguntará si desea actualizar las que están abiertas. Puede optar por conservar la versión local tal como está en lugar de actualizarla con la versión pública, o bien fusionar las actualizaciones en la biblioteca local.

## ***Publicación de bibliotecas***

Si nunca ha publicado una biblioteca determinada, la publicación implica crear una copia pública de la biblioteca local en la base de datos. Si vuelve a publicar una biblioteca, los contenidos de la biblioteca local reemplazarán los contenidos de la versión pública existente. Después de volver a publicar, puede actualizar esta biblioteca en cualquier otro proyecto para que sus versiones locales estén sincronizadas con la versión pública. Aunque pueda publicar una biblioteca, siempre se guardará una versión local en el proyecto.

**Importante:** Si realiza cambios en la biblioteca local y, mientras tanto, la versión pública de la biblioteca también se cambia, se considerará que su biblioteca está desincronizada. Se recomienda que empiece actualizando la versión local con los cambios públicos, realizar los cambios que desee y después publicar la versión local otra vez para que las dos versiones sean idénticas. Si primero realiza los cambios y publica, se sobrescribirán los cambios de la versión pública.

Figura 9-7  
Cuadro de diálogo Publicar bibliotecas



### **Para publicar bibliotecas locales en la base de datos**

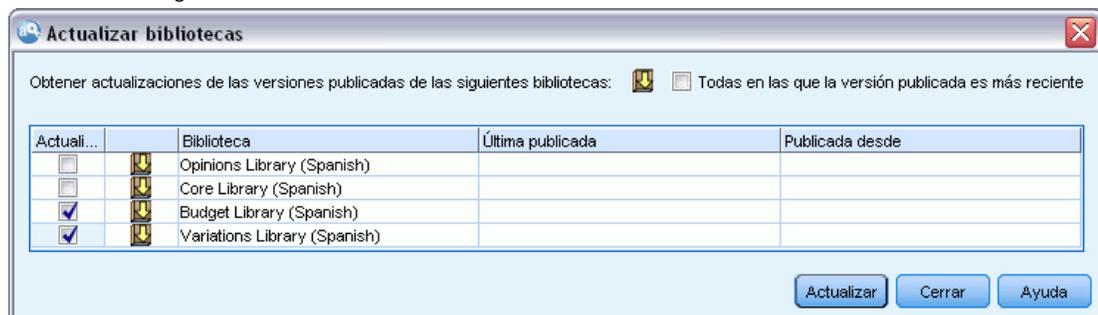
- ▶ En los menús elija Recursos > Publicar bibliotecas. Se abre el cuadro de diálogo Publicar bibliotecas con todas las bibliotecas que necesitan publicarse seleccionadas por defecto.
- ▶ Marque la casilla de verificación de la izquierda de cada biblioteca que desee publicar o volver a publicar.
- ▶ Pulse en Publicar para publicar las bibliotecas en la base de datos Administrar bibliotecas.

### **Actualización de bibliotecas**

Cuando abre o publica al cerrar un proyecto, puede actualizar o publicar las bibliotecas que ya no estén sincronizadas con las versiones públicas. Si la versión de la biblioteca pública es más reciente que la versión local, un cuadro de diálogo le preguntará si desea actualizar las que están abiertas. Puede optar entre conservar la versión local en lugar de actualizarla con la versión pública, y reemplazar la versión local con la de la pública. Si la versión pública de una biblioteca es más reciente que la versión local, puede actualizar la versión local para sincronizar su contenido con el de la versión pública. Actualizar significa incorporar en la versión local los cambios encontrados en la versión pública.

*Nota:* si siempre actualiza las bibliotecas cuando abre un proyecto o las publica cuando cierra un proyecto, será menos probable que tenga bibliotecas desincronizadas. [Si desea obtener más información, consulte el tema Compartimiento de bibliotecas el p. 210.](#)

Figura 9-8  
Cuadro de diálogo Actualizar bibliotecas



### **Para actualizar bibliotecas locales**

- ▶ En los menús elija Recursos > Actualizar bibliotecas. Se abre el cuadro de diálogo Actualizar bibliotecas con todas las bibliotecas que necesitan actualizarse seleccionadas por defecto.
- ▶ Marque la casilla de verificación de la izquierda de cada biblioteca que desee publicar o volver a publicar.
- ▶ Pulse en Actualizar para actualizar las bibliotecas locales.

## Resolución de conflictos

### Conflictos de la biblioteca local frente a la pública

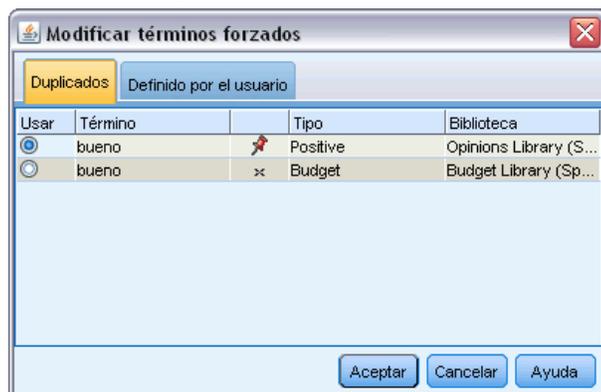
Cuando abre un proyecto, IBM® SPSS® Text Analytics for Surveys realiza una comparación de las bibliotecas locales y de las que figuran en la lista del cuadro de diálogo Administrar bibliotecas. Si alguna de las bibliotecas locales del proyecto no están sincronizadas con las versiones publicadas, se abre el cuadro de diálogo Advertencia de sincronización de biblioteca. Puede elegir entre las opciones siguientes para seleccionar las versiones de biblioteca que desea utilizar aquí:

- **Todas las bibliotecas locales del archivo.** Esta opción conserva todas las bibliotecas locales tal como están. Siempre puede volver a publicarlas o actualizarlas en otro momento.
- **Todas las bibliotecas publicadas de este equipo.** Esta opción sustituye las bibliotecas locales que se muestran por las versiones que se encuentran en la base de datos.
- **Todas las bibliotecas más recientes.** Esta opción sustituye las bibliotecas locales más antiguas por las versiones públicas más recientes de la base de datos.
- **Otro.** Esta opción permite seleccionar manualmente las versiones que desee eligiéndolas a partir de la tabla.

### Conflictos en los términos forzados

Cuando añade una biblioteca pública o actualiza una biblioteca local, pueden quedar al descubierto conflictos y entradas duplicadas entre los términos y los tipos de esta biblioteca y los términos y los tipos de las otras bibliotecas de sus recursos. Si esto ocurre, se le pedirá que verifique las resoluciones de conflictos que se proponen o que las cambie antes de finalizar la operación en el cuadro de diálogo Editar términos forzados. [Si desea obtener más información, consulte el tema Forzado de términos en el capítulo 10 el p. 222.](#)

Figura 9-9  
Cuadro de diálogo Editar términos forzados



El cuadro de diálogo Editar términos forzados contiene las parejas de términos o tipos conflictivos. Se utilizan colores de fondo alternos para distinguir visualmente cada pareja conflictiva. Estos colores pueden cambiarse en el cuadro de diálogo Opciones. [Si desea obtener más información,](#)

consulte el tema [Opciones: Pestaña Mostrar en el capítulo 2 el p. 19](#). El cuadro de diálogo Editar términos forzados contiene dos pestañas:

- **Duplicados.** Esta pestaña contiene los términos duplicados que se encuentran en las bibliotecas. Si aparece un icono de chincheta detrás de cada término, significa que esta aparición del término está forzada. Si aparece un icono de X de color negro, significa que esta aparición del término se pasará por alto durante la extracción porque está forzado en algún otro sitio.
- **Definido por el usuario.** Esta pestaña contiene una lista de los términos que se han forzado manualmente en este panel de términos del diccionario de tipo y no a través de los conflictos.

*Nota:* el cuadro de diálogo Editar términos forzados se abrirá después de que añada o actualice una biblioteca. Si cancela este cuadro de diálogo, no significa que cancele la actualización o la adición de la biblioteca.

#### ***Para resolver conflictos***

- ▶ En el cuadro de diálogo Editar términos forzados, seleccione el botón de radio de la columna Usar correspondiente al término que desea forzar.
- ▶ Cuando haya terminado, pulse en Aceptar para aplicar los términos forzados y cerrar el cuadro de diálogo. Si pulsa en Cancelar, se cancelarán los cambios realizados en el cuadro de diálogo.

## ***Acerca de los diccionarios de biblioteca***

Los recursos utilizados para extraer datos de texto se guardan en forma de plantillas y bibliotecas. Cada biblioteca se compone de tres diccionarios.

- El **diccionario de tipo** contiene una recopilación de palabras que se agrupan bajo una misma etiqueta, tipo o nombre. Cuando el motor de extracción lee los datos de texto, compara las palabras que encuentra en el texto con los términos definidos en los diccionarios de tipo. Durante la extracción, las formas declinadas de los sinónimos y los términos de tipo se agrupan bajo un término objetivo llamado concepto. Los conceptos extraídos se asignan al diccionario de tipo donde aparecen como términos. Puede administrar los diccionarios de tipo en los paneles central y superior izquierdo del editor—el árbol de bibliotecas y el panel de términos. [Si desea obtener más información, consulte el tema Diccionarios de tipo el p. 215.](#)
- El **diccionario de sustitución** contiene una recopilación de términos definidos como sinónimos o como elementos opcionales que se utilizan para agrupar términos similares bajo un término objetivo, llamado concepto en los resultados de extracción finales. Puede administrar los diccionarios de sustitución en el panel inferior izquierdo del editor utilizando la pestaña Sinónimos y la pestaña Opcional. [Si desea obtener más información, consulte el tema Diccionarios de sustitución/sinónimos el p. 225.](#)
- El **diccionario de exclusión** contiene una recopilación de términos y tipos que se eliminarán de los resultados de la extracción final. Puede administrar los diccionarios de exclusión en el panel del extremo derecho del editor. [Si desea obtener más información, consulte el tema Diccionarios de exclusión el p. 231.](#)

[Si desea obtener más información, consulte el tema Trabajo con bibliotecas en el capítulo 9 el p. 202.](#)

### ***Diccionarios de tipo***

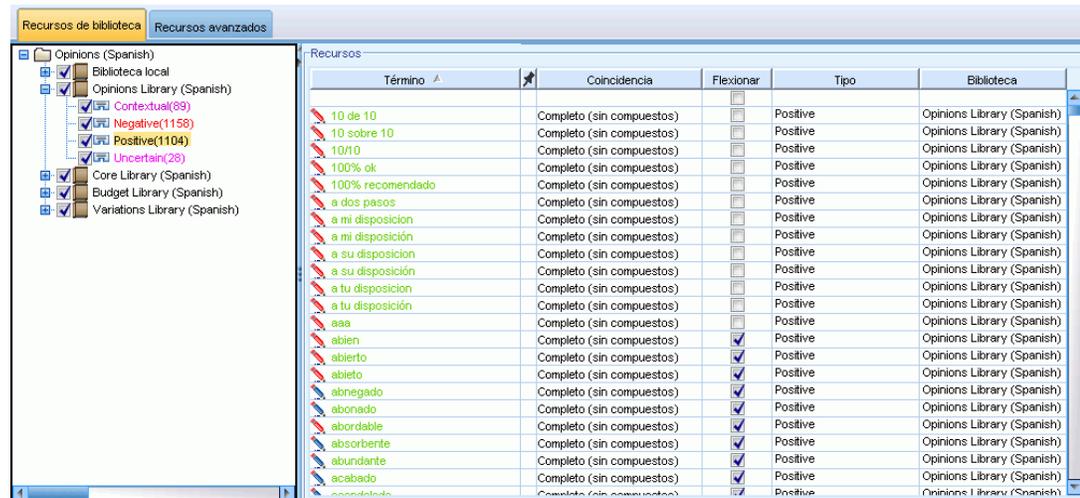
Un **diccionario de tipo** está compuesto por un nombre o etiqueta de tipo, y una lista de términos. Los diccionarios de tipo se administran en los paneles central y superior izquierdo de la pestaña Recursos de la biblioteca del editor. Puede acceder a esta vista en Ver > Editor de recursos en los menús.

Cuando el motor de extracción lee los datos de texto, compara las palabras que encuentra en el texto con los términos definidos en los diccionarios de tipo. Los términos son palabras o frases de los diccionarios de tipo de los recursos lingüísticos.

Cuando una palabra coincide con un término, se le asigna al nombre de tipo correspondiente a dicho término. Cuando los recursos se leen durante la extracción, los términos que se encontraron en el texto pasan una serie de procedimientos antes de que se conviertan en conceptos en el panel Resultados extraídos. Si el motor de extracción considera que los diferentes términos que pertenecen a un mismo diccionario son sinónimos, se agruparán bajo el término que aparezca con más frecuencia y se denominará como un **concepto** en el panel Resultados extraídos. Por

ejemplo, cuando los términos pregunta y consulta aparecen bajo el nombre de concepto pregunta al final.

Figura 10-1  
Árbol de bibliotecas y panel de términos



La lista de diccionarios de tipo se muestra en el panel del árbol de bibliotecas a la izquierda. El contenido de cada diccionario de tipo aparece en el panel central. Los diccionarios de tipo constituyen algo más que una simple lista de términos. La manera en que las palabras y las frases de los datos de texto se hacen coincidir con los términos definidos en los diccionarios de tipo está determinada por la opción de coincidencia que se haya definido. Una **opción de coincidencia** especifica de qué manera el tema se ancla a una palabra o frase candidata en los datos de texto. [Si desea obtener más información, consulte el tema Adición de términos el p. 219.](#)

También puede ampliar los términos del diccionario de tipo especificando si desea que se generen y se añadan automáticamente al diccionario formas declinadas de los términos. Al generar formas declinadas, automáticamente se añaden los plurales de los términos en singular, y los singulares de los términos en plural, y los adjetivos en el diccionario de tipo. [Si desea obtener más información, consulte el tema Adición de términos el p. 219.](#)

*Nota:* conceptos que no se encuentran en ningún diccionario de tipo pero que se extraen del texto se tipifican automáticamente como <Unknown>.

## Tipos incorporados

IBM® SPSS® Text Analytics for Surveys se entrega con un conjunto de recursos lingüísticos en forma de bibliotecas enviadas y recursos compilados. Las bibliotecas incluidas contienen un conjunto de diccionarios de tipo integrados que incluyen <Location>, <Organization>, <Person> y <Product>.

Estos diccionarios de tipo los utiliza el motor de extracción para asignar tipos a los conceptos que extrae, como la asignación del tipo <Location> al concepto parís. Aunque en los diccionarios de tipo incorporados hay un gran número de términos ya definidos, no abarcan todas

las posibilidades. Por lo tanto, puede añadir términos en los diccionarios o crear los suyos propios. Para obtener una descripción de los contenidos de un determinado diccionario de tipo enviado, lea la anotación del cuadro de diálogo Propiedades de tipo. Seleccione el tipo en el árbol y elija Editar > Propiedades en el menú contextual.

*Nota:* Además de las bibliotecas enviadas, los recursos compartidos (que también utiliza el motor de extracción) contienen un gran número de definiciones complementarias a los diccionarios de tipo incorporados, pero su contenido no está visible en el producto. Sin embargo, puede forzar que un término especificado en los diccionarios compilados se coloque en otro diccionario. [Si desea obtener más información, consulte el tema Forzado de términos el p. 222.](#)

## **Creación de tipos**

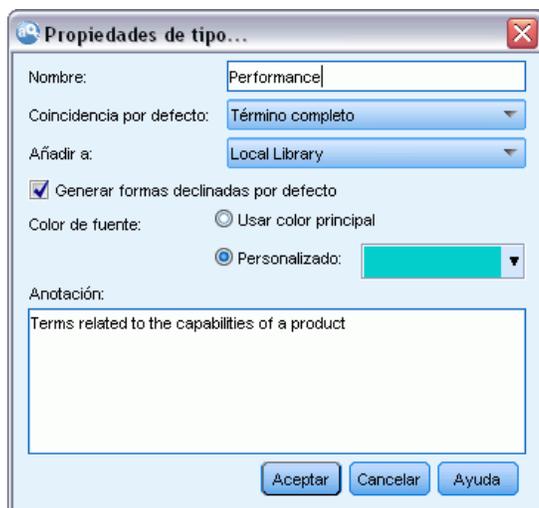
Puede crear diccionarios de tipo para facilitar la agrupación de los términos similares. Cuando durante el proceso de extracción se descubren los términos que aparecen en este diccionario, se asignarán a este nombre de tipo y se extraerán bajo un nombre de concepto. Siempre que se crea una biblioteca, se incluye una biblioteca de tipos vacía para que pueda empezar a introducir términos inmediatamente.

Si está analizando texto sobre alimentos y desea agrupar términos bajo la categoría de verduras, puede crear su propio diccionario de tipo de <Verduras>. Luego puede añadir términos como zanahoria, bróquil y espinacas si considera que son términos importantes que aparecerán en el texto. A continuación, y durante la extracción, si se encuentra alguno de esos términos, se extraen como conceptos y se asignarán al tipo <Verduras>.

No es necesario que defina cada forma de una palabra o expresión, porque puede optar por generar las formas declinadas de los términos. Al elegir esta opción, el motor de extracción reconoce automáticamente las formas singular y plural de los términos entre el resto de las formas como pertenecientes a este tipo. Esta opción resulta particularmente útil cuando el tipo contiene mayoritariamente sustantivos, puesto que es improbable que desee incluir las formas declinadas de verbos o adjetivos.

**Importante:** Se recomienda encarecidamente que no cree nuevos tipos en la biblioteca Opinions, ya que no se tomarán en cuenta durante el proceso. Los contenidos de la biblioteca Opinions se maneja de manera distinta que el resto de las bibliotecas, puesto que sirve para generar patrones. En su lugar, trabaje en los tipos que ya existen en la biblioteca o añada tipos nuevos a otra biblioteca del proyecto.

Figura 10-2  
Cuadro de diálogo *Propiedades de tipo*



**Nombre.** El nombre que asigna al diccionario de tipo que está creando. Se recomienda que no utilice espacios en blanco en los nombres de tipo, sobre todo si hay dos o más nombres de tipo que empiecen con la misma palabra.

**Coincidencia por defecto.** El atributo de coincidencia por defecto indica al motor de extracción cómo debe hacer coincidir este término con los datos de texto. Siempre que se añade un término a este diccionario de tipo, este es el atributo de coincidencia que se asigna automáticamente. Siempre puede cambiar la opción de coincidencia manualmente en la lista de términos. Entre las opciones disponibles se encuentran: Término completo, Inicio, Fin, Cualquiera, Inicio o fin, Completo e inicio, Completo y fin, Completo y (inicio o fin) y Completo (sin compuestos). [Si desea obtener más información, consulte el tema Adición de términos el p. 219.](#)

**Añadir a.** Este campo indica la biblioteca donde creará el nuevo diccionario de tipo.

**Generar formas declinadas por defecto.** Esta opción indica al motor de extracción que utilice la morfología gramatical para capturar y agrupar formas similares de los términos que añade a este diccionario, como las formas singular y plural del término. Esta opción resulta particularmente útil cuando el tipo contiene mayormente sustantivos. Cuando selecciona esta opción, todos los términos nuevos que se añaden a este tipo adoptarán automáticamente esta opción, aunque puede cambiarlo manualmente en la lista.

**Color de fuente.** Este campo permite distinguir los resultados de este tipo del resto de resultados de la interfaz. Si selecciona Usar color principal, también se utilizará el color de tipo por defecto para este diccionario de tipo. Este color por defecto se establece en el cuadro de diálogo Opciones. [Si desea obtener más información, consulte el tema Opciones: Pestaña Mostrar en el capítulo 2 el p. 19.](#) Si selecciona Personalizado, seleccione un color de la lista desplegable.

**Anotación.** Este campo es opcional y puede utilizarse para introducir comentarios o descripciones.

#### ***Para crear un diccionario de tipo***

- ▶ Seleccione la biblioteca donde desea crear un diccionario de tipo nuevo.

- ▶ En los menús elija Herramientas > Nuevo tipo. Aparecerá el cuadro de diálogo Propiedades de tipo.
- ▶ Escriba el nombre del diccionario de tipo en el cuadro de texto Nombre y elija las opciones que desee.
- ▶ Pulse en Aceptar para crear el diccionario de tipo. El nuevo tipo será visible en el panel del árbol de bibliotecas y aparecerá en el panel central. Puede empezar a añadir términos inmediatamente. Si desea obtener más información, consulte [Adición de términos](#).

*Nota:* estas instrucciones indican cómo realizar cambios en la Editor de recursos vista. Tenga en cuenta que este tipo de ajuste puede hacerlo directamente en el panel Resultados extraídos o en el panel Datos. [Si desea obtener más información, consulte el tema Refinamiento de los resultados de la extracción en el capítulo 5 el p. 86.](#)

## Adición de términos

El panel del árbol de bibliotecas muestra las bibliotecas y puede expandirse para mostrar también los diccionarios de tipo que contienen. En el panel central aparece una lista de los términos de la biblioteca o del diccionario de tipo seleccionados, según cuál sea la selección en el árbol.

Figura 10-3  
Panel de términos

Término	Coincidencia	Flexio...	Tipo	Biblioteca
malencarado	Completo (sin compuestos)	<input checked="" type="checkbox"/>	Negative	Opinions Library (Spanish)
malevolo	Completo (sin compuestos)	<input checked="" type="checkbox"/>	Negative	Opinions Library (Spanish)
malhablado	Completo (sin compuestos)	<input checked="" type="checkbox"/>	Negative	Opinions Library (Spanish)
malhecho	Completo (sin compuestos)	<input checked="" type="checkbox"/>	Negative	Opinions Library (Spanish)
malicioso	Completo (sin compuestos)	<input checked="" type="checkbox"/>	Negative	Opinions Library (Spanish)
maligno	Completo (sin compuestos)	<input checked="" type="checkbox"/>	Negative	Opinions Library (Spanish)
malintencionado	Completo (sin compuestos)	<input checked="" type="checkbox"/>	Negative	Opinions Library (Spanish)
malisimo	Completo (sin compuestos)	<input checked="" type="checkbox"/>	Negative	Opinions Library (Spanish)
malmirado	Completo (sin compuestos)	<input checked="" type="checkbox"/>	Negative	Opinions Library (Spanish)
malo	Completo (sin compuestos)	<input checked="" type="checkbox"/>	Negative	Opinions Library (Spanish)
maloliente	Completo (sin compuestos)	<input checked="" type="checkbox"/>	Negative	Opinions Library (Spanish)
malsano	Completo (sin compuestos)	<input checked="" type="checkbox"/>	Negative	Opinions Library (Spanish)
maisonante	Completo (sin compuestos)	<input checked="" type="checkbox"/>	Negative	Opinions Library (Spanish)
malufrido	Completo (sin compuestos)	<input checked="" type="checkbox"/>	Negative	Opinions Library (Spanish)
maltratado	Completo (sin compuestos)	<input checked="" type="checkbox"/>	Negative	Opinions Library (Spanish)
maltrato	Completo (sin compuestos)	<input checked="" type="checkbox"/>	Negative	Opinions Library (Spanish)
maltrecho	Completo (sin compuestos)	<input checked="" type="checkbox"/>	Negative	Opinions Library (Spanish)
maléfico	Completo (sin compuestos)	<input checked="" type="checkbox"/>	Negative	Opinions Library (Spanish)
malévolo	Completo (sin compuestos)	<input checked="" type="checkbox"/>	Negative	Opinions Library (Spanish)
malisimo	Completo (sin compuestos)	<input checked="" type="checkbox"/>	Negative	Opinions Library (Spanish)
mandón	Completo (sin compuestos)	<input checked="" type="checkbox"/>	Negative	Opinions Library (Spanish)
marginado	Completo (sin compuestos)	<input checked="" type="checkbox"/>	Negative	Opinions Library (Spanish)
marrano	Completo (sin compuestos)	<input checked="" type="checkbox"/>	Negative	Opinions Library (Spanish)
me dificulta	Completo (sin compuestos)	<input type="checkbox"/>	Negative	Opinions Library (Spanish)

En el Editor de recursos, puede añadir términos a un diccionario de tipo directamente en el panel de términos o a través del cuadro de diálogo Añadir nuevos términos. Los términos que añada pueden ser palabras simples o compuestas. En la parte superior de la lista siempre encontrará una fila en blanco para que pueda añadir un término nuevo.

*Nota:* estas instrucciones indican cómo realizar cambios en la Editor de recursos vista. Tenga en cuenta que este tipo de ajuste puede hacerlo directamente en el panel Resultados extraídos o en el panel Datos. [Si desea obtener más información, consulte el tema Refinamiento de los resultados de la extracción en el capítulo 5 el p. 86.](#)

### **Columna de términos**

En esta columna, introduzca palabras simples o compuestas en la casilla. El color en el que aparece el término depende del color del tipo donde se ha almacenado o forzado el término. Puede cambiar los colores de tipo en el cuadro de diálogo Propiedades de tipo. [Si desea obtener más información, consulte el tema Creación de tipos el p. 217.](#)

### **Columna de Forzar**

En esta columna, al colocar un icono de chincheta en esta casilla, está indicando al motor de extracción que pase por alto el resto de las apariciones de este mismo término en otras bibliotecas. [Si desea obtener más información, consulte el tema Forzado de términos el p. 222.](#)

### **Columna de Emparejar**

En esta columna, seleccione una opción de coincidencia para indicar al motor de extracción cómo debe hacer coincidir este término con los datos de texto. Consulte la tabla para ver los ejemplos pertinentes. Puede cambiar el valor por defecto editando las propiedades de tipo. [Si desea obtener más información, consulte el tema Creación de tipos el p. 217.](#) En los menús elija Editar > Cambiar coincidencia. A continuación figuran las opciones básicas de coincidencia, puesto que también pueden combinarse:

- Inicio. Si el término que se encuentra en el diccionario coincide con el principio de un concepto extraído del texto, se asigna este tipo. Por ejemplo, si especifica *tarta*, coincidirá con *tarta de manzana*.
- Fin. Si el término que se encuentra en el diccionario coincide con el final de un concepto extraído del texto, se asigna este tipo. Por ejemplo, si especifica *manzana*, coincidirá con *tarta de manzana*.
- Cualquiera. Si el término que se encuentra en el diccionario coincide con cualquier parte de un concepto extraído del texto, se asigna este tipo. Por ejemplo, si especifica *manzana*, la opción Cualquiera tipificará *tarta de manzana*, *manzana reineta* y *tarta de manzanas reinetas* de la misma forma.
- Término completo. Si el concepto completo extraído del texto coincide con el término exacto del diccionario, se asignará este tipo. Si se añade un término como Término completo, Completo e inicio, Completo y fin, Completo y todo o Completo (sin compuestos) se forzarán la extracción de un término.

Además, puesto que el tipo <Person> solo extrae nombres con dos partes, como *edith piaf* o *mohandas gandhi*, puede añadir explícitamente los nombres propios en este diccionario de tipo si intenta extraer un nombre propio sin que se mencione ningún apellido. Por ejemplo, si

desea captar todas las instancias de *edith* como nombre propio, debe añadir *edith* al tipo <Person> utilizando Término completo o Completo e inicio.

- **Completo (sin compuestos).** Si el concepto completo extraído del texto coincide con el término exacto en el diccionario, se asignará este tipo y la extracción se detendrá para impedir que la extracción haga coincidir el término con un compuesto más largo. Por ejemplo, si especifica *manzana*, la opción Completo (sin compuestos) tipificará *manzana* pero no el compuesto *zumo de manzana* a menos que se fuerce en otro lugar.

En la tabla siguiente, supongamos que el término *In manzana* está en el diccionario de tipo. En función de la opción de coincidencia, en esta tabla se muestran los conceptos que se extraerían y tipificarían si se encontraran en el texto.

Tabla 10-1  
Ejemplos de coincidencias

Opciones de coincidencia para el término:  mar	Conceptos extraídos			
	mar	tarta de manzana	madura manzana	tarta de manzana casera
Término completo	✓			
Inicio		✓		
Fin			✓	
Inicio o fin		✓	✓	
Completo e inicio	✓	✓		
Completo y fin	✓		✓	
Completo y (inicio o fin)	✓	✓	✓	
Cualquiera		✓	✓	✓
Completo y todo	✓	✓	✓	✓
Completo (sin compuestos)	✓	<i>no se extrae nunca</i>	<i>no se extrae nunca</i>	<i>no se extrae nunca</i>

### **Columna de Flexionar**

En esta columna, seleccione si el motor de extracción debe generar formas declinadas de este término durante la extracción, de manera que se agrupen juntas. El valor por defecto de esta columna está definido en Propiedades de tipo, pero puede cambiar esta opción en cada caso individual directamente en la columna. En los menús elija Editar > Cambiar flexión.

### **Columna de tipos**

En esta columna, seleccione un diccionario de tipo de la lista desplegable. La lista de tipos se filtra según la opción que haya seleccionado en el panel del árbol de bibliotecas. El primer tipo de la lista siempre es el tipo por defecto seleccionado en el panel del árbol de bibliotecas. En los menús elija Editar > Cambiar tipo.

### **Columna de bibliotecas**

En esta columna, se muestra la biblioteca en la que está almacenado el término. Puede arrastrar y soltar un término en otro tipo en el panel del árbol de bibliotecas para cambiarlo de biblioteca.

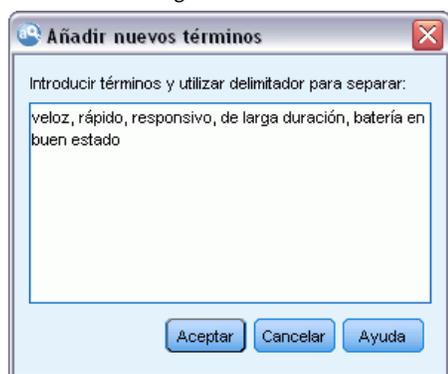
### **Para añadir un único término a un diccionario de tipo**

- ▶ En el panel del árbol de bibliotecas, seleccione el diccionario de tipo al que desea añadir el término.
- ▶ En la lista de términos del panel central, escriba el término en la primera casilla disponible y defina las opciones que desee para dicho término.

### **Para añadir varios términos a un diccionario de tipo**

- ▶ En el panel del árbol de bibliotecas, seleccione el diccionario de tipo al que desea añadir los términos.
- ▶ En los menús elija Herramientas > Nuevos términos. Aparecerá el cuadro de diálogo Añadir nuevos términos.

Figura 10-4  
Cuadro de diálogo Añadir nuevos términos



- ▶ Especifique los términos que desee añadir al diccionario de tipo seleccionado escribiéndolos o copiando y pegando el conjunto de términos. Si especifica varios términos, deberá separarlos mediante el delimitador definido en el cuadro de diálogo Opciones, o añadir cada término en una línea nueva. [Si desea obtener más información, consulte el tema Opciones de configuración en el capítulo 2 el p. 17.](#)
- ▶ Pulse en Aceptar para añadir los términos al diccionario. La opción de coincidencia se establece automáticamente en la opción por defecto para esta biblioteca de tipo. El cuadro de diálogo se cierra y los nuevos términos aparecen en el diccionario.

### **Forzado de términos**

Si desea que un término se asigne a un tipo determinado, puede añadirlo al diccionario de tipo correspondiente. Sin embargo, si hay varios términos con el mismo nombre, el motor de extracción debe conocer qué tipo se va a utilizar. Por lo tanto, se le indicará que seleccione el tipo que desea utilizar. A esta acción se le llama **forzar** un término dentro de un tipo. Esta opción

resulta especialmente útil cuando se sustituye la asignación de tipo de un diccionario compilado (interno y no editable). En general, se recomienda evitar los términos duplicados.

La acción de forzar no *eliminará* el resto de las apariciones de este término, sino que el motor de extracción las pasará por alto. Más adelante podrá cambiar qué aparición debe utilizarse mediante el forzado o la anulación del forzado de un término. Es posible que también deba forzar un término dentro de un diccionario de tipo cuando añada o actualice una biblioteca pública.

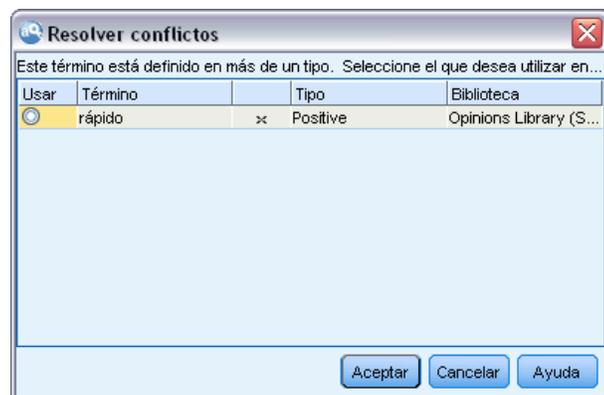
Figura 10-5  
Iconos de estado de Forzar

Término	Forzar	Coincidencia	Flexionar	Tipo	Biblioteca
rrecomienda	📌	Completo (sin compuestos)	<input checked="" type="checkbox"/>	Positive	Opinions Library (Spanish)
rápido	📌	Completo e inicio	<input checked="" type="checkbox"/>	Positive	Opinions Library (Spanish)
rápidamente		Completo e inicio	<input type="checkbox"/>	Positive	Opinions Library (Spanish)
rápido	✖	Completo (sin compuestos)	<input checked="" type="checkbox"/>	Positive	Opinions Library (Spanish)
sabroso		Completo (sin compuestos)	<input checked="" type="checkbox"/>	Positive	Opinions Library (Spanish)
sagaz		Completo (sin compuestos)	<input checked="" type="checkbox"/>	Positive	Opinions Library (Spanish)
saludable		Completo (sin compuestos)	<input checked="" type="checkbox"/>	Positive	Opinions Library (Spanish)

Puede ver qué términos están forzados o cuáles se pasan por alto en la columna de Forzar, la segunda columna del panel de términos. Si aparece un icono de chincheta, significa que esta aparición del término está forzada. Si aparece un icono de X de color negro, significa que esta aparición del término se pasará por alto durante la extracción porque está forzado en algún otro sitio. Además, cuando fuerza un término, éste aparece del color del tipo dentro del que se ha forzado. Esto significa que si fuerza un término que se encuentra en Tipo 1 y Tipo 2 dentro de Tipo 1, cada vez que vea este término en la ventana aparecerá con el color de fuente definido para Tipo 1.

Para cambiar el estado pulse dos veces con el ratón en el icono. Si el término aparece en algún otro lugar, se abre el cuadro de diálogo Resolver conflictos para que pueda seleccionar cuál de las apariciones debe utilizarse.

Figura 10-6  
Cuadro de diálogo Resolver conflictos



## Cambio de nombre de los tipos

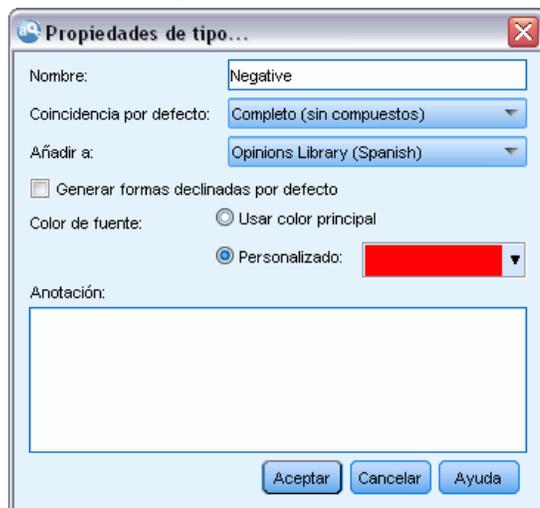
Puede cambiar el nombre de un diccionario de tipo o cambiar otras opciones de diccionario editando las propiedades de tipo.

**Importante:** Se recomienda que no utilice espacios en blanco en los nombres de tipo, sobre todo si hay dos o más nombres de tipo que empiecen con la misma palabra. También se recomienda no cambiar el nombre de los tipos en las bibliotecas Core u Opinions, ni cambiar sus atributos de coincidencia por defecto.

### **Para cambiar el nombre de un tipo**

- ▶ En el panel del árbol de bibliotecas, seleccione el diccionario de tipo al que desea cambiar el nombre.
- ▶ Pulse el botón derecho del ratón y elija Propiedades de tipo en el menú contextual. Aparecerá el cuadro de diálogo Propiedades de tipo.

Figura 10-7  
Cuadro de diálogo Propiedades de tipo



- ▶ Escriba el nombre nuevo del diccionario de tipo en el cuadro de texto Nombre.
- ▶ Pulse en Aceptar para aceptar el nombre nuevo. El nombre de tipo nuevo aparece visible en el panel del árbol de bibliotecas.

## **Cómo mover tipos**

Puede arrastrar un diccionario de tipo y soltarlo en otra ubicación de una biblioteca o en otra biblioteca del árbol.

*Nota:* Se recomienda que no mueva los tipos incorporados.

### **Para cambiar el orden de un tipo en una biblioteca**

- ▶ En el panel del árbol de bibliotecas, seleccione el diccionario de tipo que desea mover.
- ▶ En los menús elija Editar > Subir para subir el diccionario de tipo una posición en el panel del árbol de bibliotecas, o Editar > Bajar para bajarlo una posición.

### **Para mover un tipo a otra biblioteca**

- ▶ En el panel del árbol de bibliotecas, seleccione el diccionario de tipo que desea mover.
- ▶ Pulse el botón derecho del ratón y elija Propiedades de tipo en el menú contextual. Aparecerá el cuadro de diálogo Propiedades de tipo. (También puede arrastrar y soltar el tipo en otra biblioteca).
- ▶ En el cuadro de lista Añadir a, seleccione la biblioteca a la que desea mover el diccionario de tipo.
- ▶ Pulse en Aceptar. El cuadro de diálogo se cierra, y el tipo se encuentra ahora en la biblioteca que ha seleccionado.

## **Desactivación y eliminación de tipos**

Si desea eliminar temporalmente un diccionario de tipo, puede desactivarlo quitándole la marca de verificación de la izquierda del nombre del diccionario en el panel del árbol de bibliotecas. Esto indica que desea conservar el diccionario en la biblioteca, pero que desea que se pasen por alto sus contenidos durante una comprobación de conflictos y durante el proceso de extracción.

También puede eliminar permanentemente los diccionarios de tipo de una biblioteca.

*Nota:* se recomienda que no elimine los tipos incorporados de las bibliotecas Core y Opinions. Se recomienda que los desactive.

### **Para desactivar un diccionario de tipo**

- ▶ En el panel del árbol de bibliotecas, seleccione el diccionario de tipo que desea desactivar.
- ▶ Pulse la barra espaciadora. Se borra la marca de la casilla de verificación a la izquierda del nombre del tipo.

### **Para eliminar un diccionario de tipo**

- ▶ En el panel del árbol de bibliotecas, seleccione el diccionario de tipo que desea eliminar.
- ▶ En los menús elija Editar > Eliminar para eliminar el diccionario de tipo.

## **Diccionarios de sustitución/sinónimos**

Un **diccionario de sustitución** es una recopilación de términos que ayuda a agrupar términos similares bajo un término objetivo. Los diccionarios de sustitución se administran en el panel inferior de la pestaña Recursos de la biblioteca. Puede acceder a esta vista en Ver > Editor de recursos en los menús.

Puede definir dos formas de sustitución en este diccionario: **sinónimos** y **elementos opcionales**. Puede pulsar en las pestañas de este panel para conmutarlas.

Después de ejecutar una extracción en los datos de texto, puede encontrar varios conceptos que son sinónimos o formas declinadas de otros conceptos. Mediante la identificación de elementos opcionales y sinónimos, puede forzar que el motor de extracción asigne esos términos a un único término objetivo.



### Elementos opcionales

Los Elementos opcionales identifican palabras opcionales de un término compuesto que puede pasarse por alto durante la extracción para poder mantener juntos los términos parecidos aunque aparezcan con ligeras diferencias en el texto. Los elementos opcionales son palabras simples que, si se eliminan de un término compuesto, pueden crear una coincidencia con otro término. Estas palabras simples pueden aparecer en cualquier lugar dentro del término compuesto (al principio, en el centro o al final). Puede definir estos elementos opcionales en la pestaña Opcional.

Por ejemplo, para agrupar los términos `ibm e ibm corp` juntos, debe declarar que `corp` se considere un elemento opcional en este caso. En otro ejemplo, si designa el término `acceso` como elemento opcional y durante la extracción se encuentran los términos `velocidad de acceso a internet` y `velocidad de internet`, se agruparán juntos bajo el término que aparezca con más frecuencia.

Figura 10-10  
Diccionario Sustitución, pestaña Opcional

Elementos opcionales	Biblioteca
<input checked="" type="checkbox"/>	Biblioteca local
<input checked="" type="checkbox"/> al, <input checked="" type="checkbox"/> en, <input checked="" type="checkbox"/> via, <input checked="" type="checkbox"/> vía	Opinions Library (Spanish)
<input checked="" type="checkbox"/> s.l.	Core Library (Spanish)
<input checked="" type="checkbox"/>	Budget Library (Spanish)
<input checked="" type="checkbox"/>	Variations Library (Spanish)

Sinónimos **Elementos opcionales**

### Definición de sinónimos

En la pestaña Sinónimos, puede especificar una definición de sinónimo en la línea vacía de la parte superior de la tabla. Empiece definiendo el término objetivo y sus sinónimos. También puede seleccionar la biblioteca en la que desea que se guarde esta definición. Durante la extracción, todas las apariciones de los sinónimos se agruparán bajo el término objetivo de la extracción final. [Si desea obtener más información, consulte el tema Adición de términos el p. 219.](#)

Por ejemplo, si los datos de texto incluyen un gran volumen de información sobre telecomunicaciones, puede que tenga estos términos: `teléfono celular`, `teléfono inalámbrico` y `teléfono móvil`. En este ejemplo, puede definir `celular` y `móvil` como sinónimos de `inalámbrico`. Si define estos sinónimos, cada aparición extraída de `teléfono celular` y de `teléfono móvil` se considerará como el mismo término que `teléfono inalámbrico` y se mostrarán juntos en la lista de términos.

Cuando esté creando sus diccionarios de tipo, puede especificar un término y luego pensar en tres o cuatro sinónimos del mismo. En ese caso, puede escribir todos los términos y luego el término objetivo en el diccionario de sustitución, y a continuación arrastrar los sinónimos.

La sustitución de sinónimos también se aplica a las formas declinadas (como los plurales) del sinónimo. En función del contexto puede imponer límites en la forma en que se sustituyen los términos. Pueden utilizarse determinados caracteres para aplicar límites sobre el proceso de la sinonimia:

- **Signo de exclamación (!).** Si hay un signo de exclamación justo delante del sinónimo `!sinónimo`, indica que las formas declinadas del sinónimo no se sustituirán por el término objetivo. Sin embargo, un signo de exclamación justo delante del término objetivo `!término_objetivo`, significa que no desea que no se apliquen más sustituciones a ninguna parte del término objetivo compuesto ni a ninguna variante.
- **Asterisco (\*).** Un asterisco situado justo después de un sinónimo, como `sinónimo*`, significa que desea que esta palabra se sustituya por el término objetivo. Por ejemplo, si ha definido `administrar*` como sinónimo y `administración` como objetivo, el término `administradores asociados` se sustituirá por el término objetivo `administración asociada`. También puede añadir un espacio y un asterisco detrás de la palabra (`sinónimo*`), por ejemplo, `internet *`. Si ha definido el objetivo como `internet` y los sinónimos como `internet * *` y `web *`, los términos `starjeta de acceso a internet` y `portal web` se sustituirán por `internet`. No puede comenzar una palabra o una cadena con el comodín de asterisco en este diccionario.
- **Intercalación (^).** Un signo de intercalación y un espacio justo delante del sinónimo, como `^ sinónimo`, significa que la agrupación de sinónimos se aplica solamente cuando el término empieza con el sinónimo. Por ejemplo, si define `^ salario` como sinónimo e `ingresos` como objetivo, y se extraen ambos términos, se agruparán juntos bajo el término `ingresos`. Sin embargo, si se extraen los términos `subir salario` e `ingresos`, no se agruparán juntos, puesto que `subir salario` no empieza por `salario`. Debe colocarse un espacio entre este símbolo y el sinónimo.
- **Signo de dólar (\$).** Un espacio y un símbolo de dólar justo delante del sinónimo, como `sinónimo $`, significa que la agrupación de sinónimos se aplica solamente cuando el término termina con el sinónimo. Por ejemplo, si define `salario $` como sinónimo e `ingresos` como objetivo, y se extraen ambos términos, se agruparán juntos bajo el término `ingresos`. Sin embargo, si se extraen los términos `salario mínimo` e `ingresos`, no se agruparán juntos porque `salario mínimo` no termina con `salario`. Debe colocarse un espacio entre este símbolo y el sinónimo.
- **Símbolo de intercalación (^) y símbolo de dólar (\$).** Si los símbolos de intercalación y de dólar se utilizan juntos, como `^ sinónimo $`, un término solo coincidirá con el sinónimo si se trata de una coincidencia exacta. Esto significa que no puede aparecer ninguna palabra delante ni detrás del sinónimo del término extraído para que pueda realizarse la agrupación de sinónimos. Por ejemplo, puede definir `^ van $` como sinónimo e `ir` como objetivo de manera que solo `van` se agrupe con `ir`, mientras que `marie van guerin` permanecería sin cambios. Además, siempre que defina un sinónimo utilizando los símbolos de intercalación y de dólar y esta palabra aparezca en cualquier lugar de texto de origen, el sinónimo se extraerá automáticamente.

Figura 10-11  
Diccionario Sustitución, pestaña Sinónimos con ejemplo

	Objetivo	Sinónimos	Biblioteca
1	maltratado	mal atención, mal atendido, mal tratado, maltrato	Biblioteca local
2	buena calidad	buena calidad	Opinions Library (Spanish)
3	lcorto	lcorta, lcortas, lcorto, lcortos	Opinions Library (Spanish)
4	lcortés	lcortes, lcortesa, lcortesas, lcorteses, lcortés	Opinions Library (Spanish)
5	le gusta	^ gusta \$, ^ gustan \$, agusto, enamorado, encuanza, encuentro a gusto, enomorado, gustado mucho, gutan, le gustan, legusta, me gusta, me gusta mucho, me gustaba, me gustan, me gustan mucho, me gustaría, me gustaría, me guste, megusta,	Opinions Library (Spanish)

Sinónimos Elementos opcionales

### Para añadir una entrada de sinónimo

- ▶ Con el panel de sustitución visualizado, pulse en la pestaña Sinónimos en el ángulo inferior izquierdo.
- ▶ En la línea vacía de la parte superior de la tabla, escriba el término objetivo en la columna Objetivos. El término objetivo que ha escrito aparece en color. Este color representa el tipo en el que el término aparece o se fuerza, si se da el caso. Si el término aparece en negro, significa que no está en ningún diccionario de tipo.
- ▶ Pulse en la segunda casilla a la derecha del objetivo y escriba el conjunto de sinónimos. Separe cada entrada utilizando el delimitador global tal como está definido en el cuadro de diálogo Opciones. [Si desea obtener más información, consulte el tema Opciones de configuración en el capítulo 2 el p. 17.](#) Los términos que especifique aparecerán en color. Este color representa el tipo en el que aparece el término. Si el término aparece en negro, significa que no está en ningún diccionario de tipo.
- ▶ Pulse en la última casilla para seleccionar la biblioteca en la que desea almacenar esta definición de sinónimo.

*Nota:* estas instrucciones indican cómo realizar cambios en la Editor de recursos vista. Tenga en cuenta que este tipo de ajuste puede hacerlo directamente en el panel Resultados extraídos o en el panel Datos. [Si desea obtener más información, consulte el tema Refinamiento de los resultados de la extracción en el capítulo 5 el p. 86.](#)

### Definición de elementos opcionales

En la pestaña Opcional, puede definir elementos opcionales para la biblioteca que desee. Estas entradas se agrupan juntas para cada biblioteca. Tan pronto como se añade una biblioteca al panel del árbol de bibliotecas, se añade una línea vacía de elemento opcional en la pestaña Opcional.

Todas las entradas pasan a estar en minúsculas automáticamente. El motor de extracción hará coincidir las entradas en mayúsculas y minúsculas del texto.

**Figura 10-12**  
Diccionario Sustitución, pestaña Opcional

Elementos opcionales	Biblioteca
<input checked="" type="checkbox"/>	Biblioteca local
<input checked="" type="checkbox"/> al, en, via, vía	Opinions Library (Spanish)
<input checked="" type="checkbox"/> s.l.	Core Library (Spanish)
<input checked="" type="checkbox"/>	Budget Library (Spanish)
<input checked="" type="checkbox"/>	Variations Library (Spanish)

Sinónimos    Elementos opcionales

*Nota:* Los términos se separan utilizando el delimitador definido en el cuadro de diálogo Opciones. [Si desea obtener más información, consulte el tema Opciones de configuración en el capítulo 2 el p. 17.](#) Si el elemento opcional que va a especificar incluye el mismo delimitador como parte del término, deberá colocar una barra invertida delante.

#### **Para añadir una entrada**

- ▶ Con el panel de sustitución visualizado, pulse en la pestaña Opcional en el ángulo inferior izquierdo del editor.
- ▶ Pulse en la casilla de la columna de Elementos opcionales correspondiente a la biblioteca a la que desea añadir esta entrada.
- ▶ Especifique el elemento opcional. Separe cada entrada utilizando el delimitador global tal como está definido en el cuadro de diálogo Opciones. [Si desea obtener más información, consulte el tema Opciones de configuración en el capítulo 2 el p. 17.](#)

### **Desactivación y eliminación de sustituciones**

Puede eliminar una entrada de forma temporal desactivándola del diccionario. Al desactivar una entrada, ésta se pasará por alto durante la extracción.

También puede eliminar las entradas obsoletas en el diccionario de sustitución.

#### **Para desactivar una entrada**

- ▶ En el diccionario seleccione la entrada que desea desactivar.
- ▶ Pulse la barra espaciadora. Se borra la marca de la casilla de verificación a la izquierda de la entrada.

*Nota:* También puede anular la selección de la casilla de verificación a la izquierda para desactivarla.

#### **Para eliminar una entrada de sinónimo**

- ▶ En el diccionario seleccione la entrada que desea eliminar.

- ▶ En los menús elija Editar > Eliminar o pulse la tecla Supr del teclado. La entrada desaparece del diccionario.

### **Para eliminar una entrada de elemento opcional**

- ▶ En el diccionario pulse dos veces en la entrada que desea eliminar.
- ▶ Elimine manualmente el término.
- ▶ Pulse Intro para aplicar el cambio.

## **Diccionarios de exclusión**

Un **diccionario de exclusión** es una lista de palabras, frases o cadenas parciales. Se pasarán por alto o se excluirán de la extracción los términos que tengan alguna coincidencia o contengan una entrada en el diccionario de exclusión. Los diccionarios de exclusión se administran en el panel derecho del editor. Por lo general, los términos que añade a esta lista serán palabras o frases de relleno que se utilizan en el texto para conferir continuidad, pero que no aportan información relevante al texto y que además pueden cargar innecesariamente los resultados de la extracción. Si añade estos términos al diccionario de exclusión, tendrá la seguridad de que no se extraerán nunca.

Los diccionarios de exclusión se administran en el panel superior derecho de la pestaña Recursos de la biblioteca del editor. Puede acceder a esta vista en Ver > Editor de recursos en los menús.

Figura 10-13  
Panel del diccionario de exclusión

	Excluidos	Biblioteca
0		
1	<input checked="" type="checkbox"/> borroso	Core Library (Spanish)
2	<input checked="" type="checkbox"/> pqáto	Opinions Library (Spanish)
3	<input checked="" type="checkbox"/> cognoscible	Opinions Library (Spanish)
4	<input checked="" type="checkbox"/> baja	Opinions Library (Spanish)
5	<input checked="" type="checkbox"/> 100% ok	Opinions Library (Spanish)
6	<input checked="" type="checkbox"/> relacion calidad/preci	Budget Library (Spanish)
7	<input checked="" type="checkbox"/> ee.uu	Core Library (Spanish)
8	<input checked="" type="checkbox"/> ninguno en especial	Opinions Library (Spanish)
9	<input checked="" type="checkbox"/> tardio	Opinions Library (Spanish)
10	<input checked="" type="checkbox"/> tienen que mejorar	Opinions Library (Spanish)
11	<input checked="" type="checkbox"/> vulgar	Opinions Library (Spanish)
12	<input checked="" type="checkbox"/> por la tarde	Opinions Library (Spanish)

En el diccionario de exclusión puede introducir una palabra, frase o cadena parcial en la línea vacía en la parte superior de la tabla. Puede añadir cadenas de caracteres al diccionario de exclusión como una o más palabras o incluso como palabras parciales utilizando el asterisco como comodín. Las entradas declaradas en el diccionario de exclusión se utilizarán para impedir que los conceptos se extraigan Si una entrada también se declara en algún otro lugar de la interfaz, como en un diccionario de tipo, se muestra con un signo de tachado en los otros diccionarios, lo que indica que actualmente está excluida. Esta cadena no tiene que aparecer en los datos de texto ni declararse como parte de ningún diccionario de tipo que vaya a aplicarse.

*Nota:* Si añade un concepto al diccionario de exclusión que también actúe como objetivo de una entrada de sinónimo, el objetivo y todos sus sinónimos también se excluirán. [Si desea obtener más información, consulte el tema Definición de sinónimos el p. 227.](#)

**Uso de comodines (\*)**

puede utilizar el comodín de asterisco para indicar que desea tratar la entrada de exclusión como una cadena parcial. Todos los términos el motor de extracción que encuentre con una palabra que empiece o termine por una cadena especificada en el diccionario de exclusión se excluirán de la extracción final. Sin embargo, existen dos casos en los que el uso de comodines no está permitido:

- Carácter de guión (-) precedido por un asterisco, como \*-
- Carácter de apóstrofe (') precedido por un asterisco, como \*'

Tabla 10-2

Ejemplos de entradas de exclusión

Entrada	Ejemplo	Resultados
palabra	<i>siguiente</i>	No se extraerá ningún concepto (ni sus términos) si contienen la palabra <i>siguiente</i> .
frase	<i>por ejemplo</i>	No se extraerá ningún concepto (ni sus términos) si contienen la frase <i>por ejemplo</i> .
parciales	<i>copyright*</i>	Se excluirán los conceptos (o sus términos) que coincidan con, o que contengan, las variaciones de la palabra <i>copyright</i> , como <i>copyrighted</i> , <i>copyrighting</i> , <i>copyrights</i> o <i>copyright 2010</i> .
parciales	<i>*ware</i>	Se excluirán los conceptos (o sus términos) que coincidan con, o que contengan, las variaciones de la palabra <i>ware</i> , como <i>freeware</i> , <i>shareware</i> , <i>software</i> o <i>hardware</i> .

**Para añadir entradas**

- ▶ En la línea vacía en la parte superior de la tabla, escriba un término. El término que especifique aparecerá en color. Este color representa el tipo en el que aparece el término. Si el término aparece en negro, significa que no está en ningún diccionario de tipo.

**Para desactivar entradas**

Puede eliminar temporalmente una entrada desactivándola en el diccionario de exclusión. Al desactivar una entrada, ésta se pasará por alto durante la extracción.

- ▶ En el diccionario de exclusión seleccione la entrada que desea desactivar.
- ▶ Pulse la barra espaciadora. Se borra la marca de la casilla de verificación a la izquierda de la entrada.

*Nota:* También puede anular la selección de la casilla de verificación a la izquierda para desactivarla.

**Para eliminar entradas**

Puede eliminar las entradas que ya no necesite en el diccionario de exclusión.

- ▶ En el diccionario de exclusión seleccione la entrada que desea eliminar.
- ▶ En los menús elija Editar > Eliminar. La entrada desaparece del diccionario.

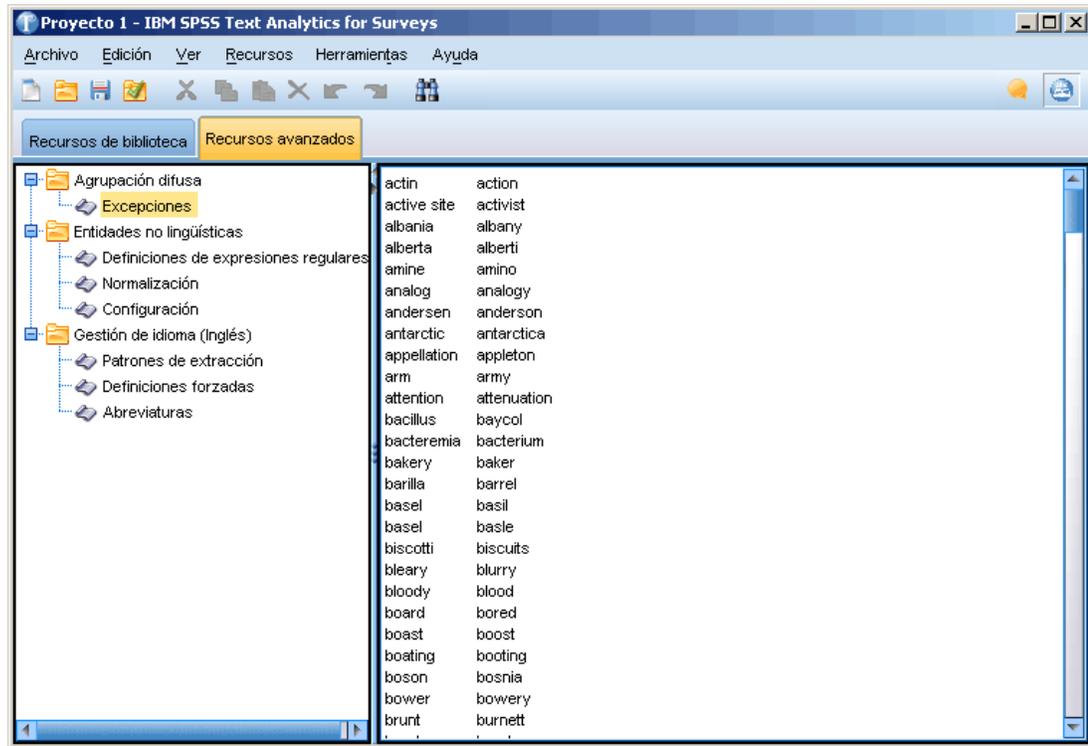
## ***Acerca de los recursos avanzados***

Además de los diccionarios de tipo, de exclusión y de sustitución, también puede trabajar con una serie de opciones de recursos avanzados como configuración de agrupación difusa o definiciones de tipo no lingüístico. Puede trabajar con estos recursos en la pestaña Recursos avanzados en la vista Editor de recursos. También puede guardar los cambios como valores por defecto para todos los proyectos, o recuperar el contenido original.

En la pestaña Recursos avanzados puede editar la siguiente información:

- **Agrupación difusa (excepciones).** Se utiliza para excluir parejas de palabras del algoritmo de agrupación difusa (corrección de error ortográfico). [Si desea obtener más información, consulte el tema Agrupación difusa el p. 236.](#)
- **Entidades no lingüísticas.** Se utiliza para activar y desactivar las entidades no lingüísticas que pueden extraerse, así como las expresiones normales y las reglas de normalización que se aplican durante la extracción. [Si desea obtener más información, consulte el tema Entidades no lingüísticas el p. 237.](#)
- **Gestión de idioma.** Se utiliza para declarar los métodos especiales para estructurar frases (patrones de extracción y definiciones forzadas) y para utilizar abreviaturas en el idioma seleccionado. [Si desea obtener más información, consulte el tema Gestión de idiomas el p. 242.](#)

Figura 11-1  
Recursos avanzados



*Nota:* Puede utilizar la barra de herramientas Buscar/Reemplazar para buscar información rápidamente o para realizar cambios globales en una sección. [Si desea obtener más información, consulte el tema Reemplazo el p. 235.](#)

#### Para Editar recursos avanzados

- ▶ Localice y seleccione la sección de recursos que desea editar. Los contenidos aparecerán en el panel derecho.
- ▶ Utilice el menú o los botones de la barra de herramientas para cortar, copiar o pegar el contenido, si es preciso.
- ▶ Edite el archivo o archivos que desea cambiar utilizando las reglas de formato de esta sección. Los cambios se guardan en cuanto los lleva a cabo. Utilice las flechas de deshacer o rehacer de la barra de herramientas para invertir la acción sobre los cambios anteriores.

## Búsquedas

En algunos casos, puede que necesite localizar información rápidamente en una sección determinada. Con la función Buscar, puede encontrar rápidamente una regla específica. Para buscar información en una sección, puede utilizar la barra de herramientas Buscar.

Figura 11-2  
Barra de herramientas Buscar



### Para utilizar la función Buscar

- ▶ Localice y seleccione la sección de recursos que desea buscar. Los contenidos aparecen en el panel derecho del editor.
- ▶ En los menús elija Editar > Buscar. La barra de herramientas Buscar aparecerá en el ángulo superior derecho del cuadro de diálogo Editar recursos avanzados.
- ▶ Escriba la cadena de palabras que desea buscar en el cuadro de texto. Puede utilizar los botones de la barra de herramientas para controlar las mayúsculas/minúsculas, la coincidencia parcial y la dirección de la búsqueda.
- ▶ Pulse en Buscar para iniciar la búsqueda. Si se encuentra una coincidencia, el texto se resalta en la ventana.
- ▶ Pulse en Buscar otra vez para buscar la siguiente coincidencia.

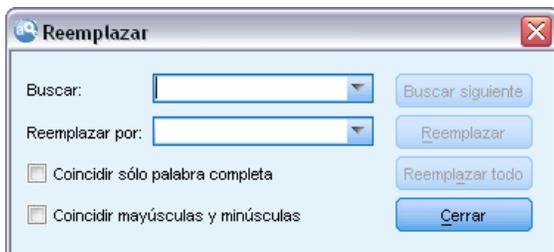
## Reemplazo

En algunos casos, puede que necesite realizar actualizaciones globales en los recursos avanzados. La función Reemplazar puede ayudarle a realizar actualizaciones uniformes en el contenido.

### Para utilizar la función Reemplazar

- ▶ Localice y seleccione la sección de recursos que desea buscar y reemplazar. Los contenidos aparecen en el panel derecho del editor.
- ▶ En los menús elija Editar > Reemplazar. Aparecerá el cuadro de diálogo Reemplazar.

Figura 11-3  
Cuadro de diálogo Reemplazar



- ▶ En el cuadro de texto Buscar, escriba la cadena de palabras que desea buscar.
- ▶ En el cuadro de texto Reemplazar por, escriba la cadena que desea utilizar en lugar del texto de búsqueda.
- ▶ Seleccione Coincidir sólo palabra completa si desea buscar o reemplazar únicamente palabras completas.

- ▶ Seleccione Coincidir mayúsculas y minúsculas si desea buscar o reemplazar únicamente palabras que coincidan con las mayúsculas y minúsculas exactamente.
- ▶ Pulse en Buscar siguiente para buscar una coincidencia. Si se encuentra una coincidencia, el texto se resalta en la ventana. Si no desea reemplazar esta coincidencia, pulse en Buscar siguiente de nuevo hasta que encuentre una coincidencia que desee reemplazar.
- ▶ Pulse en Reemplazar para sustituir la coincidencia seleccionada.
- ▶ Pulse en Reemplazar para sustituir todas las coincidencias de la sección. Se abrirá un mensaje indicando el número de reemplazos realizados.
- ▶ Cuando haya terminado de reemplazar, pulse en Cerrar. El cuadro de diálogo se cierra.

*Nota:* Si ha cometido un error de reemplazo, puede deshacer la acción cerrando el cuadro de diálogo y eligiendo Editar > Deshacer en los menús. Deberá realizar esta acción una vez por cada cambio que desee deshacer.

## Agrupación difusa

En el diálogo Configuración de extracción, si selecciona Acomodar la ortografía a un límite mínimo de caracteres raíz de:, significa que ha activado el algoritmo de agrupación difusa.

La agrupación difusa ayuda a agrupar las palabras que comúnmente se escriben mal o que tienen una ortografía similar pasando temporalmente por alto todas las vocales (excepto la primera) o consonantes dobles o triples de las palabras extraídas, y luego comparándolas para comprobar si son las mismas. Durante el proceso de extracción, la función de agrupación difusa se aplicará a los términos extraídos, y los resultados se comparan para determinar si se han encontrado coincidencias. En caso afirmativo, los términos originales se agrupan juntos en la lista de extracción final. Se agrupan bajo el término que aparece más veces en los datos.

*Nota:* Si los dos términos de comparación están asignados a tipos diferentes, excluyendo el tipo <Unknown>, la técnica de agrupación difusa no se aplicará a esta pareja de términos. En otras palabras, para poder aplicar la técnica, los términos deben pertenecer al mismo tipo o al tipo <Unknown>.

Si ha activado esta función y descubre que dos palabras con ortografía similar se han agrupado incorrectamente, puede excluir dichas palabras de la agrupación difusa. Para ello escriba las parejas mal agrupadas en la sección Excepciones de la pestaña Recursos avanzados. [Si desea obtener más información, consulte el tema Acerca de los recursos avanzados el p. 233.](#)

En el ejemplo siguiente se muestra el proceso de la agrupación difusa. Si se ha activado la agrupación difusa, estas palabras parecen ser iguales y se emparejan de la forma siguiente:

```

color -> colr          mountain -> montn
colour -> colr         montana -> montn

modeling -> modlng     furniture -> furntr
modelling -> modlng    furnature -> furntr

```

En el ejemplo anterior, probablemente quiera impedir que salida y salud se agrupen juntas. Por lo tanto, puede incluir estas palabras en la sección Excepciones de la siguiente manera:

```

salida      salud

```

**Importante:** En algunos casos, las excepciones de agrupación difusa no finalizan el proceso de emparejamiento de 2 palabras porque se han aplicado algunas reglas de sinónimos. En ese caso, puede intentar introducir sinónimos usando el comodín de signo de exclamación (!) para impedir que las palabras se hagan sinónimas en los resultados. [Si desea obtener más información, consulte el tema Definición de sinónimos en el capítulo 10 el p. 227.](#)

#### **Reglas de formato para excepciones de agrupación difusa**

- Defina solo una pareja de excepción por línea.
- Utilice palabras simples o compuestas.
- Utilice solamente caracteres en minúsculas para las palabras. Las palabras en mayúsculas se pasarán por alto.
- Utilice un `tabulador` para separar cada palabra de una pareja.

## **Entidades no lingüísticas**

Cuando trabaja con determinados tipos de datos, puede ser de gran interés extraer fechas, números de la seguridad social, porcentajes u otras entidades no lingüísticas. Estas entidades están explícitamente declaradas en el archivo de configuración, donde puede activar o desactivar las entidades. [Si desea obtener más información, consulte el tema Configuración el p. 241.](#) Con el fin de optimizar el rendimiento del motor de extracción, la entrada que se realiza en el proceso no lingüístico se normaliza para agrupar entidades similares según los formatos predefinidos. [Si desea obtener más información, consulte el tema Normalización el p. 240.](#)

*Nota:* Puede activar y desactivar la extracción de entidades no lingüísticas en la configuración de extracción.

#### **Entidades no lingüísticas disponibles**

Pueden extraerse las entidades no lingüísticas de la tabla siguiente. El nombre del tipo está entre paréntesis.

Direcciones (<Address>)	Organizaciones (<Organization>)
Aminoácidos (<Aminoacid>)	Porcentajes (<Percent>)
Monedas (<Currency>)	Productos (<Product>)
Fechas (<Date>)	Proteínas (<Gene>)
Retraso (<Delay>)	Números de teléfono (<PhoneNumber>)
Dígitos (<Digit>)	Horas (<Time>)
Direcciones de correo electrónico (<email>)	Números de la seguridad social (<SocialSecurityNumber>)
Direcciones de HTTP/URL (<url>)	Pesos y medidas (<Weights-Measures>)
Dirección IP (<IP>)	

### **Limpieza del texto para proceso**

Antes de que se realice la extracción de entidades no lingüísticas, el texto de entrada se limpia. Durante este paso, se realizan los cambios temporales siguientes para que las entidades no lingüísticas puedan identificarse y extraerse como tales:

- Cualquier secuencia de dos o más espacios se sustituye por un espacio único.
- Las tabulaciones se sustituyen por espacios.
- Los caracteres de secuencia o caracteres únicos de final de línea se sustituyen por un espacio, mientras que las secuencias múltiples de final de línea se marcan como final de un párrafo. El final de línea puede indicarse mediante retornos de carro (CR) y cambio de línea (LF), o incluso ambos a la vez.
- Los códigos HTML y XML se pasan por alto temporalmente y se ignoran.

### **Definiciones de expresiones regulares**

Cuando se extraen entidades no lingüísticas, puede editar o añadir a la expresión regular aquellas definiciones que se utilizan para identificar las expresiones regulares. Esto se hace en la sección Definiciones de expresiones regulares de la pestaña Recursos avanzados. [Si desea obtener más información, consulte el tema Acerca de los recursos avanzados el p. 233.](#)

El archivo está dividido en varias secciones. La primera sección se llama `[macros]`. Además de dicha sección, puede existir una sección adicional para cada entidad no lingüística. Puede añadir secciones a este archivo. En cada sección, las reglas están numeradas (*regex1*, *regex2*, etc.). Estas reglas deben numerarse secuencialmente desde 1–*n*. Toda interrupción en la numeración hará que el proceso de este archivo quede suspendido.

En determinados casos, una entidad puede depender del idioma. Se considera que una entidad depende del idioma si toma un valor que no sea 0 como parámetro de idioma en el archivo de configuración. [Si desea obtener más información, consulte el tema Configuración el p. 241.](#) Cuando una entidad depende del idioma, debe utilizarse el idioma como prefijo del nombre de sección, como en `[inglés/PhoneNumber]`. Esta sección contendrá reglas que se apliquen solo a los números de teléfono ingleses si se asigna el valor de 2 al idioma de la entidad `PhoneNumber`.

**Importante:** Si realiza cambios en este archivo o en cualquier otro en el editor, y el motor de extracción deja de funcionar como se esperaba, utilice la opción Restablecer originales en la barra de herramientas para restaurar el archivo con el contenido original enviado con el producto. Es necesario estar familiarizado con las expresiones regulares para trabajar con este archivo. Si necesita ayuda adicional en esta área, póngase en contacto con IBM Corp. para obtener ayuda.

### **Caracteres especiales. `[]{}()\ * + ? | ^ $`**

Todos los caracteres coinciden entre sí excepto los caracteres especiales siguientes, que se utilizan para un propósito específico en las expresiones: `. [ { ( ) \ * + ? | ^ $` Para utilizar estos caracteres como tales, deben ir precedidos por una barra inclinada invertida (`\`) en la definición:

Por ejemplo, si intentaba extraer direcciones web, el carácter de punto y aparte es muy importante en la entidad, por lo que debe colocar una barra inclinada invertida, así:

```
www\[a-z]+\.[a-z]+
```

### Operadores de repeticiones y cuantificadores ? + \* {}

Para que las definiciones sean más flexibles, puede utilizar varios comodines que sean estándares en las expresiones regulares. Son \* ? +

- *Asterisco* \* indica que hay *ceros o más* elementos de la cadena precedente. Por ejemplo, `ab*c` coincide con “*ac*”, “*abc*”, “*abbc*”, y etc.
- *El signo más* + indica que hay *uno o más* elementos de la cadena precedente. Por ejemplo, `ab+c` coincide con “*abc*”, “*abbc*”, “*abbbc*”, pero no con *ac*”.
- *El signo de interrogación* ? indica que hay *ceros o más* elementos de la cadena precedente. Por ejemplo, `model?ado` coincide con “*modelado*” y con “*modelado*”.
- *Limitar la repetición mediante llaves* {} indica los límites de la repetición. Por ejemplo:
  - ▶ `[0-9]{n}` coincide con un dígito que se repite exactamente *n* veces. Por ejemplo, `[0-9]{4}` coincide con “*1998*”, pero no con “*33*” ni con “*19983*”.
  - ▶ `[0-9]{n,}` coincide con un dígito que se repite *n* o más veces. Por ejemplo, `[0-9]{3,}` coincide con “*199*” o con “*1998*”, pero no con “*19*”.
  - ▶ `[0-9]{n,m}` coincide con un dígito que se repite entre *n* y *m* veces, inclusive. Por ejemplo, `[0-9]{3,5}` coincide con “*199*”, “*1998*” o “*19983*”, pero no con “*19*” ni “*199835*”.

### Guiones y espacios opcionales

En algunos casos, puede que necesite incluir un espacio opcional en una definición. Por ejemplo, si desea extraer monedas como “*pesos uruguayos*”, “*peso uruguayo*”, “*pesos uruguay*”, “*peso uruguay*”, “*pesos*” o “*peso*”, tendrá que afrontar el hecho de que puede haber dos palabras separadas por un espacio. En este caso, esta definición se escribirá como `(uruguayo |uruguay )?pesos?`. Puesto que *uruguayo* o *uruguay* van seguidos de un espacio cuando se utilizan con *pesos/peso*, el espacio opcional debe definirse en la secuencia opcional `(uruguayo |uruguay )`. Si no se había definido el espacio en la secuencia opcional, como `(uruguayo|uruguay)?pesos?`, no coincidiría con “*pesos*” o “*peso*”, puesto que es necesario el espacio.

Si busca una serie de objetos que incluyen guiones (-) en una lista, el guión debe estar definido al final. Por ejemplo, si busca una coma (,) o un guión (-), utilice `[, -]` pero nunca `[-, ]`.

### Orden de las cadenas en listas y macros

Siempre debe definir la secuencia más larga antes que la más corta; de lo contrario, nunca se detectará la más larga, puesto que la coincidencia se producirá en la secuencia más corta. Por ejemplo, si busca las cadenas “*teléfono*” o “*tele*”, entonces “*teléfono*” debe definirse antes que “*tele*”. Así pues, `(teléfono|tele)` y no `(tele|teléfono)`. Esto también se aplica a las macros, puesto que estas son listas de cadenas.

### Orden de las reglas en la sección de definición

Defina una regla por línea. En cada sección, las reglas están numeradas (*regexp1*, *regexp2*, etc.). Estas reglas deben numerarse secuencialmente desde 1–*n*. Toda interrupción en la numeración hará que el proceso de este archivo quede suspendido. Para desactivar una entrada, coloque un símbolo de número (#) al principio de cada línea que se utiliza para definir la expresión regular. Para activar una entrada, elimine el símbolo de número (#) del principio de la línea.

En cada sección, las reglas más específicas deben definirse antes que las más generales para asegurar un proceso apropiado. Por ejemplo, si busca una fecha con el formato “*mes año*” y con el formato “*mes*”, la regla “*mes año*” debe definirse antes que la regla “*mes*”. Este es un ejemplo de cómo debe definirse:

```
#@# Enero 1932
regex1=$(MONTH),? [0-9]{4}

#@# Enero
regex2=$(MONTH)
```

y no

```
#@# Enero 1932
regex1=$(MONTH)

#@# enero 1932
regex2=$(MONTH),? [0-9]{4}
```

### **Uso de macros en las reglas**

Cuando se utiliza una secuencia específica en varias reglas, puede utilizar una macro. Así pues, si necesita cambiar la definición de esta secuencia, deberá cambiarla solo una vez, y no es necesario que lo haga en todas las reglas a las que haga referencia. Por ejemplo, imagine que tiene la macro siguiente:

```
MONTH=((enero|febrero|marzo|abril|mayo|junio|julio|agosto|septiembre|octubre|
noviembre|diciembre)|(ene|feb|mar|abr|may|jun|jul|ago|sep|oct|nov|dic) (\.)?)
```

Siempre que haga referencia al nombre de la macro, deberá estar entre \$ ( ) , por ejemplo:  
`regex1=$(MONTH)`

Todas las macros deben definirse en la sección [macros].

## **Normalización**

Cuando se extraen entidades no lingüísticas, las entidades que se encuentran se normalizan para agrupar entidades parecidas según los formatos predefinidos. Por ejemplo, los símbolos de moneda y sus equivalentes en palabras se consideran lo mismo. Las entradas de normalización se almacenan en la sección Normalización de la pestaña Recursos avanzados. [Si desea obtener más información, consulte el tema Acerca de los recursos avanzados el p. 233](#). El archivo está dividido en varias secciones.

**Importante:** Este archivo solo deben utilizarlo usuarios experimentados. Es altamente improbable que necesite cambiar este archivo. Si necesita ayuda adicional en esta área, póngase en contacto con IBM Corp. para obtener ayuda.

### **Reglas de formato para la normalización**

- Añada solo una entrada de normalización por línea.
- Respete estrictamente las secciones de este archivo. No pueden añadirse secciones nuevas.
- Para desactivar una entrada, coloque un símbolo de número (#) al principio de la línea. Para activar una entrada, elimine el símbolo de número (#) del principio de la línea.

## Configuración

Puede activar y desactivar los tipos de entidades no lingüísticas que desee extraer en el archivo de configuración de entidades no lingüísticas. Al desactivar las entidades que no necesita, puede disminuir el tiempo de proceso necesario. Esto se hace en la sección Configuración de la pestaña Recursos avanzados. [Si desea obtener más información, consulte el tema Acerca de los recursos avanzados el p. 233](#). Si no se activa la extracción no lingüística, el motor de extracción lee este archivo de configuración durante el proceso de extracción para determinar cuáles son los tipos de entidades no lingüísticas que deben extraerse.

La sintaxis de este archivo es la siguiente:

```
#nombre<TAB>Idioma<TAB>Código
```

Tabla 11-1  
Sintaxis del archivo de configuración

Etiqueta de columna	Descripción
#nombre	Término por el que se hará referencia a las entidades no lingüísticas en los otros dos archivos requeridos para la extracción de entidades no lingüísticas. Los nombres que se utilizan aquí son sensibles a mayúsculas y minúsculas.
Idioma	El idioma de los registros. Se recomienda seleccionar el idioma específico; sin embargo, existe una opción Cualquiera. Estas son las opciones posibles: 0 = Cualquiera que se utilice cuando regexp no es específico de un idioma y puede utilizarse en varias plantillas con idiomas diferentes, por ejemplo, direcciones IP/URL/correo electrónico; 1 = Francés; 2 = Inglés; 4 = Alemán; 5 = Español; 6 = Holandés; 8 = Portugués; 10 = Italiano.
Código	Código de categoría léxica. La mayoría de las entidades tomar un valor de “s” excepto en unos pocos casos. Entre los posibles valores se incluyen: s = palabra vacía; a = adjetivo; n = sustantivo. Si está activado, las entidades no lingüísticas se extraen en primer lugar y se aplican los patrones de extracción para identificar su rol en un contexto más amplio. Por ejemplo, a los porcentajes se les asigna un valor “a.” Supongamos que se extrae 30% como entidad no lingüística. Se identificaría como un adjetivo. Así pues, si el texto contiene “30% aumento salario,” la entidad no lingüística “30%” se ajusta al patrón de categoría léxica “ann” (adjetivo sustantivo sustantivo).

### Orden en la definición de entidades

El orden en que se declaran las entidades en este archivo es relevante y afecta a la forma en que se extraerán. Se aplican en el orden de la lista. Si cambia el orden, cambiará el resultado. Las entidades no lingüísticas más específicas deben definirse antes que las más generales.

Por ejemplo, la entidad no lingüística “Aminoácido” se define mediante:

```
regexp1=($(AA)-?$(NUM))
```

donde \$(AA) corresponde a

“(ala|arg|asn|asp|cys|gln|glu|gly|his|ile|leu|lys|met|phe|pro|ser)”, que son secuencias específicas de tres letras que corresponden a aminoácidos específicos.

Por otro lado, la entidad no lingüística “Gen” es más general y se define mediante:

```
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 “Gen” se define antes que “Aminoácido” en la sección Configuración, “Aminoácido” nunca tendrá una coincidencia porque `regexp3` en “Gen” siempre coincidirá primero.

### **Reglas de formato para la configuración**

- Utilice un `tabulador` para separar cada entrada en una columna.
- No elimine ninguna línea.
- Respete la sintaxis que se muestra en la tabla anterior.
- Para desactivar una entrada, coloque un símbolo de número (#) al principio de la línea. Para activar una entidad, elimine el símbolo de número (#) del principio de la línea.

## **Gestión de idiomas**

Todos los idiomas actuales tienen maneras especiales de expresar ideas, estructurar frases y utilizar abreviaturas. En la sección Gestión de idioma, puede editar patrones de extracción, forzar definiciones para dichos patrones y declarar abreviaturas para el idioma que ha seleccionado en la lista desplegable de Idiomas.

- Patrones de extracción
- Definiciones forzadas
- Abreviaturas

## **Patrones de extracción**

Cuando se extrae información de los registros, el motor de extracción aplica un conjunto de patrones de categorías léxicas a una “pila” de palabras del texto para identificar los términos candidatos (palabras y frases) para la extracción. Puede añadir o modificar los patrones de extracción.

Las categorías léxicas incluyen elementos gramaticales, como sustantivos, adjetivos, participios, determinantes, preposiciones, conjunciones, nombres propios, iniciales y partículas. Un patrón de extracción de categorías léxicas está compuesto por una serie de estos elementos. En los productos de minería de texto de IBM Corp., cada categoría léxica está representada por un único carácter para facilitar la definición de sus patrones. Por ejemplo, un adjetivo está representado por la letra en minúscula *a*. Por defecto aparece el conjunto de códigos soportados al principio de cada sección de patrones de extracción por defecto, junto con una serie de patrones y ejemplos de cada patrón, para facilitarle la comprensión de cada código que se utilice.

### **Reglas de formato para los patrones de extracción**

- Un patrón por línea.
- Utilice # al principio de una línea para desactivar un patrón.

El orden en el que aparecen los patrones de extracción es muy importante, porque el motor de extracción lee una secuencia determinada de palabras solo una vez, y se le asigna al primer patrón de extracción para el que el motor encuentra una coincidencia.

## Definiciones forzadas

Cuando se extrae información de los registros, el motor de extracción explora el texto e identifica la categoría léxica de cada palabra que encuentra. En algunos casos, una palabra puede ajustarse a varios roles diferentes en función del contexto. Si desea forzar una palabra para que tome un rol de categoría léxica o para excluirla completamente del proceso, puede hacerlo en la sección Definición forzada de la pestaña Recursos avanzados. [Si desea obtener más información, consulte el tema Acerca de los recursos avanzados el p. 233.](#)

Para forzar un rol de categoría léxica para una palabra determinada, debe añadir una línea en esta sección empleando esta sintaxis:

*términos:código*

Tabla 11-2  
Descripción de la sintaxis

Entrada	Descripción
términos	Nombre de término.
código	Código de un solo carácter que representa el rol de categoría léxica. Puede enumerar hasta seis códigos de categorías léxicas diferentes por unitérmino. Además puede impedir que una palabra se extraiga en palabras o frases compuestas mediante el código en minúscula s, por ejemplo, adicional:s.

### Reglas de formato para definiciones forzadas

- Una línea por palabra.
- Los términos no pueden tener el carácter de dos puntos.
- Utilice el carácter en minúscula s como código de categoría léxica para impedir que una palabra se extraiga.
- Utilice un máximo de seis códigos de categorías léxicas por línea. Los códigos admitidos de categorías léxicas se muestran en la sección Patrones de extracción. [Si desea obtener más información, consulte el tema Patrones de extracción el p. 242.](#)
- Utilice un asterisco (\*) como carácter comodín al final de una cadena para conseguir coincidencias parciales. Por ejemplo, si especifica `adic*:s`, palabras como `añadir`, `adicional`, `adicionalmente`, `adictivo` y `adicción` nunca se extraerán como término o como parte de un término de palabra compuesta. Sin embargo, si se declara de manera explícita una coincidencia de palabra como término en un diccionario compilado o en las definiciones forzadas, sí que se extraerá. Por ejemplo, si especifica `adic*:s` y `adictivo:n`, `adictivo` se extraerá si se encuentra en el texto.

## Abreviaturas

Cuando el motor de extracción está procesando texto, generalmente considera los puntos como una indicación de que la frase ha terminado. Esto suele ser correcto; sin embargo, este manejo de los puntos no se aplica en el caso de las abreviaturas.

Si extrae términos del texto y detecta que no se han manejado bien determinadas abreviaturas, deberá declararlas de manera explícita en esta sección.

*Nota:* Si la abreviatura ya aparece como definición de sinónimo o está definida como término en un diccionario de tipo, no será necesario añadir su entrada aquí.

***Reglas de formato para las abreviaturas***

- Defina una abreviatura por línea.

# ***Excepciones de textos en japonés***

Aunque los textos en japonés se someten a unos procesos y una minería similares a los de otros idiomas admitidos en IBM® SPSS® Text Analytics for Surveys, hay cierto número de diferencias. Las diferencias más pequeñas se describen junto a las instrucciones del resto de idiomas en esta documentación. Sin embargo, algunas de las diferencias más grandes se cubren en este capítulo del apéndice.

## ***Extracción y categorización de textos en japonés***

Al realizar la minería en textos en japonés, el proceso es similar al de otros idiomas admitidos. [Si desea obtener más información, consulte el tema Acerca de la minería de datos de texto en el capítulo 1 el p. 3.](#) Sin embargo, hay ciertas diferencias para el japonés que se indican a continuación.

### ***Cómo funciona la extracción***

Durante la extracción de los conceptos e ideas clave de las respuestas, IBM® SPSS® Text Analytics for Surveys se basa en el análisis de texto basado en lingüística. Este método ofrece la velocidad y la rentabilidad propia de los sistemas basados en estadísticas. Pero ofrece un grado de precisión mucho mayor, y con menos intervención humana. El análisis de texto basado en lingüística se basa en el ámbito de estudio conocido como proceso de idioma natural, conocido también como lingüística computacional.

Para textos en japonés, la diferencia entre enfoques basados en estadística y basados en lingüística durante el proceso de extracción puede ilustrarse utilizando la palabra 沈む como ejemplo. Utilizando esta palabra podemos encontrar expresiones como 日が沈む, que se traduce como *el sol descende* o 気分が沈む, que se traduce como *sentirse por los suelos*. Si utiliza únicamente técnicas estadísticas, 日 (que se traduce como *sol*), 気分 (que se traduce como *sentirse*) y 沈む (que se traduce como *abajo*) se extraen por separado. Sin embargo, cuando utilizamos el verificador de datos de impresiones, que utiliza técnicas lingüísticas, no sólo se extraen 日, 気分 y 沈む, sino que 気分が沈む (que se traduce como *sentirse por los suelos*) se extrae y se asigna al tipo <悪い - 悲しみ全般>. El uso de técnicas basadas en lingüística a través del verificador de datos de impresiones permite extraer expresiones con más sentido. El análisis y la captura de emociones reducen la ambigüedad del texto, lo que convierte a la minería de texto basada en lingüística, por definición, en el método más fiable.

Comprender el funcionamiento del proceso de extracción puede ayudarle a tomar decisiones clave a la hora de ajustar los recursos lingüísticos (bibliotecas, tipos, sinónimos, etc.). Estos son los pasos del proceso de extracción:

- Conversión de datos de origen en un formato estándar
- Identificar los términos candidatos
- Identificar las clases de equivalencias y la integración de sinónimos

- Asignar un tipo
- Crear índices y, cuando se le pida, extraer patrones con un verificador de datos secundario

### **Paso 1. Conversión de datos de origen en un formato estándar**

En este primer paso, los datos que se importan se convierten en un formato uniforme que puede utilizarse para realizar otros análisis. Esta conversión se lleva a cabo internamente y no cambia los datos originales.

### **Paso 2. Identificar los términos candidatos**

Es importante comprender el rol de los recursos lingüísticos en la identificación de los términos candidatos durante la extracción lingüística. Los recursos lingüísticos se utilizan siempre que se ejecuta una extracción. Existen en forma de plantillas, bibliotecas y recursos compilados. Las bibliotecas incluyen listas de palabras, relaciones y otra información que se utiliza para especificar o ajustar la extracción. Estos recursos compilados no se pueden ver ni editar. Sin embargo, los recursos restantes sí pueden editarse en el Editor de recursos.

Los recursos compilados son los componentes internos principales del motor de extracción en SPSS Text Analytics for Surveys. Estos recursos incluyen un diccionario general que contiene una lista de formatos base con un código de categoría léxica (sustantivo, verbo, adjetivo, etc.). Los recursos también incluyen tipos incorporados reservados que se utilizan para asignar muchos términos extraídos a los tipos siguientes: <地名>, <組織> o <人名>. [Si desea obtener más información, consulte el tema Tipos disponibles para textos en japonés el p. 254.](#)

Además de los recursos compilados, se entregan varias bibliotecas con el producto y pueden utilizarse para complementar los tipos y las definiciones de conceptos en los recursos compilados, así como para ofrecer sinónimos. Estas bibliotecas—y las bibliotecas personalizadas que haya creado—se componen de varios diccionarios. Estos incluyen diccionarios de tipo, diccionarios de sinónimos y diccionarios de exclusión. [Si desea obtener más información, consulte el tema Edición de recursos para textos en japonés el p. 250.](#)

Una vez importados y convertidos los datos, el motor de extracción empezará a identificar los términos candidatos para la extracción. Los términos candidatos son palabras o grupos de palabras que se utilizan para identificar conceptos en el texto. Durante el proceso del texto, las palabras simples (**unitérminos**) y palabras compuestas (**multitérminos**) se identifican mediante extractores de patrones de categorías léxicas. Por ejemplo, el multitérmino 青森りんご, que responde al patrón de categoría léxica <地名> + <名詞>, tiene dos componentes. A continuación, las palabras clave de impresiones candidatas se identifican mediante el análisis de enlace de texto de impresiones.

Por ejemplo, supongamos que tiene el siguiente texto en japonés: 写真が新鮮で良かった. En este caso, el motor de extracción asignaría el tipo de impresión 良い - 褒め・賞賛, tras hacer coincidir (品物) + が + 良い mediante una de las reglas de enlace de texto de impresiones.

*Nota:* los términos del diccionario general compilado arriba mencionado representan una lista de todas las palabras que probablemente no resultan relevantes o que son lingüísticamente ambiguas, como los unitérminos. Estas palabras se excluyen de la extracción cuando se están identificando los unitérminos. Sin embargo, volverán a evaluarse cuando determine las categorías léxicas no cuando busque palabras compuestas candidatas más largas (multitérminos).

### **Paso 3. Identificar las clases de equivalencias y la integración de sinónimos**

Después de identificar los unitérminos y multitérminos candidatos, el software utiliza un diccionario de normalización para identificar las clases de equivalencias. Una clase de equivalencia es la forma básica de una frase o una forma simple de dos variantes de la misma frase. El propósito de asignar frases a las clases de equivalencias es asegurarse de que, por ejemplo, `efecto secundario` y 副作用 no se consideren conceptos separados. Para determinar qué concepto se va a utilizar para la clase de equivalencia (es decir, cuál de las dos opciones, `efecto secundario` o bien 副作用 se utiliza como término principal), el motor de extracción aplica las reglas siguientes en el orden que aparece:

- La forma especificada por el usuario en una biblioteca.
- La forma más frecuente, según lo definido por recursos compilados con anterioridad.

#### **Paso 4. Asignar un tipo**

A continuación, se asignan tipos a los conceptos extraídos. Un tipo es una agrupación semántica de conceptos. En este paso se utilizan tanto los recursos compilados como las bibliotecas. Los tipos incluyen elementos como conceptos de nivel más alto, palabras positivas y negativas, nombres propios, lugares, organizaciones, etc. [Si desea obtener más información, consulte el tema \*Diccionarios de tipo en el capítulo 10\* el p. 215.](#)

Los recursos del japonés tienen un conjunto de tipos distintivo. [Si desea obtener más información, consulte el tema \*Tipos disponibles para textos en japonés\* el p. 254.](#)

#### **Paso 5. Creación de índices y coincidencias de patrones con extracción de eventos**

Se crea el índice del conjunto completo de registros estableciendo un marca entre una posición de texto y el término representativo de cada clase de equivalencia. De esta manera se presupone que todos los casos de la forma declinada de un concepto candidato se indexa como forma básica candidata. Para cada forma básica se calcula la frecuencia global.

SPSS Text Analytics for Surveys puede detectar no solamente tipos y conceptos, sino también las relaciones entre ellos. Hay varios algoritmos y bibliotecas disponibles en el producto que proporcionan la capacidad de extraer patrones de relaciones de análisis de enlace de texto entre tipos y conceptos. Son especialmente útiles cuando se intentan detectar opiniones específicas (por ejemplo, reacciones ante productos).

## ***Cómo funciona la extracción secundaria***

Cuando realiza una extracción en textos en japonés, automáticamente obtiene conceptos de las palabras clave básicas y los 8 tipos básicos, incluidos 人名, 地名, 組織名, 名詞, 形容詞, 動詞, 形容動詞 y その他. Sin embargo, para aprovechar al máximo los recursos predeterminados proporcionados para textos en japonés, debe seleccionar uno de los siguientes verificadores de datos secundarios: Impresión o Dependencia.

Si selecciona un verificador de datos secundario, también podrá extraer patrones de análisis de enlace de texto y descubrir las relaciones entre los términos del texto.

**Análisis secundario.** Cuando se inicia una extracción, la extracción de palabras clave básicas tiene lugar utilizando el conjunto predeterminado de tipos. [Si desea obtener más información, consulte el tema \*Tipos disponibles para textos en japonés\* el p. 254.](#) Sin embargo, cuando selecciona un verificador de datos secundario, puede obtener muchos más conceptos o conceptos más completos dado que el extractor ahora incluirá partículas y verbos auxiliares como parte del concepto. Por

ejemplo, supongamos que tenemos la oración 肩の荷が下りた, traducida como “*Me he quitado un gran peso de encima*”. Con este ejemplo, la extracción de palabras clave básicas puede extraer cada concepto de manera separada como: 肩 (*encima*), 荷 (*peso*), 下りる (*he quitado*), pero la relación entre estas palabras no se extrae. Sin embargo, si aplica el análisis de impresiones, puede extraer conceptos más completos relacionados con un tipo de impresión como el concepto =肩の荷が下りた, que se traduce como “*quitarse un gran peso de encima*”, asignado al tipo <良い-安心>. En el caso del análisis de impresiones, también se incluye un gran número de tipos adicionales. Además, si selecciona un verificador de datos secundario, también podrá generar resultados del análisis de enlace de texto.

*Nota:* Cuando se activa un verificador de datos secundario, el proceso de extracción tarda más en completarse. [Si desea obtener más información, consulte el tema \*Cómo funciona la extracción secundaria\* el p. 247.](#)

- **Análisis de dependencias.** Si selecciona esta opción, sacará el máximo partido de las partículas extendidas para los conceptos de extracción de la extracción de tipos y palabras clave básicos. También puede obtener los resultados de patrones más completos a partir del análisis de enlace de texto (TLA) de dependencias.
- **Análisis de impresiones.** Si selecciona este verificador de datos, sacará el máximo partido de los conceptos extraídos adicionales y, cuando sea aplicable, de la extracción de resultados de patrones del TLA. Además de los tipos básicos, también puede beneficiarse de más de 80 tipos de impresiones, entre las que se incluyen 嬉しい, 吉報, 幸運, 安心, 幸福, etc. Estos tipos se utilizan para descubrir conceptos y patrones en el texto a través de la expresión de emociones, impresiones y opiniones. Hay tres opciones que dictan dónde se centra el análisis de impresiones: Todas las impresiones, Sólo la impresión representativa y Sólo conclusiones.

### ***Opciones de análisis de impresiones***

Al trabajar con textos en japonés, puede decidir extraer conceptos y tipos adicionales mediante el verificador de datos de impresiones. Este verificador de datos incluye más de 80 tipos adicionales para ayudarle a extraer opiniones, sentimientos y emociones de sus datos de texto. Asimismo, si selecciona Análisis de impresiones como verificador de datos secundario, también debe seleccionar una de las siguientes opciones, que indica al motor de extracción qué impresiones extraer:

- **Todas las impresiones**
- **Sólo la impresión representativa**
- **Sólo conclusiones**

Durante la extracción, el verificador de datos de impresiones empieza dividiendo un registro en cláusulas, cada una de las cuales contiene un predicado. Por ejemplo, el verificador de datos interpreta el texto “4月になったがまだ寒い。”, que se traduce como “*Es abril, pero sigue haciendo frío.*”, como 2 cláusulas a pesar de que sólo contiene un carácter de punto 。. A continuación, el motor de extracción examina cada cláusula para comprobar si se ajusta a la opción seleccionada.

Examinemos las tres opciones mediante el siguiente texto de muestra: “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。”. Este texto se traduce así: “*Una empleada del servicio no fue amable, pero la habitación era grande y bastante*

*satisfactoria. También estoy satisfecho con la cena.*” Durante la extracción, el texto original se divide en las siguientes cláusulas:

- 案内してくれた仲居さんは無愛想だったが、, que significa “*Una empleada del servicio no fue amable, pero*”
- 部屋は広くて申し分なかった。 , que significa “*La habitación era grande y bastante satisfactoria.*”
- 夕食も満足。 , que significa “*También estoy satisfecho con la cena.*”

### **Todas las impresiones**

Esta opción extrae todas las impresiones, opiniones y emociones que coinciden con los recursos y las reglas de enlace de texto de impresiones. Con nuestra muestra, se pudieron extraer los siguientes conceptos del texto de muestra.

Tabla A-1

Possible resultado de la muestra utilizando la opción “*Todas las impresiones*”

Concepto	Tipo
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>

*Nota:* En la tabla anterior, las filas segunda y tercera muestran cómo puede obtener el extractor dos conceptos de la misma cláusula.

### **Sólo la impresión representativa**

Esta opción sólo extrae las opiniones o emociones más representativas expresadas en cada cláusula. Si hay varias opiniones o emociones en el texto, se aplica un algoritmo. Este algoritmo intenta determinar la importancia de las impresiones encontradas, así como la posición de las palabras de una cláusula. En algunos casos en los que se encuentren dos palabras clave de impresiones con la misma importancia, se extraerá la palabra clave de impresión que esté más cerca del final de la cláusula en lugar de la que esté más al principio.

部屋は広くて, que se traduce como *la habitación era grande*, no se extrae del texto, ya que en esta cláusula 申し分なかった, que aparece en segundo lugar, se considera más importante que 部屋は広くて, que aparece en primer lugar, cuando se aplica el algoritmo interno y la posición de las palabras.

Tabla A-2

Possible resultado del texto utilizando la opción “*Sólo la impresión representativa*”

Concepto	Tipo
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>

### **Sólo conclusiones**

Esta opción obliga al extractor a identificar y extraer una palabra clave de impresión que represente la conclusión de todo el registro. No todos los textos tienen una conclusión, así que en algunos casos no se puede extraer nada de un texto determinado con esta opción. Además, cuanto mayor sea el registro, más difícil será para el verificador de datos identificar la conclusión principal. Aunque poco frecuente, todavía es posible extraer varias conclusiones.

満足, que se traduce como *satisfecho*, se considera la conclusión esencial de las impresiones expresadas en el texto.

Tabla A-3

Posible resultado del texto utilizando la opción "Sólo conclusiones"

Concepto	Tipo
満足	<満足>

### **Funcionamiento de la categorización**

Existen varias técnicas diferentes entre las que puede escoger para crear categorías. Puesto que cada conjunto de datos es exclusivo, el número de técnicas y el orden en el que las aplica puede cambiar con el tiempo. Puesto que su interpretación de los resultados puede ser diferente de la interpretación de otros, puede que deba experimentar con diferentes técnicas para comprobar con cuál de ellas se obtiene el mejor resultado para los datos del texto.

En esta guía, la **generación de categorías** hace referencia a la generación de definiciones de categoría y clasificación mediante el uso de una o más técnicas incorporadas, y **categorización** hace referencia al proceso de puntuación o etiquetaje por el que se asignan identificadores exclusivos (nombre/ID/valor) a las definiciones de categorías para cada registro.

Durante la generación de categorías, los conceptos y los tipos que se extrajeron se utilizan como los cimientos para las categorías. Cuando crea categorías, los registros se asignan automáticamente a categorías si contienen texto que coincida con un elemento de una definición de categoría.

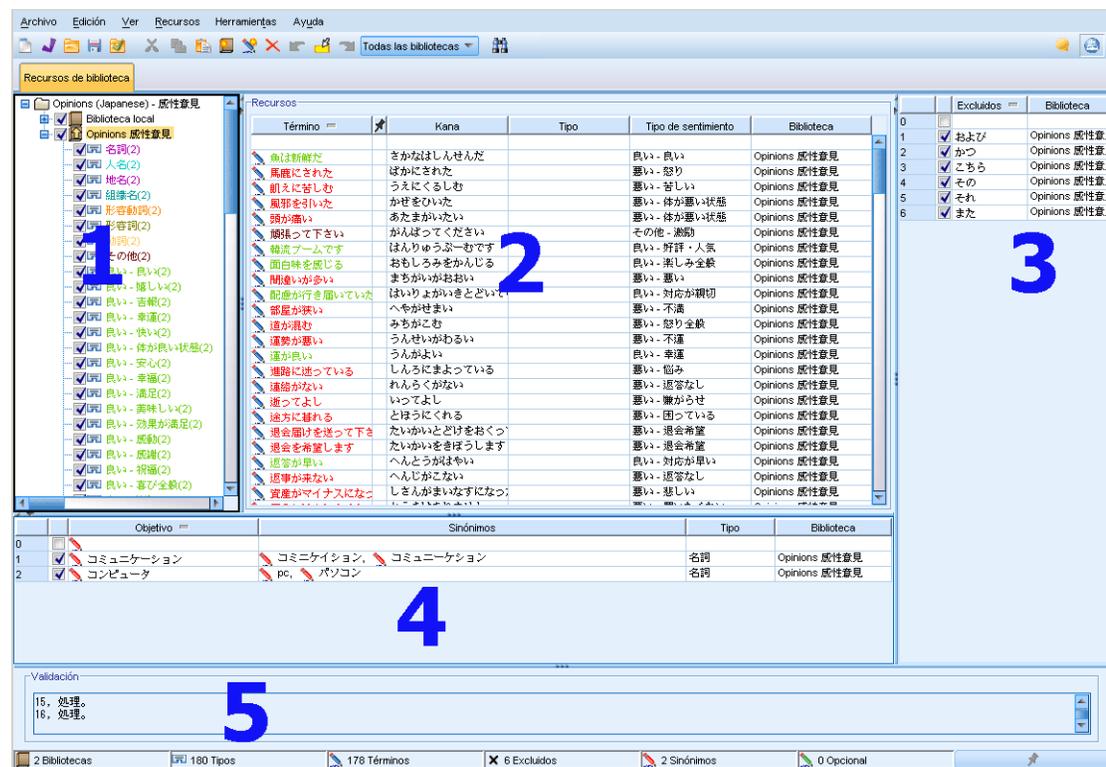
IBM® SPSS® Text Analytics for Surveys ofrece varias técnicas automáticas de generación de categorías para ayudarle a categorizar los registros rápidamente. Cada una de las técnicas resulta idónea para determinados tipos de datos y situaciones, pero a menudo conviene combinar técnicas en el mismo análisis para capturar el rango completo de registros. Puede ver un concepto en diversas categorías o detectar categorías redundantes.

### **Edición de recursos para textos en japonés**

A partir de IBM® SPSS® Text Analytics for Surveys versión 4, tiene a su disposición una nueva plantilla y un paquete de análisis de texto (TAP) para textos en japonés. Puede realizar cambios en los recursos añadiendo y editando términos para personalizarlos en sus datos. El paquete de análisis de texto también contiene un conjunto de categorías compuesto por categorías que representan impresiones positivas, negativas y contextuales/genéricas.

Puede trabajar con sus recursos en Editor de recursos. Los editores trabajan de manera similar con todos los idiomas de texto; sin embargo, hay varias diferencias significativas para los textos en japonés tal y como se describe aquí.

Figura A-1  
Editor de recursos vista para textos en japonés



Los siguientes puntos destacan algunas de las diferencias claves al trabajar con recursos para textos en japonés. Para obtener una descripción general de los cuatro paneles principales de la pestaña Recursos de la biblioteca, consulte [La interfaz del editor](#) el p. 191.

**1. Panel de bibliotecas.** Situada en la esquina superior izquierda, esta área funciona de forma muy parecida a como lo hace con otros idiomas. Sin embargo, hay varias diferencias como la imposibilidad de crear nuevos tipos o de cambiar el nombre de los tipos. [Si desea obtener más información, consulte el tema Trabajo con bibliotecas en el capítulo 9 el p. 202.](#)

**2. Panel de términos de diccionarios de tipo.** Situado a la derecha del panel del árbol de bibliotecas, este panel es bastante diferente para textos en japonés. Además de tener el nombre del término, también puede añadir el nombre en escritura kana, así como seleccionar uno o dos tipos a los que puede asociar el término. Sin embargo, no puede generar formas conjugadas de los términos o asignar opciones de coincidencia con términos en japonés a diferencia de otros idiomas distintos del japonés. [Si desea obtener más información, consulte el tema Panel de términos, de tipos y del árbol de bibliotecas para el japonés el p. 252.](#)

**3. Panel del diccionario de sustitución/sinónimos.** En los recursos para textos en japonés, encontrará la pestaña Sinónimos, en la que podrá definir todos los sinónimos para sus recursos. En la pestaña Sinónimos, hay una columna adicional denominada Tipo en la que debe designar el tipo de los sinónimos introducidos. [Si desea obtener más información, consulte el tema Uso del diccionario de sinónimos para textos en japonés el p. 259.](#) *Nota:* La pestaña Elementos opcionales no aparece porque no es aplicable a textos en japonés.

**4. Panel del diccionario de exclusión.** No hay diferencias en este panel para recursos para textos en japonés, excepto que no se admite el uso del comodín \*.

**5. Panel de validación.** Para textos en japonés, hay un panel de validación adicional utilizado para comprobar sus recursos antes de la extracción. Cuando se extrae de textos en japonés, el motor de extracción vuelve a compilar los recursos automáticamente si se detectan cambios antes de empezar el proceso de extracción. Para evitar posibles errores durante la extracción, puede volver a compilar y validar los recursos antes de la extracción de modo que pueda corregir los errores que encuentre. [Si desea obtener más información, consulte el tema Validación y compilación de recursos del japonés el p. 260.](#)

*Nota:* No hay recursos avanzados o reglas de enlace de texto editables para textos en japonés, de modo que estas pestañas no están disponibles.

### ***Panel de términos, de tipos y del árbol de bibliotecas para el japonés***

El modo en que trabaja con bibliotecas y tipos para recursos del japonés es muy parecido al de otros idiomas. [Si desea obtener más información, consulte el tema Diccionarios de tipo en el capítulo 10 el p. 215.](#)

Sin embargo, hay varias diferencias principales, entre las que se incluyen:

- Los recursos para textos en japonés tienen un conjunto de tipos distintivo. [Si desea obtener más información, consulte el tema Tipos disponibles para textos en japonés el p. 254.](#)
- Los tipos no se pueden crear ni se puede cambiar su nombre; sin embargo, sus propiedades se pueden editar. [Si desea obtener más información, consulte el tema Edición de las propiedades de tipo del japonés el p. 258.](#)
- Puede añadir y editar términos, incluida la especificación de un nombre en escritura kana para un término, así como la asignación a un tipo y un tipo de impresión secundaria. [Si desea obtener más información, consulte el tema Panel de términos, de tipos y del árbol de bibliotecas para el japonés el p. 252.](#)

El panel del árbol de bibliotecas muestra las bibliotecas y sus diccionarios de tipo. Si selecciona una biblioteca o tipo en el panel de la izquierda, un panel de términos a la derecha mostrará los términos de las bibliotecas o diccionarios de tipo seleccionados. Puede añadir términos a un diccionario de tipo directamente en el panel de términos o a través del cuadro de diálogo Añadir términos. Los términos que añade pueden ser palabras simples o compuestas. En la parte superior de la lista siempre encontrará una fila en blanco para que pueda añadir un término nuevo.

Cuando defina un término en un diccionario de tipo, se considerará un sustantivo de manera predeterminada y se asignará automáticamente al tipo <名詞>. Sin embargo, puede cambiar el tipo por otro tipo básico como <動詞>, <形容詞>, <地名>, etc. Si el motor de extracción detecta que este término tiene la misma categoría léxica que el tipo al que lo había asignado en la columna Tipo, se asignará a dicho tipo y se extraerá. También puede asignar el término a uno de los tipos de impresión en la columna Tipo de impresión. A continuación, cuando utilice el verificador de datos secundario Impresión, el texto se procesará por segunda vez para intentar encontrar términos y asignarlos a los tipos de impresión. Además, si define un tipo de impresión y un tipo básico y el motor de extracción detecta que este término coincide con ambos tipos al realizar también el análisis de impresiones secundario, el tipo de impresión tendrá preferencia y se mostrará en el

panel de resultados extraídos y en los resultados del análisis de enlace de texto. Por ejemplo, si un verbo se extrajo como un tipo de verbo <動詞> y también como un tipo positivo como “amado”, este término se mostrará como perteneciente al tipo positivo en la interfaz, ya que a menudo es más interesante captar una impresión y no una categoría léxica.

Figura A-2

Paneles de bibliotecas y de términos para recursos del japonés

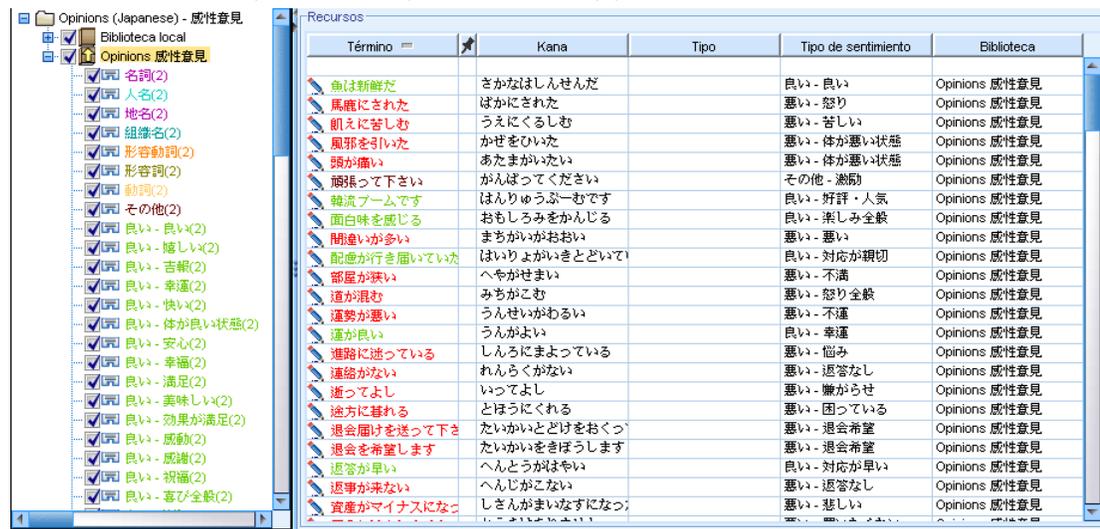


Tabla A-4

Descripciones de columnas del panel de términos

Nombre de columna	Descripción de columna
Término	Introduzca palabras simples o compuestas en la casilla. El color en el que aparece el término depende del color del tipo donde se ha almacenado o forzado el término. Puede cambiar los colores de tipo en el cuadro de diálogo Propiedades de tipo. Si desea obtener más información, consulte el tema Edición de las propiedades de tipo del japonés el p. 258. Por lo general, el término está escrito en kanji pero también puede incluir la escritura kana. <b>Importante:</b> No se admite la introducción de verbos utilizando caracteres katakana.
Forzar	Al pulsar y colocar un icono de chincheta en esta casilla, permite que el motor de extracción pase por alto el resto de las apariciones de este mismo término en otras bibliotecas. Si desea obtener más información, consulte el tema Forzado de términos en el capítulo 10 el p. 222. Esto funciona igual en todos los idiomas.
Kana	Introduzca en escritura kana el nombre del término en escritura kanji.
Tipo	Seleccione el nombre de tipo básico al que debería asignarse el término. Si desea obtener más información, consulte el tema Tipos disponibles para textos en japonés el p. 254.
Tipo de impresión	Si se va a realizar un análisis secundario, seleccione el nombre del tipo de impresión al que debería asignarse el término. Si desea obtener más información, consulte el tema Tipos disponibles para textos en japonés el p. 254.
Biblioteca	Seleccione la biblioteca en la que está almacenado su término. Puede arrastrar y soltar un término en otro tipo en el panel del árbol de bibliotecas para cambiarlo de biblioteca.

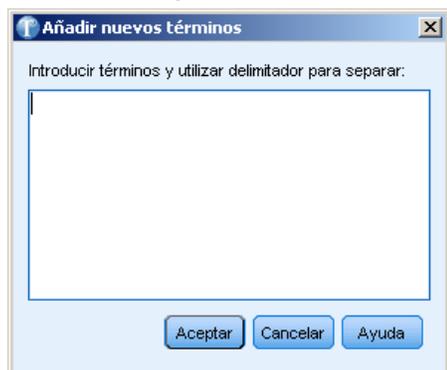
**Para añadir un único término a un diccionario de tipo**

- ▶ En el panel del árbol de bibliotecas, seleccione el diccionario de tipo al que desea añadir el término.
- ▶ En la lista de términos del panel central, escriba el término en la primera casilla disponible y defina las opciones que desee para dicho término.

**Para añadir varios términos a un diccionario de tipo**

- ▶ En el panel del árbol de bibliotecas, seleccione el diccionario de tipo al que desea añadir los términos.
- ▶ En los menús elija Herramientas > Nuevos términos. Aparecerá el cuadro de diálogo Añadir nuevos términos.

Figura A-3  
Cuadro de diálogo Añadir nuevos términos



- ▶ Especifique los términos que desee añadir al diccionario de tipo seleccionado escribiéndolos o pegando un conjunto de términos. Si especifica varios términos, deberá separarlos mediante el delimitador definido en el cuadro de diálogo Opciones, o añadir cada término en una línea nueva. [Si desea obtener más información, consulte el tema Opciones de configuración en el capítulo 2 el p. 17.](#)
- ▶ Pulse en Aceptar para añadir los términos al diccionario. El cuadro de diálogo se cierra y los nuevos términos aparecen en el diccionario.

***Tipos disponibles para textos en japonés***

No puede añadir nuevos tipos a los recursos del japonés; sin embargo, puede añadir y eliminar términos de ellos. Las siguientes tablas incluyen el conjunto de tipos del japonés disponibles actualmente.

***Tipos para la extracción básica***

Cuando se inicia una extracción, se utilizan los siguientes tipos.

Tabla A-5  
Tipos para la extracción básica

Tipos	Descripción
名詞	Palabras que hacen referencia a cosas, como “coche” y “película”. Sin embargo, los nombres personales, nombres de lugares y nombres de organizaciones tienen una categoría independiente.
人名	Nombres que se corresponden con los nombres de personas específicas, como “Tokugawa” y “Ieyasu”. Las combinaciones de nombres y apellidos, como “Tokugawa Ieyasu”, también son nombres personales.
地名	Nombres como “Tokio” y “Londres” que hacen referencia a lugares específicos.
組織名	Nombres que hacen referencia a empresas y organizaciones específicas, como “Federación de organizaciones económicas”.
形容動詞	Palabras como “tranquilo (shizuka)” que describen las características o el estado de un objeto y que pueden utilizarse en frases como “no [adjetivo] (~ de nai)” y “un objeto [adjetivo] (~ na koto)”.
形容詞	Palabras como “divertido (tanoshii)” que describen las características o el estado de un objeto y que pueden utilizarse en frases como “volverse [adjetivo] (~ku naru)” y “un objeto [adjetivo] (~i koto)”.
動詞	Palabras que describen un movimiento o acción, incluidos los verbos de tipo I (raíz consonántica), verbos de tipo II (raíz vocálica) y verbos irregulares (sagyou henkaku y kagyou henkaku).
その他	Palabras como adverbios, adjetivos antepuestos al sustantivo, conjunciones e interjecciones; algunos ejemplos incluyen “bastante”, “cualquier”, “entonces” y “gracias”.

### Tipos de análisis de impresiones

Cuando selecciona el verificador de datos secundario para la extracción de impresiones, obtiene un gran número de tipos además de los 8 tipos básicos.

Tabla A-6  
Tipos de análisis de impresiones

Tipos	Descripción
良い - 良い	Expresiones de algo generalmente positivo que puede clasificarse como “bueno”.
良い - 嬉しい	Describe un evento deseable que produce un estímulo agradable.
良い - 吉報	Describe un evento agradable que sólo puede ser posible mediante un esfuerzo considerable.
良い - 幸運	Describe un evento feliz que sólo puede ser posible mediante la casualidad o una coincidencia sorprendente.
良い - 快い	Sugiere que algo es un estímulo o entorno que produce una sensación fisiológica agradable.
良い - 体が良い状態	Describe un estado en el que el cuerpo no tiene ninguna enfermedad, lesión ni fatiga o un estado en el que el estado físico está mejorando.
良い - 安心	Sugiere que una persona está tranquila y no tiene el riesgo de sufrir daños.
良い - 幸福	Indica que una persona ha obtenido condiciones especialmente favorables o cariño mediante las acciones de dicha persona o las circunstancias de su nacimiento.
良い - 満足	Describe un evento deseable que relaja la mente.
良い - 美味しい	Indica que un alimento tiene un sabor agradable.
良い - 効果が満足	Implica que algo determinado ha producido el efecto esperado.
良い - 感動	Sugiere que la importancia, significado o valor de algo es asombrosamente bueno.

<b>Tipos</b>	<b>Descripción</b>
良い - 感謝	Sugiere que una persona reconoce las acciones de otra de manera positiva.
良い - 祝福	Expresa la opinión de que la situación de otra persona es favorable (en un grado aceptable para el orador).
良い - 喜び全般	Eventos positivos de otro tipo o eventos positivos con poca relación con el orador.
良い - 楽しい	Indica o anticipa actividades como compañerismo, diversión y entretenimiento.
良い - 可らしい	Denota que algo tiene una cualidad humorística que proporciona un estímulo agradable.
良い - 笑い	Expresa una sonrisa o risa provocada por algo bueno y/o humorístico.
良い - 期待	Predice que un evento bueno ocurrirá en el futuro.
良い - 楽しみ全般	Eventos agradables de otro tipo y/o actividades/comportamientos positivos con poca relación con el orador.
良い - 金額への賞賛	Implica que, desde el punto de vista del comprador, algo tiene un valor monetario deseable.
良い - 対応が早い	Sugiere que un servicio se ha proporcionado o completado de manera oportuna.
良い - 対応が親切	Sugiere que la actitud o comportamiento del proveedor de un servicio fue solícito.
良い - 説明が良い	Expresa la idea de que el tipo y/o cantidad de información y/o el método de suministro son adecuados.
良い - 対応への賞賛	Opiniones, distintas de las anteriores, que elogian al proveedor de un servicio.
良い - 褒め・賞賛	Opiniones, distintas de las anteriores, que elogian las características, capacidades y/o funcionamiento de algo determinado.
良い - 好き	Expresa el deseo de poseer o acercarse a algo determinado.
良い - 入会希望	Describe el deseo de pertenecer o seguir formando parte de un grupo determinado.
良い - 買いたい	Implica que una persona desea o tiene la intención de utilizar dinero para obtener algo determinado.
良い - 好評・人気	Indica que el número de personas que desean o aprecian algo determinado ha superado un objetivo determinado.
良い - 売れた	Indica la presencia de personas que adquieren algo determinado o que el número o valor de las compras ha superado un objetivo determinado.
悪い - 悪い	Expresiones de algo generalmente negativo que puede clasificarse como “malo”.
悪い - 怒り	Sensación distintiva de enfado que aparece cuando algo no sucede como se había planificado.
悪い - 批判	Expresa la idea de que otra persona no ha tomado la decisión adecuada.
悪い - お叱り	Palabras o acciones que intimidan a otra persona para ajustarse a las intenciones de una persona.
悪い - 誹謗・中傷	Palabras utilizadas para demostrar una opinión excesivamente pobre de otra persona.
悪い - 軽蔑	Implica que hay grandes carencias en el carácter, capacidades y/u otras cualidades de otra persona.
悪い - 恨み	Expresa una represalia o resentimiento por un inconveniente provocado por otra persona.
悪い - 嫌がらせ	Palabras utilizadas para inhibir la comunicación.
悪い - 不満	Sentimiento desagradable provocado por la incapacidad de obtener un estado o algo deseado.
悪い - 不味い	Indica que un alimento tiene un mal sabor.
悪い - 効果が不満	Implica que algo no ha producido el efecto esperado.
悪い - 金額が不満	Implica que, desde el punto de vista del comprador, el valor monetario de algo determinado no es deseable.

<b>Tipos</b>	<b>Descripción</b>
悪い - 対応への不満	Sugiere que el proveedor de un servicio es el responsable.
悪い - 対応が遅い	Implica que un servicio no se ha realizado/completado de manera oportuna o que el servicio todavía debe completarse.
悪い - 対応が不親切	Denota un sentimiento desagradable provocado por la actitud o comportamiento del proveedor de un servicio.
悪い - 説明が悪い	Expresa la idea de que el tipo y/o cantidad de información y/o el método de suministro no son adecuados.
悪い - 返答なし	Denota que el proveedor de un servicio no proporciona una respuesta adecuada, aunque la situación lo exija.
悪い - 不快	Sugiere que algo es un estímulo o entorno que produce una sensación fisiológica negativa.
悪い - 怒り全般	Sentimientos de enfado distintos de los anteriores. Enfado general experimentado por la organización o empresa del orador o descripciones de eventos provocados por dicho enfado.
悪い - 悲しい	Sentimiento desagradable distintivo experimentado cuando una persona pierde o no puede obtener algo.
悪い - 凶報	Expresa la idea de que no se ha podido lograr un objetivo determinado a pesar de un esfuerzo considerable.
悪い - 不運	Indica un resultado negativo provocado por una desafortunada coincidencia y/o mala suerte, no por culpa de la propia persona.
悪い - ショック	Sugiere que una persona está disgustada por un evento o algo negativo imprevisto y no es capaz de encontrar una respuesta adecuada.
悪い - 残念	Sentimiento de descontento experimentado cuando algo que se había anticipado no sucede.
悪い - 落胆	Estado en el que una persona se ve invadida por un sentimiento de descontento y desilusión.
悪い - 諦め	Sugiere que algo negativo, experimentado por el orador u otra persona, no puede mejorarse.
悪い - 後悔	Expresa la idea de que en el pasado una persona no tomó la decisión adecuada, aunque estaba disponible como opción.
悪い - 謝罪	Indica el reconocimiento por parte del orador de haber hecho daño a otra persona.
悪い - 淋しい	Expresa la idea de que el contacto con otras personas es escaso o que las personas con las que puede entrar en contacto son pocas.
悪い - 哀れみ	Expresa la idea de que la situación de otra persona es significativamente peor que la del orador.
悪い - 悩み	Indica que una persona debe tomar una decisión pero no es capaz de elegir entre las opciones disponibles.
悪い - 困っている	Expresa la idea de que no hay ningún modo eficaz de responder a una situación que exige actuar.
悪い - 苦しい	Expresa un estado psicológico desagradable en el que una persona no puede actuar con normalidad debido a motivos externos o a los errores o equivocaciones de dicha persona.
悪い - 体が悪い状態	Describe un estado en el que el cuerpo tiene alguna enfermedad, lesión y/o fatiga o un estado en el que el estado físico no está mejorando.
悪い - 不安	Expresa la idea de que algo puede no continuar en su estado deseado o no satisfacer las expectativas.
悪い - 恐怖	Sugiere que parece probable que algo determinado provoque daños o lesiones.
悪い - 悲しみ全般	Sentimientos de tristeza distintos de los anteriores, como una tristeza general sobre algo no específico.
悪い - 嫌い	Indica que una persona quiere mantener algo alejado o alejarse de algo.

Tipos	Descripción
悪い - 退会希望	Describe el deseo de abandonar o dejar de formar parte de un grupo determinado.
悪い - 買いたくない	Sugiere que una persona no quiere algo determinado o no tiene la intención de pagar por ello.
悪い - 不評・不人気	Indica que el número de personas a las que les gusta algo determinado no ha alcanzado un objetivo determinado o que hay demasiadas personas con sentimientos negativos hacia ello.
悪い - 売れていない	Indica la ausencia de personas que adquieran algo determinado o que el número o valor de las compras no ha alcanzado un objetivo determinado.
その他 - 疑問	Expresiones que exigen información que requiere un examen o consideración mayor de la otra persona.
その他 - 問い合わせ	Expresiones que exigen información que ya está en manos de la otra persona.
その他 - 要望	Expresiones que ordenan a la otra persona (cuando la otra persona es el responsable directo o tiene un rango inferior que el orador) que solucione un problema.
その他 - 提案・忠告	Expresiones que ordenan a la otra persona (cuando la otra persona es el responsable directo o tiene un rango inferior que el orador) que se comporte mejor.
その他 - お願い	Expresiones que ordenan a la otra persona (cuando la otra persona no es el responsable o no tiene un rango inferior que el orador) que haga algo.
その他 - 激励	Expresiones que animan a otra persona o descripciones de comportamiento alentador.
その他 - 勧誘	Expresiones que ordenan a otra persona que haga algo junto con el orador.
その他 - 驚き	Expresa la idea de que lo repentino o la escala de un evento trascienden el juicio/la comprensión racional.
評価なし - 評価なし	Ninguna expresión de evaluación.

## Edición de las propiedades de tipo del japonés

Aunque no puede crear tipos en recursos del japonés, puede ver y editar propiedades de tipo. Tenga en cuenta que las opciones como la opción de coincidencia y las formas declinadas no se aplican a textos en japonés.

Figura A-4

Cuadro de diálogo Propiedades de tipo para recursos para textos en japonés



**Nombre.** Nombre del diccionario de tipo.

**Añadir a.** Este campo indica la biblioteca donde creará el nuevo diccionario de tipo.

**Color de fuente.** Este campo permite distinguir los resultados de este tipo del resto de resultados de la interfaz. Si selecciona Usar color principal, también se utilizará el color de tipo por defecto para este diccionario de tipo. Este color por defecto se establece en el cuadro de diálogo Opciones. [Si desea obtener más información, consulte el tema Opciones: Pestaña Mostrar en el capítulo 2 el p. 19.](#) Si selecciona Personalizado, seleccione un color de la lista desplegable.

**Anotación.** Este campo es opcional y puede utilizarse para introducir comentarios o descripciones.

#### **Para ver o editar propiedades de tipo**

- ▶ Seleccione el tipo cuyas propiedades quiera ver.
- ▶ Pulse el botón derecho del ratón y elija Propiedades de tipo en el menú contextual. Aparecerá el cuadro de diálogo Propiedades de tipo.
- ▶ Haga los cambios necesarios.
- ▶ Pulse en Aceptar para guardar los cambios en el diccionario de tipo.

## **Uso del diccionario de sinónimos para textos en japonés**

Para textos en japonés, el diccionario de sustitución sólo contiene una pestaña para gestionar sus sinónimos: la pestaña Sinónimos. Los sinónimos asocian dos o más palabras con el mismo significado. Los sinónimos también pueden utilizarse para agrupar términos con sus abreviaturas, o para agrupar palabras que suelen escribirse mal con la ortografía correcta.

Figura A-5

Entradas de sinónimos para textos en japonés

	Objetivo	Sinónimos	Tipo	Biblioteca
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

Una definición de sinónimo se compone de dos partes. El término objetivo es el término bajo el cual desea que el motor de extracción agrupe todos los términos sinónimos. A menos que utilice este término objetivo como sinónimo de otro término objetivo o que se excluya, es probable que se convierta en el concepto que aparece en el panel Resultados extraídos. La lista de sinónimos son los términos que se agruparán bajo el término objetivo.

En la pestaña Sinónimos, puede especificar una definición de sinónimo en la línea vacía de la parte superior de la tabla. Empiece definiendo el término objetivo y sus sinónimos. También puede seleccionar la biblioteca en la que desea que se guarde esta definición. Durante la extracción, todas las apariciones de los sinónimos se agruparán bajo el término objetivo de la extracción final. [Si desea obtener más información, consulte el tema Adición de términos en el capítulo 10 el p. 219.](#)

Cuando esté creando sus diccionarios de tipo, puede especificar un término y también pensar en tres o cuatro sinónimos del mismo. En ese caso, puede escribir todos los términos y luego el término objetivo en el diccionario de sustitución, y a continuación arrastrar los sinónimos.

**Importante:** Los comodines y los caracteres especiales no son compatibles con los sinónimos para textos en japonés.

#### **Para añadir una entrada de sinónimo**

- ▶ En la línea vacía de la parte superior de la tabla en la pestaña Sinónimos del panel de sustitución, escriba el término objetivo en la columna Objetivos. El término objetivo que ha escrito aparece en color. Este color representa el tipo en el que el término aparece o se fuerza, si se da el caso. Si el término aparece en negro, significa que no está en ningún diccionario de tipo.
- ▶ Pulse en la segunda casilla a la derecha del objetivo y escriba el conjunto de sinónimos. Separe cada entrada utilizando el delimitador global tal como está definido en el cuadro de diálogo Opciones. Todos los sinónimos introducidos deberían tener el mismo tipo. [Si desea obtener más información, consulte el tema Opciones de configuración en el capítulo 2 el p. 17.](#) Los términos que especifique aparecerán en color. Este color representa el tipo en el que aparece el término. Si el término aparece en negro, significa que no está en ningún diccionario de tipo.
- ▶ En la tercera columna, la columna Tipo, designe un tipo para estos sinónimos. El objetivo, sin embargo, toma el tipo asignado durante la extracción. Sin embargo, si el objetivo no se ha extraído como un concepto, entonces el tipo que aparece en esta columna se asigna al objetivo en los resultados de extracción.
- ▶ Pulse en la última casilla para seleccionar la biblioteca en la que desea almacenar esta definición de sinónimo.

*Nota:* estas instrucciones indican cómo realizar cambios en la Editor de recursos vista. Tenga en cuenta que este tipo de ajuste puede hacerlo directamente en el panel Resultados extraídos o en el panel Datos. [Si desea obtener más información, consulte el tema Refinamiento de los resultados de la extracción en el capítulo 5 el p. 86.](#)

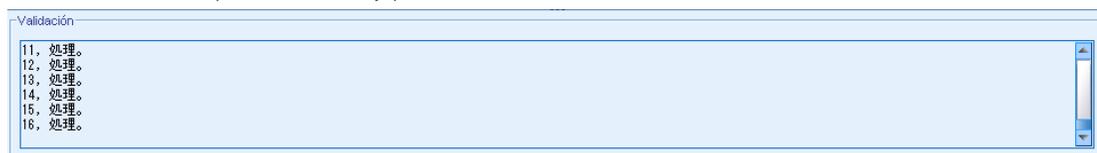
## **Validación y compilación de recursos del japonés**

Para textos en japonés, hay un panel de validación adicional utilizado para comprobar sus recursos antes de la extracción. Antes de que comience el proceso de extracción para textos en japonés, el motor de extracción vuelve a compilar los recursos automáticamente si se detectan cambios antes de empezar el proceso de extracción. Si se encuentra un error durante la extracción, puede que el proceso no se complete correctamente.

Para evitar errores de compilación, recomendamos que valide y compile sus recursos después de realizar cambios en Editor de recursos. Si aparece algún mensaje de error, puede realizar correcciones e intentar validarlo de nuevo.

**Figura A-6**

*Panel de validación para textos en japonés*



**Para validar recursos**

- ▶ En los menús elija Herramientas > Validar recursos. Se abrirá el panel de validación para mostrar los mensajes de error y compilación.

**Otras excepciones para el japonés****Recursos internos que anulan recursos definidos por el usuario**

Para textos en japonés, los recursos predeterminados incluyen algunos recursos básicos internos y compilados con anterioridad. Estos recursos internos no son editables. Por este motivo, puede utilizar Editor de recursos para realizar cambios y ajustes. En casi todos los casos, los términos, sinónimos y entradas en la lista de exclusión que defina en sus recursos tendrán preferencia sobre los recursos internos compilados con anterioridad. Sin embargo, hay varias excepciones como se indica en algunos de los siguientes ejemplos.

- Hay casos en los que añadir términos a un tipo específico no tiene ningún efecto en los resultados de extracción. Es más probable que ocurra esto cuando los datos contengan oraciones largas que incluyan varios elementos morfológicos, puntuación diversa o símbolos. Además, como los recursos para textos en japonés ya contienen un gran número de términos comunes compilados con anterioridad, hay varias palabras comunes para las que siempre se forzará una definición lingüística específica.
- Puede que no sea capaz de excluir términos como ある, いる o なる debido a que el motor de extracción siempre forzará la extracción de estos términos.
- Aunque es posible cambiar el tipo del término 東京 de <地名> a <名詞>, el motor de extracción ignorará su cambio si intenta cambiar el tipo de un término de <地名> a <動詞> o a <形容詞> mediante el diccionario (de tipo) de palabras clave.
- Puede haber casos en los que los cambios que haga en Editor de recursos o afecten a los resultados de extracción de una oración y no de otra, ya que el proceso de extracción termina haciendo referencia a las palabras con co-ocurrencia en cada oración.

**Problema de visualización de caracteres katakana de ancho medio**

Los caracteres katakana de ancho medio se convierten internamente en caracteres katakana de ancho completo durante la extracción pero siguen apareciendo como caracteres katakana de ancho medio cuando se muestran en el panel Datos que se encuentra en la interfaz de usuario. Tenga en cuenta que los caracteres katakana de ancho medio no pueden resaltarse en el panel Datos. Para evitar este problema, convierta todos sus registros a caracteres katakana de ancho completo antes de procesarlos.

**Uso de caracteres en mayúsculas y minúsculas**

Los caracteres alfabéticos en mayúsculas se convierten temporalmente en caracteres alfabéticos en minúsculas cuando los lee la aplicación. Sin embargo, el panel Datos mostrará el texto utilizando las mismas mayúsculas y minúsculas que el texto original. Los caracteres en mayúsculas y minúsculas se tratan de igual forma en este producto.

# Avisos

Esta información se desarrolló para productos y servicios ofertados en todo el mundo.

IBM puede no ofertar los productos, servicios o funciones mencionadas en este documento en otros países. Consulte con su representante local de IBM la información de los productos y servicios disponibles actualmente en su zona. Toda referencia a un producto, programa o servicio de IBM no pretende afirmar ni sugerir que haya que usar ese producto, programa o servicio de IBM. Se puede utilizar en su lugar cualquier producto, programa o servicio equivalente funcionalmente que no infrinja ningún derecho de propiedad intelectual de IBM. Sin embargo, queda bajo la responsabilidad del usuario evaluar y verificar el funcionamiento de cualquier producto, programa o servicio que no sea de IBM.

IBM puede tener patentes o aplicaciones con patente en trámite que cubran la materia descrita en este documento. La recepción de este documento no le garantiza licencia alguna sobre estas patentes. Puede enviar sus consultas de licencia, por escrito, a:

*IBM Director of Licensing, IBM Corporation, North Castle Drive, Armonk, NY 10504-1785, EE.UU.*

Para consultas de licencia relativas a la información de conjuntos de caracteres de doble byte (DBCS), póngase en contacto con el departamento de propiedad intelectual de IBM de su país o envíe su consulta, por escrito, a:

*Intellectual Property Licensing, Legal and Intellectual Property Law, IBM Japan Ltd., 1623-14, Shimotsuruma, Yamato-shi, Kanagawa 242-8502 (Japón).*

**El siguiente párrafo no afecta a Reino Unido ni a ningún otro país donde las provisiones mencionadas sean incompatibles con la legislación local:** INTERNATIONAL BUSINESS MACHINES PROVEE ESTA PUBLICACIÓN “TAL CUAL” SIN GARANTÍAS DE NINGUNA CLASE, EXPRESAS O IMPLÍCITAS, INCLUYENDO, PERO NO QUEDANDO LIMITADAS POR ESTAS, LAS GARANTÍAS IMPLÍCITAS DE NO VIOLACIÓN, COMERCIALIZACIÓN Y CONFORMIDAD PARA UN PROPÓSITO PARTICULAR. Algunos estados no permiten descargos de responsabilidad de garantías expresas o implícitas en algunas transacciones. Por tanto, puede que esta declaración no le afecte.

Esta información puede incluir imprecisiones técnicas o errores tipográficos. Se realizan modificaciones periódicas de la información aquí contenida; dichos cambios se incorporarán a las nuevas ediciones de la publicación. IBM puede realizar sin aviso mejoras y/o cambios en los productos y programas descritos en esta publicación.

Toda referencia en esta información de sitios Web que no pertenecen a IBM se proporciona solo por comodidad y de ninguna manera sirve como promoción de los citados sitios Web. Los materiales en esos sitios Web no forman parte de los materiales de este producto de IBM y el uso de esos sitios Web queda bajo su propia responsabilidad.

IBM puede usar o distribuir cualquier información proporcionada por usted del modo que considere apropiado sin incurrir en obligación alguna por su parte.

Los titulares de licencia de este programa que deseen obtener información del mismo para activar: (i) el intercambio de información entre programas creados de forma independiente y otros programas (incluido este) y (ii) el uso mutuo de la información intercambiada, deben ponerse en contacto con:

*IBM Software Group, Attention: Licensing, 233 S. Wacker Dr., Chicago, IL 60606, EE.UU..*

Esta información puede estar disponible, sujeta a los términos y condiciones apropiados, incluyendo en algunos casos el pago de una cantidad.

IBM provee el programa con licencia descrito en este documento y todo el material con licencia disponible para el mismo bajo los términos del Contrato IBM con clientes, el Acuerdo internacional de programas bajo licencia de IBM o cualquier otro contrato equivalente entre nosotros.

Los datos de funcionamiento contenidos fueron determinados en un entorno controlado. Por tanto, los resultados obtenidos en otros entornos de funcionamiento pueden variar de forma significativa. Algunas mediciones se han llevado a cabo en sistemas en fase de desarrollo y no existen garantías de que esas mediciones se correspondan con las de sistemas disponibles de forma global. Además, algunas medidas se han estimado por extrapolación. Los resultados reales pueden diferir. Los usuarios de este documento deberían verificar los datos aplicables a su entorno específico.

La información relativa a productos distintos a los de IBM se obtuvo de los proveedores de esos productos, sus anuncios publicados o cualquier otra fuente a disposición pública. IBM no ha comprobado esos productos y no puede confirmar la precisión de su funcionamiento, su compatibilidad o cualquier otra cuestión relacionada con productos que no son de IBM. Las preguntas relativas a la capacidad de productos que no son de IBM deberán ser remitidas a los proveedores de dichos productos.

Esta información contiene ejemplos de datos e informes usados en operaciones comerciales diarias. Para ilustrarlos lo mejor posible, los ejemplos incluyen nombres de personas, empresas, marcas y productos. Todos estos nombres son ficticios y cualquier parecido con los nombres y direcciones de una empresa comercial existente es mera coincidencia.

Si está viendo esta información en una copia electrónica, puede que no aparezcan las fotografías ni las ilustraciones en color.

### ***Marcas comerciales***

IBM, el logotipo de IBM, [ibm.com](http://www.ibm.com) y SPSS son marcas comerciales de IBM Corporation, registradas en muchas jurisdicciones de todo el mundo. Existe una lista actualizada de marcas comerciales de IBM en Internet en <http://www.ibm.com/legal/copytrade.shtml>.

Adobe, el logotipo Adobe, PostScript y el logotipo PostScript son marcas registradas o marcas comerciales de Adobe Systems Incorporated en Estados Unidos y/u otros países.

Intel, el logotipo de Intel, Intel Inside, el logotipo de Intel Inside, Intel Centrino, el logotipo de Intel Centrino, Celeron, Intel Xeon, Intel SpeedStep, Itanium y Pentium son marcas comerciales o marcas registradas de Intel Corporation o sus filiales en Estados Unidos y otros países.

Linux es una marca registrada de Linus Torvalds en Estados Unidos, otros países o ambos.

Microsoft, Windows, Windows NT, y el logotipo de Windows son marcas comerciales de Microsoft Corporation en Estados Unidos, otros países o ambos.

UNIX es una marca registrada de The Open Group en Estados Unidos y otros países.

Java y todas las marcas comerciales y los logotipos basados en Java son marcas comerciales de Sun Microsystems, Inc. en Estados Unidos, otros países o ambos.

Capturas de pantalla de productos de Microsoft reimpresas con permiso de Microsoft Corporation.

Otros productos y nombres de servicio pueden ser marcas comerciales de IBM u otras empresas.

- ! símbolos ^ \* \$ en sinónimos, 228
  
- abreviaturas, 242–243
- abrir proyectos, 47
- activación de entidades no lingüísticas, 241
- actualización
  - bibliotecas, 210, 212
  - gráficos, 167
  - plantillas, 193
- actualización de gráficos, 167
- adaptación
  - errores de puntuación, 84
  - errores ortográficos, 84
- adición
  - bibliotecas públicas, 204
  - conceptos en categorías, 157
  - descriptores, 104
  - elementos opcionales, 229
  - sinónimos, 87, 227
  - sinónimos para el japonés, 259
  - sonidos, 19, 21
  - términos en diccionarios de tipo, 219
  - términos en diccionarios de tipo del japonés, 252
  - términos para la lista de exclusión, 232
  - tipos, 89
- administración
  - bibliotecas locales, 206
  - bibliotecas públicas, 208
  - categorías, 156
- agrupación difusa (excepciones), 84, 233, 236
- ajuste de columna, 19
- ajustes -automáticos-, 173
- almacenamiento
  - guardar proyectos automáticamente, 18
  - proyectos, 51
  - recursos, 197
  - recursos como plantillas, 193
  - resultados de la extracción, 85
- aminoácidos (entidad no lingüística), 237
- ampliación de categorías, 125
- análisis de dependencias, 247–248
- análisis de impresiones, 247–248
  - opciones, 248
- análisis de texto, 3, 7–8, 11
- análisis secundario
  - análisis de dependencias, 247
  - análisis de impresiones, 247
- AND, operador de regla, 154
- anotaciones
  - para categorías, 108, 156
  - para proyectos, 48
- antienlaces, 117
- Archivos Excel *.xls/.xlsx*, 53, 56
- Archivos Excel Microsoft Excel *.xls/.xlsx*
  - importación de categorías predefinidas, 131
- archivos recuperados, 18
- arrastrar y soltar, 131
- asignación de nombres
  - bibliotecas, 206
  - categorías, 108, 156
  - diccionarios de tipo, 223
- asignar marcas, 74
- asterisco (\*)
  - diccionario de exclusión, 232
  - sinónimos, 228
- avisos legales, 262
  
- barras de estado, 75
- biblioteca Budget, 216
- bibliotecas, 15, 202, 215
  - actualización, 212
  - adición, 204
  - asignación de nombres, 206
  - aviso de sincronización de bibliotecas, 210
  - biblioteca Budget, 216
  - bibliotecas enviadas por defecto, 202
  - bibliotecas locales, 210
  - bibliotecas públicas, 210
  - cambio de nombre, 206
  - compartimiento y publicación, 210
  - compartir, 74
  - Core library, 216
  - creación, 203
  - desactivación, 207
  - diccionarios, 202
  - eliminación, 207, 209
  - enlace, 204
  - exportación, 209
  - importación, 208
  - Opinions library, 216
  - publicar, 51, 211
  - sincronización, 210
  - vista, 206
- bibliotecas enviadas (por defecto), 202
- bibliotecas por defecto, 202
- botón mostrar, 96
- botón recuento, 96
- buscar y reemplazar (recursos avanzados), 234–235
- búsqueda de términos y tipos, 205
  
- cambio
  - origen de datos, 61
  - plantillas, 194
- cambio de nombre
  - bibliotecas, 206
  - categorías, 130, 156
  - diccionarios de tipo, 223
  - plantillas de recursos, 195
  - proyectos, 51

- características nuevas, 1
- categorías, 26, 93–94, 107, 156
  - adición a, 157
  - ampliación, 118, 125
  - anotaciones, 108, 156
  - cambio de nombre, 130
  - copia, 164
  - creación, 101, 123, 131
  - creación de categorías vacías nuevas, 130
  - creación manual, 129
  - descriptores, 102, 104, 108
  - desplazamiento, 159
  - edición, 156, 158
  - eliminación, 166
  - estrategias, 101
  - etiquetas, 108, 156
  - exportación, 53
  - forzado de palabras, 162
  - forzado de respuestas, 161
  - fusión, 160
  - generación, 6, 110, 113, 115, 118, 127
  - gráfico de malla, 167
  - gráficos similares, 167
  - impresión, 164
  - nombres, 108, 156
  - paquetes de análisis de texto, 41, 43
  - propiedades, 108, 156
  - recuento, 96
  - refinamiento de los resultados, 9, 156
  - relevancia, 98–99
- categorías léxicas, 242–243
- categorías nuevas, 130
- categorías predefinidas, 131–132, 142
  - formato compacto, 138
  - formato con sangrado, 140
  - formato de lista plana, 137
- categorización, 6, 93, 250
  - derivación de raíz de conceptos, 113, 118
  - inclusión de conceptos, 113, 118, 120
  - manual, 129
  - métodos, 101
  - redes semánticas, 113, 118, 121
  - reglas de co-ocurrencia, 113, 118, 122
  - técnicas de frecuencia, 123
  - técnicas lingüísticas, 110, 125
  - uso de técnicas de agrupación, 113
  - utilización de técnicas, 118
- clasificación natural, 50
- coincidencia de texto, 108, 156, 162
- color de fuente, 218, 259
- colores
  - diccionario de exclusión, 232
  - establecimiento de opciones de color, 19
  - para barras de gráficos de resumen, 60
  - para tipos y términos, 218, 259
  - sinónimos, 229
- colores personalizados, 19
- columna de documentos, 95–96
- combinación de categorías, 160
- compartimiento de bibliotecas, 210
  - actualización, 212
  - adición de bibliotecas públicas, 204
  - publicar, 51, 211
- compartir proyectos, 74
- configuración, 17, 19, 21
- conglomerado, 186
- coordenadas polares, 182
- copia
  - categorías, 164
- copia de seguridad de recursos, 197
- copia de visualizaciones, 187
- Core library, 216
- correo electrónico (entidad no lingüística), 237
- creación
  - bibliotecas, 203
  - categorías, 101, 110, 131
  - categorías con reglas, 146
  - diccionarios de tipo, 217, 258
  - elementos opcionales, 229
  - entradas del diccionario de exclusión, 232
  - plantilla a partir de los recursos, 193
  - proyectos, 28
  - reglas de categoría, 145–146, 153
  - sinónimos, 86–87, 227
  - sinónimos para el japonés, 259
  - tipos, 89
- creación de plantillas a partir de recursos, 193
- datos
  - actualización, 72
  - archivos IBM SPSS Statistics *.sav*, 53
  - cambiar el origen de datos, 61
  - categorización, 93, 110, 129
  - editar propiedades de variables, 50
  - exportación, 53
  - extracción, 82
  - generación de categoría, 6, 113, 115, 118, 125
  - IBM SPSS Data Collection, 34, 54, 67
  - IBM SPSS Statistics archivos *.sav*, 30, 54, 63
  - Microsoft Excel Archivos Excel *.xls/.xlsx*, 31, 53, 56, 64
  - ODBC, 33, 66
  - ordenación, 50
  - refinamiento de los resultados, 86
  - selección de los orígenes de datos, 29, 62
  - vista, 49
- datos de encuesta, 2, 26–27, 34, 67
- definiciones, 102, 108
- definiciones forzadas, 242–243
- delimitador, 18
- delimitador global, 18
- desactivación
  - barras de estado, 75
  - bibliotecas, 207
  - diccionarios de exclusión, 232

- diccionarios de sinónimos, 236
- diccionarios de sustitución, 230
- diccionarios de tipo, 225
- entidades no lingüísticas, 241
- desactivación de entidades no lingüísticas, 241
- descriptores, 95
  - categorías, 102, 108
  - edición en categorías, 158
  - selección de los mejores, 104
- desplazamiento
  - categorías, 159
  - diccionarios de tipo, 224
- diccionario de exclusión, 202, 231–232
- diccionario de sustitución, 202, 225, 229–230
- diccionario de tipo, 202
  - adición de términos, 219
  - adición de términos para el japonés, 252
  - cambio de nombre, 223
  - creación de tipos, 217, 258
  - desactivación, 225
  - desplazamiento, 224
  - elementos opcionales, 215
  - eliminación, 225
  - forzado de términos, 222
  - sinónimos, 215
  - tipos incorporados, 216
- diccionario de tipo Budget, 216
- diccionario de tipo Location, 216
- diccionario de tipo Negative, 216
- diccionario de tipo Organization, 216
- diccionario de tipo Person, 216
- diccionario de tipo Positive, 216
- diccionario de tipo Product, 216
- diccionario de tipo Uncertain, 216
- diccionario de tipo Unknown, 216
- diccionarios, 15, 215
  - excluidos, 202, 215, 231
  - sustituciones, 202, 215, 225
  - tipos, 202, 215
- dígitos (entidad no lingüística), 237
- direcciones (entidad no lingüística), 237
- direcciones IP (entidad no lingüística), 237
- distancia máxima de búsqueda, 116, 122, 128
  
- edición
  - categorías, 156, 158
  - propiedades, 156
  - refinado de los resultados de la extracción, 86
  - reglas de categoría, 155
- edición de gráficos
  - tamaño de los elementos gráficos, 176
- edición de visualizaciones, 171
  - adición de efectos 3-D, 182
  - categorías, 180
  - colores y tramas, 174
  - combinación de categorías, 180
  - configuración automática, 172
  - contracción de categorías, 180
  - ejes, 178
  - escalas, 178
  - estilo de guión, 174
  - exclusión de categorías, 180
  - forma de puntos, 175
  - formatos de numeración, 177
  - márgenes, 177
  - ordenación de categorías, 180
  - paneles, 182
  - posición de la leyenda, 187
  - reglas, 172
  - relación de aspecto de puntos, 175
  - relleno, 177
  - rotación de puntos, 175
  - texto, 173
  - transformación de los sistemas de coordenadas, 182
  - transparencia, 174
  - transponer, 182
  - única, 172
- editor de recursos, 11, 193–194, 233
  - actualización de plantillas, 193
  - cambio de recursos, 194
  - creación de plantillas, 193
  - opción del delimitador global, 18
  - para el japonés, 250
- elementos gráficos
  - cambio, 183
  - conversión, 183
  - modificadores de colisión, 186
  - tipos, 184
- elementos opcionales, 225
  - adición, 229
  - definición de, 227
  - eliminación de entradas, 230
  - objetivo, 229
- eliminación
  - bibliotecas, 207, 209
  - categorías, 166
  - desactivación de bibliotecas, 207
  - diccionarios de tipo, 225
  - elementos opcionales, 230
  - entradas excluidas, 232
  - plantillas de recursos, 195
  - reglas de categoría, 155
  - sinónimos, 230
- entidades no lingüísticas, 85
  - activación y desactivación, 241
  - aminoácidos, 237
  - dígitos, 237
  - direcciones, 237
  - direcciones de correo electrónico, 237
  - direcciones de HTTP/URL, 237
  - direcciones IP, 237
  - expresiones regulares, *RegExp.ini*, 238
  - fechas, 237
  - horas, 237

- monedas, 237
- normalización, *NonLingNorm.ini*, 240
- números de la seguridad social, 237
- números de teléfono, 237
- pesos y medidas, 237
- porcentajes, 237
- proteínas, 237
- errores de puntuación, 84
- errores ortográficos, 27, 84, 236
- esquivar, 186
- estadísticos
  - descripciones, 184
  - edición de visualizaciones, 184
- estructuración en componentes, 119
- estructuración en componentes de los términos, 119
- etiquetas para las categorías, 108, 156
- excepciones de enlace, 117
- exclusión
  - conceptos de la extracción, 91
  - de enlaces de categoría, 117
  - de exclusión difusa, 236
  - desactivación de bibliotecas, 207
  - desactivación de diccionarios, 225, 230
  - desactivación de entradas de exclusión, 232
- exportación, 53
  - bibliotecas públicas, 209
  - categorías predefinidas, 142
  - como archivos IBM SPSS Statistics *.sav*, 54
  - como archivos Microsoft Excel *.xls/.xlsx*, 56
  - formatos de resultados, 53
  - gráficos de resumen, 58
  - para IBM SPSS Data Collection, 54
  - plantillas, 196
  - resultados de la categorización, 53–54, 56
- extracción, 3, 78, 82, 84, 202, 215, 245
  - almacenamiento de resultados, 85
  - conceptos, tipos y patrones, 78
  - entidades no lingüísticas, 85
  - forzado de palabras, 92
  - refinamiento de los resultados, 9, 86
  - resultados, 26
  - unitérminos, 4, 85, 246
- fechas (entidad no lingüística), 237
- fiabilidad, 8
- filtro de bibliotecas, 206
- forma plural de las palabras, 218
- formas declinadas, 119, 215, 217–219, 258
- formato compacto, 138
- formato con sangrado, 140
- formato de lista plana, 137
- forzado
  - extracción de conceptos, 92
  - forzar dentro, 97
  - forzar fuera, 97
  - mostrar columnas, 97
  - palabras en categorías, 162
- respuestas, 161
- términos, 222
- frecuencia, 123
- frecuencia de tipo, 123
- fusión de categorías, 160
- generación de categoría, 6, 110, 113, 250
  - excepciones de enlace de clasificación, 117
  - técnica de derivación de raíz de conceptos, 6, 115, 128
  - técnica de inclusión de conceptos, 128
  - técnica de redes semánticas, 6, 115, 128
  - técnica de reglas de co-ocurrencia, 6, 115, 128
  - utilización de técnicas, 6, 115
- generación de categorías, 6, 110–111, 113–116, 118–125, 127, 129, 250
  - técnicas de agrupación entre iguales, 127
- generar formas declinadas, 215, 217–219, 258
- gráfico de barras de categorías, 167–168
- gráfico/tabla de malla de categorías, 167, 169–170
- gráficos, 167
  - actualización, 167
  - edición, 171
  - exportación de gráficos de resumen, 58
  - gráfico de malla de categorías, 167
  - modo de exploración, 170
  - tamaño de los elementos gráficos, 176
- gráficos de barras, 167
- gráficos de malla, 167
- gráficos de resumen, 58
- guardado automático de proyectos, 18
- guardado de resultados de la extracción, 85
- horas (entidad no lingüística), 237
- HTTP/URL (entidad no lingüística), 237
- IBM SPSS Data Collection, 34, 67
  - cambiar el origen de datos, 61
  - exportación, 54
- IBM SPSS Statistics archivos *.sav*, 30, 63
  - cambiar el origen de datos, 61
  - exportación, 54
  - formato de resultados, 53
- importación
  - actualizar datos, 72
  - bibliotecas públicas, 208
  - categorías predefinidas, 132
  - datos de entrada, 34, 67
  - ODBC, 32, 65
  - plantillas, 196
  - preparación de datos, 27
  - volver a importar los datos, 61
- impresión de categorías, 164
- informes y gráficos de resumen, 58
- Japonés, 245, 250
  - Editor de recursos, 250

- propiedades de tipo, 258
- tipos, 254, 260–261
- jitter, 186
- Language Weaver, 22, 36, 70
- leyenda
  - posición, 187
  - \*.lib, 208
- marca de completo, 74
- marca de importante, 74
- marcar respuestas, 74
- marcas comerciales, 263
- marcos de código, 131–132
- Microsoft Excel Archivos Excel .xls/.xlsx, 31, 56, 64
  - cambiar el origen de datos, 61
  - exportación de categorías predefinidas, 142
  - formato de resultados, 53
  - importación de categorías predefinidas, 132
- minería de datos de texto, 3
- modificadores de colisión, 183
- modo de edición, 171
- modo de exploración, 170
- monedas (entidad no lingüística), 237
- mostrar columnas en el panel categorías, 95
- mostrar columnas en el panel de datos, 97
- nombre de categoría, 95
- normalización, 240
- NOT, operador de regla, 154
- novedades, 1
- número de la seguridad social (entidad no lingüística), 237
- número máximo de categorías para crear., 116
- números de teléfono (entidad no lingüística), 237
- opción de coincidencia, 215, 217, 219–221, 258
- opciones, 17
  - mostrar, 19
  - sistema, 18
  - sonido, 21
  - traducción, 22
- opciones de sonido, 21
- operadores booleanos, 154
- operadores de regla & | !() , 154
- Opinions library, 216
- OR, operador de regla, 154
- ordenar datos y variables, 50
- origen de datos ODBC, 32–33, 61, 65–66
- paletas
  - desplazamiento, 172
  - ocultación, 172
  - representación, 172
- panel Categorías, 94
- panel Datos, 97
  - botón mostrar, 96
- panel Visualización
  - actualización de gráficos, 167
  - gráfico de malla de categorías, 167
- paquetes de análisis de texto, 37, 41–43
- paquetes de análisis de texto \*.tap, 37, 41–43, 46
- pasar por alto conceptos, 91
- patrones, 78, 81
- patrones de extracción, 242
- permutaciones, 85
- pesos/medidas (entidad no lingüística), 237
- pila, 186
- plantillas, 4, 190, 246
  - actualización o acción de guardar como, 193
  - cambio de nombre, 195
  - cambio de recursos, 194
  - copia de seguridad, 197
  - creación a partir de recursos, 193
  - eliminación, 195
  - importación y exportación, 196
  - restauración, 197
  - TLA, 194
- plantillas de recursos, 4, 190, 246
- porcentajes (entidad no lingüística), 237
- preferencias, 17, 19, 21
- pregunta con final abierto, 2
- pregunta con final cerrado, 2
- preparación de los datos, 27
- propiedades
  - categorías, 108, 156
  - para tipos del japonés, 258
  - proyectos, 48
  - variables, 50
- proteínas (entidad no lingüística), 237
- proyectos, 26, 47
  - almacenamiento, 51
  - apertura, 47
  - barra de estado, 75
  - cambio de nombre, 51
  - compartir, 74
  - creación, 28
  - opciones para bibliotecas, 18
  - origen de datos, 29, 62
  - paquetes de análisis de texto, 37
  - propiedades, 48
  - selección de categorías y recursos, 37
  - selección de variables, 34, 67
  - traducción, 36, 70
  - volver a utilizar categorías, 164
- publicar, 51, 211
  - adición de bibliotecas públicas, 204
  - bibliotecas, 210
- recuento, 96
- recuperación de archivos, 18
- recursos
  - bibliotecas enviadas por defecto, 202
  - cambio de los recursos de una plantilla, 194

- copia de seguridad, 197
  - edición de recursos avanzados, 233
  - restauración, 197
- recursos avanzados, 233
  - buscar y reemplazar en el editor, 234–235
- recursos lingüísticos, 26, 202
  - paquetes de análisis de texto, 41, 43
  - plantillas de recursos, 190
- refinamiento de los resultados
  - adición de conceptos a tipos, 89
  - adición de sinónimos, 87
  - categorías, 9, 156
  - creación de tipos, 89
  - exclusión de conceptos, 91
  - forzado de extracción de conceptos, 92
  - resultados de la extracción, 9, 86
- registros, 97
- reglas
  - creación, 153
  - edición, 155
  - eliminación, 155
  - operadores booleanos, 154
  - sintaxis, 146
  - técnica de reglas de co-ocurrencia, 122
- reglas de categoría, 145–146, 151, 153–155
  - de la co-ocurrencia de conceptos, 6, 114–115, 118, 122, 128
  - de sinónimos, 6, 113–115, 118, 125, 128
  - ejemplos, 151
  - reglas de co-ocurrencia, 113, 118, 125
  - sintaxis, 146
- relevancia de las respuestas y categorías, 98–99
- respuestas, 97
  - forzar dentro de categorías, 161
  - marcar, 74
  - marcar como completas, 74
- respuestas en blanco, 27
- respuestas vacías, 27
- restauración de recursos, 197
  
- secciones de gestión de idiomas, 233, 242
  - abreviaturas, 242–243
  - definiciones forzadas, 242–243
  - patrones de extracción, 242
- seguimiento de respuestas, 74
- separadores, 18
- separadores de texto, 18
- signo de dólar (\$), 228
- signo de exclamación (!), 228
- silenciado de sonidos, 21
- símbolo de intercalación (^), 228
- sin categorizar, 95
- sincronización de bibliotecas, 210–212
- sinónimos, 86, 225
  - ! símbolos ^ \* \$, 228
  - adición, 87, 227, 259
  - agrupación difusa (excepciones), 84, 236
  - colores, 229
  - definición de, 226
  - eliminación de entradas, 230
    - para textos en japonés, 259
    - términos objetivo, 227, 259
- sistemas de coordenadas
  - transformación, 182
- superiores
  - adición a las categorías, 102, 108, 157
  - adición a tipos, 89
  - creación de tipos, 86
  - en categorías, 102, 108
  - exclusión de la extracción, 91
  - extracción, 78
  - forzado en la extracción, 92
  - los mejores descriptores, 104
  - supervisar, 74
- sustitución de recursos por una plantilla, 194
  
- tabla de malla, 167
- técnica de derivación de raíz de conceptos, 113, 118, 125, 128
- técnica de inclusión de conceptos, 7, 113, 115, 118, 120, 125
- técnica de redes semánticas, 7, 113, 115, 118, 121, 125, 128
- técnica de reglas de co-ocurrencia, 7, 113, 115, 118, 122, 125, 128
- técnicas
  - arrastrar y soltar, 131
  - derivación de raíz de conceptos, 113, 118, 125
  - frecuencia, 123
  - inclusión de conceptos, 113, 118, 120, 125
  - redes semánticas, 113, 118, 121, 125
  - reglas de co-ocurrencia, 113, 118, 122, 125
- técnicas lingüísticas, 3, 6, 8, 115
- términos
  - adición a tipos, 219
  - adición a tipos del japonés, 252
  - adición en el diccionario de exclusión, 232
  - búsqueda en el editor, 205
  - color, 218, 259
  - formas declinadas, 215
  - forzado de palabras dentro de categorías, 162
  - forzado de términos, 222
  - opciones de coincidencia, 215
- términos objetivo, 229
- tipos, 215
  - adición de conceptos, 86
  - búsqueda en el editor, 205
  - color por defecto, 19, 218, 259
  - creación, 217, 258
  - diccionarios, 202
  - extracción, 78
  - frecuencia de tipo, 123
  - para el japonés, 254, 258, 260–261
  - tipos incorporados, 216
- títulos para el gráfico de resumen de exportación, 60

- 
- TLA, 194
  - todos los documentos, 95
  - traducción, 22, 72
    - diálogo de configuración de traducción, 72
    - diálogo opciones, pestaña traducción, 22
    - traducción al inglés, 36, 70, 72
    - valor de precisión de la traducción, 37, 71, 73
  
  - unitérminos, 85
  
  - valor de enlace mínimo, 116
  - variables
    - actualización, 72
    - cambiar el origen de datos, 61
    - coincidencia, 69
    - editar propiedades, 50
    - exportación, 53
    - importación, 30, 32, 63, 65
    - variables de ID, 26–27, 34, 67
    - variables de referencia, 26–27, 34, 67
    - variables de texto, 26–27, 34, 67
  - variables de ID, 26–27, 34, 67
  - variables de referencia, 26–27, 34, 67
  - variables de texto, 26–27, 34, 67
  - vista
    - bibliotecas, 206
    - categorías, 167
    - datos, 49
  - vista de pregunta, 11–12
  - vista del proyecto completo, 11, 14
  - vistas
    - proyecto completo, 14
    - ventana Análisis de texto, 11
    - ventana del editor de recursos, 15
    - vista de pregunta, 12
  - visualizaciones
    - categorías, 180
    - colores y tramas, 174
    - copia, 187
    - edición, 171
    - ejes, 178
    - escalas, 178
    - estilos de guión, 174
    - forma de puntos, 175
    - formatos de numeración, 177
    - márgenes, 177
    - paneles, 180, 182
    - posición de la leyenda, 187
    - relación de aspecto de puntos, 175
    - relleno, 177
    - rotación de puntos, 175
    - texto, 173
    - transformación de los sistemas de coordenadas, 182
    - transparencia, 174
    - transponer, 180, 182
  - volver a importar los datos, 61
  - volver a utilizar
    - categorías, 164