

No te puedes perder... Estrategias de implicación del lector en un corpus de folletos turísticos españoles

Sara Piccioni
Maria Giovanna Biscu
Universidad de Bolonia

Resumen

Este trabajo tiene un doble objetivo: por un lado, presentar un corpus bilingüe (español/italiano) representativo de los géneros textuales del turismo, elaborado en el ámbito del proyecto de investigación “El lenguaje de la comunicación turística español-italiano”; por otro lado, exponer los resultados de un análisis preliminar basado en un corpus piloto de folletos turísticos españoles. Tras una presentación general de los objetivos principales del proyecto y de las motivaciones que han llevado a su ideación, se ilustran los criterios de recopilación y sistematización de los textos, el diseño del corpus y su anotación. En la segunda parte, se estudian las estrategias lingüísticas de implicación del interlocutor a partir de un análisis de listas de frecuencia y de colocados, así como de los lemas asociados con específicos modos y tiempos verbales. Por último, se presentan unas consideraciones finales sobre posibles aplicaciones didácticas y futuros desarrollos en la investigación de los géneros del lenguaje del turismo.

1. Introducción

En los últimos años, se han intensificado las reflexiones sobre las estrategias comunicativas del lenguaje del turismo, así como los trabajos de investigación sobre léxico, morfosintaxis y cuestiones inherentes a la traducción, sobre todo de elementos (inter)culturales. Dentro de esta tendencia, se coloca un proyecto llevado a cabo por cuatro universidades italianas, “El lenguaje de la comunicación turística español-italiano. Aspectos léxicos, pragmáticos e interculturales”,¹ cuyo rasgo distintivo e innovador es la aplicación de la lingüística de corpus al estudio del lenguaje del turismo desde una óptica contrastiva. El proyecto, financiado por el *Ministero dell’Università e della Ricerca* y coordinado por Maria Vittoria Calvi, reúne a centros de investigación de la Universidad de Milán, la Universidad de Bolonia (sede de Forlì), la Universidad de Trento y la Universidad de Verona.²

El presente estudio tiene una doble finalidad: por un lado, presentar los aspectos técnicos de la compilación del corpus del lenguaje del turismo creado en el marco de este proyecto; por otro, exponer los resultados de un análisis preliminar que utiliza un corpus piloto de folletos turísticos para estudiar las estrategias lingüísticas de implicación del interlocutor. En § 2, tras una presentación general de las motivaciones principales que han llevado a la

¹ *Il linguaggio della comunicazione turistica spagnolo-italiano. Aspetti lessicali, pragmatici e interculturali.*

² Las coordinadoras de los distintos centros son: Maria Vittoria Calvi (Universidad de Milán), Pilar Capanaga (Universidad de Bolonia, sede de Forlì) y Elena Maria Liverani (Universidad de Trento y Universidad de Verona).

ideación del proyecto, se prestará especial atención a sus objetivos y a las dos herramientas que se van a crear, es decir, una base de datos y un corpus lingüístico bilingüe comparable, representativo del lenguaje del turismo. Sucesivamente, se ilustrarán los criterios de recopilación de los textos seleccionados para la construcción del corpus (§ 3.1) y los métodos utilizados para su anotación (§ 3.2). En particular, se hará hincapié en la descripción del trabajo que le compete al grupo de investigación de la Universidad de Bolonia, sede de Forlì (§ 4). En la segunda parte, para examinar las estrategias de implicación del lector en los folletos turísticos, se analizan las listas de frecuencia de las partes del discurso (§ 5.1) y los colocados de los verbos (§ 5.2). Asimismo, se lleva a cabo un estudio semántico de los lemas asociados con modos y tiempos verbales específicos (§ 5.3) para detectar patrones de co-variación que sean indicativos del nivel de implicación del lector. Por último, en § 6, se apunta a posibles aplicaciones didácticas del estudio y a desarrollos futuros para la investigación.

2. La aplicación de la lingüística de corpus al estudio del lenguaje del turismo: un proyecto de investigación

El proyecto “El lenguaje de la comunicación turística español-italiano. Aspectos léxicos, pragmáticos e interculturales” responde a la necesidad de sondear las múltiples facetas de un lenguaje que está vinculado a un sector socio-económico central en España y en Italia.

La marcada heterogeneidad e interdisciplinariedad del lenguaje del turismo hace necesario llevar a cabo estudios que puedan abarcar todos sus aspectos de forma exhaustiva. Se ha apuntado que el lenguaje del turismo se sitúa a medio camino entre la lengua general y las lenguas especializadas,³ abarcando diferentes niveles de expresión, que van de la comunicación de especialista a especialista a la comunicación dirigida a semi-especialistas o a no especialistas (Calvi 2006, Gotti 2006). Se pasa, pues, de un discurso con elevada densidad conceptual y terminológica a un lenguaje menos especializado, marcado por reformulaciones y ejemplificaciones. Esta característica, intrínseca del lenguaje del turismo, unida a la falta de herramientas que permitan abordar el fenómeno en su totalidad, ha generado estudios parciales, centrados en ámbitos comunicativos específicos.

Estas limitaciones afectan tanto a los estudios “tradicionales” como a los que utilizan una metodología de corpus. En los estudios “tradicionales” se privilegian aquellos textos cuya finalidad es informar a los viajeros sobre un lugar, o bien promocionar destinos y servicios (Kelly 1997: 35), excluyendo así ámbitos estratégicos para los profesionales del turismo y de la traducción, como pueden ser los documentos legales y comerciales que regulan las transacciones en el mercado turístico. Los estudios que utilizan corpus electrónicos, a pesar de ser más sistemáticos con respecto a los que se basan en análisis manuales, tienden a describir y estudiar ámbitos limitados –sobre todo en relación al lenguaje de las guías– a partir de corpus exigüos creados para investigar cuestiones específicas en estudios individuales (Castello 2002, Nigro 2006). El Proyecto TURICOR (Corpas Pastor 2003) constituye una destacada excepción, ya que propone un corpus multilingüe de grandes dimensiones, abierto

³ A propósito de esto, algunos autores dudan a la hora de clasificar el lenguaje del turismo entre las lenguas especializadas. Por ejemplo, tal y como afirma Gotti (2006: 19), “the term ‘special language’ cannot be applied to the language of tourism because the latter may be distinguished from general language not for its use of special linguistic rules absent from general language but for its quantitatively greater and pragmatically more specific use of such conventions”.

al uso en distintas áreas, incluidas las aplicaciones de generación automática de textos. El corpus creado, sin embargo, reúne únicamente textos jurídicos, excluyendo así los demás ámbitos de la comunicación turística.

Por lo que se refiere a la dimensión contrastiva, los estudios realizados se centran en aspectos funcionales y discursivos (por ejemplo, Ramm 2000) o en el ámbito de los *translation studies*, sobre todo en lo que atañe a la traducción de culturemas (De la Cruz Trainor 2004, Cómitre Narváez 2004, Nigro 2006). Ambas vertientes privilegian el estudio contrastivo con el inglés (español-inglés o italiano-inglés), mientras que se registra una exigüidad de estudios sistemáticos comparativos entre el español y el italiano.

Desde el punto de vista didáctico, destaca la escasez de recursos para la formación, tanto de futuros mediadores lingüísticos (intérpretes y traductores) como de otros profesionales y operadores que se quieran especializar en el sector de la comunicación turística. La reconocida baja calidad de las traducciones de los textos turísticos (Kelly 2000, De la Cruz Trainor 2004, Nobs 2006) ha puesto de manifiesto con fuerza la necesidad de una formación capaz de responder a las exigencias del mercado, que pueda beneficiar tanto a los operadores del sector como a los destinatarios/usuarios finales de la oferta turística. La escasez de recursos didácticos se hace aún más patente si se considera la intensificación del uso de textos turísticos en los cursos de traducción. Esto se debe a una serie de reflexiones teóricas sobre las prerrogativas didácticas de un lenguaje tan heterogéneo como el del turismo, que garantiza la posibilidad de escoger materiales con un nivel de dificultad lingüística y de especialización adecuado para las competencias de los estudiantes (Kelly 2000). Nuestra experiencia concreta en la Facultad de Traducción e Interpretación de la Universidad de Bolonia (sede de Forlì),⁴ sobre todo en los cursos con estudiantes itálofonos que poseen un nivel avanzado de español, nos ha llevado a constatar que, trabajando variedades textuales más especializadas, se puede introducir al estudiante en la labor de documentación terminológica indispensable para el futuro profesional.

Tanto la investigación lingüística como la didáctica de L2 y la traducción se ven afectadas por la falta de herramientas que permitan observar el lenguaje del turismo en su conjunto. El proyecto “El lenguaje de la comunicación turística español-italiano” intenta responder a las carencias expuestas en este apartado a través de la creación de dos herramientas complementarias: por un lado, una base de datos formada por textos italianos y españoles que pertenezcan a diferentes géneros representativos del lenguaje del turismo; por otro, un corpus electrónico comparable que reúna los mismos textos y que se irá creando de forma paralela a la recogida de materiales. La base de datos dará acceso a los textos en un formato que permita la visualización de todas sus características formales y gráficas (por ejemplo, pdf, word, directorios de páginas web, etc.), y también en formato texto, mientras que el corpus recogerá textos anotados morfosintácticamente y lematizados, permitiendo búsquedas lingüísticas de dintinto tipo (véase § 3.1).

Desde el punto de vista práctico, el trabajo de selección de materiales para la base de datos y el corpus se ha dividido en tres macrogrupos, de los que se encargan los distintos equipos de trabajo: 1) textos legislativos y normativos / burocráticos; 2) guías y folletos; 3) páginas web, material publicitario y reportajes. De esta manera, se procurará tener una muestra representativa de la mayoría de los géneros identificados en el ámbito del lenguaje del turismo.

Las dos herramientas serán utilizadas tanto para la investigación, como para la didáctica. En lo que atañe a la investigación, la consulta de la base de datos permitirá explorar la

⁴ *Scuola Superiore di Lingue Moderne per Interpreti e Traduttori, Università di Bologna, sede di Forlì.*

dimensión multimodal de los textos turísticos, generalmente caracterizados por una abundancia de imágenes, fotografías, mapas y –en el caso de las páginas web– documentos audio o vídeo. El corpus electrónico comparable, por otro lado, dará la oportunidad de describir los aspectos lingüísticos más variados de forma sistemática y puntual, mediante datos estadísticos, listados de frecuencia, colocaciones y concordancias.

A nivel didáctico, la base de datos y el corpus electrónico serán explotados para idear materiales útiles para los estudiantes, en particular, para futuros mediadores u otro tipo de profesionales que se quieran especializar en el sector turístico. La creación de materiales didácticos abarca, por un lado, cursos que prevén la enseñanza de la lengua y de la cultura relacionada con la comunicación turística y, por otro, módulos específicos que formen parte de cursos de traducción. En este segundo caso, la explotación del corpus electrónico refuerza la autonomía de los estudiantes a la hora de profundizar en aspectos lingüísticos concretos a partir de materiales auténticos. Para fomentar el autoaprendizaje, también se planea organizar cursos e-learning.

Por último, por lo que se refiere al sector profesional, el corpus lingüístico será utilizado para realizar un glosario bilingüe italiano-español y fichas terminológicas a través de la extracción de términos, para dar cuenta de los ámbitos con un alto nivel de especialización.

Después de exponer los objetivos del proyecto, a continuación nos centraremos en la descripción del corpus lingüístico, ilustrando los criterios que rigen su diseño y su anotación.

3. El corpus lingüístico del lenguaje del turismo

3.1. Criterios de recopilación

El criterio principal para la recopilación de los textos que van a integrar el corpus es la representatividad, indispensable para abarcar la gran variedad lingüística y discursiva del lenguaje del turismo (§ 2). Otro criterio a tener en cuenta es la selección de textos originales (no traducidos) en las dos lenguas, que responde a la necesidad de analizar las peculiaridades de cada una sin las interferencias debidas a la traducción. El resultado será un corpus bilingüe, comparable con respecto al género textual de referencia y representativo de la lengua del turismo. Cuando sea posible, al corpus se integrarán textos enteros, respondiendo a la necesidad de observar los rasgos lingüísticos correspondientes a distintas partes funcionales de un mismo texto.

Definidos los criterios de recopilación, se ha realizado una exploración preliminar de los tipos de textos producidos en el ámbito de actividades relacionadas con el turismo. Se han considerado distintos medios, incluyendo –además de material editorial, prensa, publicaciones periódicas de tipo tradicional– material electrónico publicado en la red. A partir de esta exploración, el centro de la Universidad de Milán, basándose en Calvi (2006) y en el trabajo del equipo de investigación GENTT (por ejemplo, García Izquierdo y Monzó Nebot 2003), ha propuesto una hipótesis de trabajo para la clasificación de los textos. Dicha clasificación establece los siguientes niveles de categorización: familia de géneros (definida por el propósito de los textos y la comunidad profesional en la que se generan), macrogénero (definido por el canal empleado), género (definido por su propósito comunicativo, participantes, tenor y modo) y tema (por ejemplo, arte e historia, espectáculos y diversiones, transportes, seguros, etc.).

Cada texto incluido en el corpus será catalogado en base a estos criterios: así, por ejemplo, un folleto turístico publicado por la Comunidad Autónoma de Castilla-La Mancha

sobre las rutas del Qujote se considera como documento institucional (familia de género), perteneciente al macrogénero del folleto, al género itinerario, sobre el tema arte e historia.

Esta categorización asegura, por un lado, máximo control sobre el tipo de textos introducidos en el corpus (garantizando, así, la representatividad del corpus equilibrado) y, por otro, la comparabilidad entre las secciones española e italiana, ya que se intentará mantener la misma proporción de las distintas categorías para las dos lenguas.

3.2. Procesamiento y etiquetado

Los textos seleccionados se someten a procesamiento para crear un corpus lingüísticamente anotado que permita el estudio de los rasgos típicos del lenguaje del turismo.

Todos los textos se recogen en formato electrónico, utilizando las versiones digitales disponibles (bajadas de la red o proporcionadas por las propias editoriales), o –a falta de éstas– procediendo al escaneado y sucesivo reconocimiento óptico (OCR) de textos impresos.

Anotación morfosintáctica (POS-tagging) y lematización se realizan por separado para las dos lenguas. Los textos italianos se anotan utilizando el *tagger* Tree-Tagger (Schmid 1994), mientras que para los textos españoles se usa FreeLing (Carreras *et al.* 2004). Para asegurar la comparabilidad entre el subcorpus italiano y el español, las etiquetas morfosintácticas utilizadas por FreeLing (basadas en el tagset EAGLES) se convierten automáticamente en etiquetas compatibles con el *output* del TreeTagger, utilizando tablas de conversión compiladas manualmente (véase Piccioni 2005). De esta forma, tanto en el subcorpus italiano como en el español, las etiquetas se refieren a las mismas categorías sintácticas y proporcionan el mismo tipo de información morfológica adicional. Automáticamente, se introducen también etiquetas XML que marcan el inicio y el final de cada frase (<s>).

Los textos anotados morfosintácticamente y lematizados se complementan manualmente con información metatextual. En esta fase, a cada texto se asignan etiquetas XML que proporcionan información relativa a su origen y tipología a través de: un código alfanumérico identificativo del texto; un código que indica la lengua (italiano o español); eventual texto matriz (para los documentos que son subpartes de textos más extensos como, por ejemplo, en el caso de guías prácticas dentro de una guía completa); autor (una indicación de la editorial para las publicaciones impresas, el directorio para los textos sacados de sitios web); fecha (año de publicación para textos impresos, fecha de acceso para los sitios web); familia de género, macrogénero, género y tema (según lo indicado en § 3.1); y, por último, una etiqueta “multimodalidad” que indica si un texto va acompañado por contenido multimedia, como imágenes, vídeos y audio.

En una fase de postprocesamiento automático se convierten los textos en el formato compatible con los programas de interrogación elegidos por cada equipo de investigación. Hasta la fecha, la porción de corpus ya recopilada se conserva en un formato compatible con el software de interrogación AntConc (Anthony 2006) para plataforma Windows y en formato compatible con la indexación con Corpus WorkBench (CWB) y la interrogación con el Corpus Query Processor (CQP) (Christ 1994) para usuarios de Linux.

La Tabla 1 ofrece un ejemplo de texto anotado en formato compatible con la interrogación en AntConc. Las etiquetas contenidas entre corchetes angulares proporcionan información metatextual (por ej., <texto...>) y estructural (por ej., <s>). Cada frase aparece en una línea separada y cada palabra está unida, por guiones bajos, a las etiquetas que indican su parte del discurso y lema correspondientes (por ej., alza_VER:ind+pre+3+s_alzar).

Este tipo de anotación permite buscar (secuencias de) palabras, partes del discurso y lemas en las combinaciones deseadas; la búsqueda se puede realizar en el corpus entero y/o en

textos seleccionados en base a criterios metatextuales (por ej., buscando todas las combinaciones NOMBRE+ADJETIVO en el género “contrato de viaje”).

```
<texto id=ESFO0001“ leng=“ES“ origen=“—“ autor="turespaña" fecha="25062003"
fam="institucional" macro="folleto" gen="guíaDes" tema="arteHis" multimodalidad="TI">
<s>
Albacete_NOMBRE-C_albacete . _FRASE_ .
</s>
<s>
La _ART-F:s_el mayor _ADJ:c+s_mayor ciudad _NOMBRE-C-F:s_ciudad de _PRE_de
Castilla _NOMBRE-C-F:s_castilla - _PUN-GUI - La _ART-F:s_el Mancha _NOMBRE-C-
F:s_mancha se _PRO-CLI-3_él alza _VER:ind+pre+3+s_alzar en _PRE_en medio _NOMBRE-
C-M:s_medio de _PRE_de una _DET-INDEF:f+s_uno vasta _ADJ:f+s_vasto
campaña _NOMBRE-C-F:s_campaña de _PRE_de un _DET-INDEF:m+s_uno
alto _ADJ:m+s_alto valor _NOMBRE-C-M:s_valor natural _ADJ:c+s_natural . _FRASE_ .
</s>
```

Tabla 1. Ejemplo de texto anotado.

La primera etapa del proyecto, todavía en fase de realización, prevé la recopilación autónoma, por parte de los distintos grupos de investigación, de textos que se van integrando a la base de datos. A medida que se recojan los materiales, las distintas unidades podrán acceder a ellos localmente.

Para la fase final del proyecto, si se consiguen resolver las cuestiones inherentes al copyright, se considerará la posibilidad de permitir el acceso remoto tanto a la base de datos como al corpus anotado, a través de una interfaz web que permitirá búsquedas simples y avanzadas.

4. El subcorpus de guías y folletos

Como se ha indicado en § 2, las distintas unidades de investigación que participan en el proyecto se encargan de macrogéneros y temas específicos distribuidos de manera que el corpus resulte equilibrado.

Nuestra unidad (Universidad de Bolonia, sede de Forlì) se ocupa de la recopilación del subcorpus de guías y folletos turísticos, definidos por los criterios indicados por Calvi (2006). Las guías turísticas se refieren, por lo tanto, a publicaciones en forma de libro, que contienen como elementos constitutivos básicos una descripción de uno o más lugares geográficos, documentación práctica sobre cómo realizar la visita y materiales complementarios (como planos, fotos, etc.) (Calvi 2006: 33). Los folletos, en cambio, son publicaciones de tamaño más bien reducido que cumplen una doble función informativa y promocional que justifica su distribución gratuita (Calvi 2006: 37).

En detalle, los materiales que integrarán el corpus incluyen: folletos turísticos publicados por las Comunidades Autónomas españolas y por las regiones italianas (*Regioni*); guías y folletos de un parque nacional español (Ordesa y Monte Perdido) y un parque nacional italiano (Gran Sasso e Monti della Laga); guías españolas de Italia y de una ciudad de arte

italiana y una española (Anaya, Aguilar, Repsol); guías italianas de España y de una ciudad de arte italiana y una española (Touring Club, De Agostini, Mondadori).

La selección del material está motivada no sólo por la comparabilidad del macrogénero de referencia de cada texto (guías y folletos), sino también de los temas (por ejemplo, enoturismo, turismo de balnearios, ciudades de arte) y de correspondencias geográficas (por ejemplo, los parques nacionales se han escogido en base a las semejanzas geofísicas que hacen suponer cierta correspondencia, tanto en los referentes como en la terminología).

El trabajo de recopilación ha tomado como punto de partida la creación de un pequeño corpus piloto de 200 mil *tokens* de folletos turísticos “institucionales”, publicados por patronatos de turismo, consejerías y juntas autonómicas. Para una adquisición más rápida de los textos en formato electrónico se han seleccionado folletos cuyas versiones en pdf se pueden descargar de las correspondientes páginas web: el hecho de que se puedan bajar de la red no influye en la definición del macrogénero al que pertenecen, ya que se trata de textos pensados para ser distribuidos en formato impreso.

El corpus piloto está compuesto por 34 folletos, de los cuales 26 publicados por el Instituto de Promoción Turística de Castilla-La Mancha, 5 por Turespaña (Instituto de Turismo de España), 2 por el Patronato Provincial de Turismo de Córdoba, 1 por el Patronato Provincial de Turismo de Granada. A nivel temático, más de la mitad de los folletos (21) presenta información general sobre alguna ciudad o pueblo, describiendo sobre todo el patrimonio histórico-artístico, la gastronomía, la artesanía, las fiestas y otros eventos. Seis folletos ilustran itinerarios por rutas de interés artístico y/o natural, proporcionando información de tipo descriptivo y práctico. Los restantes siete folletos se centran en temas específicos, que van del enoturismo, a las casas cuevas de Granada, al turismo ecuestre, etc. La limitada variedad de textos incluidos en este minicorpus comporta una reducida representatividad; se trata, sin embargo, de un límite “provisional”, ya que se conseguirá mayor equilibrio según se vayan añadiendo textos al corpus.

Los pdf convertidos en formato texto han sido sometidos a control formal para verificar la calidad de la conversión. En esta fase, se han eliminado las numerosas páginas que contenían únicamente listas de direcciones de hoteles, restaurantes y otros establecimientos comerciales. La decisión de eliminar estas listas está motivada, por un lado, por su limitada aportación a la descripción del lenguaje del turismo, por otro, por la influencia que este material puede tener en la elaboración de estadísticas de distinto tipo en el corpus; piénsese, por ejemplo, en cómo estos elementos pueden influir en el cálculo de la longitud media de las frases. Asimismo, frases con largas secuencias de sustantivos pueden causar problemas a los programas de anotación morfosintáctica.

Los textos recogidos han sido anotados morfosintácticamente y lematizados, añadiendo, por último, la información metatextual indicada en § 3.2.

Este minicorpus será utilizado en los apartados siguientes para explorar las estrategias de implicación del interlocutor en folletos turísticos españoles.

5. Implicación del interlocutor en folletos turísticos españoles

Uno de los factores de variación dentro del lenguaje del turismo es el nivel de implicación del interlocutor en los textos. Se trata de un aspecto que atañe a la dimensión del tenor, ya que determina la distancia social y los roles de los participantes involucrados en la comunicación. En este nivel del análisis, es posible verificar si un texto dado está orientado hacia la mera comunicación de contenidos o si, en cambio, tiene como objetivo establecer una relación entre emisor y destinatario (Ramm 2000: 151). Su relevancia en el estudio del lenguaje del turismo

se debe a la gran variación que se puede constatar entre familias de géneros distintas: piénsese, por ejemplo, en la diferente implicación del interlocutor en un anuncio publicitario y en un billete de avión.

En este apartado presentamos algunos datos preliminares basados en el análisis del minicorpus de folletos turísticos descrito en § 4. Al tratarse de textos orientados a la promoción de destinos y/o productos turísticos, la hipótesis de partida es que estén marcados por un alto nivel de implicación del interlocutor. Nuestro propósito es identificar los rasgos lingüísticos que definen las estrategias de implicación. Por ello, en los subapartados que siguen, nos vamos a centrar en el estudio de los verbos, a través del análisis de listas de frecuencia, de colocados de los verbos y de los lemas típicos de las diversas formas verbales.

5.1. Frecuencia de las partes del discurso

Como acabamos de explicar, nuestro punto de partida es la observación del uso de los verbos en el corpus. La Tabla 2 contiene una lista de las formas verbales personales más frecuentes (se han eliminado las formas no personales y aquellas con frecuencia menor de 50).

Como es evidente, el uso verbal en el corpus se caracteriza por la marcada incidencia de formas de indicativo; entre éstas destacan los presentes, sobre todo de tercera persona (singular y plural). Entre las formas de indicativo, los presentes son casi siete veces más frecuentes que los tiempos pasados y 37 veces más frecuentes que los futuros; si se consideran los modos verbales en su conjunto, la tercera persona se utiliza con una frecuencia 16 veces mayor que las demás personas verbales.

Frecuencia	Forma verbal
4560	<i>Presente de indicativo, 3 pers. sing.</i>
1771	Presente de indicativo, 3 pers. plur.
611	Pret. indef., 3 pers. sing.
318	<i>Presente de indicativo, 1 pers. plur.</i>
230	Pret. indef., 3 pers. plur.
229	Imperf. de indicativo, 3 pers. sing.
207	Imperf. de subjuntivo, 3 pers. sing.
184	<i>Futuro de indicativo, 3 pers. sing.</i>
105	<i>Presente de indicativo, 2 pers. sing.</i>
57	<i>Presente de subjuntivo, 2 pers. sing.</i>

Tabla 2. Formas verbales personales más frecuentes en el minicorpus de folletos. En cursiva las formas verbales que serán objeto de discusión.

A falta de datos comparativos sobre las frecuencias relativas de las distintas formas verbales en un corpus de referencia, no es posible trazar conclusiones a partir de estos datos, aunque, en general, se observa un marcado desequilibrio a dos niveles: a nivel de los tiempos (donde los presentes son mucho más frecuentes que los demás tiempos) y a nivel de las personas verbales (donde las formas de primera y segunda persona resultan ser muy escasas frente a las de tercera). El resto del análisis se focaliza en la forma más frecuente, es decir el

presente de indicativo de tercera persona, y en una serie de formas más marginales, que se distinguen del uso más frecuente por el tiempo (por ejemplo, el futuro de indicativo) y/o la persona (las formas de primera y segunda persona).

5.2. Colocados de los verbos

El análisis de las listas de los colocados a la izquierda de los verbos presenta elementos interesantes. La Tabla 3 proporciona una lista de los veinte colocados más frecuentes que aparecen inmediatamente a la izquierda de las formas verbales señaladas arriba. Para los colocados que se comentan a continuación, se indican las frecuencias absolutas y su incidencia entre todos los colocados de las formas verbales consideradas.

Por lo que se refiere a los colocados de los verbos en presente de indicativo (tercera persona), hay que señalar que el colocado más frecuente (el pronombre personal *se*) es compatible con formas reflexivas e impersonales, como se desprende de los siguientes ejemplos:

- (1) En torno a esa fiesta *se celebra* la Semana Grande del Corpus, vivida intensamente por los toledanos.
- (2) A 30 minutos de Caudete *se encuentra* Montealegre del Castillo.

Frente a estas formas, que no comportan implicación del interlocutor, se observa la presencia de pronombres personales de primera y segunda persona, que se refieren al lector a través de las formas inclusivas de la primera persona plural (*nos*), o bien dirigiéndose a él directamente con el *tú*:

- (3) La Vía Verde de la Sierra de Alcaraz, *nos ofrece* la oportunidad de recorrer sus caminos y senderos.
- (4) Con su nueva imagen Castilla-La Mancha *te da* la bienvenida.

Pres. Ind. 3 Sing.	%	Fut. Ind. 3 Sing.	%	Pres. Ind. 1 Plur.	%	Pres. Ind. 2 Sing.	%	Pres. Subj. 2 Sing.	%
<i>se</i> (1164)	25,5	<i>se</i> (30)	16,3	,		<i>te</i> (76)	72,38	<i>No</i> (51)	89,47
que		<i>Nos</i> (27)	14,67	.		,		<i>te</i> (2)	3,51
,		<i>te</i> (22)	11,96	que		<i>no</i> (6)	5,71	<i>no</i> (2)	3,51
.		Que		<i>nos</i> (23)	7,23	<i>No</i> (6)	5,71	los	
de	2,8	<i>viajero</i> (18)	9,78	lo		Y		las	
<i>Se</i> (129)		<i>visitante</i> (15)	8,15	y		.			
Y		,		le		tampoco			
La	1,27	Donde		<i>te</i> (8)	2,51	también			
<i>nos</i> (58)	0,63	Y		donde		Tampoco			
Donde		Quién		Si		actualmente			
<i>le</i> (29)		<i>no</i> (2)	1,09	<i>No</i> (6)	1,89	Si			
también		La		la		mientras			
interior	0,48	recorrido		si		Donde			
ciudad		.		<i>no</i> (5)	1,57	Que			
<i>no</i> (22)		Corduente		<i>Te</i> (4)	1,26				
castillo		calizos		<i>Nos</i> (4)	1,26				
lo		Fuero		También					
La	0,31	<i>aventurero</i> (1)	0,54	también					
)		<i>cliente</i> (1)	0,54	Aquí					
<i>te</i> (14)		subterránea		interior					

Tabla 3. Colocados de izquierda de algunas formas verbales seleccionadas en el minicorpus de folletos. En negrita cursiva los colocados que serán objeto de discusión.

Características aparentemente similares se observan entre los colocados de las formas de futuro, entre los cuales también figuran los pronombres *se*, *nos* y *te*. Sin embargo, las proporciones de su distribución son muy distintas: en primer lugar, el pronombre *se* representa casi un tercio (28,3%) de los colocados de las formas presentes, mientras que para el futuro sólo corresponde al 16,3% de sus colocados; más evidente es el desequilibrio en el uso de los pronombres de primera y segunda persona, que –en conjunto– constituyen apenas el 1,9% en el presente frente al 27% en el futuro (tanto en este caso como en las cifras presentadas abajo, los porcentajes se refieren a las formas con inicial mayúscula y minúscula).

Otro dato de interés es la alta incidencia de elementos léxicos (es decir, palabras no funcionales) entre los colocados del futuro: *viajero* y *visitante* por sí solos representan en conjunto casi el 18% de los colocados del futuro, mientras que los del presente (*interior*, *ciudad*, *castillo*) constituyen apenas el 1,31% de todos los colocados. Es interesante comparar el contexto de uso de estos elementos:

- (5) *El castillo preside* uno de los dos cerros que encierra a la villa.
- (6) *La ciudad tiene* fama por la elaboración de grandes tinajas de barro.
- (7) En estas páginas *el viajero descubrirá* lugares repletos de historia.
- (8) *El visitante disfrutará* de excepcionales fortalezas medievales, arropadas por paisajes de extraordinaria belleza.
- (9) *El caminante aventurero encontrará* en las proximidades la posibilidad de participar en el Parque Natural de las Lagunas de Ruidera, en todo tipo de actividades relacionadas con la vela, el piragüismo [...].

En las formas presentes, los colocados léxicos se encuentran en secciones informativas de los folletos dedicadas a la descripción de un lugar. En las formas de futuro, en cambio, *el viajero* o el *caminate aventurero* hacen referencia a los lectores, proyectándolos en el texto: aun dirigiéndose al interlocutor de forma impersonal, las designaciones producidas favorecen la identificación del lector con una imagen positiva, implicándole indirectamente en el texto mismo.

Si pasamos a considerar las demás formas verbales de primera y segunda persona, notamos –como es lógico– una abundancia de pronombres personales que hacen referencia al interlocutor tratándole de *tú* o englobándole en el *nosotros* inclusivo. Lo que más llama la atención, sin embargo, es la gran incidencia de la negación (*no* y *No*), mucho más marcada aquí con respecto a los verbos hasta ahora considerados. La distribución de la negación varía de manera muy sensible según las formas verbales: se va de una incidencia mínima en las terceras personas del presente de indicativo (0,48%) hasta la asombrosa incidencia en la segunda persona del presente de subjuntivo (92,98%), pasando por la segunda persona del presente de indicativo (11,42%), la primera persona plural (3,46%) y el futuro (1,09%). Los siguientes ejemplos ilustran mejor esta variación:

- (10) Además, la ciudad dispone de muy buenas vías de comunicación, de manera que el visitante *no encontrará* dificultades para desplazarse hasta ella.
- (11) *No podemos abandonar* la ciudad sin pasear a la sombra de los árboles del Parque de Abelardo Sánchez y visitar el Museo de Albacete.
- (12) *No te puedes perder*.
- (13) *No olvides*.

Todos los ejemplos presentan cierto grado de implicación del interlocutor: el futuro lo hace a través de sus colocados léxicos (*visitante*, en este ejemplo), mientras que las demás formas se dirigen directamente al interlocutor. Cabe destacar, además, que los últimos dos

ejemplos son títulos recurrentes de subapartados en los folletos publicados por el Instituto de Promoción Turística de Castilla-La Mancha: por un lado, esto explica la gran incidencia de la negación entre los colocados de las formas verbales interesadas (subrayando, además, la ya mencionada desequilibrada recopilación del corpus); por otro, es evidente que las formas verbales de segunda persona y las correspondientes negaciones tienden a usarse en esas partes de los textos cuya función es llamar la atención del lector. La fuerte asociación entre la segunda persona del presente de subjuntivo y la negación *no* tiene, además, raíces en la morfología del sistema verbal del español, en el cual estas formas expresan lo que en otras lenguas es el imperativo negativo (por ejemplo, *no olvides*).

Los datos aquí presentados arrojan luz sobre algunos de los elementos lingüísticos utilizados en los folletos para implicar al interlocutor. En primer lugar, se ha señalado cierta asociación del futuro de indicativo con elementos lingüísticos que apuntan al interlocutor, correspondientes a pronombres personales y a elementos léxicos que se refieren indirectamente al lector (*viajero, visitante, etc.*), favoreciendo su identificación con figuras con valoración positiva (Calvi 2006: 67-68). En segundo lugar, se ha observado que la negación presenta una asociación más fuerte con las formas verbales que indican implicación del interlocutor, con incidencias muy altas en la primera y segunda persona, intermedias en el futuro y muy bajas en la tercera persona del presente de indicativo. Tottie (1991) y Biber (1989) han señalado que la negación es un rasgo lingüístico propio de los registros orales y – más en general– de los registros con un alto grado de implicación: los datos provisionales hasta aquí recogidos parecen, por lo tanto, indicar que el lenguaje de los folletos se sirve de mecanismos típicos de la oralidad para acercarse a sus lectores y cumplir con su función persuasiva.

5.3. *Análisis de los lemas verbales*

Para verificar de qué manera las distintas formas verbales contribuyen a implicar al lector en el discurso, tomamos en consideración el tipo de verbos que tienden a utilizarse con cada forma. Con este fin, se ha calculado el grado de asociación entre las formas verbales analizadas en § 5.1 y los lemas verbales que les corresponden en el corpus; el objetivo es responder a preguntas como, ¿qué verbos se utilizan más en presente de indicativo y qué verbos en futuro? El grado de asociación se ha establecido calculando la medida estadística *log-likelihood* (Dunning 1994) para parejas compuestas por un lema verbal y la(s) forma(s) verbal(es) en la(s) que dicho lema se utiliza en el corpus: la Tabla 4 señala, para cada forma verbal, los primeros diez lemas que tienden a asociársele en el corpus; cada lema va acompañado por una indicación del valor de *log-likelihood* calculado con UCS toolkit (Evert 2004).

(18) Enclave estratégico para todos los moradores que la poblaron, Chinchilla *conserva* vestigios de su época romana.

En (14) *estar* se limita a señalar la existencia de un edificio concreto en un lugar dado. En (15-18) los verbos *alzarse* y *conservar* hacen hincapié en una propiedad específica del objeto al que se refieren: *alzar* subraya la idea de majestuosidad (como lo señalado en (16) por los colocados *castillo, orgullo, belleza, domina*), mientras que *conservar* se usa más bien para enfatizar algún rasgo positivo del objeto (*magnífico* en (17)) o su antigüedad (18).

Otro ejemplo significativo es el verbo *ofrecer*, cuya estructura argumental prevé un participante beneficiario. Como se indica en los ejemplos que siguen, en algunos casos el lector queda implicado directamente (en (19), a través del *nosotros* inclusivo); en otros casos, aunque no se le mencione de manera explícita, la presencia del beneficiario queda implícita en la estructura argumental del verbo (20):

(19) Para endulzar el paladar la repostería de la provincia nos *ofrece* los dulces llamados de sartén [...].

(20) Debido al recorrido y a la oscuridad, la carrera *ofrece* una imagen nocturna muy espectacular.

Característica común a todas las demás formas es el uso de verbos que se refieren a procesos materiales relacionados con el espacio (*pasar, partir, escapar, acercarse, llevar*, etc.), utilizados para dar instrucciones o trazar un itinerario en el que se supone la participación activa del lector:

(21) Desde el campo de fútbol *seguimos* el carreterín hasta encontrarnos a los pocos metros con la N 420, en donde *cruzamos* la carretera y *giramos* a la derecha.

(22) La visita no será completa si no te *escapas* a descubrir los parajes que nos ofrece el Parque Natural de la Serranía de Cuenca [...].

(23) Una pista de tierra en buen estado que le *llevará* al parque.

Por lo que se refiere al futuro de indicativo (tercera persona singular) y al presente de subjuntivo (segunda persona singular), destaca el uso de verbos referidos a procesos mentales (por ejemplo, *descubrir, sorprender, olvidar, observar*), que se refieren a la experiencia emotiva y sensorial del lector/viajero:

(24) Pasar por Malpica del Tajo, donde su castillo a orillas del Tajo, te *sorprenderá* por su espectacular imagen.

(25) [...] formando un conjunto de singular belleza paisajística, teñido con un colorido particular en cada estación del año y con innumerables rincones que aguardan en silencio a que los *descubras*.

Por último, cabe destacar el uso de verbos modales de posibilidad (*poder, permitir*) y obligación (*deber*), sobre todo en relación con el presente de indicativo (primera y segunda

persona singular), donde la categoría modal aleja el discurso de la pura denotación para invitar al lector a visitar y contemplar un sitio o a probar una especialidad gastronómica:

(26) En toda la región *podemos* saborear exquisitos platos llenos de recuerdos literarios de El Quijote como el pisto y las migas.

Para resumir, los verbos en presente de indicativo de tercera persona singular, aún respondiendo a una función informativa/descriptiva, implican al receptor llamando su atención a través de verbos con fuerte carga connotativa y, en otros casos, presentándole como beneficiario implícito o explícito de la visita/viaje. Las demás formas verbales –de por sí asociadas a un nivel más alto de implicación del lector en virtud de su morfología y de sus colocados– responden a dos estrategias principales: por un lado, proporcionan al viajero instrucciones sobre cómo realizar el viaje, principalmente en relación con su desplazamiento; por otro, utilizan verbos mentales y modales para involucrar al hipotético viajero en una experiencia que apela a sus sentidos, proyectándole hacia un viaje atractivo y apetecible.

6. Consideraciones finales

El proyecto “El lenguaje de la comunicación turística español-italiano” promete proporcionar dos herramientas muy valiosas para investigadores, profesores y estudiantes de lengua y traducción, profesionales de la traducción y de la comunicación turística. En particular, el corpus bilingüe responde a la carencia de corpus de este tipo que representen el lenguaje del turismo en sus múltiples vertientes. La cuidada sistematización de los géneros dentro del corpus se perfila como elemento –además de innovador– imprescindible para quien se acerque a este lenguaje, con fines tanto descriptivos como prácticos. La anotación lingüística abre la posibilidad de explorar en profundidad aspectos específicos de la comunicación turística.

Nuestro estudio piloto, a pesar de estar basado en un corpus exiguo de textos, ha permitido evidenciar algunos de los elementos lingüísticos que favorecen la implicación del lector y definen un registro orientado a la persuasión. Los datos sugieren que hay patrones de co-variación entre distintos elementos: en el análisis de los colocados se han señalado los patrones que atañen a la asociación entre formas verbales, pronombres y negación; los lemas verbales han puesto de manifiesto la asociación entre formas verbales específicas y tipo de verbos (por ejemplo, entre verbos en tercera persona del presente de indicativo y verbos existenciales, o entre las formas de primera y segunda persona y verbos mentales). Sin embargo, tanto las listas de colocados como las concordancias apuntan a ulteriores patrones de asociación, por ejemplo, entre formas verbales y conectores de distinto tipo (adverbios, pronombres relativos, etc.). Un análisis estadístico de los patrones de co-variación de estos elementos puede arrojar luz sobre las propiedades metadiscursivas de los textos turísticos y las estrategias de cohesión utilizadas en distintos registros. Disponer de un corpus representativo en el que los textos estén categorizados por género abre la puerta a estudios sistemáticos de los registros del lenguaje del turismo, siguiendo, por ejemplo, las pautas marcadas por el análisis multidimensional de Biber *et al.* (1998). Por último, el corpus constituye una valiosa herramienta para aplicaciones en el ámbito de la categorización automática de los géneros.

Los datos presentados también se perfilan como buen punto de partida para aplicaciones didácticas en la clase de español L2: analizando listas de frecuencia y colocados y accediendo a las correspondientes concordancias, los estudiantes pueden reflexionar sobre los mecanismos lingüísticos y las estrategias comunicativas típicas de los textos con función

persuasiva. De esta manera, además de enriquecer su repertorio lingüístico de cara a la redacción de sus propios textos, los discentes aprenden métodos y estrategias aplicables a la investigación autónoma de los más variados aspectos lingüísticos. Cuando se disponga del correspondiente corpus italiano, las unidades didácticas podrán centrarse también en aspectos contrastivos y se podrá verificar, por ejemplo, si el italiano presenta los mismos fenómenos observados en español o bien si las dos lenguas utilizan recursos distintos. En la clase de traducción, estas consideraciones pueden impulsar la reflexión sobre la necesidad de respetar el registro del original, animando a los estudiantes a buscar y aplicar estrategias lingüísticas para conseguir este objetivo.

Referencias bibliográficas

Anthony, L. (2006): “Developing a Freeware, Multiplatform Corpus Analysis Toolkit for the Technical Writing Classroom”, *IEEE Transactions on Professional Communication*, 49, 3, 275-286.

Biber, D. (1989): *Variation across Speech and Writing*, Cambridge, CUP.

Biber, D., S. Conrad y R. Reppen (1998): *Corpus Linguistics: Investigating Language Structure and Use*, Cambridge, CUP.

Calvi, M.V. (2006): *Lengua y comunicación en el español del turismo*, Madrid, ArcoLibros.

Carreras, X. *et al.* (2004): “Freeling: An open-source suite of language analyzers”, en M.T. Lino *et al.* (eds.), *Proceedings of the 4th International Conference on Language Resources and Evaluation (LREC '04)*, 26th-28th May 2004, Lisbon / Paris, ELRA, 239-242.

Castello, E. (2002): *Tourist-Information Texts. A Corpus-Based Study of Four Related Genres*, Padova, UNIPRESS.

Christ, O. (1994): “A modular and flexible architecture for an integrated corpus query system”, en *Proceedings of the 3rd Conference on Computational Lexicography and Text Research, Budapest, Hungary, 7th-10th July 1994*, 23-32.

Cómitre Narváez, I. (2004): “La traducción de culturemas en publicaciones del sector turístico. Un estudio empírico”, en J.A. Gallegos Rosillo y H. Benz Busch (eds.), *Traducción y cultura. El papel de la cultura en la comprensión del texto original*, Málaga, Libros ENCASA, 115-138.

Corpas Pastor, G. (2003): “Turicor: Compilación de un corpus de contratos turísticos (alemán, español, inglés, italiano) para la generación textual multilingüe y la traducción jurídica”, en E. Ortega Arjonilla (dir.) y E. Echeverría Pereda, E. Alarcón Navío y C. Mata Pastor (coords.),

Panorama actual de la investigación en traducción e interpretación. Vol. II, Granada, Atrio, 373-384.

De la Cruz Trainor, M.M. (2004): “Traducción al inglés de términos culturales en textos turísticos”, en J.A. Gallegos Rosillo y H. Benz Busch (eds.), *Traducción y cultura: el papel de la cultura en la comprensión del texto original*, Málaga, Libros ENCASA, 83-114.

Dunning, T. (1994): “Accurate Methods for the Statistics of Surprise and Coincidence”, *Computational Linguistics*, 19(1), 61-74.

Evert, S. (2004): *The Statistics of Word Cooccurrences: Word Pairs and Collocations*, Dissertation, Institut für maschinelle Sprachverarbeitung, University of Stuttgart. <http://www.collocations.de/>.

García Izquierdo, I. y E. Monzó Nebot (2003): “Corpus de géneros GENTT. Una enciclopedia para traductores”, *Traducción & Comunicación*, 4, 31-53.

Gotti, M. (2006): “The Language of Tourism as Specialized Discourse”, en O. Palusci y S. Francesconi (eds.), *Translating Tourism. Linguistic-Cultural Representations*, Trento, Università degli Studi di Trento, 15-34.

Kelly, D. (1997): “The translation of texts from the tourist sector: textual conventions, cultural distance and other constraints”, *Trans, Revista de traductología*, 2, 33-42.

Kelly, D. (2000): “Text Selection for Developing Translator Competence: Why Texts From The Tourist Sector Constitute Suitable Material”, en C. Schäffner y B. Adab (eds.), *Developing Translation Competence*, Amsterdam/Philadelphia, John Benjamins, 157-167.

Nigro, M.G. (2006): *Il linguaggio specialistico del turismo: aspetti storici, teorici e traduttivi*, Roma, Aracne.

Nobs, M.-L. (2006): *La traducción de folletos turísticos: ¿qué calidad demandan los turistas?*, Granada, Comares.

Piccioni, S. (2005): “The Lorca Corpus at the crossroads of philology and Corpus Linguistics”, en P. Danielsson y M. Wagenmakers (eds.), *Proceedings of Corpus Linguistics 2005, volume 1 of The Corpus Linguistics Conference Series*. <http://www.corpus.bham.ac.uk/PCLC/>.

Ramm, W. (2000): “Textual Variation in Travel Guides”, en E. Ventola (ed.), *Discourse and Community: Doing Functional Linguistics*, Tübingen, Gunter Narr, 147-167.

Schmid, H. (1994): "Probabilistic part-of-speech tagging using decision trees", en D. Jones (ed.), *International Conference on New Methods in Language Processing, Manchester, UK, September 1994*, Manchester, University of Manchester Institute of Science and Technology (UMIST), 44-49.

Tottie, G. (1991): *Negation in English Speech and Writing: A Study in Variation*, San Diego, Academic.