

# TRADUCCIÓN Y NUEVAS TECNOLOGÍAS: EL RESUMEN AUTOMÁTICO DE PÁGINAS WEB Y EL PROCESO DE DOCUMENTACIÓN PARA EL DISCURSO JURÍDICO-TURÍSTICO<sup>1</sup>

María Cristina Toledo Báez  
Universidad de Málaga

## Resumen

En el presente artículo nos aproximamos a las nuevas tecnologías aplicadas a la Traducción, en concreto a una aplicación de la Lingüística Computacional: el resumen automático de páginas web. Analizaremos su impacto en la fase documental del traductor mediante tres objetos de estudio: a) el macrocorpus multilingüe *Turicor*; b) un sistema de resumen automático gratuito en línea denominado *SweSum* y c) una página web en español de la compañía *Pullmantur*. De esta forma, una vez estudiado un ejemplo en español de un extracto automático de condiciones generales de contratos de viajes combinados, específicamente en la modalidad de cruceros, destacaremos las ventajas que aporta en los planos terminológico, temático, sintáctico y documental para la traducción del discurso jurídico-turístico en el par de lenguas inglés-español.

**Palabras clave:** Traductología, tecnologías de la información y las comunicaciones (TIC), lingüística computacional, traducción asistida por ordenador, sistemas de resumen automático, documentación, discurso jurídico-turístico, condiciones generales, contratos de viajes combinados.

## Abstract

This paper is concerned with new technologies applied to translation, focusing specifically on an application of computational linguistics: automatic webpage summarisation. The impact of automatic extraction on the translator's documentation process is explored using three sets of study data: a) the multilingual macrocorpus *Turicor*; b) a multilingual automatic extraction system available free-of-charge on the web, called *SweSum*; and c) a webpage pertaining to the Spanish company *Pullmantur*. After analysing an example in Spanish of automatically extracted terms and conditions of package tours, specifically cruises, we highlight the advantages of automatic webpage summarisation in legal and tourist translation between Spanish and English, from the point of view of terminology, subject content, syntax and documentation.

1. El presente artículo ha sido realizado en el seno del proyecto *La contratación turística electrónica multilingüe como mediación intercultural: aspectos legales, traductológicos y terminológicos* (Ref. n.º HUM-892, 2006-2009. Proyecto de Excelencia, Junta de Andalucía) y en el marco del Programa de Formación del Profesorado Universitario (FPU) (Ref. AP2005-2792).

**Keywords:** translation studies, information and communication technologies (ICT), computational linguistics, computer-aided translation, automatic extraction systems, documentation, legal and tourist translation, terms and conditions, package tour contracts.

## 1. Introducción

En el marco de la sociedad de la información y de las TIC, Internet se presenta como una herramienta de trabajo imprescindible para el traductor ya que le ofrece de forma conjunta nuevas formas de comunicación y transmisión del conocimiento científico, además de facilitarle y mejorar sensiblemente sus labores de búsqueda documental. De hecho, como apunta Pinto Molina (2000: 2), la «revolución documental» permite al traductor abordar una mayor cantidad de información en un menor período de tiempo, lo cual contribuye a aumentar sensiblemente su rendimiento. Sin embargo, el traductor ha de ser capaz de discernir la información fiable de la no fiable, pues, si bien la Red constituye una herramienta valiosísima, adolece de una serie de inconvenientes, entre los que destaca la credibilidad o fiabilidad de la información (Toledo 2006).

En este sentido, el resumen automático constituye una ayuda muy atractiva para el traductor profesional, puesto que, en su labor documental, un resumen orientativo e indicativo de páginas web le lleva a seleccionar la información más pertinente y adecuada. Habida cuenta del escaso tiempo del que normalmente disponen los traductores, el resumen aparece como una solución que, al reducir la información de documentos electrónicos, ofrece múltiples ventajas y facilita su labor traductológica en gran medida.

De esta forma, en el presente artículo, nuestro objetivo es aproximarnos al resumen automático de páginas web en el par de lenguas inglés-español con una metodología específica: simular el proceso de documentación para un encargo de traducción al español de unas condiciones generales o cláusulas de contratos de viaje combinado de la modalidad cruceros. Para ello, se han empleado tres objetos de estudio:

1. Un ejemplo de condiciones generales o cláusulas de contratos de viaje combinado de la modalidad cruceros en inglés, extraído del macrocorpus multilingüe *Turicor*, el cual se elaboró en el seno del proyecto *TURICOR: Compilación de un corpus de contratos turísticos (alemán, español, inglés, italiano) para la generación textual multilingüe y la traducción jurídica*<sup>2</sup> (MCYT, BFF2003-04616, 2003-2006), dirigido por la Dra. Corpas Pastor. Este documento de condiciones generales sirvió como texto origen para el proceso de traducción de inglés a español.

2. Es posible acceder al proyecto *Turicor* a través de su URL: <http://www.turicor.com> [Consulta: 06-02-08].

2. Un sistema de resumen automático de páginas web denominado *SweSum*, creado en la Universidad de Estocolmo.
3. Una página web en español de la compañía *Pullmantur* con las condiciones generales de un viaje combinado de la modalidad de cruceros y que sirvió como texto de entrada para el sistema *SweSum*.

Una vez realizado el proceso de documentación, en concreto de búsquedas de textos paralelos, y de posterior resumen automático, analizamos los resultados en aras de estudiar las ventajas que aporta en los planos terminológico, temático, sintáctico y documental.

## **2. La sociedad de la información: el traductor como documentalista y tecnólogo**

En una sociedad marcada por la información y las comunicaciones, el traspase de lenguas y culturas desempeña un papel fundamental. Resulta evidente que la lengua materna constituye un elemento clave en nuestra identidad nacional y cultural y es preciso destacar que, en el contexto europeo, la pluralidad de lenguas es una realidad y, más si cabe, una seña de identidad (Toledo 2006: 5).

Como resulta obvio, las TIC, con Internet y sus servicios como máximos exponentes, pueden contribuir de diversas maneras al mantenimiento y desarrollo de una auténtica sociedad plurilingüe y, si bien resultan de suma importancia para el usuario medio, para el traductor o intérprete profesional han pasado a convertirse en imprescindibles, sobre todo en la fase de documentación previa a la traducción, donde el traductor hace las veces de documentalista. Esta exigencia de emplear Internet para documentarse se hace aún más evidente en el caso de la traducción especializada, como es la traducción jurídico-turística, pues el traductor ha de lidiar con lenguajes restringidos o sublenguajes. En consecuencia, como señala Gonzalo García (Gonzalo García y García Yebra 2004: 275) cualquier tipo de traducción, y en particular la traducción especializada, está marcada por la importancia que ejerce el conocimiento del campo y por la necesidad de la documentación.

No obstante, cabe apuntar que no sólo de Internet vive el profesional de la traducción, sino que también encontramos otros medios que facilitan la labor traductológica. Nos referimos a las tecnologías del lenguaje humano o tecnologías lingüísticas (Lavid 2005), cuyo aspecto más relevante es que, en su vertiente aplicada y práctica, han dado lugar a las «industrias de la lengua», es decir, productos o servicios basados en el tratamiento del lenguaje por personas o por máquinas o por una combinación de unas y otras. En este contexto, surgen las Tecnologías de la Traducción, que se podrían definir como una subdisciplina de las Tecnologías Lingüísticas, a medio camino entre la Traductología, el Procesamiento de Lenguaje Natural y la Lingüística Computacional.

En este campo de las Tecnologías de la Traducción, merecen especial atención los proyectos desarrollados en el panorama nacional, entre los que destacan los proyectos de gestión terminológica, lexicografía y corpus que se realizan en el Institut Universitari de Lingüística Aplicada (IULA) de la Universidad Pompeu Fabra (Barcelona); los diversos proyectos de terminología y terminografía del Instituto de Terminología Bilingüe y Traducción Especializada (ITBYTE) de la Universidad de Valladolid; los proyectos de terminología y ontologías de la Universidad de Granada, o los proyectos desarrollados en la Universidad de Málaga, a saber, *Turicor*, anteriormente señalado y *La contratación turística electrónica multilingüe como mediación intercultural: aspectos legales, traductológicos y terminológicos* (Ref. n. ° HUM-892, 2006-2009. Proyecto de Excelencia, Junta de Andalucía).<sup>3</sup>

Por tanto, queda patente que en los proyectos mencionados se aúnan las nuevas tecnologías, en este caso la Generación del Lenguaje Natural, con el Derecho turístico y, al mismo tiempo, con la Traducción. En el presente trabajo, se abordará un ámbito que se podría encuadrar como una modalidad de generación textual multilingüe (GLN) y, por extensión, perteneciente a la Lingüística Computacional (LC) y al Procesamiento del Lenguaje Natural (PLN),<sup>4</sup> que también puede aplicarse a otros campos, en especial a la Traducción, puesto que goza de un carácter totalmente interdisciplinar: la generación automática y multilingüe de resúmenes (GR). Dicha GR goza de un auge indiscutible debido a la cantidad ingente de información que encontramos en Internet, lo cual va estrechamente vinculado a las aportaciones documentales que le suponen al traductor.

En el ámbito nacional, la investigación más representativa acerca de la generación automática de resúmenes posee dos focos principales: por un lado, la Universidad de Alicante, donde destacan los trabajos de Acero et al. (2001) y, por otro, el IULA-UPF, con da Cunha (2005). De esta forma, es notorio que el resumen automático constituye un campo de interés; por otra parte, en lo que concierne al resumen automático de páginas web, cabe resaltar que no existen proyectos centrados en este tipo de resumen, si bien gran parte de los sistemas que existen en Internet son multilingües y, por tanto, resumen en varias lenguas, entre ellas en español, con

3. La URL del proyecto HUM-892 es la siguiente: <http://www.uma.es/hum892/> [Consulta: 05-02-2009].

4. La línea que diferencia el Procesamiento del Lenguaje Natural (PLN) de la Lingüística Computacional (LC) es más que delgada, pues, para muchos autores, no existe. Como muestra, Lavid (2005: 61) las considera como dos disciplinas diferentes que comparten el mismo objetivo; el grupo de investigación *Natural Language Processing* de Sheffield (<http://nlp.shef.ac.uk/>) defiende que el PLN engloba a la LC y, además, hace referencia a la historia de la Inteligencia Artificial y al estudio general de la función cognitiva de los procesos computacionales. Por su parte, el profesor Ruslan Mitkov, en una conferencia impartida en la Universidad de Málaga el 5 de mayo de 2008, aclaró que, a su juicio, ambos conceptos van unidos pero que el PLN se centra en la parte teórica mientras que la LC tiene el objetivo final de trabajar la parte práctica, la aplicación real, de forma que, gracias a esta última, se desarrollan las tecnologías lingüísticas. Por este motivo, en nuestro caso nos centraremos más en la LC, aunque es inevitable que ambas disciplinas vayan de la mano.

lo que su uso puede llevarse al terreno de la Traducción, en general, y al proceso de documentación, en particular, y convertirse en una útil herramienta que forme parte de las tecnologías de la *Computer-Aided Translation Technology* (CAT) o traducción asistida por ordenador.

### 3. El resumen automático

#### 3.1. Definición y tipos de resumen

En un principio, cabría pensar que la generación de lenguaje natural (GLN) y la generación de resúmenes (GR) constituyen una misma actividad, pues ambas aluden al hecho de generar lenguaje de forma automática y, de hecho, varios autores consideran que la GR trata de una disciplina dentro de la GLN (Lavid 2005: 216). Con el fin de observar las diferencias entre ambas, procederemos a leer una definición de cada una. La GLN, de acuerdo con Reiter y Dale (1997: 57) es:

*Natural language generation is the subfield of artificial intelligence and computational linguistics that is concerned with the construction of computer systems that can produce understandable texts in [...] human languages from some underlying non-linguistic representation of information.*

Por su parte, la GR es definida por Acero et al. (2001: 282) de la siguiente forma:

*Por generación automática de resúmenes de texto entendemos el proceso por el cual se identifica la información sustancial proveniente de una fuente (o varias) para producir una versión abreviada destinada a un usuario particular (o grupo de usuarios) y una tarea (o tareas).*

Por tanto, la GLN y la GR comparten el mismo objeto, la generación de un texto, pero se diferencian en la fuente de información, pues mientras en GLN se parte de una representación abstracta del conocimiento, tal y como apuntaban Reiter y Dale, en generación de resúmenes se toma como base un texto en lengua natural. En cualquier caso, consideraremos que la GR constituye una modalidad de la GLN y, por ende, entra de lleno dentro del campo de las tecnologías de la traducción (TT).

Por otra parte, los sistemas de resúmenes automáticos, como señalan numerosos autores (cf. Hahn y Mani 2000; Mani y Maybury 2001; Mani 2001), se dividen en dos grandes categorías: *extract* y *abstract*. Por un lado, como señala Teufel y Moens (2002) un *extract* o extracto es un conjunto de pasajes (desde palabras sueltas a párrafos enteros) extraídos del texto de entrada literalmente, que unidos forman un resumen. La mayoría de las aproximaciones al resumen automático toman esta concepción (Luhn 1959; Edmundson 1969), ya que resulta bastante útil en el campo

de la recuperación de información para hacerse una idea del contenido de un texto, aunque los resultados del texto como tal no sean de una gran calidad.

Por otro lado, un *abstract* o resumen es un texto generado nuevamente, producido por alguna representación interna que resulta después del análisis del texto de entrada. Esta idea surge con el motivo de mejorar la calidad final del texto del resumen, y para ello se lleva a cabo un proceso posterior, acertando, fusionando o revisando el material (Mani 2001).

Respecto a los tipos de resúmenes, la mayoría de los autores (cf. Mani 2001; Hahn y Mani 2000; Teufel y Moens 2002, entre otros) presentan una tipología de resúmenes similar. Para el presente trabajo tomaremos la clasificación de Acero et al. (2001) como modelo. Ésta es, quizá, una de las propuestas más completas, en la cual se clasifican los distintos tipos de resúmenes de acuerdo con tres variables: el propósito, el enfoque y el alcance.

Variable	Tipo de resumen		
Propósito	Indicativo (anticipa el contenido)	Informativo (incorpora información nueva)	Crítico (con opiniones o comentarios)
Enfoque	Genérico (recoge temas principales del documento)		Adaptado al usuario
Alcance	Un único documento		Un conjunto de documentos

**Tabla 1.** Clasificación de resúmenes (Acero et al. 2001)

Por ende, comprobamos que existen distintos tipos de resúmenes de acuerdo con el uso que se haga de él. En nuestro caso, creemos más oportuno optar por resúmenes indicativos y genéricos de varios documentos, si bien de un sólo documento resulta también de suma utilidad, pues, en este caso, resumiremos una página web en aras de comprobar la posible aplicación del resumen automático al proceso de documentación del traductor.

### 3.2. El resumen automático de páginas web

De acuerdo con Jatowt e Ishizuka (2004: 344), son dos las aproximaciones al resumen de páginas web: por un lado, la aproximación basada en el contenido y, por otro, la basada en el contexto. El primer tipo emplea el contenido textual de los documentos web en cuestión y no difiere mucho del resumen automático tradicional, si bien, dadas las características de los documentos HTML, esta aproximación tropieza con numerosos escollos, a saber: el lenguaje de marcado, la flexibilidad de la estructura o la existencia de elementos gráficos o multimedia.

La aproximación al resumen de páginas web basada en el contexto, la cual toma como modelo la estructura de hipertexto de Internet, no sólo resume el contenido

textual que aparece en la página que deseamos resumir, sino que también resume el contenido textual de las páginas web que enlazan con la página en cuestión (cf. Amitay y Paris 2000 y Sun et al. 2005). De esta forma, se emplean para el resumen los párrafos y otras unidades textuales que se encuentran en los enlaces activos. Sin duda, esta aproximación resulta más completa y exhaustiva que la basada en el contenido, aunque, no obstante, adolece de ciertos inconvenientes como, por ejemplo, el que se resuma información no del todo pertinente.

En cualquier caso, el resumen de páginas web constituye un campo de investigación con gran proyección de futuro y, de hecho, varios buscadores en línea como *Yahoo!*, *LookSmart* o *Google* ofrecen un listado de las páginas web como resultados junto con un breve extracto que describe el contenido (Sun et al. 2005: 194). Por tanto, su interés resulta patente para cualquier usuario pero, en particular, para aquellos que encuentran en Internet su principal fuente de información, de comunicación y de formación, como es el caso de los traductores.

### 3.3. *SweSum, un sistema de resumen automático de páginas web*

Tal y como señala Da Cunha (2005: 15), existen numerosos sistemas de generación de resúmenes automáticos de pago<sup>5</sup> para páginas web y para otros tipos de documentos, pero, en nuestro caso, nos centraremos en *SweSum*,<sup>6</sup> un programa desarrollado por Martin Hassel y Hercules Dalianis, de la Universidad de Estocolmo, porque, en primer lugar, se trata de un sistema gratuito y en línea y porque, en segundo lugar, sus excelentes resultados han quedado patentes en diversas publicaciones (Hassel y Dalianis 2005; De Smedt et al. 2005). *SweSum* se diseñó en un principio para trabajar con lenguas escandinavas (sueco, danés y noruego), pero posteriormente se le añadieron otras lenguas como inglés, francés, español, alemán y farsi, haciendo un total de ocho lenguas.

La arquitectura del sistema se basa en una combinación de métodos lingüísticos, estadísticos y heurísticos para ofrecer como resultado un extracto de las oraciones principales del texto en cuestión. De Smedt et al (2005: 3-5) señalan que el proceso de elaboración de resúmenes de *SweSum* consta de tres fases:

1. En la primera fase, se realizan dos acciones: la *tokenization* o fragmentación del texto en oraciones y la extracción de palabras claves.

5. En la Red hallamos numerosos sistemas de generación de resumen automático de pago, ya que se trata de productos comerciales. Entre dichos sistemas destacan los siguientes: *Brevity Document Summarizer* (disponible en <http://www.lextek.com/brevity/>), *Intelligent Miner for Text* (disponible en <http://www-306.ibm.com/software/data/iminer/>), *BusinessObjects Text Analysis* (disponible en [http://www.businessobjects.com/product/catalog/text\\_analysis/](http://www.businessobjects.com/product/catalog/text_analysis/)) y *Sinope Summarizer* (disponible en <http://www.sinope.info/en/Download>) [Consulta: 01-02-09].

6. La URL del sistema *SweSum* es la siguiente: <http://swesum.nada.kth.se/index-eng.html> [Consulta: 01-10-09].

2. En la segunda fase, se elabora un ranking con las oraciones que aparecen con mayor frecuencia.
3. En la última fase, se resume el texto.

Estas tres fases se corresponden a los tres pasos de elaboración de *abstracts* o resúmenes propiamente dichos, a saber, comprensión del texto, extracción de las partes más importantes y, por último, la generación del resumen. En la fase de extracción de los elementos de mayor relevancia, Danialis et al. (2003) especifican que *SweSum* asigna mayor índice de importancia a los títulos, oraciones en las que aparezcan las palabras claves obtenidas en la primera fase, oraciones con nombres propios o con cifras, etc. De hecho, esta forma de selección de información es típica en el ámbito de la recuperación de la información, pues, al igual que *SweSum*, la mayoría de los sistemas se basa en métodos estadísticos, entre los que destacan la posición en el texto (Edmundson 1969), la frecuencia de palabras y oraciones (Luhn 1959) o determinadas oraciones claves (por ejemplo, «es importante tener en cuenta») (Edmundson 1969). Asimismo, dichos sistemas recurren con frecuencia a la medida *tf.idf* («*Term-frequency times inverse document frequency*») (Hovy y Lin 1997; Mani y Bloedorn 1999). Comprender el método de selección de información es clave para determinar la importancia de los extractos obtenidos en el posterior proceso de documentación pues determinará por qué se han seleccionado ciertos elementos frente a otros.

En cuanto a la interfaz del sistema (cf. Figura 1), comprobamos que *SweSum* permite al usuario especificar la URL de la página que se desea resumir así como el idioma y el porcentaje del texto original que determina la longitud del extracto; además, el usuario tiene la posibilidad de proporcionar al sistema una serie de palabras claves con el fin de producir resúmenes, en parte, adaptados al usuario y de seleccionar la opción de mostrar las palabras claves y diversas estadísticas.

Una vez analizado el sistema de resumen automático *SweSum*, procederemos a analizar el documento de entrada para, a continuación, comprobar la utilidad de esta herramienta de la Lingüística Computacional en la fase de documentación del traductor.

#### **4. El inglés jurídico de las cláusulas de contratos turísticos: categorización del texto origen**

Como se ha especificado previamente, el texto origen es un ejemplo de condiciones generales o cláusulas de contratos de viaje combinado de la modalidad cruceros (cf. Figura 2), de la compañía *Princess Cruises*,<sup>7</sup> obtenido de *Turicor*. Estamos, por

7. La URL de la empresa *Princess Cruises* y de la cual se ha obtenido el texto origen es la siguiente: <http://www.princess.com/> [Consulta: 01-02-09].



tanto, ante un texto del Derecho turístico, el cual es definido por Auriolos Martín (2002: 30) como el «Derecho objetivo cuyas normas regulan las relaciones jurídicas nacidas de la actividad turística».

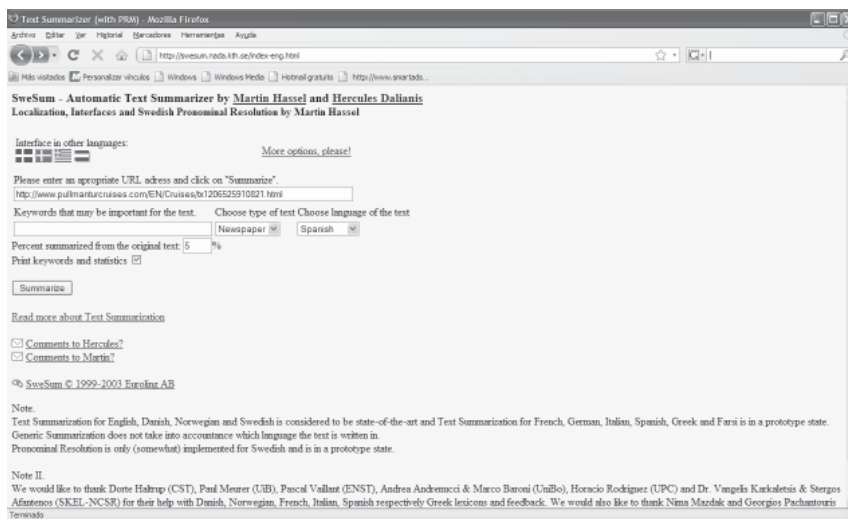


Figura 1. Captura de pantalla del sistema de resumen automático SweSum.

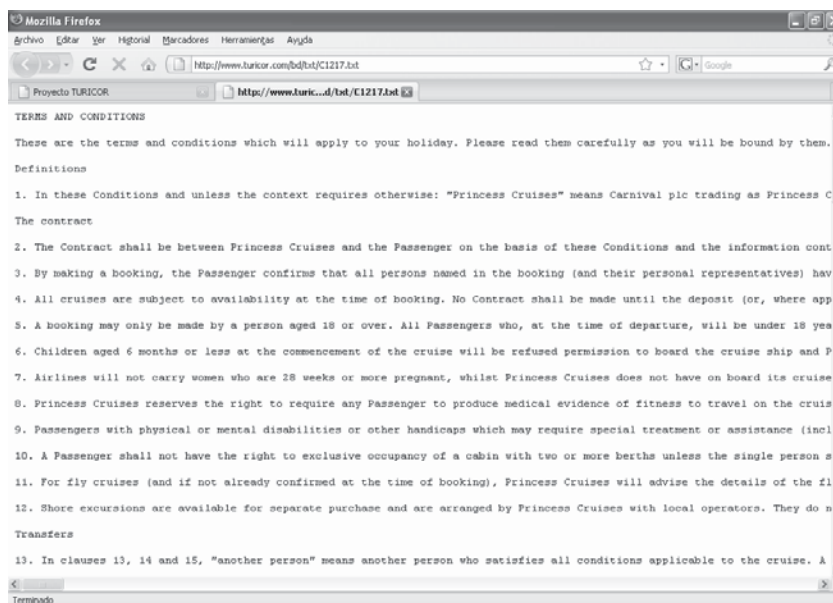


Figura 2. Captura de pantalla de las condiciones generales en inglés extraídas de Turicor.

Respecto a los contratos de viaje combinado en general y, por extensión, de los cruceros, la contratación de este tipo de viajes suele caracterizarse por la existencia de cláusulas o condiciones generales que se incorporan al contrato (Márquez Lobillo 2000: 272) y cuya presencia se hace especialmente necesaria debido a que el usuario ha de conocer al detalle las características del producto que contrata y compra.

#### ***4.1. El discurso jurídico de las condiciones generales de viaje combinado en lengua inglesa***

Tras categorizar el texto origen, no cabe duda de que nos encontramos ante un documento especializado del ámbito jurídico, para más señas, del Derecho turístico. En lo que concierne al discurso jurídico en inglés, hemos de destacar que, tal y como afirma Borja Albí (2000: 24), los textos legales ingleses presentan un léxico particular y limitado junto con unas características morfosintácticas y textuales propias que los distinguen claramente de cualquier otro tipo de texto.

Existe mucha literatura en torno al discurso jurídico en inglés, entre cuyos autores destacan Borja Albí (2000), Alcaraz Varó y Hughes (2001) y Alcaraz Varó (2002). Tras recopilar las características más frecuentes señaladas por los distintos investigadores, cabe destacar que los rasgos léxicos que aparecen en el texto jurídico que nos ocupa son las siguientes:

- Vocablos de origen francés o normando, como es el caso de *force majeure* (fuerza mayor), que aparece con gran frecuencia en los contratos de viajes combinados y en otros documentos jurídicos del turismo. Además de este término, otros vocablos de origen francés que sobresalen son aquellos acabados en *-age* y en el contrato aparecen, entre otros, *damage, damages*,<sup>8</sup> *coverage*, etc.
- Preposiciones sufijadas exclusivas del lenguaje legal tan características como *hereinafter, thereunder, thereafter, therewith*, etc.
- La redundancia expresiva, mediante la cual se repiten términos que son sinónimos parciales tanto en dobles (*null and void, sole and exclusive*) como en tripletes (*solemnly, sincerely and truly*). En este caso, encontramos un doblete en la oración siguiente: «To the extent that any provision in these Conditions is made null and void by the Warsaw Convention, the Athens Convention or any legislation compulsorily applicable or is otherwise unenforceable, it shall be void to that extent but no further».

8. La diferencia entre *damage* y *damages* es que la primera suele referirse a daños y perjuicios mientras que la principal acepción de la segunda es «indemnización por daños y perjuicios» (Alcaraz Varó 2002: 76).

—Los adjetivos de uso jurídico, pertenecientes muchos de ellos al lenguaje estándar, como es el caso de *absolute*, que, de acuerdo con Alcaraz Varó (2002: 78), se aplica en el discurso jurídico a lo «definitivo, categórico, firme» y en el texto encontramos el ejemplo de *absolute right* («Princess Cruises reserves the absolute right to decide whether or not to omit any such port(s) and/or to call at additional ports»).

En cuanto a los rasgos sintácticos, señalamos los siguientes:

- Frecuencia de la pasiva. Tal y como expone Borja Albi (2000: 40), la voz pasiva se utiliza sobre todo en acuerdos, contratos y documentos en los que lo importante es el resultado de la acción y no tanto los actores de la misma. En el texto de *Turicor*, al tratarse precisamente de cláusulas de contrato, hallamos numerosos ejemplos de voz pasiva, como, por ejemplo, el siguiente: «In the case of a booking by or on behalf of any such Passenger made before it could reasonably have been known that the Passenger would not be able to join the cruise by reason of this clause».
- Empleo de determinantes especiales con premodificación. Borja Albi (2000: 41) apunta que es una característica única del lenguaje legal el empleo de *such* como determinante sin ir acompañado del artículo indefinido, ejemplo de lo cual es la siguiente oración del texto: «Children aged 6 months or less at the commencement of the cruise will be refused permission to board the cruise ship and Princess Cruises shall have no liability whatsoever for any consequences of such refusal».
- Uso de grupos verbales peculiares. Se observa una elevada proporción de no finitos (formas impersonales) y de finitos del tipo *modal auxiliary* (normalmente *shall + be + past participle*), que es la fórmula más habitual para expresar obligación en los acuerdos legales en los que una parte se obliga («The Contract shall be between Princess Cruises and the Passenger on the basis of these Conditions and the information contained in the brochure, and shall be governed by English law»). Para expresar el derecho de las partes a hacer o no hacer determinados actos, se utiliza normalmente la construcción *may + infinitivo*, que aparece también con frecuencia en el texto origen: «Passengers should note that airlines may not allow transfers on scheduled flights and that a flight booking may have to be cancelled and rebooked».

A la vista de lo anterior, queda patente que el discurso jurídico en inglés emplea un lenguaje altisonante y arcaizante así como una sintaxis complicada. No obstante, cabe destacar que, en el caso de los contratos de viaje combinado, el lenguaje jurídico no resulta tan complicado y opaco como en otros documentos legales de mayor envergadura como sentencias, testamentos o leyes. La razón con más peso

que lleva a esta aparente simplicidad pueda deberse al origen de los documentos, ya que se trata de textos que aparecen en páginas web relacionadas con el mundo del Turismo y, por tanto, los destinatarios principales serán los propios consumidores. A pesar de ello, las características léxicas y sintácticas son propias del inglés del Derecho turístico, como ha quedado patente en la descripción previa realizada y, por tanto, habrán de ser extrapoladas a un discurso turístico-jurídico en español.

## **5. El proceso de documentación con ayuda del resumen automático y la traducción de condiciones generales**

### ***5.1. El proceso de documentación a través de Internet***

Tal y como hemos señalado anteriormente, la documentación suple en parte las carencias conceptuales, en este caso del campo del Derecho turístico, y ayuda al traductor a adquirir la competencia comunicativa necesaria para afrontar con éxito el encargo en cuestión. Esta capacidad para documentarse, también denominada «subcompetencia heurística» (Corpas Pastor 2001: 39), «competencia profesional e instrumental» (Rabadán y Fernández 2002: 396) o «infocompetencia» (Sales Salvador 2006: 13), ocupa, de acuerdo con Hurtado Albir (1996: 62), un lugar central en el conjunto de competencias, ya que permite al traductor adquirir conocimientos sobre el campo temático, sobre la terminología y sobre las ramas de funcionamiento textual del género en cuestión. Por tanto, el proceso documental se perfila como la piedra angular de la tarea previa a la labor traductológica.

En este sentido, debido al auge de las nuevas tecnologías, cabe remarcar que la documentación entendida en su concepción más tradicional está dejando paso a la denominada documentación electrónica en tanto aplicación de las tecnologías de la información y las comunicaciones. En este sentido, Internet ha surgido como un nuevo continente para la información, la comunicación y el intercambio (Seghiri Domínguez 2007: 24) y, gracias a ello, podemos alcanzar un gran caudal de información de diversa naturaleza.

No obstante, antes de proceder a documentarnos, es preciso determinar las necesidades informativas de nuestro texto en cuestión. En el caso de las cláusulas de contrato de viaje combinado de la modalidad crucero se podrían señalar las siguientes:

1. Temáticas y conceptuales, es decir, información, en este caso, sobre la legislación relacionada con los viajes combinados de la modalidad de crucero, explicación de su significado, aspectos legales;
2. Terminológicas, es decir, explicación del léxico y de la fraseología especializados;
3. Estilísticas, *i.e.*, conocimiento de cómo se redacta un contrato de esta tipología, tiempos y modos idiomáticos, el uso frecuente de ciertas construcciones lingüísticas, etc.;

4. Gramaticales, esto es, que se ajuste a las reglas gramaticales propias de estos textos especializados; y
5. Formales, es decir, estructura habitual, características ortotipográficas, etc.

En concreto, nos centraremos en las necesidades formales y estilísticas mediante la búsqueda de textos paralelos de acuerdo con la definición de Merlo Vega (2004: 330), esto es, «documentos, bien en lengua origen o en lengua meta, que traten de temas similares que el texto que se debe traducir», ya que son útiles tanto para la comprensión del texto como para la contextualización concreta de términos y conceptos. En nuestro caso, haremos búsquedas para textos en lengua meta, es decir, en español y, para más señas, de cláusulas de contrato de viaje combinado de la modalidad crucero.

Una vez determinada la información que se necesita, el traductor como documentalista ha de diseñar el plan de documentación y, al mismo tiempo, establecer en qué fuentes se puede acceder a los datos que precisa. En nuestro caso, optaremos por la búsqueda por palabras clave mediante buscadores, más concretamente el archiconocido motor de búsqueda *Google*. La búsqueda por palabras clave suele ser la más recurrida por su flexibilidad, aunque cuenta con el inconveniente de presentar un exceso de documentos no pertinentes o, como se ha venido denominando, «ruido documental» (Seghiri Domínguez 2007: 31-32). Por este motivo, la selección de palabras claves es esencial y, en nuestro caso, emplearemos la siguiente ecuación de búsqueda: cláusulas + contrato + viaje+ combinado + crucero. *Google* ofrece 1.530 resultados y, como es de esperar, la inmensa mayoría son condiciones generales para los cruceros. No obstante, consultar cada uno de los resultados obtenidos en la consulta supondría un tiempo considerable y es precisamente el tiempo uno de los problemas del traductor profesional y, además, una dificultad a la que deben aprender a enfrentarse los traductores en formación. Por ello, llegados a este punto, podría ser de gran utilidad el resumen automático de páginas web y es precisamente su aportación al proceso de documentación lo que analizaremos a continuación.

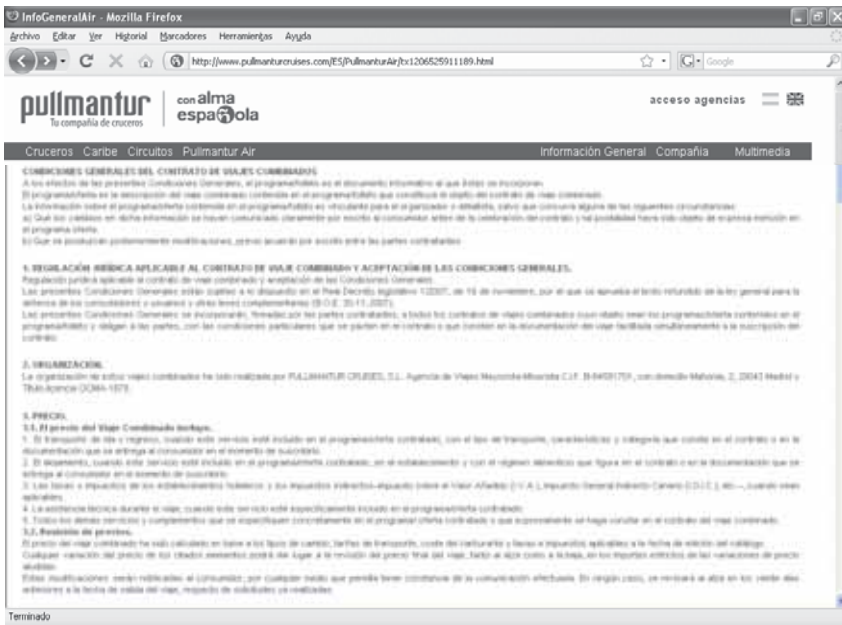
## 5.2. Empleo de *SweSum* para la documentación electrónica

Con el fin de comprobar y analizar la utilidad del resumen automático ante un número elevado de resultados, nos centraremos en un resultado de nuestra búsqueda, el ofrecido por la empresa *Pullmantur*,<sup>9</sup> una compañía de cruceros española, en el que aparecen detalladas todas las condiciones generales (cf. Figura 3) y que va a servir de documento de entrada para el sistema de resumen automático *SweSum*.

9. La URL de la empresa *Pullmantur* y de la cual se ha obtenido el texto objeto de análisis es la siguiente: <http://www.pullmanturcruises.com/ES/PullmanturAir/tx1206525911189.html> [Consulta: 02-02-09].

Una vez especificado el texto objeto de análisis para el resumen automático de páginas web, concretamos una serie de variables necesarias y ya especificadas en la descripción de *SweSum* (cf. Apartado 3.3):

- Porcentaje resumido del texto original: 20%
- Posibilidad de que aparezcan palabras claves y estadísticas: sí
- Lengua del texto de entrada: español



**Figura 3.** Captura de pantalla de las condiciones generales en español extraídas de Pullmantur.

Tras concretar las variables, procedimos a resumir de forma automática la página web en cuestión y el resultado fue el extracto que aparece en la figura 4 (cf. Figura 4), el cual pasamos a describir a continuación:

- El formato de la página web ha cambiado y no encontramos imágenes ni figuras en el extracto realizado por *SweSum*.
- Al tratarse de extractos y no de resúmenes propiamente dichos, se han seleccionado sin modificaciones diversos párrafos de la página web de entrada y cabe destacar la tendencia a optar por las oraciones iniciales y finales de cada apartado, debido en parte al método estadístico empleado que se basa en la posición en el texto (Edmundson 1969). Un ejemplo de ello lo

- encontramos en el apartado 3.2, sobre la revisión de precios donde se ha seleccionado la última oración de las cuatro que aparecían. «En ningún caso, se revisará al alza en los veinte días anteriores a la fecha de salida del viaje, respecto de solicitudes ya realizadas».
- Se han respetado todos los títulos del texto de entrada, pues como se ha especificado anteriormente, Danialis et al. (2003) indican que *SweSum* asigna mayor índice de importancia a los títulos en la fase de extracción de palabras claves.
  - Aparecen al final una serie de palabras claves, a saber, «viaje, consumidor, servicio, contrato, organizador», cuya utilidad se analizará más adelante.
  - Apenas encontramos errores ortotipográficos ni de gramática, aunque sí cierta incoherencia en diversas oraciones, como es el caso, de nuevo, del apartado sobre el precio (apartado 3.2.), donde hallamos el siguiente fragmento: «El precio del Viaje Combinado incluye. Revisión de precios. En ningún caso, se revisará al alza en los veinte días anteriores a la fecha de salida del viaje, respecto de solicitudes ya realizadas». Como puede comprobarse, al tratarse de extractos, las oraciones no guardan coherencia entre sí, si bien puede deducirse el sentido global del ejemplo.
  - El extracto es bastante extenso, si bien este problema puede solventarse reduciendo el porcentaje del texto de entrada en las variables de *SweSum*.

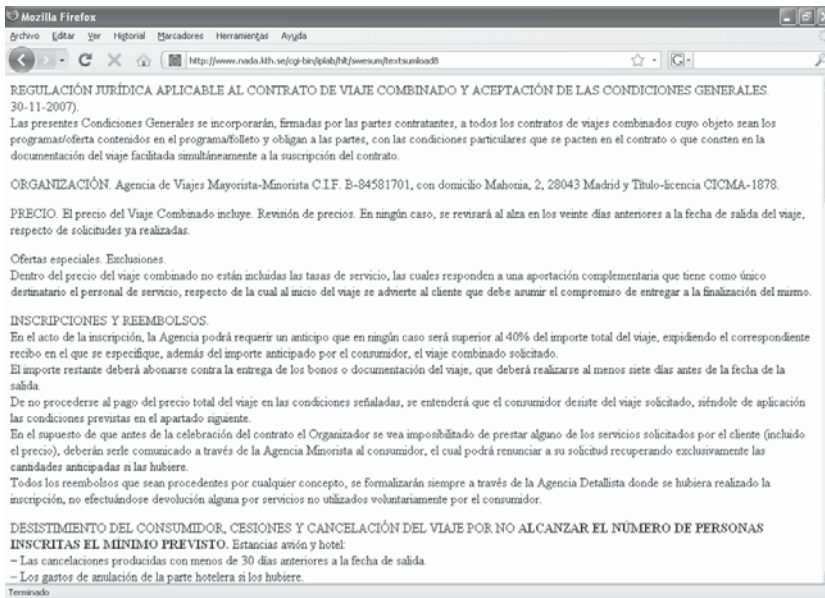


Figura 4. Captura de pantalla del resumen automático de la página de Pullmantur.

A grandes rasgos, podemos destacar que el resumen automático de la página web de *Pullmantur* resulta adecuado para hacerse una idea global de los contenidos de la página web completa y, por tanto, gracias al extracto de *SweSum*, puede determinarse que esa página web resulta de interés y merece la pena consultarla durante el proceso de documentación. A continuación, procedemos a analizar las aportaciones de este programa en concreto al proceso de documentación previo a la labor traductológica.

### ***5.3. Aportaciones del empleo de SweSum en el proceso de documentación previo a la traducción***

El empleo del sistema de resumen automático *SweSum* aporta, a nuestro juicio, las siguientes ventajas a la labor documental necesaria para nuestro encargo de traducción:

- El hecho de que *SweSum* seleccione los títulos del documento de entrada supone un beneficio para el traductor pues se mantiene el esquema del texto y, de este modo, sirve como modelo para la macroestructura que ha de seguirse a la hora de traducir. En este sentido, la consulta del extracto en español determina que, por ejemplo, no se suele emplear un primer apartado de definiciones como en las condiciones generales en inglés, donde encontramos el apartado *Definitions*, si bien sería preciso consultar un corpus representativo de condiciones generales de viaje combinado de la modalidad cruceros para poder extraer conclusiones sobre convenciones textuales de este tipo de documentos.
- Las palabras claves que aparecen al final del extracto («viaje, consumidor, servicio, contrato, organizador») tienen un gran valor terminológico y de referencia porque permiten al traductor:
  - a) conocer el contenido general de un documento;
  - b) acercarse a la terminología más característica del texto en cuestión, en este caso, del viaje combinado de la modalidad crucero;
  - c) partir de dichas palabras claves para otras búsquedas documentales, ya que sirven como ecuación de búsqueda;
  - d) emplear las palabras claves como equivalentes en español de las palabras claves en inglés del texto origen.
- De nuevo desde el punto de vista terminológico contrastivo, la comparación del extracto de *SweSum* con el texto original permite conocer la terminología empleada en la lengua española y compararla con la utilizada en inglés. Un ejemplo de ello lo encontramos en la tabla siguiente:



Término en lengua inglesa del documento original	Término en lengua española del extracto realizado por SweSum
<i>Terms and conditions</i>	Condiciones generales
<i>Contract</i>	Contrato
<i>Cruise</i>	Crucero
<i>Package Holiday/ Package Tour/ Package Travel</i>	Viaje combinado
<i>Booking</i>	Reserva
<i>Shore excursions</i>	Excursiones opcionales

**Tabla 2.** *Equivalentes terminológicos en las lenguas española e inglesa*

Como puede comprobarse, el extracto supone un acercamiento a la terminología empleada en el texto origen, con lo que la ayuda que ofrece a la hora de traducir queda patente. Además, este acercamiento terminológico indica ciertas particularidades de, en este caso, el inglés turístico, como es el hecho de que para expresar «viaje combinado» se empleen tres equivalentes en lengua inglesa (cf. Tabla 1).

- Desde el punto de vista sintáctico, el extracto permite hacer paralelismos entre los rasgos de la sintaxis del español y del inglés. Teniendo en cuenta las características del inglés jurídico detalladas anteriormente (cf. Apartado 4.2.), es posible comparar cómo se han traducido a la lengua española:
- Frecuencia de la voz pasiva: Como se ha visto anteriormente, la voz pasiva se utiliza sobre todo en acuerdos, contratos y documentos en inglés jurídico y, en el texto que nos ocupa, hallamos un ejemplo que se ha traducido al español mediante la voz activa:

«Shore excursions and shuttle services are arranged by Princess Cruises with local operators».	«La organización de estas visitas y excursiones dependen de prestatarios locales».
--	--

- Uso de grupos verbales peculiares: El empleo de construcciones impersonales con *shall* y *may* en inglés jurídico queda demostrado en la gran cantidad de oraciones que aparecen en el texto origen. En cuanto a su traducción al español, ambos grupos verbales suelen traducirse con el futuro simple de indicativo, tal y como podemos comprobar a través de estos dos ejemplos:

«Princess Cruises shall at all times be responsible for the full and proper performance of the Contract».	«Los Organizadores y los Detallistas de viajes combinados responderán de los daños sufridos por el consumidor como consecuencia de la no ejecución o ejecución deficiente del contrato».
«The Passenger may cancel the Contract at any time prior to the commencement of the cruise».	«En todo momento el consumidor y usuario podrá dejar sin efecto los servicios solicitados o contratados».

- Otro de los usos más interesantes de los extractos automáticos, y que ha quedado demostrado con los ejemplos anteriores, es su empleo como texto paralelo porque sirve para familiarizarse con la estructura de un determinado tipo de texto, en este caso, del discurso jurídico-turístico y, además, de dos lenguas, la española y la inglesa. De este modo, es posible comprobar en este texto cómo el inglés es más sintético frente al español a la hora de expresar ciertas ideas:

«These are the terms and conditions which will apply to your holiday. Please read them carefully as you will be bound by them».	«Las presentes Condiciones Generales se incorporarán, firmadas por las partes contratantes, a todos los contratos de viajes combinados cuyo objeto sean los programas/oferta contenidos en el programa/folleto y obligan a las partes, con las condiciones particulares que se pacten en el contrato o que consten en la documentación del viaje facilitada simultáneamente a la suscripción del contrato».
---	---

En definitiva, a la vista de lo anterior, queda demostrado que *SweSum* aporta una serie de ventajas al proceso de documentación y posterior consulta de resultados en los planos terminológico, temático, léxico, sintáctico y macroestructural.

#### 5.4. Utilidad del resumen automático en traducción

Una vez comentadas las ventajas que supone el empleo en concreto de un sistema de resumen automático de páginas web, resultaría de interés centrarse en los beneficios que aporta en general este tipo de programas a la labor documental del traductor, entre los cuales destacan los siguientes:

- Facilita la tarea de hallar información pertinente y fiable ante un maremágnum de resultados tras una búsqueda en Internet.
- Orienta al traductor a la hora de comprobar si una página web es de su interés para su labor de documentación, ya que podrá seleccionar dicha URL en función de la adecuación o no del resumen automático y, de este modo, decidir si desea consultar la página web completa.
- Ofrece una lista de conceptos claves que se corresponde, en general, con los términos de mayor relevancia del texto, con lo que supone una fuente de ayuda terminológica.
- Permite conocer la temática general del texto en cuestión.
- Sirve como acceso temático en la recuperación de información que, por otra parte, es multilingüe.
- Supone una fuente de actualización de conocimientos para el traductor acerca de las novedades en su campo de especialización.

Además de estas ventajas, los sistemas de resumen automático pueden serle útil al traductor en las ocasiones en las que, por imperativos técnicos, tenga que saber elaborar resúmenes de los documentos o páginas web que traduce, bien para organizar su propia memoria de traducción en beneficio propio o por necesidades de la misma traducción, «pues a veces el autor delega en éste la elaboración y traducción del resumen que acompaña a su trabajo» (Pinto Molina y Cordón 1999: 126).

En suma, consideramos que, gracias a un extracto realizado a partir de un sistema de resumen automático, el usuario, en este caso el traductor, puede tener una idea bastante aproximada del tema principal del documento HTML, pues se le ofrecen palabras claves y los fragmentos más relevantes. Entonces, en consonancia con la tipología de resúmenes de Acero et al. (cf. Apartado 3.1.), si el traductor busca un resumen crítico o informativo, los extractos llevados a cabo por *SweSum* o por cualquier otro sistema no sirven de gran ayuda. En contraste, si únicamente pretende encontrar una ayuda para discernir entre documentos adecuados o no para su labor documental, estimamos que los extractos automáticos pueden serle muy prácticos y convenientes, tal y como ha quedado demostrado.

## 6. Conclusiones

En el presente artículo se ha profundizado en la importancia de las nuevas tecnologías aplicadas a la Traducción, una disciplina que, por su carácter interdisciplinar, goza de las aportaciones de otras disciplinas y que, de este modo, consolida su presencia y relevancia tanto en el ámbito nacional como internacional.

De esta forma, basándonos en el proyecto de investigación *Turicor* desarrollado en la Universidad de Málaga, nos hemos aproximado al Procesamiento del Lenguaje Natural, y, más concretamente, a la Generación del Lenguaje Natural y a la Generación de Resúmenes. Mediante el sistema de resumen automático de páginas web *SweSum*, hemos procurado mostrar la aportación del resumen automático al proceso documental del traductor, el cual, ante el acopio informativo, presente particularmente en Internet, precisa cribar las fuentes de documentación de las que dispone. Por tanto, el traductor debe poseer conocimientos de fuentes de información y técnicas de selección, evaluación, búsqueda y recuperación documental.

En este contexto, cabe señalar que, tras emplear *SweSum* en el proceso documental previo a la traducción de condiciones generales de viaje combinado de la modalidad crucero, el resumen automático de páginas web supone una ayuda al traductor en diversos aspectos:

- En cuanto a la macroestructura textual, el extracto automático permite familiarizarnos con la estructura empleada en el tipo de texto en cuestión, con lo que actúa, entre otras cosas, como texto paralelo.

- Respecto a la terminología, el extracto automático, gracias sobre todo a su lista de palabras claves, supone un acercamiento a los términos del Derecho y del Turismo de mayor frecuencia y, por extensión, de mayor relevancia.
- En lo que concierne a la sintaxis, el extracto permite hacer paralelismos entre los rasgos sintácticos del español y del inglés en función de las características del discurso jurídico-turístico en ambas lenguas.

Por consiguiente, a la vista de los resultados, consideramos que, dentro del proceso de documentación previo a la labor traductora, la inclusión del resumen automático de páginas web y, por extensión, de documentos con otros formatos, constituye una herramienta adicional y muy recomendable que, de hecho, abre nuevas vías de investigación sumamente interesantes, entre las cuales destaca que el resumen automático pase a formar parte de las tecnologías de la *Computer-Aided Translation Technology* (CAT) y, por ende, del entorno conocido como el *translator's workbench*. No obstante, se trata de un terreno prácticamente ignoto aunque no exento de interés, en particular para la práctica de la traducción profesional.

Asimismo, otra vía de investigación en torno al resumen automático es la aplicación de esta herramienta en el campo de la didáctica de la Traducción, pues su uso puede beneficiar en gran medida a los traductores en ciernes que precisan agilizar sobremanera el proceso de documentación. Este uso cobra aún más sentido en la enseñanza de la traducción especializada, donde la labor documental ha de ser más exhaustiva en aras de adquirir la competencia comunicativa necesaria para afrontar con éxito el encargo en cuestión.

Como conclusión, el resumen automático de páginas web y de otros documentos, aparte de una novedosa aplicación de la Lingüística Computacional y el Procesamiento del Lenguaje Natural a la Traductología, constituye una herramienta en potencia a cuyos resultados se ha aproximado este artículo pero que aún han de ser explorados y, a la postre, aplicados con éxito a la Traducción en toda su extensión.

## Bibliografía

- Acero, Ignacio, Matías Alcojor, Alberto Díaz, José María Gómez y Manuel Maña (2001). Generación automática de resúmenes personalizados. *XVII Congreso de la Sociedad Española de Procesamiento de Lenguaje Natural (SEPLN'2001)*, 12-14 septiembre 2001, Jaén (España).
- Alcaraz Varó, Enrique (2002). *El inglés jurídico* (5ª ed.). Barcelona: Ariel Derecho.
- Alcaraz Varó, Enrique y Brian Hughes (2001). *El español jurídico*. Barcelona: Ariel Derecho.
- Amitay, Einat y Cécile Paris (2000). Automatically Summarising Web sites: is there a way around it? En *Proceedings of the ninth International Conference on In-*

- formation and Knowledge Management*, 173-179, McLean, Virginia, Estados Unidos.
- Aurioles Martín, Adolfo (2002). *Introducción al derecho turístico: derecho privado del turismo*. Madrid: Tecnos, D.L.
- Borja Albí, Anabel (2000). *El texto jurídico inglés y su traducción al español*. Barcelona: Ariel Lenguas Modernas.
- Corpas Pastor, Gloria (2001). La competencia traductora. A propósito del texto médico especializado. En *Traducción y nuevas tecnologías. Herramientas auxiliares del traductor. Encuentros en torno a la Traducción 4*, Valero Garcés, Carmen e Isabel De la Cruz Cabanillas (eds.), 11-26. Alcalá de Henares: Universidad de Alcalá de Henares, Servicio de Publicaciones.
- Da Cunha Fanego, Iria. (2005). *Hacia un modelo lingüístico de resumen automático de artículos médicos en español*. Proyecto de investigación, Universidad Pompeu Fabra, Instituto Universitario de Lingüística Aplicada.
- de Smedt, Koenraad, Anja Liseth, Martin Hassel y Hercules Dalianis (2005). How short is good? An evaluation of automatic summarization. En *Nordisk Sprogteknologi 2004. Årbog for Nordisk Språkteknologisk Forskningsprogram*, H. Holmboe (ed.), 267-287, Museum Tusulanums Forlag.
- Edmundson, H. P. (1969). New Methods in Automatic Abstracting. *Journal of the ACM*, 16(2), 264-285.
- Gonzalo García, Consuelo y Valentín García Yebra (eds.) (2004). *Manual de documentación y terminología para la traducción especializada*. Madrid: Arco/ Libros.
- Hahn, Udo e Inderjeet Mani (2000). The Challenges of Automatic Summarization. *IEEE Computer*, 33(11), 29-36.
- Hassel, Martin y Hercules Dalianis (2005). Generation of Reference Summaries. En *Human Language Technologies as a Challenge for Computer Science and Linguistics*, Abril 21-23, Poznan, Polonia.
- Hovy, Eduard y Chin Yew Lin (1997). Automated Text Summarization in SUMMARIST. En *Advances in Automatic Text Summarization*, I. Mani y M. Maybury (eds.), 18-24. Cambridge: Massachusetts Institute of Technology.
- Hurtado Albir, Amparo (1996). La enseñanza de la traducción directa 'general'. Objetivos de aprendizaje y metodología. En *La enseñanza de la traducción*, Amparo Hurtado Albir (ed.), 31-56. Castellón: Universitat Jaume I. Col.lecció Estudis sobre la Traducció, 3.
- Jatowt, Adam y Mitsuru Ishizuka (2004). Web Page Summarization Using Dynamic Context. En *Proceedings of the 13th international World Wide Web*, 344-345. Nueva York, Estados Unidos.
- Lavid, Julia (2005). *Lenguajes y nuevas tecnologías. Nuevas perspectivas, métodos y herramientas para el lingüista del siglo XXI*. Madrid: Cátedra.

- Luhn, H. P. (1959). The Automatic Creation of Literature Abstracts. *IBM Journal of Research and Development*, 2(2), 159-162.
- Mani, I. y Bloedorn, E. (1999). «Summarizing Similarities and Differences Among Related Documents». *Information Retrieval* 1(1), 35-67.
- Mani, Inderjeet y Mark T. Maybury (2001). *Advances in Automatic Text Summarization*. Cambridge: Massachusetts Institute of Technology.
- Mani, Inderjeet (2001). *Automatic Summarization*. Philadelphia: John Benjamins B.V.
- Márquez Lobillo, Patricia (2000). La celebración del contrato de viaje combinado a través de Internet. En *Derecho y turismo. III Jornadas de Derecho turístico de Adolfo Auriolés Martín* (coord.) Sevilla: Consejería de Turismo y Deporte.
- Merlo Vega, José Antonio (2004). Uso de la documentación en el proceso de traducción especializada. En *Manual de documentación y terminología para la traducción especializada*. Consuelo Gonzalo García y Valentín García Yebra (eds.), 309-336. Madrid: Arco Libros, S. L.
- Pinto Molina, María (2000). Documentación para la Traducción en la sociedad de la información. *XV Colloque Association Internationale de Bibliologie*. Salamanca, AIB.
- Pinto, María y José Antonio Cordón (eds.) (1999). *Técnicas documentales aplicadas a la traducción*. Madrid: Editorial Síntesis.
- Rabadán, Rosa y Purificación Fernández Nistal (2002). *La traducción inglés-español: fundamentos, herramientas, aplicaciones*. León: Universidad de León, Servicio de Publicaciones.
- Reiter, Ehud y Robert Dale (1997). Building Applied Natural-Language Generation Systems. *Journal of Natural-Language Engineering*, 3:57- 87.
- Sales Salvador, Dora (ed.) (2005). *La Biblioteca de Babel: documentarse para traducir*. Albolote (Granada): Comares.
- (2006). *Documentación aplicada a la traducción: presente y futuro de una disciplina*. Gijón: Ediciones Trea.
- Seghiri Domínguez, Míriam (2007). *Compilación de un corpus trilingüe de seguros turísticos (español-inglés-italiano): aspectos de evaluación, catalogación, diseño y representatividad*. Málaga: Universidad de Málaga.
- Sun, Jian-Tao, Dou Shen, Hua-Jun Zeng, Qiang Yang, Yuchang Lu y Zheng Chen (2005). Web-page summarization using clickthrough data. En *Proceedings of the 28th Annual International ACM SIGIR conference on Research and Development in Information Retrieval*, 194-201.
- Teufel, Simone y Marie-Francine Moens (2002). Summarizing scientific articles: Experiments with relevance and rhetorical status. *Computational Linguistics*, 28(4), 409-445.
- Toledo Báez, María Cristina. (2006). *Aproximación a la generación automática multilingüe de resúmenes*. Trabajo de investigación, Universidad de Málaga.