



En este artículo presentaremos los aspectos teóricos y prácticos de las investigaciones en traducción realizadas sobre corpus bilingües informatizados. En concreto revisaremos el proceso de construcción y explotación del corpus *TECTRA* (*Textos para Estilística Comparada e TRAducción*), que cuenta en la actualidad con 14 textos en inglés y sus correspondientes traducciones al gallego (más de 1.100.000 palabras) y que sirve como principal herramienta y fuente de datos a una serie de estudios sobre Estilística comparada que llevamos a cabo en la Universidade de Vigo.

# *TECTRA*: teoría y práctica de las investigaciones con corpus en el marco de los Estudios de Traducción

*This paper introduces the practical and theoretical implications of the use of computerized bilingual corpora in Translation Studies. We review the guidelines for the construction and exploitation of the corpus TECTRA (Textos para Estilística Comparada e TRAducción – Texts for Comparative Stylistics and Translation), which at present contains 14 English texts and their Galician translations (more than 1,100,000 words). This corpus is the main tool and data source for a Comparative Stylistics project being developed at the University of Vigo.*

ALBERTO ÁLVAREZ LUGRÍS  
*Universidade de Vigo.*

o INTRODUCCIÓN



186

Según la concepción de los Estudios de traducción en general y de la Estilística comparada en particular como disciplinas empírico-descriptivas y explicativas (cf. Toury 1995, Álvarez Lugrís 1998, 1999a, 1999b, en prensa-a), cualquier trabajo de investigación realizado dentro de estos campos que pretenda tener validez epistemológica y ontológica así como posibilidades de aplicación futura tanto en el desarrollo del aparato teórico de la disciplina como en didáctica, tendrá que basarse por fuerza en el estudio de una gran cantidad de datos empíricos reales. Esto quiere decir que las investigaciones relacionadas con los Estudios de traducción tienen que recurrir al uso de corpus como objeto de estudio y como fuente de: a) hipótesis de trabajo desarrolladas a partir de la observación de hechos de traducción; y b) ejemplos y contraejemplos que sirvan para confirmar o falsear las primeras hipótesis, siguiendo así un método de trabajo que podemos denominar helicoidal, basado en la retroalimentación continua entre los componentes descriptivo y teórico (Toury 1995).

En las páginas que siguen expondremos los criterios de diseño y construcción del corpus *TECTRA* y los métodos de explotación del mismo, desarrollando las cuestiones metodológicas que afectan al tratamiento de los fenómenos microestructurales estudiados en Álvarez Lugrís (1999a, en prensa-a). Para terminar, incluimos como apéndices datos sobre la composición y tamaño del corpus, proceso de alineamiento, marcación de los textos y programas de análisis textual.

I EL CORPUS *TECTRA*

Las investigaciones de base empírica realizadas por medio de corpus informatizados son de especial utilidad para la Estilística comparada por lo que suponen de ruptura con la tradicio-

nal tendencia de los estudiosos de la traducción a la atomización del objeto de estudio y a la introspección como única fuente de datos. Y si es importante evitar estas dos distorsiones por reducción, no lo es menos superar el sesgo que supone estudiar un texto traducido (TT) de un único traductor, con sus particularidades e incluso manías. Por otra parte, el dato resultante de una observación solamente se puede convertir en hipótesis de trabajo cuando se trata de la descripción de un fenómeno recurrente a lo largo de una serie de análisis; a su vez, la hipótesis sólo podrá ser falseada, o por el contrario elevada a la categoría de norma (es decir, de comportamiento normal) del proceso traductivo, cuando es comprobada en un corpus más amplio de casos semejantes y sujetos a las mismas condiciones. Quiere esto decir que la parcialidad y la secuencialidad que caracterizan toda comparación (Álvarez Lugrís en prensa-a, 82) exigen el uso de un corpus relativamente amplio en nuestras investigaciones si pretendemos llegar a resultados válidos y útiles. Por eso, debido a motivos prácticos, el trabajo con corpus de cualquier tipo sólo es operativo cuando lo realizamos con la ayuda de un ordenador. Sería imposible enfrentarse manualmente a un corpus de tan sólo unos cientos de miles de palabras sin una herramienta como ésta. La Lingüística informática y la Lingüística de corpus<sup>1</sup> —disciplinas relativamente bien desarrolladas en los últimos años— ofrecen por este motivo a los Estudios de traducción una herramienta de trabajo de uso casi obligado.

La adopción de corpus como método y herramienta de trabajo, sin embargo, no fue fácil ni inmediata. En un primer momento el uso de corpus contó con numerosos detractores, como atestiguan las descalificaciones

<sup>1</sup> Mantenemos la denominación lingüística de corpus por ser la más extendida, pero pensamos que no es «a very good name; it is an odd discipline that is called by the name of its major research tool and data source» (Aarts 1998:1).



hechas por Chomsky y la escuela generativista, más interesados en la búsqueda de reglas de transformación para la lengua inglesa que en los estudios extensivos de cualquier fenómeno lingüístico. A pesar de todo, en los últimos treinta años son cada vez más frecuentes las investigaciones basadas en corpus y/o textos electrónicos dentro de cualquier rama de las Humanidades (cf. Hockey 1994:5) —especialmente en Lingüística descriptiva, estudios de autoría, Lingüística forense y traducción, particularmente la traducción automática— de forma que es relativamente fácil encontrar corpus adecuados para casi cualquier tipo de investigación o, si no, construir uno para fines específicos, como en nuestro caso. En realidad, hoy en día contamos ya con una metodología muy sólida en Lingüística de corpus que incluso podemos calificar como refinada en el caso de los Estudios de traducción (Sinclair 1992:395), lo que nos permite desarrollar colecciones de textos que respondan a nuestros objetivos e intereses.

### 1.1 *Diseño y composición de TECTRA*

En su estudio sobre los aspectos cualitativos y cuantitativos que se deben tener en cuenta cuando se selecciona un corpus de textos para investigaciones en traducción, van Doorslaer (1995:251) comenta que, por lo general, los distintos autores utilizan corpus formados por textos buscados al azar, sin un criterio teórico ni metodológico definido. Es posible que la simple búsqueda de ejemplos que confirmen o desmientan una hipótesis sí se pueda realizar de tal forma, pero el estudio sistemático de un determinado fenómeno, y sobre todo el estudio de las normas de traducción, requiere la elaboración de unos sólidos criterios de diseño del corpus y de selección de los textos que lo compongan. Si partimos de la idea de que las normas de traducción son una función del contexto histórico-social al que pertenecen el TT y el traductor (Toury 1995), tendremos, por cohe-

rencia, que aceptar que debemos controlar las distintas variables que puedan afectar a nuestro objeto de estudio y que influyan en las manifestaciones específicas de las normas. Esto implica que nuestro corpus debe ser *representativo* del *universo* que queremos estudiar, ya que «unless the corpus is representative, it is *ipso facto* unreliable as a means of acquiring knowledge» (Summers 1993:186).

Por *universo* entendemos el conjunto de todos los individuos que comparten la característica o características estudiadas y para su delimitación tendremos que atender a parámetros como lengua, época histórica (del TO y/o del TT), género, traductor, autor, etc. La noción de *representatividad*, en cambio, resulta más complicada de definir, ya que, en primer lugar, dependerá del número y de la clase de fenómenos que se quiera estudiar. En nuestro caso, nos decidimos por adoptar el punto de vista de Della Summers cuando describe el corpus Longman/Lancaster, que considera que la representatividad no es

the full range of variability of the target population... What we mean by *representative* is covering what we judge to be the typical and central aspects of the language and providing enough occurrences of words and phrases for the lexicographers, and other students of language, to believe that they have sufficient evidence from the corpus to make accurate statements about lexical behaviour (1993:186-187; cursiva en el original),

definición que adaptada a los Estudios de traducción podría reformularse de la siguiente forma: *un corpus representativo es aquel que contiene lo que el investigador considera los aspectos típicos básicos del universo de traducciones estudiado y que presenta casos suficientes para extraer de él deducciones sobre el comportamiento traductivo* (Álvarez Lugrís en prensa-a:104). La representatividad total es, desde luego, un ideal utópico e inalcanzable pero que, sin embargo, debemos



tener como objetivo al que tender. La noción de representatividad lleva asociada siempre la de *balance o equilibrio*, que podemos definir como un punto deseable de compensación entre los distintos componentes del corpus, de forma que no predomine en él ni un género, ni un determinado autor o traductor, ni una época histórica, etc., dependiendo del fenómeno estudiado.

La representatividad y el equilibrio fueron, por lo tanto, los criterios básicos que guiaron la selección de los textos que forman parte de *TECTRA* en la actualidad. El universo investigado es el formado por las traducciones de novelas inglesas y estadounidenses al gallego publicadas entre los años 1989 y 1994.

### 1.2 Criterios de selección de textos

Una de las debilidades teóricas y metodológicas que con más insistencia se le reprochan a los Estudios descriptivos de traducción que Toury diseña en *Descriptive Translation Studies and Beyond*, es la indefinición de la noción de traducción (como producto), noción que Tirkonen-Condit (1997:56) califica como demasiado borrosa. En efecto, si tomamos al pie de la letra la definición de Toury (1995:26-29), según la cual traducción es todo texto que sea aceptado como tal dentro de la cultura receptora, encontraremos no pocos problemas para seleccionar TTs válidos para nuestro análisis: ¿se les debe suponer la misma representatividad a todas las traducciones? ¿podemos/debemos estudiar traducciones deficientes o pseudotraducciones o adaptaciones de la misma forma? ¿debemos otorgar la consideración de traducciones —y, por lo tanto, de objeto de nuestro estudio— a los productos de los estudiantes de traducción? La propuesta de Toury presenta demasiados problemas operativos, por lo que preferimos adoptar la definición de traducción de Chesterman (1997:59-63), según la cual es la cultura receptora la que decide y acepta qué es traducción y qué no lo es, pero sólo entre un número limitado de textos (y por lo tanto no

cualquier texto): los que los traductores presentan como traducciones. Según Chesterman, los traductores son personas a las que su sociedad les reconoce autoridad y conocimientos suficientes para traducir y en los que la sociedad confía, por lo que

the target culture designates certain texts as translations largely on the implicit say-so of translators: the relation between the target culture and its translators is therefore fundamentally one of trust ... it is the target culture which nevertheless *confirms* translation status (1997:59; cursiva en el original).

De esta forma se llega a una solución intermedia: las traducciones se siguen concibiendo, como mantiene Toury desde una perspectiva descriptiva, como fenómenos exclusivos de la cultura receptora, pero su definición ya no es tan escurridiza, por lo que se evitan así sesgos y deficiencias en los estudios: una postura exclusivamente descriptivista «necessarily overlooks much of the motivation for studying translation behaviour in the first place and inevitably leads to a rather one-legged theory» (Chesterman 1993:3-4).

Para *TECTRA* escogimos textos de naturaleza no cuestionable, textos totalmente aceptados como traducciones y que no crean dudas sobre su naturaleza de textos traducidos. Su aceptabilidad está fuera de toda duda, ya que se trata de títulos de éxito (varios de los cuales, como *A chammada da selva*, *A perla* y *O vixía no centeo*, han visto ya la cuarta o quinta edición, algo relativamente infrecuente en las traducciones al gallego) publicados por las editoriales gallegas con más volumen de publicación y venta: Galaxia, Xerais, Sotelo Blanco, Edicións Positivas. Además, nueve de los catorce títulos que componen *TECTRA* están publicados en distintas colecciones de literatura infantil y juvenil, géneros que acaparan la inmensa mayoría de las traducciones editadas en la última década en nuestro país.

A pesar de que ciertos autores como van



Leuven-Zwart afirman que «since in most cases it is neither possible nor indeed necessary to compare the two texts in their entirety, the microstructural analysis will be limited to a few passages chosen at random» (1989:155), creemos, por el contrario, que un aspecto básico de todo estudio basado en un corpus es la selección del mismo. La validez de nuestras hipótesis y del resultado de nuestras investigaciones depende en gran medida de los textos elegidos para el análisis; en consecuencia, el corpus se debe elegir siempre en función de la naturaleza y de los objetivos de nuestra investigación. Esto no supone, sin embargo, falsear la investigación escogiendo únicamente aquellos textos que sirvan para confirmar hipótesis establecidas *a priori*; muy al contrario, lo que se pretende es que los TTs (y sus respectivos TOs) seleccionados sean *relevantes* para el tipo de estudio que se va a realizar. Un corpus elegido al azar y sin criterios que controlen las distintas variables probablemente no dará como resultado más que una serie de datos inconexos e incluso contradictorios.

El proyecto *TECTRA* nació con el objetivo de estudiar las traducciones al gallego de novelas en lengua inglesa realizadas entre los años 1989 y 1994, para lo cual escogimos catorce traducciones y sus correspondientes originales. La selección de los textos responde a una serie de criterios que repasamos enseguida.

### 1.2.1 Elección de textos literarios

Luc van Doorslaer nos advierte que

Although literary texts are certainly influenced by collective norms too, we cannot reach the same prediction level, as our objects of study will always remain, precisely because of their specific literariness, more individual acts or products (1995:250; la cursiva es nuestra),

es decir, que el texto literario, a pesar de estar sujeto también a normas socio-culturales, supo-

ne una desviación con respecto a la lengua estándar no literaria, precisamente por su carácter literario. Esto es cierto, pero sólo si tomamos el estándar no literario como único modelo e invariante de una lengua en un momento histórico determinado; o si pretendemos hacer de ese estándar no literario el único objeto de estudio de una investigación. Sin embargo, hay ocasiones en las que el recurso a los textos literarios no sólo es inevitable desde un punto de vista operativo, sino que es deseable desde un punto de vista metodológico.

Dada la especial situación de lengua minorizada en la que se encuentra el gallego y teniendo en cuenta el panorama de las traducciones a nuestra lengua<sup>2</sup> (que está aún en un doble proceso de normalización-normativización y de fortalecimiento de la variedad literaria) es prácticamente imposible construir un corpus relativamente extenso de traducciones no literarias. A pesar de que contamos ya con la traducción gallega de varios libros de ensayo y con la traducción mensual de *O Correo da Unesco* (con un número de artículos relativamente grande traducidos directamente del inglés), el conjunto de los textos no supone, hoy en día, un volumen extenso, y mucho menos en el año que comenzamos a elaborar *TECTRA*. Por otra parte, tampoco la traducción de noticias o artículos de prensa supone un número significativo de textos que permita realizar un estudio serio. Se traducen aún pocos manuales o libros de texto/divulgación, por lo que tampoco aquí podríamos encontrar terreno para nuestro trabajo.

El sesgo que supone la extraordinariamente grande proporción de producción literaria, y concretamente de narrativa, en relación con el número total de textos editados en gallego es también evidente si tomamos la producción original. El *CORGA* (*Corpus de referencia do galego*

<sup>2</sup> Sobre el papel de las traducciones en el desarrollo de la evolución de la gallega, véase Álvarez Lugris en prensa-c).



*actual*), que está elaborando el Centro Ramón Piñeiro de Investigacións en Humanidades, cuenta con un 44% de textos de narrativa, porcentaje que se eleva al 66% si incluimos otros géneros literarios (Álvarez Lugris 1997:8).

La literatura, y en concreto la narrativa, parece ser en este momento la única vía de estudio de la traducción del inglés al gallego. Y no hacemos esta afirmación con desánimo y resignación, como si de un mal inevitable se tratase, ya que estamos de acuerdo con diversos autores como Baker (1993:238) o Koller (1995:203) en que la literatura es un sistema textual que merece ser investigado por derecho propio. Es más, coincidimos totalmente con la afirmación del profesor Eugenio Coseriu de que la lengua literaria no es en absoluto una desviación con respecto a la lengua *corriente*; en todo caso, según nos dice el profesor rumano, la lengua corriente sería una desviación, una *reducción*, de la totalidad del lenguaje y de su capacidad expresiva:

el lenguaje poético [= literario] resulta ser, no un uso lingüístico entre otros, sino lenguaje simplemente (sin adjetivos): realización de todas las posibilidades del lenguaje como tal... Por un lado, el lenguaje poético no presenta una reducción del lenguaje; por otro, no se añade propiamente ninguna función, ya que las diferentes posibilidades que en tal lenguaje se actualizan pertenecen ya al lenguaje como tal... el lenguaje poético representa la plena funcionalidad del lenguaje (1971:203).

Por otra parte, y como defiende János Kohn (1996:47), el estudio de los textos literarios traducidos, junto con la perspectiva que proporciona el uso de grandes corpus, contribuiría a reducir las tradicionales diferencias existentes entre los estudios de base lingüística y los de base literaria, integrando los dos dentro del campo general de los Estudios de traducción. De esta forma, las radicales diferencias entre las perspectivas lingüística y literaria, recogidas

por ejemplo en Klaudy (1991:94-95), dejarían de tener sentido y de crear unas oposiciones que nunca existieron en realidad o, mejor dicho, que nunca debieron existir: el estudio de traducciones literarias, por ejemplo, no tiene por qué centrarse exclusivamente en la función estética; no tiene por qué limitarse a hacer juicios de valor sobre las traducciones, o a establecer recomendaciones sobre cómo se debe traducir, etc.

Tres son, por lo tanto, los motivos que determinaron el uso de traducciones literarias en *TECTRA*: a) un motivo epistemológico, ya que la lengua literaria es tan válida como cualquier otra variedad para llevar a cabo, desde la óptica de la Estilística comparada, análisis microestructurales sobre los modos de traducir; b) un motivo práctico: los literarios son el tipo de textos más abundantes dentro del repertorio de traducciones al gallego; c) del anterior motivo práctico se deriva otro de alcance epistemológico: si los textos literarios son los más abundantes, entonces lo más relevante para el estudio de las normas de traducción del inglés al gallego serán, precisamente, los textos literarios. Esta afirmación tautológica viene a confirmar que la literatura *qua* literatura no debe ser descartada de ningún tipo de estudio que tenga como objeto la lengua y sus usos o representaciones. En el caso de las traducciones al gallego, de este tipo de textos<sup>3</sup> podremos extraer más conclusiones y más fiables sobre pautas de comportamiento traductivo —e insistimos, no sólo literario—, criterios de traducción, políticas editoriales, etc., que de ningún otro tipo de texto.

### 1.2.2 Criterios de selección cronológica

Una de las pautas que nos fijamos en nuestro estudio fue la progresión del análisis: partiendo de una observación particular pretendemos lle-

<sup>3</sup> En la bibliografía figura la relación completa de textos que componen *TECTRA*.



gar a la formulación de una hipótesis mediante la repetición de las observaciones. Tal es también el fundamento del método helicoidal de Toury (1995:36-38), que garantiza el avance del conocimiento gracias a la *reducción* del objeto de estudio en un primer momento y a la *repetición* del análisis en fases sucesivas. La reducción del campo de estudio guió nuestra selección de textos para *TECTRA*: decidimos limitarnos al estudio de las traducciones publicadas en un período de cinco años. Se trata de un período corto para intentar buscar una cierta unidad en los comportamientos de los traductores; pensamos que la dispersión temporal podría suponer deformaciones en el resultado final. En todo caso, sería preciso que en el futuro se lleven a cabo estudios sobre otros períodos igualmente cortos para comprobar si existen o no diferencias entre las soluciones adoptadas por los distintos traductores. Pretendemos hacer así distintas calas (cada cinco años) en las formas de traducir más actuales dentro de un panorama editorial en ebullición y en continuo cambio, en el que surgen cada vez más traductores e incluso editoriales. La idea inicial era que al terminar esta investigación terminase también un nuevo período de cinco años (1994-1999) con el que poder comenzar a trabajar para comprobar si se repiten las pautas de comportamiento o si por el contrario aparecen otras nuevas. En este momento estamos comenzando a ampliar *TECTRA* con nuevas traducciones publicadas en ese segundo período.

Uno de los textos escogidos, sin embargo, no entra en el período 1989-1994. Se trata de *A chamada da selva*, publicado en 1982. Con la inclusión de este texto en *TECTRA* pretendemos someter nuestra investigación a un control mediante el contraste del resto del corpus con una traducción de un período anterior. Podemos avanzar que *A chamada da selva* ofrece ciertas diferencias dignas de mención con respecto a los fenómenos estudiados en las otras trece novelas. ¿Por qué esta novela y no otra?

En primer lugar porque unos meses antes de su publicación, en octubre de 1982, la aparición de las *Normas ortográficas e morfológicas do idioma galego* (cf. RAG -ILGA 1982) marca un hito en la historia de nuestra lengua; es la primera vez que el gallego tiene un estándar oficial, una norma culta unificada para la lengua, sobre todo la escrita. *A chamada da selva*, publicada en diciembre de 1982 es, por lo tanto, una de las primeras traducciones que utiliza el gallego normativo, del que sabemos que el traductor Gonzalo Navaza es un magnífico conocedor. Por otra parte, esta novelita de Jack London fue uno de los libros más reeditados en el ámbito gallego en la década de 1980, tanto en traducción como en producción original, lo que demuestra su buena aceptación por parte del público lector y su utilidad como representante de las traducciones realizadas en esa época. Se trata de un único texto de control, pero nos parece suficiente, ya que se trata de una traducción entre catorce, unas treinta mil palabras entre quinientas mil.

En cuanto a los textos originales en inglés, el período estudiado es mucho más amplio, reflejando así la realidad del mercado: las editoriales gallegas publican sobre todo traducciones de textos editados originalmente como media unos diez o quince años antes que la versión gallega<sup>4</sup>. Nuestra selección comprende desde 1768 (*A Sentimental Journey* de Laurence Sterne) a 1982 (*The Golem* de Isaac B. Singer). Con la elección de este amplio abanico temporal pretendemos demostrar que la fecha de edición del texto original no influye en la microestructura de los textos traducidos al gallego; es decir, la distancia temporal de los distintos TOs con respecto a un mismo punto (el lustro 1989-94)

<sup>4</sup> Con la única excepción de la literatura para los más pequeños, de la que se suelen traducir, aparte de los textos que podríamos denominar clásicos (*Pinocho*, los cuentos de Perrault, etc.), los principales autores modernos, sobre todo los de lengua alemana (cf. Viceversa 1995, 1996; Costenla 1997).



no es un factor determinante de las soluciones a las que los distintos traductores recurrieron en su labor (sobre la cuestión de la temporalidad, cf. Fernández Ocampo 2000).

### 1.2.3 *El tamaño de TECTRA*

La dimensión de un corpus suele ser la mayor de las preocupaciones iniciales de los investigadores, pero la cuestión del tamaño demuestra tener poca importancia a largo plazo. Efectivamente, se ha demostrado que un corpus más grande no es necesariamente mejor. Aunque Sinclair (1991) recomienda que construyamos el corpus más grande que podamos y que lo hagamos crecer con periodicidad, reconoce también que hay muy poca diferencia, a efectos estadísticos, entre, por ejemplo, un corpus de un millón de palabras y otro de diez millones. Las palabras que más se repiten en un texto o corpus, como demostró Zipf ya en el año 1935 (apud Sinclair 1991), son unas quince o veinte palabras de significado gramatical (artículos, preposiciones, conjunciones, etc.); por otra parte, aproximadamente la mitad del vocabulario de un texto o corpus son palabras que aparecen solamente una o a lo sumo dos veces. Por lo tanto, multiplicar el número de palabras de un corpus por diez supone sólo que encontraremos unas cuantas palabras nuevas con una única aparición y que la frecuencia de las palabras más usuales aumenta proporcionalmente al número de palabras del corpus.

Las consideraciones anteriores pueden ser extrapoladas a las investigaciones de traducción basadas en corpus: el tamaño en sí mismo no va a ser determinante de los resultados obtenidos; sí lo serán, en cambio, la selección y la representatividad de los textos.

*TECTRA* contiene veintiocho textos con un total de 1.127.044 palabras, de las cuales 551.878 corresponden a las catorce traducciones gallegas y 575.166 a los catorce originales ingleses. Los textos que forman *TECTRA* son unidades completas, tanto en la sección gallega como en

la inglesa, ya que pensamos, como Sinclair (1991:19), que merece la pena incluir textos enteros en lugar de fragmentos para asegurar la representatividad del corpus; además, esto posibilita que en el futuro se puedan hacer, sobre el mismo corpus, estudios macroestructurales sobre, por ejemplo, la traducción de conectores oracionales, la organización textual en los TTs y TOs, etc. Por otra parte, la extensión de cada uno de los textos no nos pareció una variable digna de entrar en nuestra consideración.

### 1.3 *Descripción intratextual de TECTRA*

Nuestra concepción de los Estudios de traducción en general y de la Estilística comparada en particular como disciplinas empírico-descriptivas ha determinado desde el primer momento el método de análisis utilizado para el estudio de *TECTRA*. Concebimos la Estilística comparada como la disciplina que tiene por objeto de estudio las relaciones que en el acto de traducción y en textos concretos se verifican entre dos lenguas (cf. Álvarez LUGRIS 1998, 1999a, en prensa-a, en prensa-b). La Estilística comparada se dedica en concreto a estudiar las relaciones o correspondencias suboracionales que se establecen entre dos lenguas durante el proceso de traducción y que se manifiestan a través de las decisiones concretas del traductor.

El método utilizado en nuestro análisis estilístico del corpus *TECTRA* se puede definir como la combinación entre el denominado método helicoidal de Toury (1995) y el método de análisis transléxico propuesto por Rabadán (1991) y que ya describimos en Álvarez LUGRIS (1999a, en prensa-a). Defendemos la necesidad de que todo análisis realizado dentro de los Estudios de traducción, y por lo tanto, también dentro de la Estilística comparada, parta del TT en dirección al TO para intentar reconstruir el proceso traductivo mediante el uso de las mismas unidades (o fragmentos) utilizadas por el traductor en el desarrollo de su trabajo. Dejamos dicho





también que previamente a la comparación efectiva del TT y su correspondiente TO debíamos proceder a una minuciosa descripción intratextual de los dos textos con el objetivo de a) establecer y caracterizar la invariante o *tertium comparationis* del proceso traductivo mediante la caracterización del TT; y b) comprobar si las funciones textuales y los estándares de la textualidad responden a los del TO.

Estas descripciones de los TTs y TOs, como propone Rabadán (1991), se basan en el estudio de las propiedades de los textos analizados atendiendo a los estándares de la textualidad: coherencia, cohesión, intencionalidad, aceptabilidad, informatividad, situacionalidad e intertextualidad. Dado que uno de los pilares de nuestra metodología es que todo análisis debe comenzar por centrarse en un corpus relativamente reducido y por el estudio de un número relativamente limitado de fenómenos o parámetros para después aumentar el alcance de las investigaciones, decidimos comenzar nuestro estudio con unos textos que fuesen muy semejantes en cuanto a los mencionados parámetros.

Podemos decir que las *normas preliminares* que guiaron las catorce traducciones (y ediciones) estudiadas eran básicamente idénticas: se trata de ediciones dirigidas a un público juvenil y adulto que pretenden dar a conocer en gallego los principales autores de lengua inglesa. La lengua utilizada por todos los traductores es el gallego estándar normativo, lo que contribuye a mejorar la comprensión y distribución de los libros, que tienen además un precio medio a pesar de ser ediciones muy cuidadas.

La *intencionalidad* de los traductores es también constante: reproducir el TO de la forma más fiel posible, lo que se manifiesta, por ejemplo, en el mantenimiento del mismo punto de vista de los autores originales, en el respeto por el formato del original (especialmente en el caso de *A Sentimental Journey*, que en su versión gallega mantiene todos los juegos

tipográficos; o *Retallos do diario de Adán y Diario de Eva*, que intentan reproducir la presentación de las ediciones originales), en la reproducción en los TTs de elementos culturales de los TOs, etc.

La *aceptabilidad* de los textos gallegos está fuera de toda duda. El uso del gallego normativo estándar posibilita la identificación entre el lector y la lengua del texto. Por otra parte, los distintos referentes culturales que encontramos en las novelas son conocidos por los receptores y rápidamente identificados gracias al cine, otras obras literarias, etc., incluso los que *a priori* pueden parecer más lejanos en el tiempo (*A Sentimental Journey*) o en el espacio (*The Call of the Wild*). Además, aunque *Unha viaxe sentimental* tiene una presentación del texto desconcertante —por la abundancia de guiones, asteriscos y otros efectos tipográficos— no creemos que su aceptabilidad se vea excesivamente afectada ya que este especial formato es una característica del TO.

Las *funciones* de los TOs (por lo general la expresiva y en menor medida la fática) se mantienen en los TTs, contribuyendo a conservar la *cohesión*. La organización interna del discurso traducido sigue bastante de cerca la del original aunque, eso sí, respetando las peculiaridades léxicas y gramaticales del gallego. Por otra parte, los cambios de registro (diafásicos o diastráticos) del original se intentan reflejar en la traducción en la medida de lo posible, a pesar de que nuestra lengua no tiene aún definidos con claridad los distintos niveles y dialectos.

La *coherencia* del TT se consigue, básicamente, mediante el uso del mismo punto de vista que en el TO. Así, por ejemplo, tenemos un autor que transcribe el diario de un narrador no omnisciente protagonista en primera persona en *Retallos do diario de Adán*; un narrador omnisciente aparentemente neutral en *A perla*; o un narrador protagonista en tercera persona en *Retrato do artista cando novo*.

Ninguno de los textos presenta relaciones



intertextuales relevantes, exceptuando, claro está, *Retallos do diario de Adán y Diario de Eva*, que en realidad casi forman una unidad.

Finalmente, en cuanto a la *situacionalidad* de los TTs de *TECTRA*, la variedad lingüística en la que están escritos es, como ya queda dicho, el gallego estándar normativo.

## 2 CONSTRUCCIÓN DEL CORPUS

Una vez realizada la labor previa de selección de textos para *TECTRA* según los criterios citados en los apartados anteriores, el paso siguiente fue la construcción del corpus propiamente dicho. Los pasos para la creación de un corpus están descritos pormenorizadamente en Sinclair (1991:13-26) y Álvarez Lugris (1996) y se pueden resumir en los siguientes párrafos.

### 2.1 Preparación y digitalización de los textos

Dado que vamos a trabajar con un corpus paralelo informatizado, el primer paso es elaborar los textos electrónicos que lo van a componer. Como *TECTRA* es por el momento un corpus de uso privado, no fue preciso solicitar el permiso de los editores, traductores y autores para utilizar sus textos; en el futuro, cuando *TECTRA* sea un corpus de uso público, habrá que solucionar este aspecto legal.

Para comenzar a crear los distintos textos electrónicos fue preciso recurrir a la exploración digital de los textos base, ya que fue imposible conseguir las versiones electrónicas utilizadas por las editoriales. Por otra parte, picar manualmente todos y cada uno de los textos no era una alternativa operativa dada la extensión del corpus. Un explorador Scanner Hewlett Packard 4P de última generación (cuando lo comenzamos a utilizar) y la última versión del programa de ROC (Reconocimiento Óptico de Caracteres) OmniPage Professional posibilitaron la digitalización relativamente rápida y fácil de los veintiocho textos. Determinados problemas prácticos como la mala calidad del

papel y de la tinta de algunas ediciones, tipos de letra poco nítidos, etc., se fueron solucionando gracias a la versatilidad del digitalizador y a la capacidad de *entrenamiento* del programa de ROC utilizado.

Otros problemas específicos de las lenguas del corpus precisaron mayor atención para solucionarlos. En el caso de los textos en inglés, el uso de varios tipos de comillas (simples, dobles, invertidas) y apóstrofes (simples, tipográficos, invertidos) que no siempre corresponden a un carácter ASCII<sup>5</sup> provocó no pocos problemas al manipular los textos con programas que no leen más que ASCII puro. Por este motivo, y contra nuestra voluntad inicial de no retocar ni corregir los textos en absoluto, nos vimos obligados a hacer la pequeña modificación de unificar las distintas comillas con el carácter ASCII #34 y el apóstrofo con el #39. De la misma forma fue preciso eliminar las tabulaciones de comienzo de párrafo, ya que el carácter utilizado por nuestro procesador (Microsoft Word) no es ASCII y, por lo tanto, tampoco servía.

En los textos gallegos, por su parte, resultó bastante problemática la abundancia de guiones utilizados para la segunda forma del artículo<sup>6</sup>, los diálogos, separación de palabras, como alternativa a los paréntesis, etc. Hasta cuatro tipos diferentes encontramos en varios textos; por supuesto, sólo uno de esos guiones es un carácter ASCII, así que procedimos a unificarlos todos. Donde fue preciso marcar la existencia de un guión largo, como en los diálogos, utilizamos dos guiones (—).

En cuanto a la posible confusión en el trata-

<sup>5</sup> ASCII: American Standard Code for Information Interchange, el juego básico de 256 caracteres sobre el cual operan todos los ordenadores.

<sup>6</sup> El artículo determinado presenta en gallego dos morfemas: *o/-lo, a/-la, os/-los, as/-las*. Las segundas formas se emplean tras formas verbales conjugadas terminadas en *-r* o *-s*, tras pronombres clíticos y tras ciertas preposiciones y conjunciones y siempre se contraen con la palabra precedente: *vou coller os cartos* → *vou colle-los cartos*, *douvos a boneca* → *dóuvo-la boneca*.



miento de los textos gallegos entre la segunda forma del artículo y la división de palabras a final de línea, decidimos, por una parte, conservar las palabras completas en la línea siguiente y, por otra, definir dentro del programa de análisis OCP (*vid. infra*) un alfabeto especial en el que las cadenas «a-, e-, i-, o-»<sup>7</sup> fuesen interpretadas como vocales simples ordenadas en la misma posición que las correspondientes «a, e, i, o».

## 2.2 *Marcación del corpus*

Con la unión de todos los textos electrónicos en un único archivo en formato ASCII terminamos la elaboración de lo que podemos llamar un corpus básico. Para adecuar este corpus limpio o de texto puro a nuestros intereses es necesario *marcarlo* o *codificarlo* según describiremos enseguida.

«L'objectiu bàsic de tenir un corpus codificat és *descriure* totes les peces que són [...] rellevants per a la seva anàlisi» (Vivaldi et al. 1996:10; la cursiva es nuestra). Efectivamente: si queremos extraer de un corpus toda la rentabilidad que posee en potencia, tendremos que recurrir a marcar o codificar la información que nos interese. Marcar o codificar quiere decir señalar dentro del corpus por medio de algún método el tipo de fenómenos que nos interesa estudiar y el lugar en el que aparecen.

Mediante la marcación de un corpus seremos capaces de depurar y localizar con más rapidez y exactitud la información relevante para nuestro estudio y, lo que es más importante, podremos utilizar herramientas informáticas que hagan más llevadero el trabajo de analizar grandes corpus de textos, labor que desde siempre se ha calificado como laboriosa, abu-

rrida y cansada (Vanderawera 1985:6). En nuestro caso, además, la codificación nos permitió utilizar un programa de alineamiento de textos paralelos que facilitó enormemente el análisis y, posteriormente, la búsqueda de ejemplos y contraejemplos para verificar nuestras hipótesis.

Existen varias formas de marcar un texto o corpus, desde las primitivas marcas individuales para uso privado hasta los modernos estándares para intercambio de información, como describen Pérez Guerra (1999), Llisterri (1999) y Magán (1996). De entre las varias posibilidades, elegimos SGML<sup>8</sup> (Standard Generalized Markup Language, Lenguaje de marcación estándar generalizado). SGML es, antes que nada, una forma de concebir y describir documentos mediante la *descripción de su estructura* e independientemente de cualquier plataforma y, sobre todo *independientemente de cualquier formato*. Una de las ventajas fundamentales de utilizar SGML como lengua de marcación es que nos permite diseñar y describir nuestras propias etiquetas o marcas; es decir, no es un sistema cerrado con un repertorio limitado de etiquetas que tengamos que utilizar obligatoriamente. Se trata más bien de un *metalenguaje* de marcación, por lo que se adapta a nuestras necesidades con total versatilidad. Para nuestra investigación decidimos establecer una marcación mínima que facilitase los análisis estadísticos y que además no supusiese una gran inversión de tiempo. La información que decidimos codificar fue: obra, autor, traductor, capítulo, sección, página, párrafo y oración, con el siguiente juego de etiquetas:

<o> obra

<sup>7</sup> Las vocales que pueden aparecer inmediatamente antes de la segunda forma del artículo. La u nunca ocupa esa posición: *canta-la canción, bebe-lo leite, senti-la música, collémo-lo coche*.

<sup>8</sup> Sobre SGML véanse también Sperberg-McQueen & Burnard (eds.) (1994) y Hockey, Murphy, Horowitz & Bohlen (eds.) (1994).



<a>	autor
<t>	traductor
<c>	capítulo
<s>	sección
<px>	página
<pr>	párrafo
<f>	oración

La codificación de más niveles estructurales no redundaría en una mayor comodidad de análisis pero sí en un mayor trabajo de preparación de los textos. Por otra parte, este conjunto concreto de etiquetas es *entendible* por programas como OCP (*vid. infra*) debido a su parecido con el estándar CÓCOA —el único que OCP acepta.

### 2.3 *Alineamiento del corpus*

El último paso en la confección de *TECTRA* fue la alineación de los textos paralelos; es decir, hacer alternar en un único archivo las oraciones y párrafos de un TT y su correspondiente TO. En esta fase lo que hicimos en realidad fue crear una segunda versión de *TECTRA*, ya que la primera, con los textos sin alinear, también la conservamos: aún hay muy pocos programas de análisis que admitan un texto alineado como entrada y, además, para determinadas investigaciones puede resultar útil conservar las dos secciones del corpus por separado.

En nuestro caso utilizamos las dos versiones de *TECTRA*: la versión sin alinear fue especialmente útil (y mucho más rápida) para hacer concordancias, índices y estadísticas de los TTs y TOs. La versión alineada se utilizó fundamentalmente para buscar directamente ejemplos y contraejemplos de traducción que apoyasen o desmintiesen nuestras hipótesis de trabajo.

Para el alineamiento de los textos probamos distintos programas como ParaConc o WordSmith, pero todos ellos demostraron tener importantes limitaciones: al estar basados únicamente en la puntuación de las oraciones, la existencia de un desfase entre los textos (una oración del TO traducida como dos en el TT o

viceversa, puntos suspensivos en un texto y no en el otro, etc.) provoca errores que se van acumulando a lo largo del proceso.

Sin embargo tuvimos la posibilidad de utilizar, o mejor dicho probar, una versión provisional del programa de alineamiento de textos que el profesor Anthony McEnery desarrolla en la Universidad de Lancaster. Este programa está basado, por una parte, en la puntuación de párrafos y oraciones, pero también, a diferencia de los anteriores, en algoritmos que buscan semejanzas formales y funcionales entre segmentos de los dos textos. El único requisito imprescindible era que los textos tuviesen marcas de comienzo y final de párrafo y oración, algo que nosotros ya habíamos codificado con antelación en *TECTRA*. Los resultados fueron sorprendentes, con casi un 90% de aciertos en el alineamiento de oraciones y casi un 100% en el de párrafos. En el apéndice 3.4. podemos ver el resultado final del alineamiento de una pareja de textos.

### 2.4 *Explotación del corpus*

No existe aún un ordenador ni un programa informático que puedan hacer ningún tipo de análisis que requiera pensar, sopesar una serie de variables, tomar decisiones, comparar, etc., como son los análisis de Estilística comparada (el destino inicial de *TECTRA*) o en general el estudio comparativo de traducciones y originales. Todo lo que un ordenador puede hacer es, básicamente, contar elementos, relacionar variables numéricas y ordenar. Esto quiere decir que los ordenadores no nos van a solucionar nunca (por lo menos no en un futuro imaginable) otros problemas más que los puramente operativos: las tareas más o menos rutinarias y automatizables de localizar, contar y distribuir elementos de cualquier tipo. Y de todas formas esto no es poco: sin la ayuda de un ordenador, una investigación como la que estamos llevando a cabo con *TECTRA* sería inviable.

La primera fase de análisis de los textos fue

básicamente artesanal y consistió en la lectura paralela de los TTs y TOs. Durante esta fase tomamos nota de los distintos fenómenos que nos parecía que podrían ser relevantes para nuestro estudio de las microestructuras<sup>9</sup>. Las variables, comentarios, etc. se iban introduciendo en una base de datos a medida que surgían, hasta totalizar casi cuatro mil registros. Lógicamente, la lectura personal e individual de casi cuatro mil páginas de texto no puede reparar en todas las ocasiones en que se repite un fenómeno; es ahora cuando el corpus entra en juego: a partir de las entradas registradas en la base de datos, y después de depurar los fenómenos que nos interesaban para el estudio, el corpus nos sirvió para localizar todas las apariciones de cada uno de los casos analizados.

El programa utilizado fue OCP (Oxford Concordance Programme), en su versión Micro OCP de 1988. OCP es un programa de análisis textual que puede generar listas de palabras (*wordlists*), índices (*indices*) y concordancias (*concordances*) y sus respectivas estadísticas a partir de archivos de texto que pueden estar escritos en una variedad de alfabetos, todos ellos romances. Una *lista de palabras* es una relación de todas las palabras que aparecen en un texto junto con el número de apariciones de cada una de ellas. Un *índice* es una lista de palabras con su número de apariciones y su localización; para establecer la localización son útiles las etiquetas que utilizamos en la marcación del corpus: <c> capítulo, <px> página, <pr> párrafo, etc. Una *concordancia*, por último, es una lista de palabras con su número de apariciones, la localización y el contexto en el que aparece cada una.

La exactitud y la utilidad de la informa-

ción que nos proporciona el programa depende de la precisión con la que se construya el archivo de órdenes (*commands file*) con el que trabaja OCP<sup>10</sup>. En este archivo de órdenes especificamos:

a) las características del texto que vamos a utilizar como entrada (*text file*): si contiene comentarios y/o etiquetas o no y de qué tipo;

b) el alfabeto en que está escrito el texto. OCP tiene la ventaja de que podemos *inventar* nuestro propio alfabeto para solucionar problemas como el ya comentado de la segunda forma del artículo;

c) el tipo de información que queremos obtener: índice, concordancia, etc. Podemos también especificar si queremos resultados de todo el texto o de una selección de palabras, cuánto contexto necesitamos, etc.;

d) el formato que preferimos para el archivo de salida de la información (*output file*).

Seguidamente presentamos un ejemplo de archivo de órdenes en el que podremos observar cómo el tipo de corpus que utilizamos y la clase de información que buscamos determinan el uso del programa de análisis textual:

```
*input
references COCOA «<» to «>».
comments between «<[» to «[>».
*words
alph «A=Ā=Á=À=A=-`A=a=á=ä=à=a=-`a
Æ=æ B=b C=c=Ç=ç CH=Ch=ch D=d
E=É=È=Ē=Ĕ=-`E=e=é=è=ê=e=-`e=è=é F=f
G=g H=h I=Í=Ì=Î=-`I=i=í=î=ï=i=-`i J=j
K=k L=l LL=Ll=ll M=m N=n Ñ=ñ
O=Ó=Ò=Ô=-`O=o=ó=ô=ò=-`o P=p Q=q R=r
RR=Rr=rr S=s T=t
U=Ú=Û=Ü=-`U=u=ú=ù=ü=-`u V=v W=w X=x
Y=y Z=z -
o 1 2 3 4 5 6 7 8 9».
```

<sup>9</sup> Para una descripción detallada de los estudios que realizamos con TECTRA, véase Álvarez Lujrís 1999a y en prensa-a.

<sup>10</sup> Se podrá encontrar un análisis más detallado de OCP y su aplicación al análisis textual de corpus electrónicos en Álvarez Lujrís (1996).





diac «<sup>ˆ</sup>Á=Ä=Å=A-É=È=Ë=Ê=Ï=Î=Í-  
=Ó=Ô=Õ=Ü=Û=  
á=à=ä=a-é=è=ë=e-í=i=ï=i-ó=ò=ö=ú=ù=ü»  
zero length.  
punc «.,;:;!>».  
padd «( ) [ ]». {Tódo los diacríticos}  
\*action  
do concordance.  
pick words «\*ing, \*ment, \*ty, \*tion, \*ance,  
\*ence».  
references P=3, O=2.  
\*format  
layout length 78 and depth 70.  
titles «Substantivos abstractos na sección  
ING de *TECTRA*» centre and page 1 right.  
\*go

En la primera sección (\*input) de este archivo de órdenes de OCP especificamos qué tipo de etiquetas de referencia contiene el texto que se va a analizar; en este caso, como ya dijimos, las etiquetas SGML utilizadas coinciden con el estándar COCOA y se encuentran siempre entre ángulos (< >). Además, los comentarios que pueda haber —es decir, cualquier anotación personal que queramos introducir en el texto pero que no va a ser analizada— figuran entre los símbolos <[ y ]>.

En la sección \*words definimos el alfabeto en el que está escrito el texto y los diacríticos y signos de puntuación utilizados. En esta sección decidimos incluir todas las letras, grafías y diacríticos posibles, ya que de otro modo la existencia tanto en los TOs como en los TTs de palabras en francés, latín, alemán, etc., podría dar lugar a errores en el funcionamiento del programa. De esta forma, el programa de análisis *sabe* que tanto *ä* como *à* son grafías que puede encontrar en un texto que en principio está escrito en inglés o gallego y que además tiene el mismo valor (*a* = *ä* = *à* = etc.) a la hora de ordenar palabras alfabéticamente (*zero length*). Nótese además que especificamos que *a-* puede ser una grafía más de la vocal *a*, para solucionar el problema ya comentado de la segunda forma del artículo

gallego.

El apartado \*action recoge el tipo de información que solicitamos; en este caso una concordancia de todas las palabras que terminen en *-ing*, *-ment*, *-ty*, *-tion*, *-ence* como parte de nuestra búsqueda de sustantivos abstractos. Definimos también el tipo y la longitud de las referencias que indiquen la procedencia de cada ejemplo: hasta tres caracteres (es decir, hasta tres dígitos, de 001 a 999) para la página y dos caracteres para la obra, por ejemplo 008AF significa página 8 de *Animal Farm*.

En la sección \*format definimos el formato que deseamos para el archivo de salida de la información. En este ejemplo: 78 caracteres por línea y 70 líneas por página; cada página lleva el encabezamiento «Substantivos abstractos na sección ING de *TECTRA*».

Por último, la expresión \*go le indica al programa que el archivo de órdenes está completo y que puede empezar a trabajar.

Hasta aquí hemos revisado los criterios teóricos para la construcción de *TECTRA*, un corpus bilingüe de originales ingleses y traducciones gallegas destinado a investigaciones de traducción y Estilística comparada; hemos repasado también algunas de las cuestiones prácticas que hay que tener en cuenta para la confección y explotación de los corpus. En los apéndices con los que termina este artículo encontraremos más información sobre *TECTRA* y los textos que lo componen.



## 3 APÉNDICES

3.1 *Textos traducidos gallegos, traductor, año de publicación y tamaño en número de palabras*

TT gallego (clave), traductor	año	tamaño
<i>A perla</i> (AP), Benigno F. Salgado	1990	24.907
<i>A revolta dos animais</i> (AR), X. Antón L. Dobao	1992	26.215
<i>Cara ó faro</i> (CF), Manuela Palacios & Xavier Castro	1993	69.015
<i>A chamada da selva</i> (AC), Gonzalo Navaza	1982	29.053
<i>Retallos do diario de Adán</i> (RD), Benigno F. Salgado	1992	4.513
<i>Diario de Eva</i> (DE), Benigno F. Salgado	1991	6.771
<i>Galicia vista por un inglés</i> (GV), X. M. Gómez Clemente	1994	45.554
<i>O Golem</i> (OG), Anxo Romero Louro	1989	12.027
<i>Novo contos</i> (NC), X. Antón L. Dobao	1994	55.579
<i>O vixía no centeo</i> (OV), X. Ramón F. Rodríguez	1990	74.757
<i>Retrato do artista cando novo</i> (RA), Vicente Araguas	1994	82.398
<i>O señor das mascas</i> (OS), X. M. Gómez Clemente	1993	62.732
<i>O terceiro home</i> (OT), M <sup>a</sup> Dolores M. Torres	1994	30.833
<i>Unha viaxe sentimental</i> (UV), Manuel Outeiriño	1992	42.524
Tamaño total		551.878

3.2 *Textos originales ingleses, autor, año de publicación y tamaño en número de palabras*

TO inglés (clave), autor original	año	tamaño
<i>The Pearl</i> (TP), John Steinbeck	1945	26.476
<i>Animal Farm</i> (AF), George Orwell	1945	30.533
<i>To the Lighthouse</i> (TL), Virginia Wolf	1927	70.836
<i>The Call of the Wild</i> (CW), Jack London	1903	31.960
<i>Extracts from Adam's Diary</i> (EA), Mark Twain	1893	4.596
<i>Eve's Diary</i> (ED), Mark Twain	1906	7.036
<i>Spanish Galicia</i> (SG), Aubrey F.G. Bell	1922	40.543
<i>The Golem</i> (TG), Isaac B. Singer	1982	12.834
<i>Nine Stories</i> (NS), J.D. Salinger	1948	55.917
<i>The Catcher in the Rye</i> (TC), J.D. Salinger	1945	75.329
<i>A Portrait of the Artist as a Young Man</i> (AP), James Joyce	1916	83.641
<i>Lord of the Flies</i> (LF), William Golding	1954	62.052
<i>The Third Man</i> (TM), Graham Greene	1950	31.793
<i>A Sentimental Journey</i> (AS), Laurence Sterne	1768	42.620
Tamaño total		575.166
Tamaño total de TECTRA		1.127.044



### 3.3 *Ejemplo de marcación de un texto con SGML*

— Marcación de obra, autor, traductor, página, párrafo, oración.

— El lugar de las ilustraciones está marcado por un comentario entre corchetes y ángulos.

— Se han conservado las páginas en blanco para mantener la paginación del original.

— Se considera que una palabra forma parte siempre de una oración y de un párrafo.

```

<o>Retallos do diario de Adán
<a>Mark Twain
<t>Benigno Fernández Salgado
<px>000
<pr><f>Retallos do Diario de Adán
<pr><f>TRADUCIDO DO MANUSCRI-
TO ORIXINAL
<pr><f>por Mark Twain
<pr><f>ILUSTRADO POR
<pr><f>F. STROTHMANN
<pr><f>TRADUCIDO PARA O GALE-
GO POR
<pr><f>BENIGNO FERNÁNDEZ SAL-
GADO
<px>001
<pr><f>[NOTA. - <f>Traducín unba boa
parte deste diario hai algúns anos e un amigo
de meu editou uns poucos exemplares de
forma incompleta pero o público nunca che-
gou a velos. <f>Dende aquela descifrei outros
cantos xeroglíficos máis e coído que Adán
adquiriu agora a suficiente entidade como
personaxe público para xustificar esta publica-
ción. - <f>M. T.]
<px>002
<[ilustración]>
<px>003
<pr><f>Diario de Adán
<pr><f>Traducido do m. s. orixinal
<s>001
<pr><f>Luns
<pr><f>Esta nova criatura de cabelo longo
non anda máis ca a marearme. <f>Sempre
zangoneando ó meu redor e seguíndome a
tódolos sitios. <f>Non me gusta que o faga.
<f>Non estou afeito a ter compañía. <f>Ogallá
se quedara cos outros animais...

```

```

<pr><f>Hoxe está cuberto, con vento do leste.
<f>Penso que teremos auga... <f>¿Nós?
<f>¿De onde saquei eu esta palabra? <f>Xa
recordo. <f>Úsaa a nova criatura.

```

```

<px>004
<[ilustración]>
<px>005

```

### 3.4 *Ejemplo de textos TO y TT alineados*

```

_____paragraph 9_____
subs d=59_____ o -99 o -
99&
Today I am getting better ideas about distan-
ces.
_____&
Hoxe xa teño unha idea máis clara do que son
as distancias.
expa d=300_____ 1 2 1 -
99&

```

I was so eager to get hold of every pretty thing that I giddily grabbed for it, sometimes when it was too far off, and sometimes when it was but six inches away but seemed a foot —alas, with thorns between!

\_\_\_\_\_&  
 Onte estaba tan ansiosa por apalpar tódalas cousas bonitas que me cegaba de mala maneira por elas; unhas veces estaban lonxe de máis e outras aínda estando a seis polgadas parecía que estaban a un pé. ¡Ai, e con espiñas de por medio!

```

expa d=336_____ 3 4 2 -
99&
I learned a lesson; also I made an axiom, all out of my own head —my very first one; The scratched Experiment shuns the thorn.
_____&

```

Deprendín unha lección e tamén fixen un axioma, e todo saíu da miña cabeceira. Foi o meu primeiro axioma: O experimento espiña-do vai evita-las espiñas.

```

subs d=46_____ 5 -99 3 -
99&
I think it is a very good one for one so young.
_____&

```



Coido que e boísimo para alguén tan novo.

\_\_\_\_\_paragraph 10\_\_\_\_\_

subs d=195\_\_\_\_\_ 0 -99 0 -  
99&

I followed the other Experiment around, yesterday afternoon, at a distance, to see what it might be for, if I could.

\_\_\_\_\_&

Para ve-lo que podía ser, onte pola tarde seguín de lonxe o outro experimento.

subs d=23\_\_\_\_\_ 1 -99 1 -  
99&

But I was not able to make out.

\_\_\_\_\_&

Non dei reparado no que era.

subs d=6\_\_\_\_\_ 2 -99 2 -  
99&

I think it is a man.

\_\_\_\_\_&

Penso que é un home.

expa d=246\_\_\_\_\_ 3 4 3 -  
99&

I had never seen a man, but it looked like one, and I feel sure that that is what it is.

\_\_\_\_\_&

Eu nunca vira antes un, pero si, parece-me que o é. Estou case segura de que eso é o que é.

subs d=0\_\_\_\_\_ 5 -99 4 -  
99&

I realize that I feel more curiosity about it than about any of the other reptiles.

\_\_\_\_\_&

Doume conta de que sinto máis curiosidade por el ca por ningún dos outros réptiles.

expa d=446\_\_\_\_\_ 6 7 5 -  
99&

If it is a reptile, and I suppose it is; for it has frowsy hair and blue eyes, and looks like a reptile.

\_\_\_\_\_&

Si, imaxino que é un réptil, pois ten o cabelo enguedellado, os ollos azuis e o aspecto dun

réptil.

expa d=292\_\_\_\_\_ 8 9 6 -  
99&

It has no hips; it tapers like a carrot; when it stands, it spreads itself apart like a derrick; so I think it is a reptile, though it may be architecture.

\_\_\_\_\_&

Non ten cadeiras e é aparafusado coma unha cenoria. Cando se pon de pé a súa figura álzase coma un guindastre. Coido que e un réptil, ora ben tamén podería ser unha obra de arquitectura.

RECIBIDO EN DICIEMBRE DE 2000

#### REFERENCIAS BIBLIOGRÁFICAS DEL CORPUS

Bell, Aubrey Fitz Gerald (1922) Spanish Galicia. Londres: John Lane The Bodley Head.

Bell, Aubrey Fitz Gerald (1994) Galicia vista por un inglés. Vigo: Editorial Galaxia. Traducción de Xosé María Gómez Clemente. Colección Galicia no espello, n.º 1.

Golding, William (1954) Lord of the Flies. Londres: Faber & Faber. 1962.

Golding, William (1993) O Señor das moscas. Santiago de Compostela: Sotelo Blanco Edicións. Traducción de Xosé María Gómez Clemente. Colección Medusa Narrativa.

Greene, Graham (1950) The Third Man. Londres: Heinemann Educational Books. Terceira edición, 1974.

Greene, Graham (1994) O terceiro home. Vigo: Editorial Galaxia. Traducción de María Dolores Martínez Torres. Colección Árbore, n.º 71.

Joyce, James (1916) A Portrait of the Artist as a Young Man. Londres: Grafton Books. Undécima edición, 1986.

Joyce, James (1994) Retrato do artista cando novo. Santiago de Compostela: Edicións Laiovento. Traducción de Vicente Araguas. Colección Narrativa, n.º 43.

London, Jack (1903) The Call of the Wild. Londres: Heinemann Educational Books. Décimo sexta edición, 1975.

London, Jack (1982) A chamada da selva. Vigo: Edi-





- cións Xerais de Galicia. Traducción de Gonzalo Navaza. Colección Xabarán, n.º 4. Segunda edición, 1983.
- Orwell, George (1945) *Animal Farm*. Londres: Penguin Books. Quincuaxésimo oitava edición, 1987.
- Orwell, George (1992) *A revolta dos animais*. Santiago de Compostela: Edicións Positivas. Traducción de Xosé Antón L. Dobao.
- Salinger, J.D. (1948) *For Esmé —with Love and Squalor*. Nova York: Penguin Books. 1986.
- Salinger, J.D. (1994) *Nove contos*. Santiago de Compostela: Sotelo Blanco Edicións. Traducción de Xosé Antón L. Dobao. Colección Medusa Narrativa.
- Salinger, J.D. (1951) *The Catcher in the Rye*. Londres: Penguin Books in association with Hamish Hamilton.
- Salinger, J.D. (1990) *O vixía no centeo*. Vigo: Edicións Xerais de Galicia. Traducción de Xosé Ramón Fernández Rodríguez. Colección Grandes do noso tempo, n.º 9. Cuarta edición, 1992.
- Singer, Isaac Bashevis (1982) *The Golem*. Londres: Penguin Books.
- Singer, Isaac Bashevis (1989) *O Golem*. Vigo: Edicións Xerais de Galicia. Traducción de Anxo Romero Louro. Colección Grandes do noso tempo, n.º 4.
- Steinbeck, John (1945) *The Pearl*. Nova York: Penguin Books. Vixésimo primeira edición, 1986.
- Steinbeck, John (1990) *A perla*. Vigo: Editorial Galaxia. Traducción de Benigno Fernández Salgado. Colección Árbore, n.º 25. Quinta edición, 1993.
- Sterne, Lawrence (1768) *A Sentimental Journey Through France and Italy*. Hertfordshire: Wordsworth Editions. 1995.
- Sterne, Lawrence (1992) *Unha viaxe sentimental por Francia e Italia*. Santiago de Compostela: Sotelo Blanco Edicións. Traducción e notas de Manuel Outeiriño. Colección Biblioteca de Traducións, n.º 1.
- Twain, Mark (Samuel Langhorne Clemens) (1893) *Extracts from Adam's Diary*. Filadelfia: Courage Books. 1993.
- Twain, Mark (Samuel Langhorne Clemens) (1992) *Retallos do diario de Adán*. Santiago de Compostela: Edicións Positivas. Traducción de Benigno Fernández Salgado.
- Twain, Mark (Samuel Langhorne Clemens) (1906) *Eve's Diary*. Filadelfia: Courage Books. 1993.
- Twain, Mark (Samuel Langhorne Clemens) (1991) *Diario de Eva*. Santiago de Compostela: Edicións Positivas. Traducción de Benigno Fernández Salgado.
- Wolf, Virginia (1927) *To the Lighthouse*. Londres: Grafton Books. Décimo oitava edición, 1977.
- Wolf, Virginia (1993) *Cara ó faro*. Santiago de Compostela: Sotelo Blanco Edicións. Limiar, traducción e notas de Manuela Palacios e Xavier Castro. Colección Biblioteca de Traducións, n.º 3.

#### BIBLIOGRAFÍA

- Aarts, Jan (1998) «Corpora: first use of the term 'corpus linguistics'». Lista electrónica de discusión *Corpora*. corpora@hd.uib.no. Luns, 6 de xullo de 1998.
- Álvarez Lugrís, Alberto (1996) «Informática e análise textual», en Gómez Guinovart & Lorenzo Suárez (eds.) (1996), pp. 87-152.
- Álvarez Lugrís, Alberto (1997) *Corga. Corpus de Referencia do Galego Actual*. Documentos de traballo n.º 2. Santiago de Compostela: Centro Ramón Piñeiro de Investigacións en Humanidades.
- Álvarez Lugrís, Alberto (1998) «Notas para a definición dunha Estilística comparada da traducción», en *Viceversa. Revista galega de traducción*, n.º 4, 1998, pp. 25-40.
- Álvarez Lugrís, Alberto (1999a) *A traducción do inglés ó galego. Proposta de Estilística comparada*. Tesis doctoral. Universidade de Vigo.
- Álvarez Lugrís, Alberto (1999b) «Notes on the Place of Comparative Stylistics within the Field of Descriptive Translation Studies», en Álvarez Lugrís & Fernández Ocampo (eds.) (1999), pp. 139-147.
- Álvarez Lugrís, Alberto (en prensa-a) *Estilística comparada da Traducción. Proposta metodolóxica e aplicación práctica ó estudio do corpus TECTRA de traducións do inglés ó galego*. Vigo: Servicio de Publicacións da Universidade de Vigo.
- Álvarez Lugrís, Alberto (en prensa-b) «Bases para el desarrollo de una metodología descriptiva en Estilística comparada», en *DEA. Documentos de Español Actual*, n.º 2.

- Álvarez LUGRÍS, Alberto (en prensa-c) «Notas sobre o papel da traducción na configuración do sistema literario galego», en *Mi\_dzy orygina\_em a przek\_adem*, n.º 6, 2001.
- Álvarez LUGRÍS, Alberto & Fernández Ocampo, Anxo (eds.) (1999) *Anovar/anosar. Volume I*. Colección anovar/anosar estudios de traducción e interpretación. Vigo: Servicio de Publicacións da Universidade de Vigo.
- Baker, Mona (1993) «Corpus Linguistics and Translation Studies. Implications and Applications», en Baker, Francis & Tognini-Bonelli (eds.) (1993), pp. 233-250.
- Baker, Mona, Francis, Gill & Tognini-Bonelli, Elena (eds.) (1993) *Text and Technology. In honour John Sinclair*. Filadelfia/Amsterdam: John Benjamins Publishing Company.
- Constenla Bergueiro, Gonzalo (1997) «Traduccions, 1996» en *Viceversa. Revista galega de traducción*, n.º 3, 1997, pp. 167-172.
- Coseriu, Eugenio (1971) «Thesen zum Thema Sprache und Dichtung» en Stempel (ed.) (1971), pp. 183-188. Empregámo-la traducción española de Marcos Martínez Hernández: «Tesis sobre el tema 'lenguaje y poesía'», en Coseriu (1985), pp. 201-207.
- Coseriu, Eugenio (1985) *El hombre y su lenguaje. Estudios de teoría y metodología lingüística*. Biblioteca Románica Hispánica, II Estudios y ensayos, n.º 272. Madrid: Editorial Gredos.
- Chesterman, Andrew (1993) «From 'Is' to 'Ought': Laws, Norms and Strategies in Translation Studies», en *Target*, n.º 5, vol. 1, pp. 1-20.
- Chesterman, Andrew (1997) *Memes of Translation. The Spread of Ideas in Translation Theory*. Benjamins Translation Library, vol. 22. Amsterdam/Filadelfia: John Benjamins Publishing Company.
- van Doorslaer, Luc (1995) «Quantitative and Qualitative Aspects of Corpus Selection in Translation Studies», en *Target*, n.º 7, vol. 2, pp. 245-260.
- Fernández Ocampo, Anxo (2000) *Un achegamento á traducción intertemporal no eido das traduccions modernas da Materia de Bretaña*. Tesis doctoral. Universidade de Vigo.
- Gómez Guinovart, Javier & Lorenzo Suárez, Anxo M. (eds.) (1996) *Lingüística e informática*. Santiago de Compostela: Tórculo Edicións.
- Gómez Guinovart, Javier, Lorenzo Suárez, Anxo M., Pérez Guerra, Javier & Álvarez LUGRÍS, Alberto (eds.) (1999) *Panorama de la investigación en lingüística informática*. Volumen monográfico de la *Revista Española de Lingüística Aplicada*, 1999. Logroño: Revista Española de Lingüística Aplicada.
- Hockey, Susan (1994) «The Center for Electronic Texts in the Humanities», en Hockey, Murphy, Horowitz & Bohlen (eds.) (1994), pp. 11-23.
- Hockey, Susan, Murphy, Gregory, Horowitz, Lisa & Bohlen, Christine (eds.) (1994) *Electronic Texts in the Humanities: Methods and Tools. Course Handbook*. Princeton: Princeton University.
- Hunyadi, L., Klaudy, K. et al. (eds.) (1991) *Könyv Papp Ferencnek. Tanulmánygyűjtemény Papp Ferenc 60. születésnapjára*. Debrecen: KLTE.
- Klaudy, Kinga (1991) «Linguistic Theory of Translation», en Hunyadi, Klaudy et al. (eds.) (1991), pp. 93-118.
- Klaudy, Kinga, Lambert, José & Sohár, Anikó (1996) *Translation Studies in Hungary*. Budapest: Scholastica Kiadó.
- Kohn, János (1996) «What Can (Corpus) Linguistics Do for Translation?», en Klaudy, Lambert & Sohár (1996), pp. 39-52.
- Koller, Werner (1995) «The Concept of Equivalence and the Object of Translation Studies», en *Target*, n.º 7, vol. 2, pp. 191-222.
- van Leuven-Zwart, Kitty M. (1989) «Translation and Original: Similarities and Dissimilarities I», en *Target*, n.º 1, vol. 2, pp. 151-181.
- Llisterri, Joaquim (1999) «Transcripción, etiquetado y codificación de corpus orales» en Gómez Guinovart, Lorenzo Suárez, Pérez Guerra & Álvarez LUGRÍS (eds.) (1999), pp. 53-82.
- Magán Muñoz, Fernando (1996) «Estándares de representación textual e multimedia», en Gómez Guinovart & Lorenzo Suárez (eds.) (1996), pp. 153-186.
- Ocp (1988) *Oxford Concordance Programme*. Oxford: Oxford University Computing Service.
- Pérez Guerra, Javier (1999) «Estándares de anotación en lingüística de corpus» en Gómez Guinovart, Lorenzo Suárez, Pérez Guerra & Álvarez LUGRÍS (eds.) (1999), pp. 25-52.
- Rabadán, Rosa (1991) *Equivalencia y traducción. Problema de la equivalencia transléctica inglés-*





- español. León: Servicio de publicaciones de la Universidad de León.
- Real Academia Galega (R.A.G.) & Instituto Da Lingua Galega (I.L.G.A.) (1995) *Normas ortográficas e morfolóxicas do idioma galego*. Vigo: R.A.G. & I.L.G.A.
- Sinclair, John (1991) *Corpus, Concordance, Collocation*. Describing English Language Series. Oxford: Oxford University Press.
- Sinclair, John (1992) «The Automatic Analysis of Corpora», en Svartvik (ed.) (1992), pp. 379-399.
- Sperberg-McQueen, C. Michael & Burnard, Leo (eds.) (1994) *Guidelines for Electronic Text Encoding and Interchange*, borrador n.º3, documento TEI P3. Chicago/Oxford: ACH-ACL-ALLC.
- Summers, Della (1993) «Longman/Lancaster English Language Corpus — Criteria and Design», en *International Journal of Lexicography*, vol. 6, n.º 3, pp. 181-209.
- Svartvik, Jan (ed.) (1992) *Directions in Corpus Linguistics. Proceedings of the Nobel Symposium 82, Stockholm 4 August 1991*. Berlín/Nova York: Mouton de Gruyter.
- Thelen, Marcel & Lewandowska-Tomaszczyk Barbara (eds.) (1997) *Translation and Meaning. Part 4*. Maastricht: Hoogeschool Maastricht.
- Tirkkonen-Condit, Sonja (1997) «Towards a Prototypical Definition of Translation», en Thelen & Lewandowska-Tomaszczyk (eds.) (1997), pp. 89-96.
- Toury, Gideon (1995) *Descriptive Translation Studies —and Beyond*. Benjamins Translation Library, vol. 4. Amsterdam/Filadelfia: John Benjamins Publishing Company.
- Vanderawera, H.J. (1985) *Dutch Novels Translated into English: The Transformation of a 'Minority' Language*. Amsterdam/Atlanta (GA): Editions Rodopi BV.
- Viceversa (1995) «Traduccions, 1993-94», en *Viceversa. Revista galega de traducción*, n.º 1, 1995, pp. 209-220.
- Viceversa (1996) «Traduccions, 1995», en *Viceversa. Revista galega de traducción*, n.º 2, 1996, pp. 259-264.
- Vivaldi, J., de Yzaguirre, Ll., Solé, X. & Cabré, M.T. (1996) *Marcaje Estructural i Morfosintàctic del Corpus Tècnic amb l'estàndard SGML*. Sèrie Informes, 1. Barcelona: Institut Universitari de Lingüística Aplicada, Universitat Pompeu Fabra.
- Zipf, G.K. (1935) *The Psychobiology of Language*. Boston: Houghton Mifflin. Apud Sinclair (1991).