

Estudios de la traducción basados en corpus. Aspectos metodológicos para el análisis de estilo

[Corpus-based Translation Studies. Methodological aspects for analysis of style]

ARIEL NOVODVORSKI

Universidad Federal de Uberlândia (MG/Brasil)

arivorski@gmail.com

Resumen

El crecimiento de las investigaciones en el ámbito de los Estudios de la Traducción basados en Corpus (ETBC) ha aumentado en las últimas décadas (Laviosa, 2002); sin embargo, aún se hace necesaria la definición de un cuadro teórico-metodológico más preciso, para la investigación de los perfiles estilísticos de la traducción (Saldanha, 2011b). Este trabajo presenta el cuadro metodológico desarrollado para el análisis de estilo, en un corpus paralelo de textos literarios, en el par lingüístico español/portugués. El corpus está integrado por tres obras del escritor argentino Ernesto Sabato, traducidas al portugués brasileño por Sergio Molina. Se describirán todos los procedimientos implicados para la compilación, preparación y análisis estilístico del corpus de estudio, con apoyo de las herramientas y utilitarios del programa *WordSmith Tools*®, 5.0 (Scott, 2008). Son parte integrante de los procedimientos de análisis la identificación de temáticas y el análisis contrastivo de semejanzas y diferencias en el plano léxico-gramatical.

Abstract

Despite the fact that the research in the field of Corpus-based Translation Studies (CTS) has been increasing in recent decades (Laviosa, 2002), a more precise theoretical and methodological framework remains necessary for the investigation of stylistic profiles regarding translation (Saldanha, 2011b). This study reports a methodological framework developed for style analysis using a parallel corpus of literary texts in the language pair Spanish-Portuguese. The corpus consists of three works of the Argentine writer Ernesto Sabato, translated into Brazilian Portuguese by Sergio Molina. All research procedures involved in the compilation, preparation and stylistic analysis of the corpus—the latter carried out with the aid of *WordSmith Tools*®, 5.0—are described in detail. They are an integral part of the procedures for identification of text's aboutness and contrastive analysis of similarities and differences in lexical-grammatical level.

Palabras clave

Estudios de la traducción basados en corpus, procedimientos metodológicos, estilo en traducción, corpus paralelo

Key words

Corpus-based translation studies, methodological procedures, style in translation, parallel corpus



Esta obra se publica con una licencia Creative Commons **Reconocimiento – NoComercial – SinObraDerivada (by-nc-nd)**: No se permite un uso comercial de la obra original ni la generación de obras derivadas.

1. Introducción

Este trabajo está situado en el ámbito de los Estudios de la Traducción basados en Corpus (ETBC) y, particularmente, se orienta hacia los estudios sobre Estilo de los textos traducidos. Asumimos estilo, aquí, como atributo textual, en el análisis de un corpus paralelo de textos literarios, en el par lingüístico español/portugués.

Pese al crecimiento de las investigaciones en el ámbito de los ETBC, que viene aumentando en las últimas décadas (Laviosa, 2002), la definición de un cuadro teórico-metodológico más preciso, para la investigación de los perfiles estilísticos de la traducción (Saldanha, 2011b), aún se hace necesario. Por otro lado, la representación de las lenguas española y portuguesa, tanto en las investigaciones nacionales como internacionales, en el ámbito de los ETBC, también es escasa.

De ese modo, el objetivo principal de este estudio es presentar detalladamente algunos de los procedimientos metodológicos desarrollados mi tesis doctoral (Novodvorski, 2013), circunscrita a la investigación del perfil estilístico en la traducción literaria entre esas lenguas. En dicha oportunidad, establecí un vínculo entre diferentes teorías y establecí una triangulación de los datos y resultados obtenidos en la investigación.

El corpus de estudio está formado por tres obras literarias del autor argentino Ernesto Sabato, traducidas al portugués brasileño por Sergio Molina. Los procedimientos metodológicos utilizan, principalmente, conceptos provenientes de la Lingüística de Corpus, en especial por la utilización de herramientas y utilitarios del programa para análisis lingüísticos *WordSmith Tools*® (WST), 5.0 (Scott, 2008).

Serán presentados en este trabajo los diferentes procedimientos metodológicos adoptados, adaptados y/o creados para la identificación: de las temáticas (*aboutness*) del corpus, por medio de la clavicidad (*keyness*), entendida como cualidad textual (Scott, 2010; Stubbs, 2010; Berber Sardinha, 2009); de semejanzas y diferencias en el plano léxico-gramatical, a partir de un análisis contrastivo, vinculadas a las (sub)categorías de la presentación del discurso (Leech y Short, 2007; Semino y Short, 2004); y de significados culturales vehiculados en las traducciones, a través de los usos de la letra cursiva y de los signos de puntuación (Saldanha, 2005; 2011c; May, 1997) y de elementos paratextuales (Genette, 2009).

La próxima sección presenta una breve delimitación del campo de estudios teóricos al que se circunscribe este trabajo.

2. Un acercamiento teórico

El origen e inspiración de los *Estudios de la Traducción basados en Corpus* (ETBC) derivan tanto de la Lingüística de Corpus (LC) como de los Estudios Descriptivos de la Traducción (EDT). Aunque las primeras influencias de la LC en los estudios de la traducción hayan surgido a partir de la búsqueda por constituir una metodología de investigación más coherente y efectiva, Laviosa (2002) señala que ya hay suficientes elementos para afirmar que los ETBC pasan a constituirse como un «nuevo paradigma». La autora entiende que se trata de un abordaje que va más allá de una metodología innovadora. En ese sentido, entre las contribuciones más destacadas procedentes de investigaciones desarrolladas en el área, Laviosa destaca la formulación de hipótesis y de fundamentaciones teóricas, la definición de herramientas para análisis empíricas, además de las innumerables aplicaciones.

Ya en el ámbito de los estudios acerca de estilo de los textos traducidos, la metodología de análisis propuesta por Malmkjaer (2004) es la Estilística Traductoria. Dicho abordaje metodológico consiste en el estudio de patrones recurrentes en las relaciones entre el texto de partida y la traducción, o entre la obra más vasta de un autor y un conjunto de traducciones de esa obra. Malmkjaer (2004: 14) apunta que «*estilo* puede ser definido como una regularidad consistente y estadísticamente significativa de ocurrencia en el texto de ciertos ítems y estructuras, o tipos de ítems y estructuras, entre aquellos que la lengua ofrece como un todo» (traducción propia).

En la estilística traductoria se busca responder a la siguiente cuestión: ¿qué hace que una traducción signifique de un modo determinado y no de otro? En el análisis estilístico monolingüe, por otro lado, Malmkjaer (2004) observa que es posible realizar afirmaciones acerca del modo como cualquier texto provoca respuestas en la mente lectora, independientemente del modo como el texto haya surgido. Malmkjaer define claramente su objeto de atención en el estilo de los textos traducidos.

La próxima sección presenta un conjunto de aspectos metodológicos, desarrollados para el análisis de estilo de los textos traducidos, aplicados a un corpus paralelo literario.

3. Una aproximación metodológica

La metodología de los ETBC adoptada para la realización del presente trabajo consistió, resumidamente, en los siguientes procedimientos:

- compilación del corpus de estudio;
- preparación para la lectura con las diversas herramientas y utilitarios del programa para análisis lingüísticos WST (Scott, 2008);



- alineación de los tres subcorpus paralelos;
- etiquetado del corpus con (sub)categorías y marcadores, propios para el análisis del discurso referido (Leech y Short, 2007; Semino y Short, 2004), en términos de la Presentación del Habla, Escritura y Pensamiento (PHE&P);
- compilación de un corpus de referencia, para el estudio de la clavicidad (*keyness*), a partir de la extracción de las palabras-clave (Scott, 2010; Stubbs, 2010; Berber Sardinha, 2009);
- etiquetado para la identificación de algunas marcas tipográficas, para el análisis contrastivo de la letra cursiva (itálicos) en textos originales y traducidos;
- itemización, para análisis de la puntuación; y
- preparación para análisis de elementos paratextuales, como notas del traductor, además de otros materiales considerados peritextos y epitextos del corpus (Genette, 2009).

La perspectiva de análisis adoptada se resume en la siguiente ilustración:

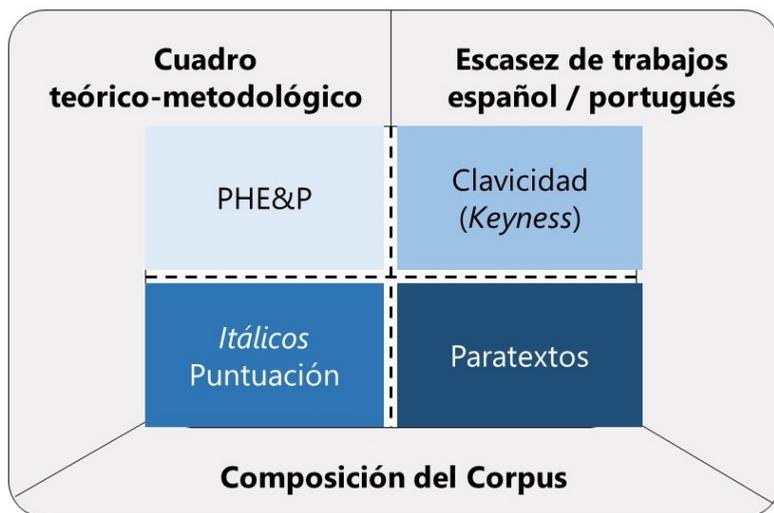


Ilustración1. Perspectiva de análisis

De ese modo, los trabajos consistieron en una triangulación de técnicas, procedimientos y datos, según: (1) la identificación de las temáticas del corpus, por medio de *clavicidad*, entendida como cualidad textual (Scott, 2010; Stubbs, 2010; Berber Sardinha, 2009); (2) el análisis contrastivo de semejanzas y diferencias en el plano léxico-gramatical, a partir de las (sub)categorías de la *presentación del discurso* [habla, escritura y pensamiento – PHE&P] (Leech y Short, 2007; Semino y Short, 2004), vinculado al *punto de vista narrativo*

(Schiavi, 1996; Hermans, 1996; May, 1994); y (3) el análisis de significados culturales vehiculados por medio de intervenciones explícitas: *itálicos* y *puntuación* (Saldanha, 2005; 2011c; May, 1997) y de *elementos paratextuales* en las traducciones (Genette, 2009).

3.1. Corpus de estudio

La composición del corpus de estudio reúne tres obras literarias del escritor argentino Ernesto Sabato, escritas en lengua español, en su variante rioplatense, y sus respectivas traducciones al portugués brasileño, realizadas por el traductor literario Sergio Molina. Los títulos de los textos originales, de las traducciones, de las editoriales y años de publicación se pueden apreciar en la siguiente ilustración:

Obras	Autor/Traductor	Editoriales	Año	1ª Publicación
<i>El túnel</i>	Ernesto Sabato	Sudamericana-Planeta	1982	1948
<i>O túnel</i>	Sergio Molina	Companhia das Letras	2008	2000
<i>Antes del fin: memorias</i>	Ernesto Sabato	Seix Barral	1999	1998
<i>Antes do fim: memórias</i>	Sergio Molina	Companhia das Letras	2008	2000
<i>La resistencia</i>	Ernesto Sabato	Seix Barral	2000	2000
<i>A resistência</i>	Sergio Molina	Companhia das Letras	2008	2008



Ilustración 2. Corpus de Estudio

3.2. Compilación y preparación del corpus de estudio

La compilación del corpus de estudio consistió en las etapas de: digitalización, por medio del programa *ABBYY FineReader®*, versión 10; aplicación del motor OCR (*Optical Character Recognition*), para reconocimiento óptico de caracteres; y posterior revisión, para corrección de problemas identificados en el proceso de digitalización del corpus. De las tres etapas, el proceso de revisión fue el que requirió mayor esfuerzo y mucha concentración, además de ocupar un tiempo de dedicación prolongado. Calculo que fue necesario destinar alrededor



de 60 horas de trabajo, distribuidas en un mes y medio. Las siguientes ilustraciones 3 y 4 muestran un resumen de esos procedimientos:

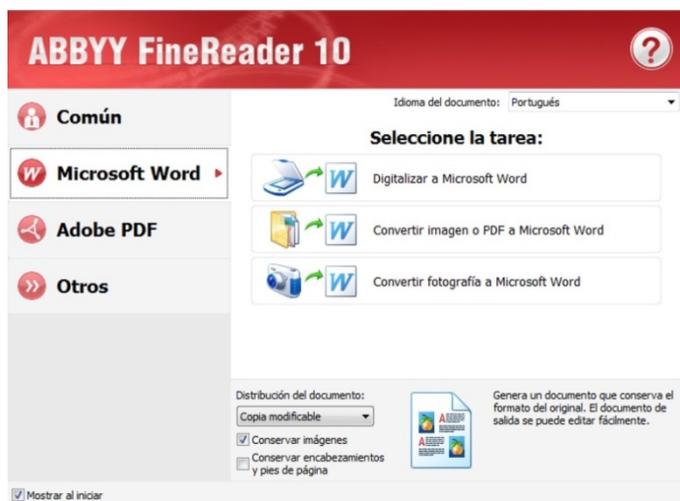


Ilustración 3. Programa ABBY FineReader®, versión 10

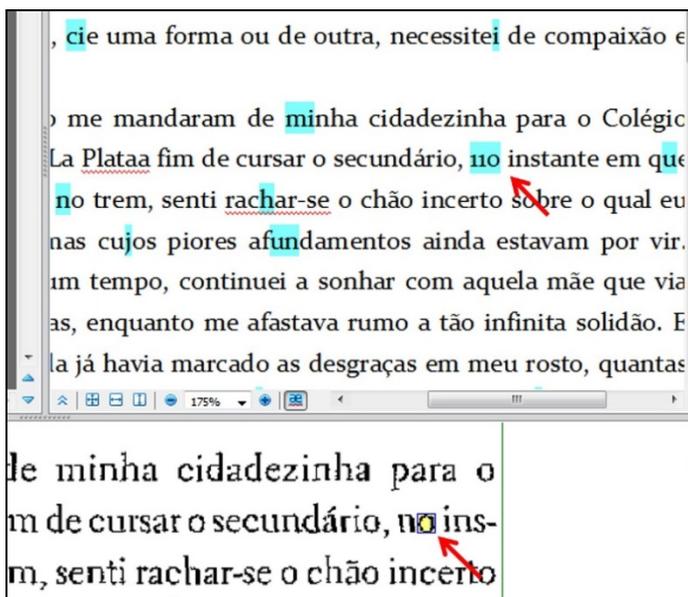


Ilustración 4. Proceso de revisión del corpus digitalizado

3.3. Extensión del corpus de estudio

Una vez que el corpus de estudio ya se encontraba digitalizado y revisado, el procedimiento siguiente consistió en convertir los archivos de formato *DOC* a

texto plano *TXT*, para posterior lectura con las herramientas y utilitarios del programa WST. Cada uno de los textos originales (TOs) y de los textos traducidos (TTs) fue archivado en ambos formatos. Para la obtención de los datos estadísticos del corpus se utilizó la herramienta *Wordlist*. El número total de ítems (*tokens*), palabras totales, se obtuvo a partir de la lectura del corpus de estudio completo, TOs y TTs; en el caso de las *formas* (*types*), palabras diferentes, se realizó una lectura separada para los textos en español y otra para los textos en portugués.

Diferentemente de los *ítems*, la totalidad de las *formas* (21.028) proporcionada por la herramienta del programa (ver Ilustración 5) no se puede tomar en cuenta, una vez que el corpus de estudio general es compuesto por textos escritos en dos lenguas diferentes. Tampoco se podría alcanzar el valor total de las *formas* en cada lengua por medio de la suma de los resultados de cada uno de los textos, una vez que una gran cantidad de *formas* de un texto también ocurre en los otros textos. De ese modo, el número total de *formas*, en cada subcorpus de TOs y TTs, fue calculado separadamente.

Además de esas informaciones más generales sobre la extensión del corpus y de cada uno de los subcorpus, también se calculó la significancia estadística de las diferencias entre las proporciones (Cf. Ilustración 5, columna p-valor), en la relación entre cada TO y TT y, aun, entre el conjunto de TOs y TTs. Para el cálculo de la significancia estadística se realizó el *test de la binomial para diferencias entre proporciones*. A seguir, la Ilustración 5 (v. pág. 124) reúne los datos generales del corpus de estudio.

3.4. Alineación del corpus de estudio

Tras las etapas de digitalización, preparación y extracción de datos más generales sobre la extensión del corpus de estudio, el paso siguiente fue el de alineación de los TOs con sus respectivos TTs. Para procesar el corpus paralelo con el utilitario *Viewer & Aligner*, del programa WST, se observó la necesidad de organizar paralelamente, lado a lado, en dos columnas, cada TO con su correspondiente TT. Esa necesidad fue observada después de muchos experimentos e intentos en los que la alineación difería considerablemente.

La finalidad de ese procedimiento, por un lado, era que el inicio de cada párrafo del TO correspondiera al del TT. Tal correspondencia hizo posible que, una vez para guardados en documentos independientes en formato *DOC* y, por último, convertidos al formato *TXT*, se pudiera procesar los TOs y TTs con el utilitario *Viewer & Aligner*, manteniéndose encontrados en los inicios de cada párrafo. Por otro lado, estando lado a lado los textos originales y sus respectivas traducciones, los procedimientos de observación, comparación y anotación de particu-



laridades, en la etapa de etiquetado, se pudieron realizar simultáneamente en ambos textos, sin pérdida de la alineación.

	N	Overall	1	2	3	4	5	6
text file		Overall	1 el túnel.txt	o túnel.txt	s del fin.txt	s do fim.txt	istência.txt	istência.txt
file size	988.934		189.293	187.741	189.946	184.538	119.983	117.433
tokens (running words) in text	163.643		31.741	30.635	31.379	29.815	20.474	19.599
tokens used for word list	163.188		31.732	30.585	31.301	29.730	20.360	19.480
sum of entries								
types (distinct words)	21.028		5.183	5.259	7.099	7.348	4.643	4.811
type/token ratio (TTR)	12.89		16,33	17,19	22,68	24,72	22,80	24,70

Corpus	Tokens	Types	TTR %	p-valor
ET_Sabato	31.741	5.183	16,33	0,0040
ET_Molina	30.635	5.259	17,19	
ADF_Sabato	31.379	7.099	22,68	0,0000
ADF_Molina	29.815	7.348	24,72	
LR_Sabato	20.474	4.643	22,80	0,0000
LR_Molina	19.599	4.811	24,70	
TOTALES	83.594 [es] 80.049 [pt] 163.643	12.120 [es] 12.568 [pt]	14,53 [es] 15,75 [pt]	0,0000

Ilustración 5. Extensión del Corpus de Estudio

La Ilustración 6 presenta una vista parcial del procesador de texto *Word*®, en la que se puede apreciar un fragmento alineado de TO y TT, con ajuste de párrafo. Además, es posible observar también parte de las etiquetas empleadas para anotación de la PHE&P:

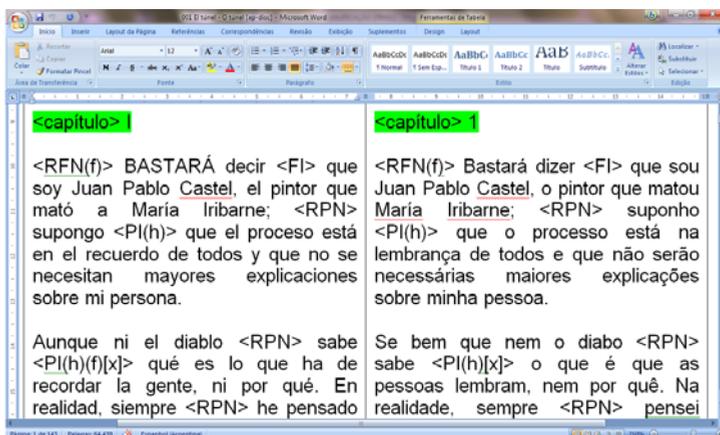


Ilustración 6. Alineación del corpus

La próxima ilustración muestra el procedimiento realizado para el ajuste de párrafo:

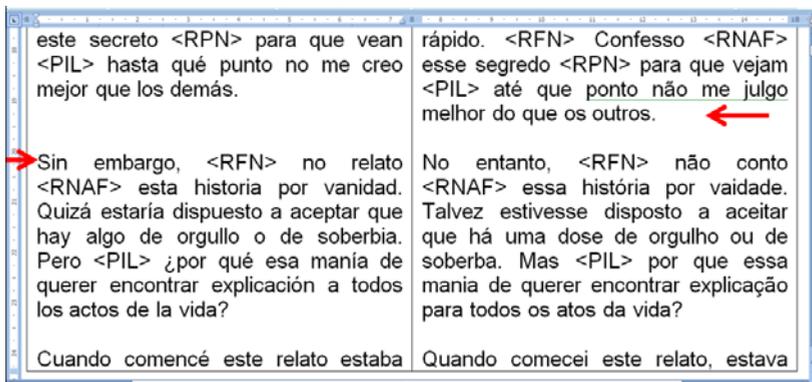


Ilustración 7. Alineación del corpus con ajuste de párrafo

Con los tres subcorpus paralelos alineados y convertidos individualmente cada TO y su respectivo TT al formato *TXT*, se pudo procesar la alineación con el utilitario *Viewer & Aligner*. Según la elección del investigador, se puede optar por la visualización del corpus alineado por frases (Cf. Ilustración 8) o alineado por párrafos (Cf. Ilustración 9):

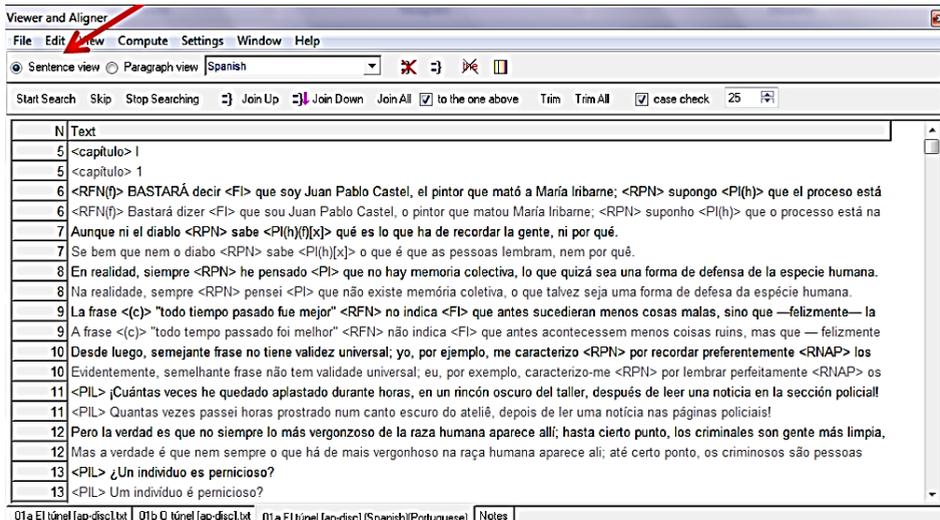


Ilustración 8. Corpus alineado por frases en el utilitario *Viewer & Aligner*

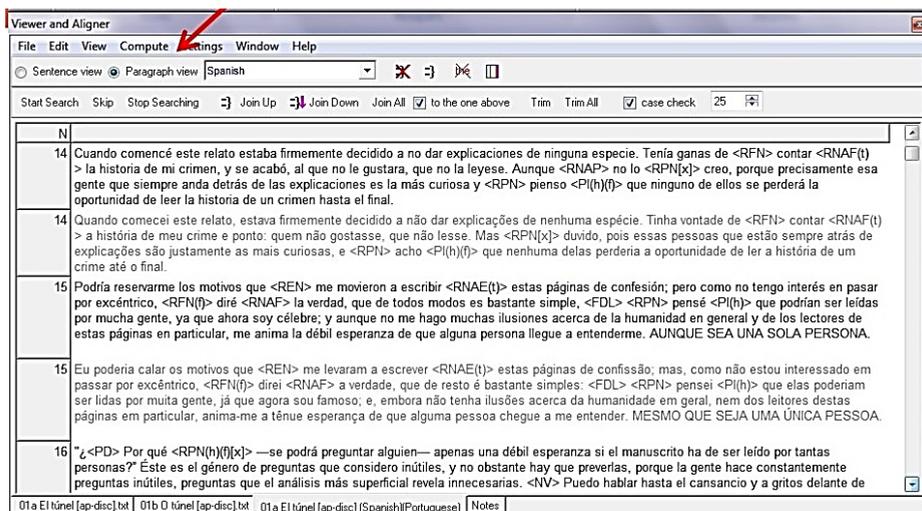


Ilustración 9. Corpus alineado por párrafos en el utilitario *Viewer & Aligner*

A partir de cada uno de esos dos modos de visualización del corpus alineado, por frases o por párrafos, el utilitario da la opción de guardar los resultados en archivo con formato *TXT*. Con eso, se pueden obtener y guardar dos archivos en dicho formato, uno que contendrá el corpus paralelo alineado por frase (Cf. Ilustración 10, pág. 127) y otro alineado por párrafo (Cf. Ilustración 11, pág. 127). Ese procedimiento es muy oportuno para el establecimiento posterior de las líneas de concordancia, con la herramienta *Concord* del WST. Las dos próximas ilustraciones muestran una vista parcial del corpus paralelo alineado en formato *TXT*:

Es importante observar que, al término de todos los procedimientos relatados para la alineación del corpus paralelo de cada uno de los tres pares de TOs y TTs, obtuve la siguiente organización por formatos del corpus:

- cada TO con su correspondiente TT, alineado lado a lado, en dos columnas, en un mismo archivo en *DOC*;
- cada TO y TT en *DOC* y en *TXT*, separadamente, sin etiquetas;
- cada TO y TT en *DOC* y en *TXT*, separadamente, con etiquetas, según el aspecto a ser analizado;
- cada TO y TT en *VWR* (*Viewer*), separadamente;
- cada TO con su correspondiente TT, alineado por frases, en *ALI* (*Aligner*);
- cada TO con su correspondiente TT, alineado por párrafos, en *ALI* (*Aligner*);
- cada par de TO/TT, alineado paralelamente por frases, en *TXT*; y

- cada par de TO/TT, alineado paralelamente por párrafos, en *TXT*.

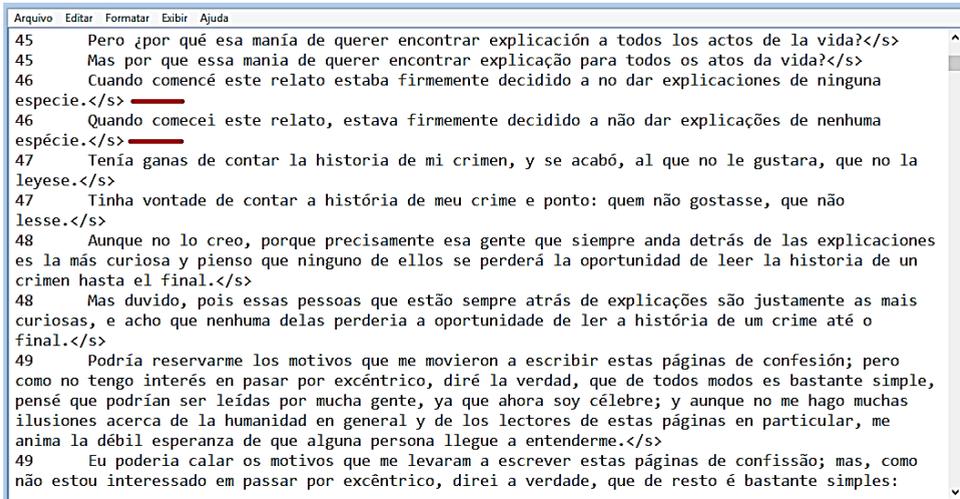


Ilustración 10. Corpus paralelo alineado por frases en archivo formato TXT

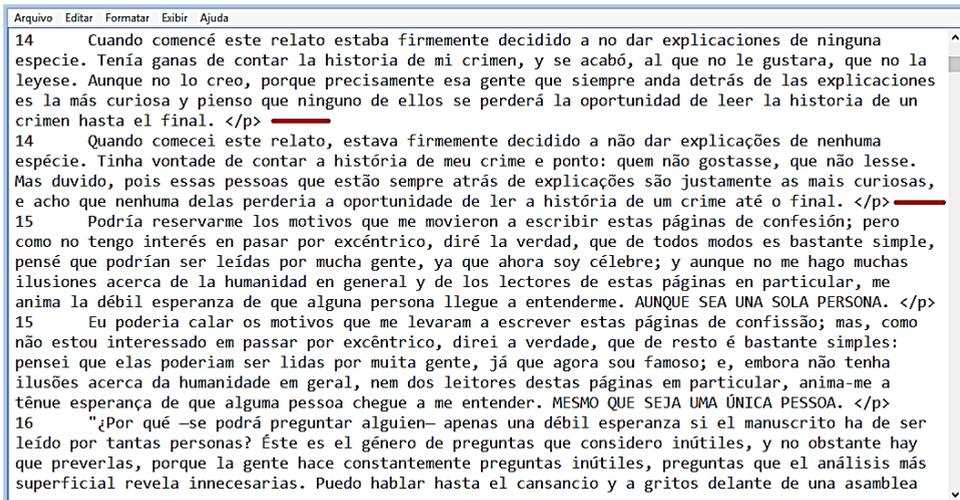
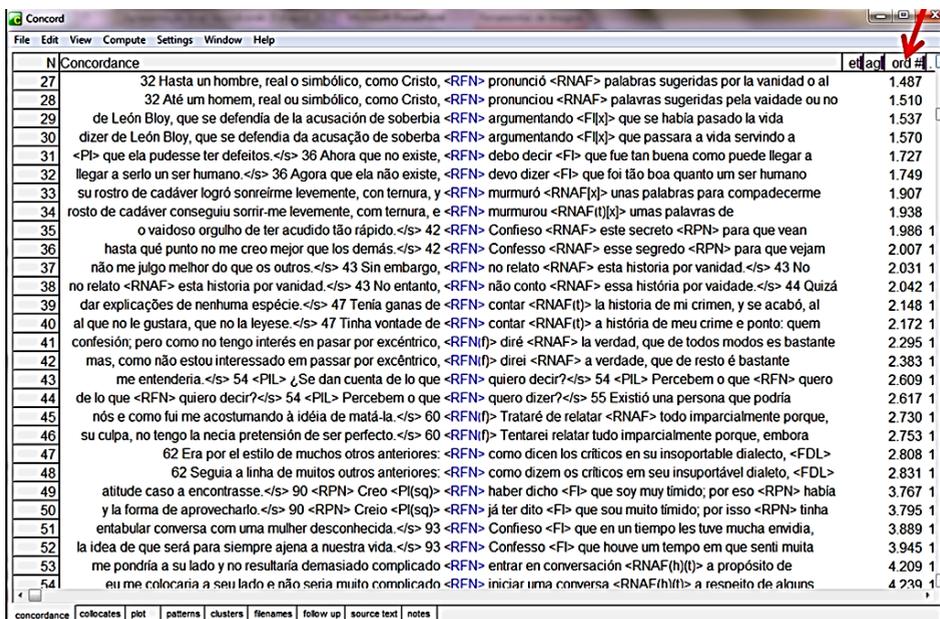


Ilustración 11. Corpus paralelo alineado por párrafos en archivo formato TXT

Con el corpus etiquetado y alineado, en formato *TXT*, el procedimiento posterior fue realizar búsquedas con la herramienta *Concord*, para establecimiento de las líneas de concordancia, posibilitando la cuantificación de los resultados, el establecimiento de tipologías y análisis cualitativo a partir de ejemplos. La próxima ilustración presenta una vista parcial de la herramienta *Concord*, con las líneas de concordancia establecidas a partir de la búsqueda por la etiqueta RFN,



una de las categorías utilizadas en el análisis del discurso referido, y con el orden secuencial dado por la columna N° de Palabra (*Word #*). Teniendo en cuenta que el corpus ya se encuentra alineado y en formato TXT, al realizar las búsquedas a partir de las etiquetas, dicha organización guiada por el número de palabra posibilita la visualización de la traducción debajo del texto original:



N	Concordance	et	aq	ord #
27	32 Hasta un hombre, real o simbólico, como Cristo, <RFN> pronunció <RNAF> palabras sugeridas por la vanidad o al			1.487
28	32 Até um homem, real ou simbólico, como Cristo, <RFN> pronunciou <RNAF> palavras sugeridas pela vaidade ou no			1.510
29	de León Bloy, que se defendía de la acusación de soberbia <RFN> argumentando <F x> que se había pasado la vida			1.537
30	dizer de León Bloy, que se defendia da acusação de soberbia <RFN> argumentando <F x> que passara a vida servindo a			1.570
31	<PI> que ela pudesse ter defeitos. </s> 36 Ahora que no existe, <RFN> debo decir <F> que fue tan buena como puede llegar a			1.727
32	llegar a serlo un ser humano. </s> 36 Agora que ela não existe, <RFN> devo dizer <F> que foi tão boa quanto um ser humano			1.749
33	su rostro de cadáver logró sonreírme levemente, con ternura, y <RFN> murmuró <RNAF x> unas palabras para compadecerme			1.907
34	rostro de cadáver conseguiu sorrir-me levemente, com ternura, e <RFN> murmurou <RNAF(t) x> umas palavras de			1.938
35	o vaidoso orgulho de ter acudido tão rápido. </s> 42 <RFN> Confieso <RNAF> este secreto <RPN> para que vean			1.986 1
36	hasta qué punto no me creo mejor que los demás. </s> 42 <RFN> Confesso <RNAF> esse segredo <RPN> para que vejam			2.007 1
37	não me julgo melhor do que os outros. </s> 43 Sin embargo, <RFN> no relato <RNAF> esta historia por vanidad. </s> 43 No			2.031 1
38	relato <RNAF> esta história por vanidade. </s> 43 No entanto, <RFN> não conto <RNAF> essa história por vaidade. </s> 44 Quizá			2.042 1
39	dar explicações de nenhuma espécie. </s> 47 Tenía ganas de <RFN> contar <RNAF(t)> la historia de mi crimen, y se acabó, al			2.148 1
40	ai que no le gustara, que no la leyesse. </s> 47 Tinha vontade de <RFN> contar <RNAF(t)> a história de meu crime e ponto: quem			2.172 1
41	confesión, pero como no tengo interés en pasar por excéntrico, <RFN(t)> diré <RNAF> la verdad, que de todos modos es bastante			2.295 1
42	mas, como não estou interessado em passar por excêntrico, <RFN(t)> direi <RNAF> a verdade, que de resto é bastante			2.383 1
43	me entendería. </s> 54 <PIL> ¿Se dan cuenta de lo que <RFN> quiero decir? </s> 54 <PIL> Percebem o que <RFN> quero			2.609 1
44	de lo que <RFN> quiero decir? </s> 54 <PIL> Percebem o que <RFN> quero dizer? </s> 55 Existió una persona que podría			2.617 1
45	nós e como fui me acostumbrando à idéia de matá-la. </s> 60 <RFN(t)> Trataré de relatar <RNAF> todo imparcialmente porque,			2.730 1
46	su culpa, no tengo la neicia pretensión de ser perfecto. </s> 60 <RFN(t)> Tentarei relatar tudo imparcialmente porque, embora			2.753 1
47	62 Era por el estilo de muchos otros anteriores: <RFN> como dicen los críticos en su insoportable dialecto, <FDL>			2.808 1
48	62 Seguia a linha de muitos outros anteriores: <RFN> como dizem os críticos em seu insuportável dialecto, <FDL>			2.831 1
49	atitude caso a encontrasse. </s> 90 <RPN> Creo <P(sq)> <RFN> haber dicho <F> que soy muy tímido, por eso <RPN> habla			3.767 1
50	y la forma de aprovecharlo. </s> 90 <RPN> Creio <P(sq)> <RFN> já ter dito <F> que sou muito tímido; por isso <RPN> tinha			3.795 1
51	entabular conversa com uma mulher desconhecida. </s> 93 <RFN> Confieso <F> que en un tiempo les tuve mucha envidia,			3.889 1
52	la idea de que será para siempre ajena a nuestra vida. </s> 93 <RFN> Confesso <F> que houve um tempo em que senti muita			3.945 1
53	me pondría a su lado y no resultaría demasiado complicado <RFN> entrar en conversación <RNAF(h)(t)> a propósito de			4.209 1
54	eu me colocaria a seu lado e não seria muito complicado <RFN> iniciar uma conversa <RNAF(h)(t)> a respeito de alguns			4.239 1

Ilustración 12. Concordancias de RFN con el corpus alineado por frases

A partir de la búsqueda con cada una de las categorías, según cada uno de los niveles de análisis realizados, guardé en formato *CNC* (*Concord*), en carpetas específicas, los resultados obtenidos. La ilustración 13 (pág. 129) muestra una vista parcial de la organización de los resultados.

Teniendo en cuenta que los ETBC se configuran como un área interdisciplinario, tras haber recurrido a los estudios sobre estilística, principalmente a aquellos que abordan aspectos vinculados a os diversos puntos de vista, apliqué al corpus de estudio el conjunto de categorías para el análisis de la presentación del discurso (PHE&P) propuesto por Leech y Short (2007) y por Semino y Short (2004), como se puede observar en las últimas dos ilustraciones. Sin embargo, tal abordaje *basado en corpus* no se mostró muy productivo, por lo menos a ese nivel de análisis. No obstante eso, el corpus mostró diversos aspectos que sería interesante investigar. De ese modo, decidí adoptar una perspectiva de análisis más *guiado por el corpus*. Tal cambio de orientación condujo al estudio de la tematicidad, a partir de la extracción de palabras clave.

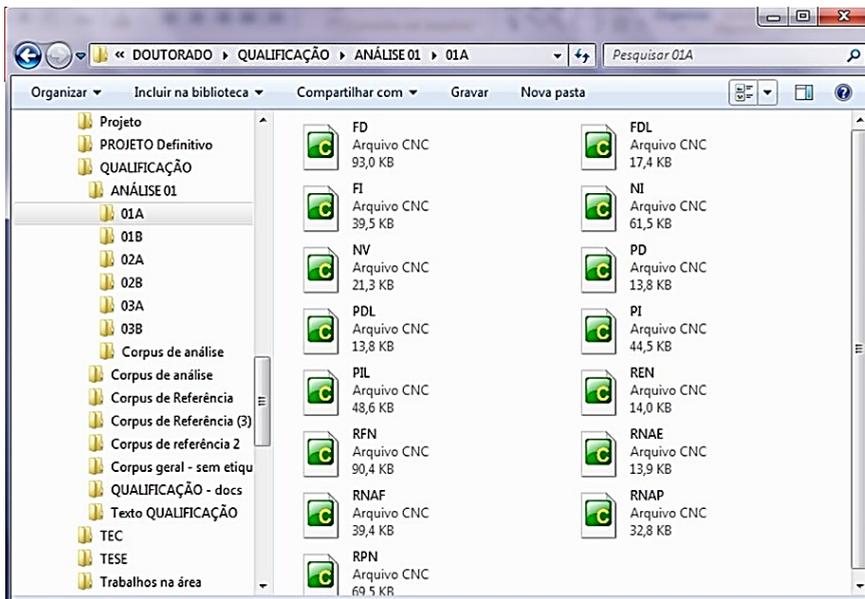


Ilustración 13. Organización de los resultados por categorías para el discurso referido

3.5. Corpus de referencia y análisis de la clavicidad

Según Mike Scott (2010), aunque aún no se comprende muy bien, el análisis de la *clavicidad* (*keyness*) empieza a despertar el interés de investigadores, como una cualidad textual que daría fuertes indicios sobre la temática del texto, junto a indicadores de estilo. En ese sentido, otro procedimiento metodológico que desarrollé, específico para la extracción de palabras clave y análisis de la tematicidad contrastiva del corpus, fue la compilación equilibrada de un Corpus de Referencia. En ese sentido, compilé un corpus de referencia compuesto por textos periodísticos, académicos y literarios, originalmente escritos en lengua española, que denominé CorREF01[es], y otro en lengua portuguesa denominado CorREF01[pt].

Durante la compilación, tuve el cuidado de equilibrar por ítems, número de palabras totales (*tokens*), los dos corpus y cada uno de los subcorpus que los componen. Además, en términos cuantitativos, también observé que el corpus de referencia fuera, por lo menos, de extensión cinco veces mayor que el corpus de estudio. Según Berber Sardinha (2004: 102), ese es el tamaño recomendado para el corpus de referencia. Otra recomendación para evitarse el filtrado de las palabras clave es que el corpus de referencia esté integrado por tipologías y géneros textuales distintos del corpus de estudio. Con eso, busqué que los textos



que integraran el corpus de referencia, en general, fueran variados lo suficiente, para no filtrar los resultados, principalmente los textos literarios. La próxima ilustración muestra la extensión y distribución de los corpus de referencia en ambas lenguas:

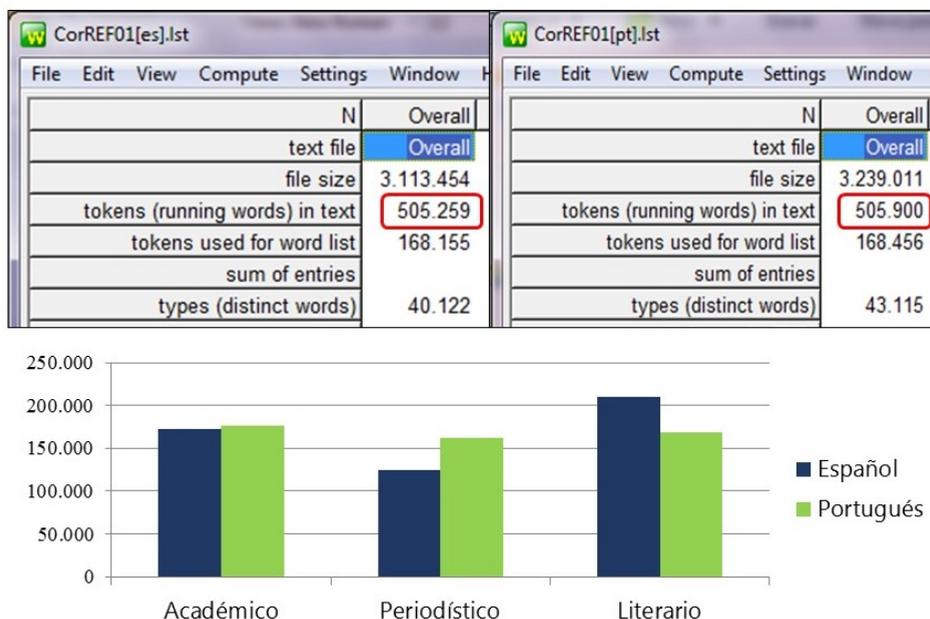


Ilustración 14. Extensión y distribución del corpus de referencia

Con los corpus de referencia en formato *TXT*, el paso siguiente fue extraer las palabras clave por medio de la herramienta *KeyWords* del programa *WST*. El resultado se obtiene a partir de la comparación entre las listas de palabras del corpus de estudio con el corpus de referencia, respectivamente en cada lengua. De ese modo, fue necesario tener archivadas en la computadora las listas de palabras correspondientes de cada corpus en formato *LST*, para extracción posterior de las palabras clave, que también fueron archivadas en formato *KWS*. A partir de los resultados, pude identificar áreas semánticas que apuntaron hacia el establecimiento de las temáticas del corpus de estudio. La siguiente ilustración es una vista parcial de las palabras clave, específicamente de los sustantivos clave, en lengua española y portuguesa, organizadas por clavicidad.

N	Key word	Freq	%	RC	Freq	RC	%	Keyness
1	VIDA	206	0,25	385	0,08	158,18	000	
2	ESTANCIA	41	0,05	8		118,93	000	
3	CARTA	78	0,09	72	0,01	118,93	000	
4	HOMBRES	117	0,14	177	0,04	115,86	000	
5	HOMBRE	195	0,23	436	0,09	114,70	000	
6	SOLEDAD	40	0,05	19		87,86	000	
7	AMOR	77	0,09	105	0,02	84,85	000	
8	EXISTENCIA	56	0,07	54	0,01	82,75	000	
9	HUMANIDAD	37	0,04	18		80,44	000	
10	MOMENTOS	48	0,06	49		67,97	000	
11	SERES	44	0,05	43		64,38	000	
12	TRISTEZA	36	0,04	27		62,79	000	
13	MUCAMA	15	0,02	0		58,57	000	
14	MUJER	69	0,08	121	0,02	57,50	000	
15	ALMA	57	0,07	85	0,02	57,30	000	
16	ESPIRITU	34	0,04	27		57,27	000	
17	ABSOLUTO	24	0,03	10		55,58	000	
18	ANGUSTIA	19	0,02	5		51,15	000	
19	MUNDO	149	0,18	439	0,09	50,58	000	
20	TEMPO	165	0,20	509	0,10	49,92	000	

Ilustración 15. Palabras clave (sustantivos) organizadas por clavicidad

3.6. Itálicos, puntuación y elementos paratextuales

Así como las palabras clave pueden ser asumidas como «indicios» (*pointers*), una vez que apuntarían hacia áreas que valdría la pena investigar (Scott, 2010; Stubbs, 2010), Saldanha (2005; 2011c) también observa que los itálicos pueden indicar áreas de interés para el investigador. La autora observa factores como la disminución del nivel de formalidad en las traducciones al inglés y una consecuente facilitación en la interpretación, vinculados a los efectos estilísticos que derivan del uso enfático del itálico. Los itálicos son recursos extralingüísticos, según Saldanha (2005), que pueden llamar la atención del lector para determinadas formas particulares, y que tanto el autor del texto como el traductor poseen control sobre el uso de los itálicos.

Sin profundizar demasiado en las funciones comunicativas que describe Saldanha (2005) para el uso del itálico, por no tratarse específicamente de la finalidad del presente trabajo, apuntamos que de los tres tipos identificados por la investigadora, el que corresponde a los usos enfáticos del itálico sería el de mayor interés para el análisis de la significancia estilística. Saldanha (2005: 95) también observa que los manuales de español y de portugués que consultó son poco consistentes o que ni siquiera abordan el asunto, cuando se trata de la distinción de palabras, letras o expresiones, en las referencias que realiza un autor, por ejemplo, a la pronunciación o al modo como se utiliza un término en determinados contextos. Es decir, el uso del itálico cuando podría apuntar hacia aspectos de variación lingüística.



Considerando los trabajos desarrollados por Saldanha (2005; 2011c) y a partir de la presencia significativa de *itálicos* en el corpus de estudio, durante los procedimientos de digitalización y preparación del corpus, percibí la importancia de anotar de algún modo las ocurrencias de *itálicos*, porque se perderían en el proceso de conversión a *TXT*. En ese sentido, otro procedimiento metodológico desarrollado en la investigación, para el análisis de lo que serían esas intervenciones posiblemente más explícitas del traductor, fue el etiquetado en el corpus de estudio de la letra cursiva (*itálicos*). Para eso, inserí las etiquetas <itálico> e </itálico> para indicar el punto de inicio y término de expresiones y frases enteras escritas en cursiva. Ese paso se hizo necesario, como ya apuntado, una vez que al convertir los textos de formato *DOC* a *TXT* se pierden tales detalles. La siguiente ilustración presenta una búsqueda para extracción de *itálicos* con la herramienta *Concord*, en el corpus alineado:

N	Concordance	et
1	pronto .</s> 15 Eso es lo que yo llamo una <itálico> buena acción </itálico> .</s> 15 Isso é	
2	.</s> 15 Isso é o que eu chamo uma <itálico> boa ação </itálico> .</s> 16 Piensen	
3	y entonces lo obligaron a comerse una rata , <itálico> pero viva </itálico> .</s> 19 Em todo	
4	fome e foi obrigado a comer uma ratazana , <itálico> só que viva </itálico> .</s> 20 No es de	
5	Einstein o gente por el estilo ; respuesta : <itálico> es fácil ser modesto cuando se es	
6	se es célebre </itálico> ; quiero decir <itálico> parecer modesto </itálico> .</s> 29	
7	de Einstein ou de gente da laia ; resposta : <itálico> é fácil ser modesto quando se é	
8	quando se é célebre </itálico> ; quer dizer , <itálico> parecer modesto </itálico> .</s> 30	
9	una persona que podría entenderme . <itálico> Pero fue , precisamente , la persona	
10	uma pessoa que poderia me entender . <itálico> Mas foi , justamente , a pessoa que	

Ilustración 16. Concordancias para extracción de *itálicos*

Además de los *itálicos*, otro nivel de análisis efectuado en la investigación consistió en la cuantificación contrastiva de todos los signos de puntuación. Con la herramienta *Concord*, realicé la búsqueda por el punto, la coma, el punto y coma, paréntesis, etc. Sin embargo, pude observar que los primeros resultados no parecían ser coherentes, dada la extensión del corpus (Cf. Ilustración 17). De ese modo, experimenté itemizar todos los signos de puntuación que iría a investigar. Ese procedimiento de itemización (Berber Sardinha, 2004) consistió en separar los signos de puntuación de las palabras que los precedían o seguían, insiriendo un espacio. Tal ajuste fue realizado con auxilio del recurso *localizar* y *sustituir*, en el procesador de texto *Word*®, sustituyendo el signo de puntuación específico por un espacio seguido o antecedido por el signo. Las siguientes ilustraciones muestran la búsqueda por el PUNTO, en el corpus con o sin itemización. En el ángulo inferior izquierdo de las imágenes se puede observar la

diferencia en los resultados: 367 ocurrencias del PUNTO, sin itemización y 3.588 ocurrencias con itemización.

N	Concordance	Set	Tag	Word #	t
137	600 — Tengo que cortar —me interrumpió de pronto—.</s> 600 — Vou ligar amanhã cedo — cheguei a dizer, com			18.985	2
138	Deus escuta <itálico> A paixão segundo são Mateus </itálico>.</s> 616 Entré en el café Marzotto.</s> 616 <capitulo>			19.694	2
139	en Dios oye <itálico> La pasión según San Mateo </itálico>.</s> 617 Na manhã seguinte, por volta das dez, telefonei.</s>			19.733	2
140	<itálico> isso provava que ela era capaz de fingir </itálico>.</s> 640 ¿Quiénes eran esas gentes que "entran y salían" y			20.246	2
141	<itálico> eso probaba que ella era capaz de simular </itálico>.</s> 641 Mas uma frase acima de tudo se gravara em mim			20.286	2
142	"Quando fecho a porta, sabem que não devem me incomodar".</s> 642 ¿Y por qué vació esa mujer cuando pregunté por la			20.309	2
143	uno.</s> 677 <itálico> Eu também penso no senhor </itálico>.</s> 678 Saqué la carta; decía una sola frase:</s> 678			21.258	2
144	</s> 679 <itálico> Yo también pienso en usted </itálico>.</s> 679 Quando o cego ouvi dobrar o papel, perguntou:</s>			21.278	2
145	686 — Así es María —dijo, como pensando para sí—.</s> 686 De fato, Maria faz com rapidez coisas que não			21.401	2
146	716 — Caramba, cómo lo lamento —comentó Allende—.</s> 716 — Espero que voltemos a nos ver.</s> 717 Espero			21.969	2
147	raiva para mim mesmo: "Que abominável comédia é essa?"</s> 721 Mientras bajaba en el ascensor, me repetía con			22.038	2
148	apenas: "Já não lhe disse que tenho pensado em tudo?"</s> 747 Observei con amargura que cuando yo le pregunté si			22.915	2
149	leer:</s> 751 <itálico> Eu também penso no senhor </itálico>.</s> 752 <itálico> Yo también pienso en usted </itálico>.</s>			23.077	2
150	752 <itálico> Yo también pienso en usted </itálico>.</s> 752 MARÍA </s> 753 MARÍA </s> 753 A letra era			23.085	2
151	multísimo a assinatura:</s> <itálico> Maria </itálico>.</s> 756 Sea como sea, me emocionó muchísimo la firma:			23.183	2
152	me emocionó muchísimo la firma:</s> <itálico> Maria </itálico>.</s> 756 Simplesmente <itálico> Maria </itálico>.</s> 757			23.195	2
153	<itálico>.</s> 756 Simplesmente <itálico> Maria </itálico>.</s> 757 Simplemente <itálico> Maria </itálico>.</s> 757			23.200	2
154	</itálico>.</s> 757 Simplemente <itálico> Maria </itálico>.</s> 757 Essa simplicidade me dava uma vaga idéia de que eu			23.205	2

Ilustración 17. Concordancia del PUNTO sin itemización

N	Concordance	Set	Tag	Word #	t
124	de minha tela.</s> 71 La observé todo el tiempo con ansiedad.</s> 71 Observei-a o tempo todo com ansiedade.</s> 72			3.423	
125	con ansiedad.</s> 71 Observei-a o tempo todo com ansiedade.</s> 72 Después desapareció en la multitud, mientras yo vacilaba			3.431	
126	entre un miedo invencible y un angustioso deseo de llamarla.</s> 72 Depois ela desapareceu na multidão, enquanto eu			3.452	
127	entre um medo invencível e um desejo angustiante de chamá-la.</s> 73 ¿Miedo de qué?</s> 74 Quizá,			3.473	
128	jugará todo el dinero de que se dispone en la vida a un solo número.</s> 74 Talvez fosse um pouco como o medo de apostar todo o			3.505	
129	todo o dinheiro de que se dispõe na vida em um único número.</s> 75 Sin embargo, cuando desapareció, me sentí irritado,			3.529	
130	entre los millones de habitantes anónimos de Buenos Aires.</s> 75 Entretanto, quando ela desapareceu, senti-me irritado,			3.560	
131	perdida entre os milhões de habitantes anónimos de Buenos Aires.</s> 76 Esa noche volví a casa nervioso, descontento, triste			3.590	
132	.</s> 76 Esa noche volví a casa nervioso, descontento, triste.</s> 76 Essa noite voltei para casa nervoso, descontente, triste			3.602	
133	.</s> 76 Essa noite voltei para casa nervoso, descontente, triste.</s> 77 Hasta que se clausuró el salón, fui todos los días y me			3.614	
134	para reconocer a las personas que se detenían frente a mi cuadro.</s> 77 Até o encerramento do salão, fui todos os dias,			3.644	
135	para reconhecer as pessoas que paravam diante de meu quadro.</s> 78 Pero no volvió a aparecer.</s> 78 Mas ela não apareceu			3.670	
136	paravam diante de meu quadro.</s> 78 Pero no volvió a aparecer.</s> 78 Mas ela não apareceu mais.</s> 79 Durante los meses			3.677	
137	78 Pero no volvió a aparecer.</s> 78 Mas ela não apareceu mais.</s> 79 Durante los meses que siguieron, sólo pensé en ella, en			3.684	
138	siguieron, sólo pensé en ella, en la posibilidad de volver a verla.</s> 79 Durante os meses que se seguiram, só pensei nela, na			3.704	
139	que se seguiram, só pensei nela, na possibilidade de revê-la.</s> 80 Y, en cierto modo, sólo pinté para ella.</s> 80 E, de			3.721	
140	revê-la.</s> 80 Y, en cierto modo, sólo pinté para ella.</s> 80 E, de certo modo, só pintei para ela.</s> 81 Fue como			3.733	
141	sólo pinté para ella.</s> 80 E, de certo modo, só pintei para ela.</s> 81 Fue como si la pequeña escena de la ventana empezara a			3.745	
142	ventana empezara a crecer y a invadir toda la tela y toda mi obra.</s> 81 Foi como se a pequena cena da janela tivesse começado			3.769	
143	começado a crescer e a invadir toda a tela e toda a minha obra.</s> 82 <capitulo> IV </s> 82 <capitulo> 4 </s> 83 UNA TARDE			3.795	
144	82 <capitulo> 4 </s> 83 UNA TARDE, por fin, la vi por la calle.</s> 83 Uma tarde, por fim, vi-a na rua.</s> 84 Caminaba por la			3.817	
145	, por fin, la vi por la calle.</s> 83 Uma tarde, por fim, vi-a na rua.</s> 84 Caminaba por la outra vereda, en forma resuelta, como			3.828	

Ilustración 18. Concordancia del PUNTO con itemización

Para el ámbito de los elementos paratextuales (Genette, 2009), específicamente en el caso de los peritextos, aquellos elementos que se encuentran en el perimetro del texto, utilicé la anotación del corpus para la recuperación posterior de las



Notas del Traductor. Las etiquetas utilizadas fueron <NT+nº>, para indicar el inicio, y </NT+nº> para el término de una nota. La numeración fue secuencial, según el orden de aparición. La siguiente ilustración trae las líneas de concordancia, a partir de la búsqueda por las <NT*>:

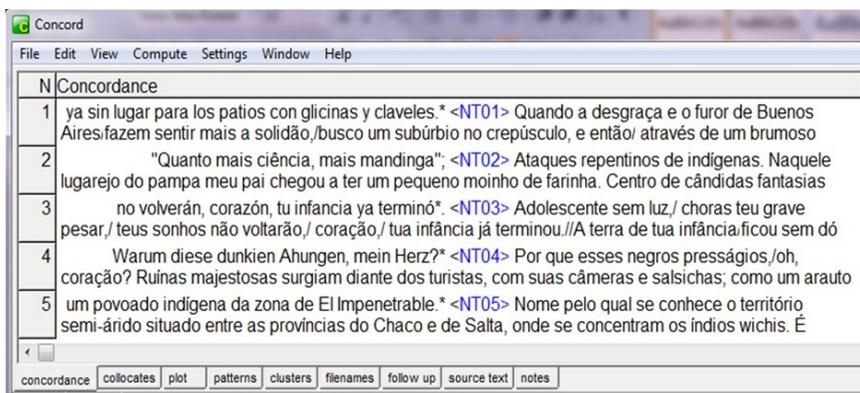


Ilustración 19. Notas del traductor

Para concluir esta presentación de procedimientos metodológicos, vinculados al análisis de estilo de TTs en el ámbito de los ETBC, cabe apuntar que también se aplicaron criterios detallados de organización para todos los elementos paratextuales, tanto para los *epitextos* (paratextos que están más allá del texto) como para los *peritextos* (paratextos que se circunscriben al espacio físico del texto). La ilustración 20 (pág. 135) exhibe una vista parcial de la organización del corpus de estudio, en carpetas independientes, a nivel de análisis y de formato.

La próxima sección concluye este trabajo, con algunas consideraciones a las que se arriba, tras la descripción minuciosa de los procedimientos metodológicos expuestos aquí.

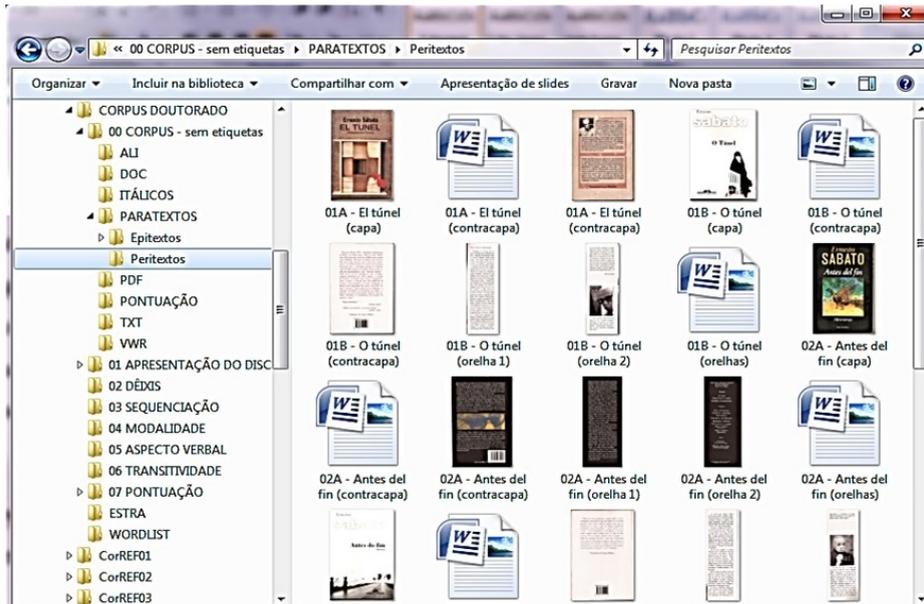


Ilustración 20. Organización del corpus de estudio y elementos paratextuales

4. Consideraciones

El objetivo principal de este trabajo consistió en presentar, detalladamente, algunos de los procedimientos metodológicos adoptados, adaptados y también creados, en una investigación del perfil estilístico circunscrito a la traducción literaria (Novodvorski, 2013), en un corpus paralelo español/portugués. Tras un breve acercamiento teórico, describí las técnicas y recursos que integraron la propuesta de triangulación de los datos y resultados obtenidos en la investigación, a partir del diseño de un cuadro metodológico. Según Saldanha (2011b), aunque se pueda hablar del crecimiento de las investigaciones en el ámbito de los ETBC, aún es necesaria la definición de un cuadro teórico-metodológico más preciso, para la investigación de los perfiles estilísticos de la traducción. En ese sentido, este trabajo también buscó cubrir, por un lado, esa carencia apuntada.

Se describieron todos los pasos implicados para la compilación, preparación y análisis estilístico del corpus de estudio, con apoyo de las herramientas y utilitarios del programa WST (Scott, 2008), para la identificación de temáticas del corpus, por medio de la clavicidad; para el análisis contrastivo de semejanzas y diferencias en el plano léxico-gramatical, vinculado a las (sub)categorías de la presentación del discurso; y para el estudio de significados culturales vehicula-

dos en las traducciones, a través de los usos de la letra cursiva, de los signos de puntuación y de elementos paratextuales.

Destaco la perspectiva de construcción de un cuadro teórico-metodológico, aplicado al análisis de estilo de los textos traducidos y, principalmente, la posibilidad de replicación de los procedimientos adoptados aquí, en futuros trabajos, dando continuidad a la investigación.

5. Bibliografía

- Berber Sardinha, Tony. 2004. *Lingüística de corpus*. Barueri: Manole. ISBN 8520416764.
- Berber Sardinha, Tony. 2009. *Pesquisa em Lingüística de Corpus com WordSmith Tools*. Campinas, SP: Mercado das Letras. ISBN 9788575911068.
- Baker, Mona. 2000. Towards a methodology for investigating the style of a literary translator. @ *Target* 12 / 2: 241-266.
- Genette, Gérard. 2009. *Paratextos Editoriais*. Tradução de Álvaro Faleiros. Cotia, SP: Atelié Editorial. ISBN 9788574804583.
- Hermans, Theo. 1996. The Translator's voice in translated narrative. @ *Target* 8 / 1: 23-48.
- Laviosa, Sara. 2002. *Corpus-Based Translation Studies: Theory, Findings, Applications*. Amsterdam/New York: Editions Rodopi. ISBN 9042014873.
- Leech, Geoffrey y Mick Short. 2007. *Style in Fiction*. Harlow: Pearson. ISBN 9780582784093.
- Malmkjaer, Kristen. 2004. Translational stylistics: Dulcken's translations of Hans Christian Andersen. @ *Language and Literature* 13 / 1: 13-24. DOI: 10.1177/0963947004039484.
- May, Rachel. 1997. Sensible Elocution: how translation works in & upon punctuation. @ *The Translator* 3 / 1: 1-20. ISSN 1355-6509.
- Novodvorski, Ariel. 2013. *Estilo das traduções de Sergio Molina de obras de Ernesto Sabato: um estudo de corpora paralelos espanhol / português*. Tesis (Doctorado en Estudios Lingüísticos) Belo Horizonte: Universidade Federal de Minas Gerais (UFMG).
- Saldanha, Gabriela. 2005. *Style of Translation: An exploration of stylistic patterns in the translations of Margaret Jull Costa and Peter Bush*. Tesis (Doctorado en Filosofía). Dublin: School of Applied Language and Intercultural Studies, Dublin City University.
- Saldanha, Gabriela. 2011b. Translator Style – Methodological considerations. @ *The Translator* 17 / 1: 25-50. ISSN 1355-6509.



- Saldanha, Gabriela. 2011c. Emphatic Italics in English Translations: Stylistic Failure or Motivated Stylistic Resources? @ *Meta: Translators' Journal* 56 /2: 424-442.
- Scott, Mike. 2008. *WordSmith Tools*, version 5.0. Liverpool: Lexical Analysis Software.
- Scott, Mike. 2010. Problems in investigating keyness, or clearing the undergrowth and marking out trails... @ M. Bondi y M. Scott, eds. *Keyness in Texts*. Amsterdam/Philadelphia: John Benjamins Publishing Company. ISBN 9789027223173, pp 43-57.
- Schiavi, Giuliana. 1996. There is always a teller in a tale. @ *Target* 8 / 1: 1-21.
- Semino, Elena y Mick Short. 2004. *Corpus Stylistics*. New York: Routledge. ISBN 0415286697.
- Stubbs, Michael. 2010. Three concepts of keywords. @ M. Bondi y M. Scott, eds. *Keyness in Texts*. Amsterdam/Philadelphia: John Benjamins Publishing Company. ISBN 9789027223173, pp. 21-42
- .

