



Under the Net

Silvia Maria Gomes Caldeira

ADVERTIMENT. La consulta d'aquesta tesi queda condicionada a l'acceptació de les següents condicions d'ús: La difusió d'aquesta tesi per mitjà del servei TDX (www.tdx.cat) i a través del Dipòsit Digital de la UB (diposit.ub.edu) ha estat autoritzada pels titulars dels drets de propietat intel·lectual únicament per a usos privats emmarcats en activitats d'investigació i docència. No s'autoritza la seva reproducció amb finalitats de lucre ni la seva difusió i posada a disposició des d'un lloc aliè al servei TDX ni al Dipòsit Digital de la UB. No s'autoritza la presentació del seu contingut en una finestra o marc aliè a TDX o al Dipòsit Digital de la UB (framing). Aquesta reserva de drets afecta tant al resum de presentació de la tesi com als seus continguts. En la utilització o cita de parts de la tesi és obligat indicar el nom de la persona autora.

ADVERTENCIA. La consulta de esta tesis queda condicionada a la aceptación de las siguientes condiciones de uso: La difusión de esta tesis por medio del servicio TDR (www.tdx.cat) y a través del Repositorio Digital de la UB (diposit.ub.edu) ha sido autorizada por los titulares de los derechos de propiedad intelectual únicamente para usos privados enmarcados en actividades de investigación y docencia. No se autoriza su reproducción con finalidades de lucro ni su difusión y puesta a disposición desde un sitio ajeno al servicio TDR o al Repositorio Digital de la UB. No se autoriza la presentación de su contenido en una ventana o marco ajeno a TDR o al Repositorio Digital de la UB (framing). Esta reserva de derechos afecta tanto al resumen de presentación de la tesis como a sus contenidos. En la utilización o cita de partes de la tesis es obligado indicar el nombre de la persona autora.

WARNING. On having consulted this thesis you're accepting the following use conditions: Spreading this thesis by the TDX (www.tdx.cat) service and by the UB Digital Repository (diposit.ub.edu) has been authorized by the titular of the intellectual property rights only for private uses placed in investigation and teaching activities. Reproduction with lucrative aims is not authorized nor its spreading and availability from a site foreign to the TDX service or to the UB Digital Repository. Introducing its content in a window or frame foreign to the TDX service or to the UB Digital Repository is not authorized (framing). Those rights affect to the presentation summary of the thesis as well as to its contents. In the using or citation of parts of the thesis it's obliged to indicate the name of the author.

Under the Net



Silvia Maria Gomes Caldeira

Tesis presentada para optar al grado de Doctora en Linguística en el programa de doctorado Ciencia Cognitiva y Lenguaje Departamento de Linguística, Universidad de Barcelona,

bajo la supervisión de
Dr. Faustino Diéguez-Vide

Universidad de Barcelona
Enero 2013

Under the Net

Silvia Maria Gomes Caldeira

Tesis presentada para optar
al grado de **Doctora en Lingüística**
en el programa de doctorado *Ciencia Cognitiva y Lenguaje*,
Departamento de Lingüística,
Universidad de Barcelona,

bajo la supervisión de

Dr. Faustino Diéguez-Vide
Universidad de Barcelona



Enero de 2013

Este trabajo de investigación tuvo el apoyo de la
Generalitat de Catalunya (2009-SGR-1428).

Dedicatória

À minha mãe que me apoia e ao meu filho que me motiva.
Ao meu filho que me apoia e à minha mãe que me motiva.

Poeminha do Contra

*Todos esses que aí estão
Atravancando meu caminho,
Eles passarão...
Eu passarinho!*

Mario Quintana
(Prosa e Verso, 1978)

Agradecimientos

A la Universidade Federal de Bahia por el apoyo financiero.

A mi amigo, hermano y compañero de saber, profesor José Garcia Vivas Miranda, un agradecimiento especial. Tengo una deuda eterna con él por haberme presentado este maravilloso mundo que es la complejidad del lenguaje humano. Y, sobre todo, por apoyarme cuando los caminos se volvieron oscuros.

Al profesor Ramon Ferrer i Cancho por el incentivo inicial en realizar esta tesis y por los comentarios valiosos sobre el trabajo.

A mi director de tesis, profesor Faustino Dieguéz-Vide, por su paciencia, dedicación y empeño. Inolvidable.

Al profesor Ricard Gavaldá Mestre, por el apoyo incondicional que me ha dado.

Al Grupo LARCA por la acogida.

A Gemma Boleda y Lluís Padró por el apoyo técnico.

Al profesor Olival Freire Junior por su cumplicidade silenciosa.

A los profesores Denise Maria Barreto Coutinho y Naomar de Almeida Filho por incentivar mis caminos futuros.

A los profesores Fernando Luiz Trindade Rêgo y Ligia Jacobsen Alvarez, mis jefes, por su apoyo.

Al profesor Clemente Tanajura.

A mis colegas de trabajo, mi familia y mis amigos.

Un agradecimiento especial a Luisa Oswaldo Cruz Lehner y Sonia Vicente por el soporte necesario en la realización de la tesis.

A mis primeros “pupilos” Pamela Pitágoras y Augusto Junior, que ahora accedieron al estatus de amigos.

Sumario

Introducción.....	13
Capítulo 1	19
1.1. Enfoque cuantitativo del lenguaje	19
1.2. Métodos cuantitativos	23
1.3. Covarianza y correlación	24
1.4. Regresión lineal	25
1.5. Trayectoria en espacio de fases o estados.....	28
Capítulo 2	31
2.1. La producción lingüística	31
2.1.1. Modelo básico de adquisición del lenguaje oral	34
2.1.2. La escritura.....	37
2.1.3. Papel de la puntuación	44
2.1.4. Aprendizaje de la escritura y puntuación.....	49
2.2. Trastornos del lenguaje y el envejecimiento	51
2.2.1. Características clínicas de la EA.....	52
2.2.2. Características verbales de la EA.....	54
2.2.3. Trastornos de la escritura en la EA.....	56
Capítulo 3	59
3.1. Material.....	59
3.1.1. Autores.....	59
3.1.1.1. Lengua inglesa	60
3.1.1.2. Lengua portuguesa.....	63
3.1.2. Datos.....	65
3.1.3. Consideraciones generales sobre el muestreo	66
3.1.4. Los signos de puntuación en inglés y portugués.....	67

3.2. Método.....	69
3.2.1. Tratamiento previo de los textos.....	70
3.2.2. Lematización	71
3.2.3. Cálculo de las métricas.....	75
3.2.4. Representación gráfica de las métricas	81
3.2.5. Implementación.....	82
3.2.5.1. Tratamiento previo.....	82
3.2.5.2. Lematización.....	83
3.2.5.3. Cálculo de las métricas	84
Capitulo 4	85
4.1. Trabajos anteriores.....	85
4.2. Iris Murdoch	89
4.2.1. Aspectos lingüísticos.....	90
4.2.2. Aspectos metodológicos	95
4.2.2.1. Tratamiento de los textos.....	96
4.2.2.2. Métodos de correlación.....	98
4.2.2.3. Longitud de los textos.....	99
4.2.2.4. Type/token ratio (TTR).....	101
4.2.2.5. WTIR	103
4.2.3. Análisis de los puntos considerados	104
4.2.3.1. Riqueza del vocabulario 1 (TTR55)	104
4.2.3.2. Riqueza del vocabulario 2 (WTIR).....	112
Capítulo 5	115
5.1. Los resultados	115
5.2. Iris Murdoch	117
5.2.1. TTR55	117
5.2.2. Longitud promedio de las oraciones (WPS).....	119

5.2.3. Longitud promedio de las pausas lingüísticas (WPP).....	120
5.2.4. Cantidad promedio de pausas lingüísticas por oración (PPS).....	123
5.2.5. Comparando WPS, WPP e PPS	125
5.3. Escritores en lengua inglesa.....	128
5.3.1. Longitud del texto.....	128
5.3.2. TTR55	130
5.3.3. WTIR.....	131
5.3.4. Longitud promedio de las oraciones (WPS).....	135
5.3.5. Longitud promedio de las pausas lingüísticas (WPP).....	139
5.3.6. Cantidad promedio de pausas lingüísticas por oración (PPS).....	141
5.3.7. Comparando parámetros (WPS, WPP e PPS).....	143
5.4. Escritores en lengua portuguesa.	150
5.4.1. Longitud del texto.....	150
5.4.2. Riqueza del vocabulario 1 (TTR55).....	151
5.4.3. WTIR.....	155
5.4.4. Longitud promedio de las oraciones (WPS).....	157
5.4.5. Longitud promedio de las pausas lingüísticas (WPP).....	162
5.4.6. Pausas lingüísticas por oraciones (PPS)	165
5.4.7. Comparando parámetros (WPS, WPP y PPS)	168
Capítulo 6	175
6.1. Conclusiones.....	177
Referencias	187
Anexos	197

Lista de figuras

Figura 2.1 Gráfico basado en un experimento de adquisición de palabras y construcción de combinaciones (frases) (O'Grady, 2005).	36
Figura 3.1 Reproducción del WTIR de Garrard et al. (2005).	77
Figura 3.2 Ejemplo de un gráfico de trayectoria de estados.	82
Figura 4.1 Crecimiento de la longitud del texto en relación con la edad de todos los libros (a) y tras eliminar el libro <i>Jackson's Dilemma</i> (b).	¡Error! Marcador no definido.
Figura 4.2 Relación entre las métricas TTR55 y WT.	103
Figura 4.3 Reproducción del trabajo de Garrard (a) y de Xuan Le (b),	105
Figura 4.4 Reproducción del gráfico del artículo de Xuan Le (2011) escaneado.	108
Figura 4.5 Crecimiento de los libros de Iris Murdoch (a), Crecimiento en detalle (b); log-log de crecimiento en detalle (c).	114
Figura 5.1 TTR55 representado por RLS (a) y por SPT (b) para Iris Murdoch.	118
Figura 5.2 WPS representado por RLS (a) y por SPT (b) para Iris Murdoch.	120
Figura 5.3 WPP representado por RLS (a) y por SPT (b) para Iris Murdoch.	122
Figura 5.4 PPS representado por RLS (a) y por SPT (b) para Iris Murdoch.	124
Figura 5.5 Comparación de SPT del WPS, WPP y PPS para Iris Murdoch.	127
Figura 5.6 SPT en TTR55 y WT para Wilkie Collins.	131
Figura 5.7 WTIR para cinco autores en lengua inglesa.	135
Figura 5.8 SPT del WPS para cuatro escritores en lengua inglesa.	138
Figura 5.9 SPT del WPP para escritores en lengua inglesa.	140
Figura 5.10 SPT del PPS para los escritores en lengua inglesa.	142
Figura 5.11 Regresión lineal de WPS x WPP (a), WPS x PPS (b) y WPP x PPS (c) para todos los autores de lengua inglesa.	147
Figura 5.12 Comparación de WPP y WPS (a) y WPP y PPS (b) para Frances Hodgson Burnett.	149
Figura 5.13 Correlación de la edad con la longitud de los últimos libros para EV (a) y JS (b).	151
Figura 5.14 SPT en el espacio de TTR55 x WT para Jorge Amado.	154
Figura 5.15 WTIR para José Saramago y Machado de Assis.	156
Figura 5.16 SPT de WPS para los escritores en lengua portuguesa.	158
Figura 5.17 SPT del WPP para escritores en lengua portuguesa.	163
Figura 5.18 SPT del PPS para escritores en lengua portuguesa.	167
Figura 5.19 Regresión lineal de WPS x WPP (a), WPS x PPS (b) y WPP x PPS (c) para todos los autores de lengua portuguesa.	171
Figura 5.20 Comparación de WPS y WPP (a) y WPP y PPS (b) para Machado de Assis.	172

Lista de tablas

Tabla 3.1 Muestreo de los libros en portugués e inglés.....	65
Tabla 3.2 Signos de puntuación principales del portugués e inglés analizados.....	68
Tabla 3.3 Ejemplos de las correcciones para los textos en inglés y portugués.....	71
Tabla 3.4: Código y descripción de las métricas analizadas.....	76
Tabla 4.1 El muestreo de Le (2011) y de este trabajo	106
Tabla 4.2 Resultados de correlaciones entre la edad y el TTR55 para Iris Murdoch.	109
Tabla 5.1 Correlación entre la edad y la longitud final de los textos en inglés. ¡Error! Marcador no definido.	
Tabla 5.2 Resumen estadístico del TTR55 de los libros de los diez autores en lengua inglesa..... ¡Error! Marcador no definido.	
Tabla 5.3. Resumen estadístico del WPS de los libros de los autores en lengua inglesa.....	138
Tabla 5.4. Resumen estadístico del WPP para los libros de los autores en lengua inglesa.	141
Tabla 5.5. Resumen estadístico del PPS para los libros de los autores en lengua inglesa.	143
Tabla 5.6 Correlaciones entre las tres métricas: WPS, WPP y PPS.	144
Tabla 5.7. Correlaciones entre la edad y la longitud final de los textos en portugués.....	150
Tabla 5.8. Resumen estadístico del TTR55 de los libros de los siete autores en lengua portuguesa.....	152
Tabla 5.9. Resumen estadístico de WPS de los libros de los autores en lengua portuguesa.	161
Tabla 5.10. Resumen estadístico del WPP de los libros de los autores en lengua portuguesa.	164
Tabla 5.11. Resumen estadístico del PPS de los libros de los autores en lengua portuguesa.	168
Tabla 5.12 Correlaciones entre WPS, WPP y PPS para autores en lengua portuguesa..	168

Introducción

Este trabajo nace con la finalidad de demostrar –o rebatir– una hipótesis de partida. Según esta hipótesis, de la misma forma que se adquiere un conocimiento lingüístico total en etapas que se van sobreponiendo –como si fueran capas–, al envejecer (y si acontece una alteración neurodegenerativa) se sufre un deterioro de algunas de esas capacidades lingüísticas en sentido inverso. Como el aprendizaje de la escritura es posterior a la adquisición de la lengua oral, y como la puntuación se aprende una vez que se domina la propia escritura, en un cuadro clínico neurodegenerativo se podría aventurar la idea de que acontecerá, primero, las alteraciones en la puntuación, después de la escritura y, por último, de la oralidad, en concomitancia con el avance progresivo de la enfermedad.

Este supuesto lleva a pensar que el análisis de la capacidad de utilizar la puntuación podría ser un marcador temprano de un deterioro cognitivo relacionado con la edad, ya que es lo último que se aprende y, si la hipótesis es correcta, lo primero que se altera. La demostración de la hipótesis –el objetivo de este trabajo– podría llevar a la formulación de diagnósticos tempranos de la enfermedad de Alzheimer partiendo, únicamente, de muestras escritas en diversas etapas: premórbida, prodrómica y presintomática.

Esta hipótesis, a su vez, está basada en la observación de Jakobson y Halle (1973) de que toda unidad lingüística sirve a la vez como contexto para las unidades más simples y/o encuentra su propio contexto en una unidad lingüística más compleja (una vez más, como si la adquisición del lenguaje se da en capas). Los autores caracterizan el proceso de desintegración de esas capas, en una enfermedad como la afasia, de la siguiente forma:

[Lo que se nota es] la desintegración de la trama sonora. Esta disolución sigue un orden temporal de gran regularidad. La regresión afásica ha resultado ser un espejo de la adquisición de los sonidos del habla por parte del niño, mostrando el desarrollo de este a la inversa. [...] Esta búsqueda del orden de adquisiciones y pérdidas no puede limitarse a la estructura fonemática, sino que debe extenderse al sistema gramatical. Tan sólo se ha hecho un pequeño número de intentos en esta dirección, y estos esfuerzos merecen continuarse (Jakobson & Halle, 1973, p. 102).

Aunque los postulados de Jakobson y Halle hacen referencia al componente sonoro (fonológico), y a pesar de que en algunos estudios se ha demostrado que no todos los componentes lingüísticos siguen el mismo patrón (por ejemplo, De Villiers en 1978 observó que el patrón de adquisición de morfemas en inglés no sigue un patrón inverso en la patología), aquí se recoge la extensión al sistema gramatical y, en concreto, a la puntuación, en busca de signos de desintegración del lenguaje relativos al envejecimiento.

Para alcanzar este objetivo, en este trabajo se propone desarrollar los subobjetivos específicos que se citan a continuación:

1. Identificar limitaciones en los métodos relacionados con la tasa *type-token*, que es una métrica para determinar la riqueza del vocabulario utilizado en una producción de lenguaje, sea oral o escrita.
2. Proponer dos nuevas métricas, el WPP (longitud promedio de las pausas lingüísticas) y el PPS (promedio de la cantidad de pausas lingüísticas por oración)¹ para identificar las alteraciones en el uso del lenguaje relacionadas con la puntuación.
3. Proponer un nuevo método de análisis de los resultados que permita observar el comportamiento de las métricas mencionadas a lo largo del tiempo.

¹ Estos conceptos serán aclarados más adelante.

Para llevar a cabo estos propósitos, se han elegido como objeto de estudio algunas obras literarias de escritores de lengua inglesa y de lengua portuguesa (son obras que escribieron en un período amplio a lo largo de sus vidas). Se aplicarán métodos cuantitativos y descriptivos de la trayectoria del uso de las palabras y la puntuación, principalmente para identificar –cuando existan, si existen– las disfunciones o alteraciones del uso de estos componentes del lenguaje cuando se trabaja en escritos en edad tardía. La identificación de estas alteraciones se realiza por medio de la comparación con la escritura de un sujeto diagnosticado con EA, la escritora Iris Murdoch.

En el capítulo 1 se presenta el campo teórico relacionado con los métodos cuantitativos de análisis del lenguaje, campo en el que se inserta el objeto de estudio. Se realiza un breve recorrido histórico y una discusión sobre los métodos de covarianza, correlación y Regresión Lineal, así como su interpretación, pues son métodos utilizados frecuentemente para detectar el deterioro cognitivo que ocurre con el envejecimiento. Se presenta el método de la trayectoria en espacio de fases o estados y se discute su adecuación para identificar alteraciones en la producción escrita del lenguaje dentro del ámbito literario.

En el capítulo 2 se presenta el campo teórico relacionado con el objeto de estudio de este trabajo: se realiza una presentación general sobre algunos aspectos del lenguaje y de la patología en el envejecimiento. Como la hipótesis de partida se basa en la pérdida de algunas capacidades cognitivas en sentido inverso a la adquisición/aprendizaje del lenguaje, se mostrarán algunos modelos teóricos relacionados con la adquisición del lenguaje oral y con el aprendizaje del lenguaje escrito, con la finalidad de evaluar las etapas por las que pasan los niños cuando utilizan el lenguaje escrito: desde una mera representación del lenguaje escrito a una estructura más elaborada de este lenguaje y finalizando con la utilización de la puntuación.

En este mismo capítulo se presentarán también algunas características del lenguaje interno o pensamiento verbal, demostrando que su estructura es diferente a la estructura del lenguaje oral y escrito: el lenguaje interno es altamente confuso, limitado, estenográfico, y fragmentado, mientras que los enunciados escritos son simples, bien estructurados y desarrollados. Esta diferenciación es relevante porque permite deducir, en la línea de diversos trabajos psicolingüísticos sobre las pausas que realizamos los seres humanos para planificar el mensaje, que una producción escrita con una mayor

cantidad de pausas lingüísticas se imbrica claramente con esos procesos de planificación relacionados con el lenguaje interno.

El capítulo 2 se cierra con una revisión de los trastornos del lenguaje observados en el envejecimiento y, más concretamente, los trastornos de la escritura en los pacientes afectados de la enfermedad de Alzheimer. Así, se presentarán diversos trabajos sobre este tema.

El capítulo 3 puntualiza todas las definiciones del presente estudio referentes al material analizado y a los métodos empleados. Para comprobar la hipótesis de que las alteraciones relacionadas con el uso de la puntuación ocurren en un periodo anterior a las disfunciones relacionadas con la escritura y el habla, en este trabajo se ha utilizado como referencia la producción literaria de la escritora inglesa Iris Murdoch, diagnosticada post-mortem con enfermedad de Alzheimer. Además del análisis y descripción del comportamiento lingüístico de la autora, se intentará mostrar si ese comportamiento se reproduce en otros autores de lengua inglesa (10 autores) y, por extensión hacia una posible detección temprana –o diagnóstico precoz–, en otros autores de lengua portuguesa (7 autores). Se han estudiado 450 libros, 328 de lengua inglesa y 122 de lengua portuguesa, que representan la evolución cronológica de la escritura de los autores analizados. El capítulo finaliza con la presentación de todos los procesos desarrollados para el tratamiento previo de los textos, la etapa de lematización de las palabras, el cálculo de las métricas, la representación gráfica de las métricas y la construcción de un programa informático para realizar las tareas descritas.

En el capítulo 4 se realizará una revisión bibliográfica más específica sobre el deterioro del lenguaje oral y escrito en el envejecimiento, así como en la patología relacionada con la enfermedad de Alzheimer. El criterio de selección de esta revisión tiene que ver con el dar a conocer las informaciones relevantes que se relacionan con la hipótesis de este trabajo, así como las conclusiones sobre las alteraciones en la estructura de las oraciones producidas detectadas en estudios longitudinales.

Una atención detallada la tendrán los trabajos anteriores que analizaron la producción literaria de la escritora Iris Murdoch. Los trabajos de Peter Garrard (2005), Serguei Pakhomov (2010) y Xuan Le (2010 y 2011), junto con sus colaboradores, serán analizados en detalle en lo que se refiere a aspectos lingüísticos y a aspectos metodológicos. En el primer caso, se trata de la relación entre el empobrecimiento del

vocabulario y el análisis desfavorable de los críticos literarios. En el segundo, son diversas las conclusiones que se discutirán que son, entre otras, las siguientes: el trabajo con porciones reducidas de algunos libros, extrapolando los resultados para el libro completo; la eliminación de diálogos; algunas críticas sobre la adecuación de ciertas métricas para aportar informaciones sobre la influencia del envejecimiento en la producción de la escritura de la autora; o un análisis sobre si la longitud del texto influye en los valores de las métricas elegidas por estos investigadores.

En el capítulo 5 se muestra el conjunto de resultados encontrados para los datos analizados de la escritora Iris Murdoch, así como para los autores en lengua inglesa y en lengua portuguesa. Se utilizará el método de la trayectoria en espacio de estados para representar la evolución de las métricas WPP (longitud promedio de las pausas lingüísticas) y PPS (promedio de la cantidad de pausas lingüísticas por oración) para cada escritor. Estas métricas se proponen en este estudio para medir la fragmentación del discurso, referenciando el libro entero.

El objetivo de la comparación entre la producción literaria en dos idiomas diferentes fue el de verificar si la alteración detectada en la producción textual de la escritora Iris Murdoch, al analizar la evolución respecto a la propia estructura de las oraciones construidas por la autora en edad más temprana respecto a sus últimos libros, acontece en el ciclo evolutivo de la escritura de otros escritores, independientemente del idioma de producción.

El capítulo 6 presenta las discusiones sobre los resultados con el objetivo principal de demostrar –o rechazar– la hipótesis propuesta de partida. Tras verificar las ideas expuestas y los resultados obtenidos, se cerrará el trabajo con unas breves conclusiones.

Por último, se presentan siete anexos. El primero recoge la Escala de Deterioro Global de Reisberg (vid. Reisberg, Ferris, de Leon, & Crook, 1982) en su versión en español. El segundo muestra la lista de todos los libros analizados, con la longitud y la edad de los autores en el momento de su producción, así como las fuentes de datos sobre los libros y los autores. El anexo número tres muestra los códigos de las etiquetas de los signos de puntuación utilizados. El anexo número cuatro recoge el algoritmo usado para el tratamiento previo de los libros. El quinto anexo muestra un ejemplo de configuración del programa *Freeling 2.0*, que será presentado en el capítulo 3, utilizado

para la lematización de las palabras. En el anexo número seis se puede consultar un cuadro comparativo de las reglas de puntuación para el inglés y para el portugués. Por último, el séptimo anexo incluye todos los gráficos producidos en este trabajo para los autores en lengua inglesa y portuguesa que no aportaron informaciones relevantes en relación con el objetivo del trabajo.

Capítulo 1

*Una idea entera se cambia porque
una palabra se trasladó de sitio.
(Pablo Neruda)*

1.1. Enfoque cuantitativo del lenguaje

Los estudios que imbrican el uso del lenguaje con la ciencia cognitiva trabajan cada vez más con métodos cuantitativos, unos métodos que se pueden encontrar tanto en lingüística general como en algunas aplicaciones de la lingüística (lingüística computacional, procesamiento de lenguaje natural o investigaciones de inteligencia artificial orientadas al lenguaje). Son diversos los métodos empleados para este fin: se encuentran desde las redes neuronales y la lógica difusa (con aplicaciones en ámbitos como la psicolingüística o la neurolingüística) hasta métodos más simples como los análisis estadísticos de los datos de un corpus (Kohler y Rieger, 1993).

Los primeros registros de aplicaciones del enfoque cuantitativo en estudios estadísticos elementales se sitúan dentro de la antigua Grecia y se basaban en determinar el número de unidades en una lengua (i.e., sílabas o letras) para cada poesía o prosa (Rendel Harris, 1883; Kennedy, 2010). Era costumbre de griegos y de romanos estimar la longitud de sus obras literarias mediante la medición de la cantidad y la longitud de las líneas.

En la poesía se computaba el número de versos, pero en la prosa era necesario determinar un patrón de longitud medio, pues era difícil que dos escribas pudieran producir líneas de longitud similar de un modo natural. Así, por ejemplo, en el caso de los griegos, la longitud estándar de una línea de texto de prosa era de 36 letras o 16 sílabas. Esta técnica de medir las líneas de acuerdo con la cantidad de letras o sílabas llevó a los escritores a definir sus escritos por estas medidas y a este procedimiento se le denominó *esticometría* (Rendel Harris, 1883; Kennedy, 2010).

Había varios motivos para utilizar la esticometría. En primer lugar, los escribas a menudo eran pagados por la cantidad de líneas que transcribían. Sus precios se establecieron por decretos legales y eran calculados utilizando las informaciones esticométricas (al igual que nuestros números de página, los totales esticométricos eran informaciones importantes para los clientes y se registraban en los márgenes de los pergaminos). En segundo lugar, los registros esticométricos realizados en la hoja del pergamino ayudaron a identificar las obras (en esa época los textos no siempre tenían títulos) y eran una garantía de que las copias estaban completas. Por último, algunos investigadores creen que los rollos de pergamino literarios clásicos estaban fabricados en longitudes estandarizadas y que eran muy caros, de forma que los autores se vieron abocados a producir sus textos respetando esos parámetros como una manera de ahorrar (y evitando los espacios en blanco).

En resumen, las columnas y las líneas en los papiros literarios clásicos eran, por lo general, bastante uniformes y fue más fácil para los escribas (y para sus lectores) mantener los recuentos precisos (Kennedy, 2010).

La técnica de la esticometría fue creada por Callimachus, de la biblioteca de Alexandria, cuando registró las informaciones de los libros de algunos catálogos de acuerdo con el nombre del autor, el título del libro, la primera línea de su contenido y el número de líneas. Hay registros esticométricos en textos clásicos, como en los de Hipócrates o Platón (Rendel Harris, 1883; Kennedy, 2010).

De igual forma, también se han analizado el Viejo y el Nuevo Testamento de la Biblia (Rendel Harris, 1883), pero los textos bíblicos se transcribieron en un sistema diferente de la esticometría. Se trata de otro sistema, denominado colometría y que consiste en la división o ruptura del texto en líneas o frases cortas de acuerdo con el sentido, de forma que se conseguía una mejor comprensión del significado y una mejor presentación al público lector. Este cambio comenzó primero en los libros poéticos del Viejo Testamento y se difundió después para todo el texto bíblico. En un principio, coexistían los dos métodos de construcción y medida (líneas de longitudes fijas y líneas divididas basadas en el significado), pero este último ha suplantado al primero, originando las frases que conocemos hoy.

En la actualidad, son muchos los enfoques del lenguaje desde el punto de vista cuantitativo. Más allá de las frases –que es lo que se trabajó en un inicio– casi todos los

aspectos del lenguaje –tanto en la modalidad oral como escrita– pueden ser analizados con métodos cuantitativos. Además de las aplicaciones ya comentadas, la propia lingüística general ya se beneficia de algunos de estos métodos: por citar algunos, la fonología, la morfología, la lexicografía, la sintaxis, la semántica del discurso, la pragmática y el diálogo (Crystal, 1994).

De las investigaciones recientes, la más importante fue la de George Kingsley Zipf, lingüista estadounidense que vivió en la primera mitad del siglo pasado. En sus estudios en inglés, latín, chino y otras lenguas, Zipf (1968; 1972) encontró un patrón de comportamiento relacionado con la frecuencia de uso de las palabras y su rango. En concreto, demostró que la palabra más frecuente en una muestra ocurría una vez de cada 10 palabras como promedio; la segunda palabra más frecuente ocurría una vez de cada 20 palabras; la tercera palabra más frecuente una vez de cada 30 palabras ... la centésima palabra más frecuente ocurrirá una vez de cada 1000 palabras; etcétera. Sugirió que su distribución de frecuencia se parecía a una serie armónica y tenía una gran precisión.

Esta evidencia apuntaba hacia la existencia de una condición fundamental de equilibrio entre la forma y la función de los hábitos de habla, o patrones de habla, en las lenguas. Según Zipf, y a partir de sus datos experimentales, preservar o recuperar esta condición de equilibrio es la causa principal de los cambios lingüísticos que ocurren frecuentemente (Ferrer i Cancho, 2003).

A partir de estos datos, se planteó, como objetivo principal de su investigación, la observación, la medida y, si era posible, la formulación de leyes preliminares que explicasen las fuerzas que rigen las manifestaciones lingüísticas, construidas sobre una base estadística. A pesar de no ser el primero en realizar estudios estadísticos relacionados con el lenguaje, Zipf presentó de forma sistemática una nueva manera de abordar los fenómenos lingüísticos, fundando un nuevo campo de estudio, la lingüística cuantitativa moderna. Este campo se diferencia de la lingüística cuantitativa tradicional en sus objetivos, métodos e intereses y, principalmente, por el empleo de la matemática y de la estadística como herramientas, más allá de la lógica y la formalización.

La lingüística cuantitativa, entonces, fue el foco de interés de aquellos investigadores que entendían el lenguaje como un sistema auto-organizado, con algunas propiedades peculiares, compuesto de elementos distinguibles que están conectados

entre sí a través de relaciones más o menos conocidas (Altmann, 1978; Köhler, 1987; Nowak, Plotkin, y Jansen, 2000; Dorogovstev y Mendes, 2001; Ferrer-i-Cancho y Solé, 2001; Ferrer-i-Cancho, Solé y Köhler, 2004; Motter, Moura, Lai y Dasgupta, 2002; Albert y Barabasi, 2002; Montemurro y Pury, 2002).

Profundizar en los estudios del lenguaje, ya sea por la vía tradicional, ya sea por la vía cuantitativa, es importante para poder describir el objeto estudiado y para explicarlo por medio de los patrones encontrados en estas investigaciones. Zipf destacó la importancia del papel de los estudios lingüísticos “tradicionales” al referirse a las frecuentes críticas de ciertos lingüistas sobre el enfoque cuantitativo del lenguaje, comentando esta cuestión en su libro:

...it is difficult to believe that linguistics has been entirely mistaken in the direction it gave to language study. Certainly no student of speech-dynamics can for moment regret the stringency of the historical and comparative disciplines which have provided him with immediately available material. Furthermore, he cannot forget that these same censured linguists were the ones who proved conclusively, as we shall see at a more opportune time, that phonetic development, whatever its amorphous and random appearance may be, and is essentially an orderly process with a high degree of precision. Without this previous auspicious knowledge that phonetic development is orderly, few persons would today dare to undertake an investigation of the causal laws behind speech-activity (Zipf, 1968, pp. 6-7).

Con el desarrollo de los ordenadores, las investigaciones lingüísticas han aumentado considerablemente, tanto desde la perspectiva tradicional como cuantitativa. En ambos casos, la capacidad de manipular grandes cantidades de datos permitía una mayor descripción y robustez en las conclusiones de las investigaciones. Además, por la naturaleza del fenómeno lingüístico, se formó una unión de varios campos de conocimiento (filósofos, lingüistas, matemáticos, físicos, biólogos, psicólogos, neurólogos, neuropsicólogos e informáticos), con el objetivo de describir el funcionamiento y las leyes que rigen el uso del lenguaje. Estos estudios

interdisciplinarios interesados en comprender cómo la información se representa y se transforma en la mente/cerebro dieron origen a la ciencia cognitiva (Gardner, 1988).

En la próxima parte del trabajo se definirán los métodos cuantitativos que serán utilizados en este trabajo.

1.2. Métodos cuantitativos

Los métodos cuantitativos presentados en este trabajo pueden ser clasificados como estadísticos clásicos y de representación dinámica. Los primeros son los más utilizados en lingüística y se destinan a cuatro objetivos generales básicos (Johnson, 2008):

1. *Reducción de los datos*: un resumen de las tendencias, la captura de los aspectos del conjunto de observaciones, como la media, la desviación estándar y las correlaciones entre variables.
2. *Inferencia*: generalizar a partir de un conjunto representativo de observaciones a un amplio universo de posibles observaciones, utilizando las pruebas de hipótesis, tales como *t*-test o el análisis de la varianza.
3. *Descubrimiento de relaciones*: encontrar patrones descriptivos o causales en los datos que se puedan describir con modelos de regresión múltiple o con análisis factorial.
4. *Exploración de procesos que puedan tener una base en la probabilidad*: proponer modelos teóricos como, por ejemplo, en la teoría de la información o en contextos prácticos (como el análisis probabilístico de oraciones).

En el presente estudio, utilizamos las reducciones de datos y la investigación de relaciones entre las variables a partir de métricas sobre los libros y la edad en que fueron escritos. Los métodos empleados son, por un lado, la covarianza y la correlación y, por otro, la Regresión Lineal. A continuación, se describirán estos métodos.²

² Se seguirá la descripción realizada por Keith Johnson (2008) para los métodos estadísticos clásicos (covarianza y correlación, y Regresión Lineal) porque se trata de una descripción clara y concisa.

Aparte de las métricas tradicionales (covarianza y correlación), utilizaremos la representación dinámica del uso del lenguaje escrito por los autores analizados, desde una perspectiva longitudinal, como una trayectoria en espacio de fases o estados. Este método servirá para analizar el comportamiento de utilización de componentes del lenguaje escrito, como la puntuación, en los libros escritos a lo largo de la vida de los escritores.

1.3. Covarianza y correlación

La idea clave de una medida de correlación entre dos variables esta en medir la desviación entre x_i ³ y la media \bar{x} . Para tener una idea general de lo fuerte que es una relación entre dos variables se deben multiplicar las desviaciones de x e y , para después sumar esos productos en todas las observaciones.

$$\sum_{i=0}^n (x_i - \bar{x})(y_i - \bar{y})$$

El producto de las desviaciones será mayor a medida que el conjunto de datos sea cada vez mayor, por lo que se necesita normalizar este valor para la longitud del conjunto de datos (el número de puntos de x e y) tomando el promedio de las desviaciones pareadas. El promedio del producto de las desviaciones se llama la covarianza de x e y ; esta covarianza se define como

$$\frac{\sum_{i=0}^n (x_i - \bar{x})(y_i - \bar{y})}{n}$$

Por su parte, la correlación es la *correlación normalizada* –es decir, corregida para el número total de valores (dividida por n)–, mediante el producto de las desviaciones estándar de x e y indica la fuerza y la dirección de una relación lineal y de proporcionalidad entre dos variables estadísticas. Se considera que dos variables cuantitativas están correlacionadas cuando hay una sistemática variación de una con

³ x_i representa una variable estadística. Es una característica que, al ser medida en diferentes individuos/objetos, puede adoptar diferentes valores. Así, por ejemplo, el x_i puede representar la cantidad de palabras utilizadas en los libros escritos por un autor; de esta forma, para cada libro tendremos un valor de x_i que lo caracteriza.

respecto a la otra. Si tenemos dos variables (x y y) existe correlación positiva si al aumentar los valores de x lo hacen también los de y .

El resultado principal aquí es que el coeficiente de correlación es simplemente una versión a escala de la suma del producto de las desviaciones utilizando la idea de que este valor será mayor cuando x e y se desvíen de sus promedios en una magnitud comparable. ‘Correlación’ es idéntico a ‘covarianza’, excepto por el hecho de que la correlación es escalada por las desviaciones estándar. Así, la covarianza puede tener cualquier valor, mientras que las correlaciones varían en el rango de 1 a -1 (una correlación positiva perfecta es 1 y una correlación negativa perfecta es -1; Johnson, 2008).

1.4. Regresión lineal

Cuando representamos la correlación lineal entre dos variables en un gráfico podemos representar esta relación por medio de una recta: se trata de representar el resultado del cálculo con la línea recta que mejor se ajuste a los datos. Típicamente, la línea de mejor ajuste se obtiene por el método de los *mínimos cuadrados*. La estimación de la pendiente de la recta de regresión de mínimos cuadrados se relaciona con la estimación de correlación de Pearson de modo directo, o sea, si el coeficiente de la correlación de Pearson es positivo, la recta de regresión de mínimos cuadrados tiene una pendiente positiva y viceversa (Wilcox, 2002).

Dado que existe un método para medir la asociación entre dos variables continuas, que nos da la correlación de Pearson (r_{xy}), y una manera de utilizar esa medida para determinar la pendiente y la intersección de la recta de mejor ajuste que relaciona x e y (suponiendo que una relación lineal es la correcta), es posible predecir el valor de y , dado un valor de x . La medida estadística r_{xy} que varía en el rango de -1 a 1 indica el grado de asociación entre las dos variables. Pero ahora la pregunta es: ¿esto es un modelo o es solo una línea? En este caso la variable y se denomina ‘dependiente’ (o de respuesta) y la variable x se denomina ‘independiente’ (la que predice).

Podremos decir que este es un *modelo de regresión* sí es razonable pensar que x puede causar y . En esta hipótesis podemos pensar en el modelo como una función lineal de relación causal. Así, por ejemplo, tenemos una relación causal lineal

indiscutible cuando relacionamos la cantidad de calor sometida a un objeto de hierro y la temperatura medida.

Por otra parte, si una relación de causalidad no parece razonable, entonces vamos a hablar de que existe una correlación entre las dos variables sin mencionar la causalidad (Johnson, 2008). Es el caso, por ejemplo de que con el paso del tiempo un escritor escriba libros cada vez mayores. Es un comportamiento no trivial, y que, hasta ahora no hay una explicación causal para que esto ocurra. Entonces podremos decir que hay una correlación entre estas dos variables, pero no hay causalidad entre ellas.

En diversos trabajos anteriores que investigaron la evolución de los libros de Iris Murdoch (y que se describirán en el capítulo 4), se utiliza, aparte de los coeficientes de Pearson y de Spearman⁴, una metodología propuesta por Sir Ronald Fisher y, genéricamente, designada como Análisis de Variancia (ANOVA⁵) (Fisher, 1935). La idea que hay detrás de este análisis es la de verificar si la variancia residual (aquella que está asociada a los errores de medida) es significativamente inferior a la variancia detectada entre las variables analizadas, indicando que existe un efecto de una variable sobre la otra y que este efecto no se relaciona con los errores de medida.

Como ejemplo, tomemos las variables WT⁶ (longitud del texto medido en palabras) y la edad en que esos libros fueron escritos. Si queremos saber si existe un comportamiento de crecimiento o decrecimiento de esta variable, medida para cada libro y que se relaciona con el paso del tiempo, se puede hacer un análisis de Regresión Lineal de estas variables por el método ANOVA⁷. Este análisis nos daría el valor de $F = 7,08$ y p (p -value) = 0,02. Esto significa que, con una probabilidad de error de 5% ($\alpha = 0,05$) podemos concluir que la relación entre estas dos variables es estadísticamente significativa, pues el p -value encontrado (p) = 0,02 es menor que el $\alpha = 0,05$.

⁴ La correlación por el método de Spearman no se abordará en este estudio, pues la métrica no se aplica para la métrica analizada en este trabajo.

⁵ Abreviatura de su sigla en inglés **Analysis of Variance.**

⁶ Las citas de variables y métricas, como el WT y TTR55 se han utilizado aquí como un recurso lingüístico para poder comprender los métodos, pero se detallarán mejor en los capítulos siguientes.

⁷ La explicación matemática de la ANOVA se puede mirar en Maroco, 2007: 154.

Pero hay ocasiones en que la regresión simple no es el método más apropiado. Es el caso de cuando tenemos dos variables independientes que se correlacionan con una variable dependiente como, por ejemplo, si se quiere verificar si el comportamiento del TTR55 (variable dependiente) depende de la longitud del texto (WT) y de la Edad (dos variables independientes). En este caso, la Regresión Lineal simple nos daría el valor de correlación individual de cada una de las métricas con la variable dependiente, pero no sería posible identificar cuál es el grado de contribución de cada una en el comportamiento estudiado. Así, la pregunta es la siguiente: ¿el vocabulario en las porciones iniciales de los libros crece (o disminuye) porque el autor ha escrito libros mayores (o menores), o porque el paso del tiempo es el factor que influye en este comportamiento?; o ¿el comportamiento de la variable dependiente puede ser explicado, a la vez, por las dos variables? En estos casos, se debe utilizar la Regresión Lineal Múltiple (RLM).

La RLM identifica si las variables independientes juntas explican el comportamiento de una variable dependiente (cuando existe una correlación entre ellas) de un modo significativo; además, cuando existe una correlación significativa entre las variables independientes y la variable dependiente, se identifica cuál es la contribución de cada una de ellas para el comportamiento identificado, analizando los Coeficientes de Regresión Estandarizados (también conocidos como los *Coeficientes Beta*), los valores de la estadística de teste y los p-value de estas pruebas⁸.

Los métodos de correlación simple, múltiple y el método ANOVA poseen una gran importancia científica en el análisis de comportamientos en el ámbito psicológico y en otras ramas de la ciencia. En este estudio se utilizará esta metodología para intentar describir algunos aspectos importantes para sostener –o rechazar– la hipótesis propuesta. Pero, en función de la naturaleza de la investigación realizada aquí, que se caracteriza por identificar momentos en la vida de escritores en los que podrían acontecer disfunciones cognitivas asociadas a su escritura, se ha optado por utilizar otro método científico, la trayectoria en espacio de fases o estados, para poder extraer conclusiones sobre ese desarrollo cognitivo. Se trata de un método descriptivo del

⁸ En este estudio no haremos la explicación matemática de los coeficientes Beta y del método de estadística de teste, pero se puede mirar en Maroco, 2007: 577.

comportamiento de un determinado fenómeno basado en el paso del tiempo y que será explicado en detalle en el siguiente apartado.

1.5. Trayectoria en espacio de fases o estados

Todo proceso en que hay cambios (ya sea movimiento, ya sea variaciones de estados) puede ser considerado como un sistema dinámico. Los fenómenos que evolucionan en el tiempo son sistemas dinámicos. En general, un sistema dinámico es una transformación en un espacio de estados posibles, lo que también es conocido como *espacio de fases* (Talman, 2008).

Las trayectorias de los espacios de fases son resultados de un conjunto de variables en el tiempo $N_1(t_1)$, $N_2(t_2)$, ..., $N_n(t_n)$, lo que describe una dinámica de un sistema con grado de libertad n^9 . Las teorías físicas los han considerado como instrumentos que permiten predecir, a partir de ciertos hechos observados, los resultados de futuras observaciones. En este estudio, con esta representación, producimos un análisis con grado de libertad igual a dos (variable 1 *versus* variable 2 en el tiempo). Por lo tanto, un grado más que si estuviéramos trabajando con las correlaciones (variable 1 *versus* tiempo).

De esta forma, por ejemplo, si se quiere observar el comportamiento de la longitud promedio de las oraciones en los libros con el paso del tiempo (con el avance de la edad del escritor), por el método de la correlación se haría el análisis confrontando la edad con los valores de la longitud promedio de las oraciones (grado de libertad uno). Apenas una variable sería analizada. Por el método de la trayectoria de estados o fases se confrontarían dos variables (como, por ejemplo, la longitud promedio de las oraciones y la longitud de los libros, estando representada la edad en el gráfico en los puntos donde estas variables fueron medidas).

Un ejemplo clásico de una representación de trayectoria es la de un cuerpo que sigue una trayectoria, descrita por el par de ecuaciones $S(T)$ y $V(t)$, donde:

S = posición de un cuerpo en movimiento;

V = velocidad del cuerpo;

⁹ El grado de libertad representa la cantidad de variables analizadas a la vez.

T = tiempo transcurrido del movimiento.

En la Figura 1.1, tenemos dos representaciones de la trayectoria, donde podremos analizar las 3 variables (S, T y V).

En el grafico (a), podemos constatar que hasta el tiempo =3 la velocidad del cuerpo era constante e igual a 2, $V(3)=2$. En el tiempo = 4 la velocidad asume el valor = 1, $V(4)=1$, regresando a la misma velocidad anterior en el tiempo = 7.

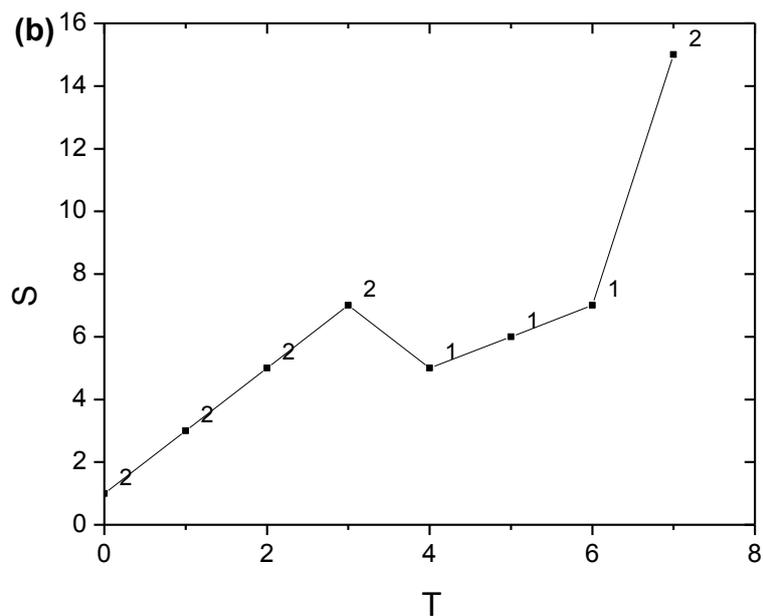
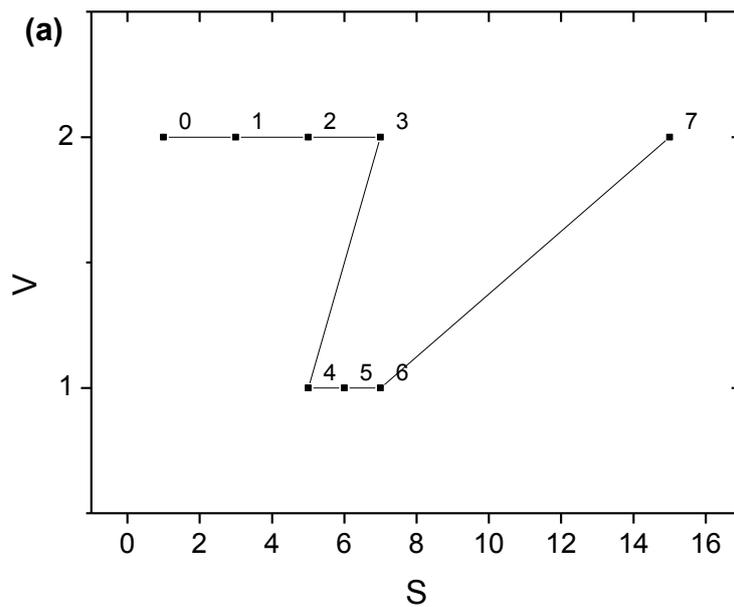


Figura 1.1 Representación de la trayectoria de la ecuación del movimiento.

En el gráfico (b), acompañamos las 3 variables desde otra perspectiva, en que se relaciona la posición y el tiempo y marcamos en el gráfico los cambios en la velocidad. Es fácil la visualización de que hasta el tiempo = 3 el desplazamiento se produjo a una velocidad constante y tras este tiempo la velocidad ha cambiado, pero el cuerpo continua desplazándose, aunque en otra trayectoria.

La principal ventaja de la utilización de la representación gráfica de la trayectoria en espacio (o estados) de fases es la visualización del cambio de comportamiento de una variable a lo largo del tiempo, como, por ejemplo, la longitud promedio de las oraciones de un libro y la edad en que se ha escrito el libro analizado.

Capítulo 2

Escribo para descubrir en que estoy pensando.

(Edward Albee)

Como ya se ha comentado en la introducción a este trabajo, en este capítulo se realizará una breve presentación del campo teórico en el que se inserta esta investigación. Este campo teórico es doble. Primero, y dado que la hipótesis de partida se relaciona con algunas capacidades cognitivas relacionadas con el lenguaje escrito, se realizará un muy breve recorrido por las diversas etapas que los seres humanos recorreremos en nuestra vida lingüística, desde la primera adquisición de la lengua hasta el aprendizaje de la escritura. Aun cuando pueden acontecer diversas causas, se planteará también la idea general de que un cambio en las pausas lingüísticas realizadas implicará un cambio en la forma de planificación, lo que indicará una alteración o disfunción cognitiva.

Segundo, y dado que la hipótesis de partida se relaciona con la potencial pérdida de esas capacidades comentadas en sujetos con enfermedad de Alzheimer, se delinearán algunos trastornos del lenguaje observados en el envejecimiento. En particular, aquellos trastornos de la escritura que se pueden observar en los pacientes con enfermedad de Alzheimer.

2.1. La producción lingüística

El lenguaje es un sistema formado por varios niveles (Crystal, 1994). En un primer nivel, asemántico, los sonidos y las letras son los elementos fundamentales del lenguaje oral y escrito, respectivamente. Estas pequeñas unidades, cuando se abstraen,

se convierten en fonemas (oral) y grafemas (escrito). Por su parte, la unión de fonemas y grafemas conforman sílabas. (Dieguéz-Vide y Peña-Casanova, 2012).

En el siguiente nivel entramos en el significado. Las mínimas unidades semánticas son los morfemas y las palabras. Los primeros poseen un significado semántico o gramatical, mientras que las palabras se forman a partir de un único morfema o mediante la concatenación de varios morfemas.

El tercer nivel, formado a partir de la combinación de palabras, se relaciona con la sintaxis y se encuentran desde unidades menores (sintagma) a unidades medias (cláusula y oración) y mayores (frase).

Por último, existiría el nivel del discurso y el texto, caracterizados por ser unidades continuas y finitas (Kohler y Rieger, 1993).

Por supuesto, todos estos niveles se individualizan para poder estudiarlos sincrónicamente, pero en toda conversación, en toda expresión y en toda producción y comprensión se vinculan unos a otros como si fueran enormes redes comunicativas.

La adquisición de estos constituyentes también acontece en etapas, poseyendo cada una de ellas un nivel creciente de complejidad agregada. La primera producción lingüística que realizamos los seres humanos es la producción de sonidos. Los bebés pronuncian los primeros sonidos básicos, también llamados “ruidos biológicos básicos” (Crystal, 1994: 236) en el periodo de 0-8 meses. Estos sonidos vocales reflejan estados biológicos (hambre o dolor) y actividades (comer o respirar). Pero antes de concluir la adquisición de los componentes de un nivel, ya comienza la adquisición de la próxima etapa. El desarrollo cognitivo relacionado con el uso del lenguaje es acumulativo y, en varios momentos, sus diversas etapas coexisten (McCune, 2008).

Existen diversas aproximaciones sobre cómo se produce la adquisición del lenguaje. Si bien existen diversas teorías, ninguna de ellas ha sido verificada en la actualidad. De entre todas las teorías postuladas, cuatro son las que han tenido –y tienen– más “seguidores”. Estas cuatro, que se presentan muy resumidas a continuación (no es este el lugar para una descripción detallada de las mismas), son las siguientes (Owens Jr., 2000; Lust y Foley, 2004):

1. *Enfoque conductista (Skinner)*: es un enfoque empirista, es decir, considera que todo comportamiento se adquiere/aprende exclusivamente a partir de estímulos

externos exclusivamente. Así, el lenguaje se define como una conducta verbal y, al igual que todas las conductas, se adquiere gracias a las contingencias del ambiente.

2. *Enfoque innatista (Chomsky)*: se trata de un enfoque que se fija en los procesos mentales y en cómo ellos se reflejan en el lenguaje. Se fundamenta en que los seres humanos nacemos con un constructo interno innato, el LAD (*Language Acquisition Device* o Dispositivo de Adquisición del Lenguaje), que permite que entendamos y produzcamos el lenguaje de un modo universal gracias a que la mente humana, innatamente, posee una Gramática Universal que “modela” (parametriza) el ambiente en una gramática particular.

3. *Enfoque cognitivo (Piaget)*: aunque nunca formalizó una teoría sobre el desarrollo del lenguaje, Piaget plantea una teoría genética y formal del conocimiento, según la cual el niño debe dominar la estructura conceptual del mundo físico y social para adquirir el lenguaje. La inteligencia era su objeto de estudio, pero, para él, el lenguaje se podría equiparar con la inteligencia.

4. *Enfoque sociocultural (Vygotsky)*: Vygotsky tampoco formalizó una teoría sobre la adquisición del lenguaje, pero, para él, el lenguaje, en principio, es la principal herramienta de comunicación y esta se desarrolla dentro de un determinado contexto histórico-social, o sea, el lenguaje poseería un carácter totalmente social. Así, Vygotsky define la adquisición del lenguaje como un mecanismo para la comunicación con los demás; después, cuando se desarrolla el lenguaje egocéntrico, el lenguaje pasa a ser una herramienta para la resolución de problemas; y, finalmente, con la evolución y el uso, el lenguaje se convierte en un pensamiento verbal con una forma interna (Daniels, 2005).

De acuerdo con lo expuesto, es fácil apreciar las diferencias de enfoques en relación con el inicio de la producción del lenguaje. En principio, las teorías expuestas se aplican principalmente a la adquisición del lenguaje, que es un proceso diferente del aprendizaje. Así, la adquisición es el proceso por el cual se consigue la competencia en la lengua materna oral, proceso que es sin esfuerzo y de modo inconsciente. También es un proceso de adquisición el dominio de una segunda lengua oral en la etapa infantil (entre tres y cinco años, dependiendo de los autores). Por el contrario, cuando se consigue la competencia de una lengua después del período infantil, o cuando se domina la lectoescritura de la propia lengua, se trata de procesos de aprendizaje, procesos que son costosos y conscientes (Dieguéz-Vide y Peña-Casanova, 2012).

Aunque el objetivo del trabajo se relaciona, de esta forma, con procesos de aprendizaje, vale la pena echar una ojeada a un modelo básico y general sobre la adquisición del lenguaje oral, pues según la hipótesis de la regresión presentada (cfr. *supra*) serían estos procesos los primeros en adquirirse y los últimos en perderse tras una patología neurodegenerativa.

2.1.1. Modelo básico de adquisición del lenguaje oral

Para abordar con mayor facilidad el estudio del lenguaje, los lingüistas dividen en partes sus varios componentes: sonidos, palabras, oraciones, significados, etc –y se ocupan de ellos por separado–. Esto puede sugerir que la adquisición del lenguaje ocurre de modo secuencial, pero como ya se ha explicado los niños van adquiriendo las capacidades de forma simultánea.

En el primer año el lenguaje del bebé pasa por varios cambios en su capacidad de producción de sonidos, hasta que pronuncia su primera palabra. Según Crystal (1994: 236), son cinco los estadios en el desarrollo inicial de las capacidades de producción del lenguaje y que describiremos a continuación:

1. *Estadio I (0-8 semanas) – Ruidos biológicos básicos:* el bebé emite sonidos asociados a actividades y/o estados, como respiración o dolor, respectivamente. Estos sonidos iniciales no tendrían ningún rasgo específico del lenguaje.

2. *Estadio II (8-20 semanas) – Arrullo y risa:* el arrullo aparece entre las 6 y 8 semanas, se desarrollan de manera concomitante al llanto y se hacen más frecuentes como respuestas a las sonrisas y al habla de los progenitores. En este estadio los bebés empiezan a realizar las primeras actividades necesarias a la posterior producción del habla.

3. *Estadio III (20-30 semanas) – Juego vocal:* los sonidos de este estadio son más estables que los del estadio anterior. El bebé produce una secuencia de sonidos, similares a consonantes y vocales, que se repiten. Al final del periodo los sonidos se combinan en secuencias cada vez más largas, convirtiéndose en las primeras emisiones balbuceadas.

4. *Estadio IV (25-50 semanas) – Balbuceo:* el bebé utiliza un conjunto pequeño de sonidos, pero con mayor frecuencia y estabilidad, produciendo el *bababa*. El ritmo de la emisión y la longitud de la sílaba son más próximos de lo que sucede en el habla adulta.

5. *Estadio V (9-18 meses) – Emisión melódica:* las variaciones en la melodía, el ritmo y el tono de voz se convierten en un rasgo fundamental de las emisiones del bebé. Las sílabas individuales presentan una melodía fija, produciendo *protopalabras* en las que los sonidos son claros, aunque sin significados definidos.

Tras el primer año, los niños, en general, empiezan a presentar los primeros signos reales de desarrollo del lenguaje; sin embargo, en realidad, los niños comienzan a producir fonemas, a utilizar palabras y a adquirir los significados antes de su dominio de todos los sonidos del idioma. Además, comienzan a construir oraciones, por lo general, después de adquirir apenas unas pocas docenas de palabras. Por lo tanto, es necesario reiterar que no existen distintos períodos de tiempo durante los cuales los niños están trabajando en los sonidos, palabras, significados y oraciones por separado, sino que todo ocurre al mismo tiempo y con niveles de complejidad y profundidad evolutivos.

En relación con las frases y, en concreto, con la construcción de las mismas, se ve favorecida por la complejidad del desarrollo de todos los niveles del lenguaje al mismo tiempo. No obstante, existen también dos fases:

1. La primera fase, que comienza a la edad de dieciocho meses, se caracteriza por la aparición de oraciones con una o dos palabras, en un patrón relativamente simple. Estas primeras y primitivas pequeñas frases son incompletas (por ejemplo, frases sin verbos como *Mama comida*, que representa la comunicación a su mamá de que tiene hambre), pero marcan el comienzo de algo grande. Los niños comienzan, entonces, a combinar palabras en patrones muy cortos y sencillos.

2. Durante la segunda fase (Estadio III según la clasificación realizada en el año 1994 por Crystal), que comienza alrededor de dos años de edad o menos, las piezas que faltan se llenan y se produce un rápido crecimiento en la capacidad de producir una gran variedad de construcciones complejas, con más de dos palabras (O'Grady, 2005). Así, las oraciones en esta fase son más elaboradas, del tipo negaciones, preguntas (cómo

dónde, cómo, por qué y cuándo), preguntas del tipo sí/no y cláusulas subordinadas. Dominar la construcción de estos diversos tipos de oraciones es una de las principales tareas en el proceso de la adquisición del lenguaje. Y esto sucede a una velocidad increíble: un niño de tres años de edad posee ya un firme control sobre las herramientas de construcción de oraciones de su lenguaje, presentando un vocabulario de 900 a 1.000 palabras y produciendo cerca de 12.000 palabras al día. No obstante, la mayoría de oraciones producidas son sencillas y muchas siguen el formato de sujeto-verbo-objeto (Owens Jr., 2000).

William O’Grady (2005) describe un experimento realizado en la Universidad de Harvard en que se registra la evolución lingüística de un niño (Damon) medido por semanas: en la primera semana el niño tenía un poco más de 12 meses. En el gráfico de la Figura 2.1, reproducida del libro de O’Grady, vemos que a los 15 meses (semana número 13), en relación con la implementación de su vocabulario –es decir, el número de palabras diferentes que ha aprendido (“Words”)– el niño alcanzó el hito de 50 palabras. En esta época, el número de combinaciones de palabras (“Combinations”: representado por la línea continua) comenzó primero lentamente para, posteriormente, avanzar con mucha más rapidez.

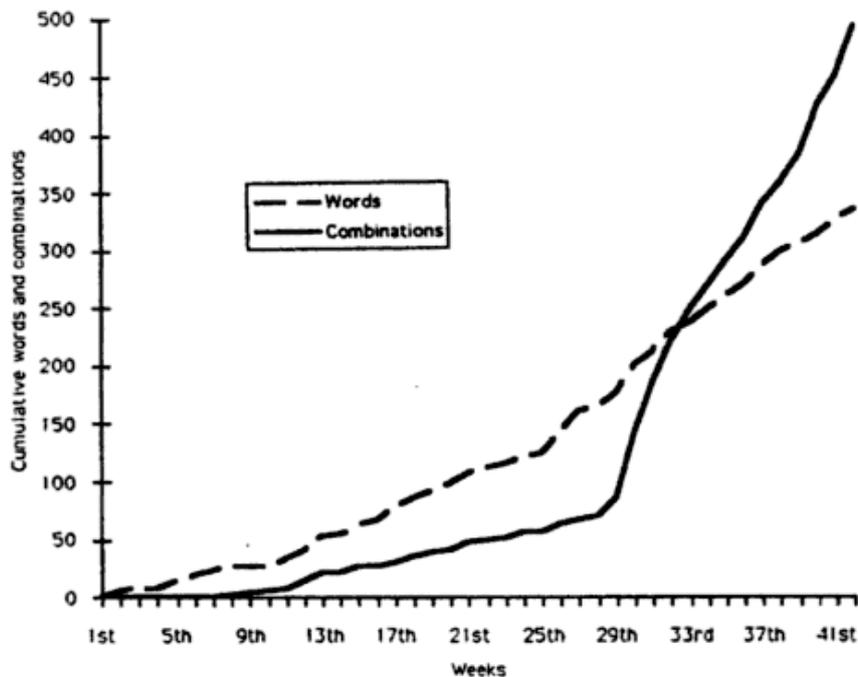


Figura 2.1 Gráfico basado en un experimento de adquisición de palabras y construcción de combinaciones (frases) (O’Grady, 2005).

Al final del estudio de nueve meses, Damon (con casi dos años) producía alrededor de 500 frases diferentes. Nótese que las dos etapas en la adquisición del lenguaje continúan evolucionando. Y lo mismo ocurre con los demás niveles de componentes del lenguaje. ¿Con el lenguaje escrito pasa lo mismo? Ahora miremos un poco sobre el papel de la escritura.

2.1.2. La escritura

Probablemente, los sistemas de escritura hayan evolucionado de forma independiente, en épocas distintas y en varios lugares del mundo: Mesopotamia, China, América Central... Aunque no existan pruebas materiales sobre el momento y el modo en que surgió la escritura, se puede conjeturar que, primitivamente, todas estas comunidades utilizaron dibujos que evolucionaron a caracteres más abstractos (Crystal, 1994; Halliday, 1990).

De acuerdo con Halliday (1990: 14), «[el lenguaje escrito] ha evolucionado de la junción de dos sistemas semióticos independientes: el lenguaje oral de un lado, y el imaginario visual del otro. El lenguaje escrito tuvo inicio cuando los dibujos fueron interpretados como lenguaje, siendo leídos en voz alta. Los símbolos de todos los lenguajes naturales escritos empezaron como dibujos, sea silábico, alfabético o ideográfico». En este sentido, Crystal referencia que los primeros ejemplos de uso convencional de símbolos escritos se encuentran en las tablillas de arcilla descubiertas en varias partes del Oriente Medio y sureste de Europa, datados de 3500 AC.

Ferdinand de Saussure (1916) y otros lingüistas, cuando comenzaron sus investigaciones sobre el lenguaje, consideraron que el lenguaje escrito necesariamente se constituía como un sistema derivado del lenguaje oral y, por esto, inferior a este. Una de las razones era que el lenguaje escrito surgió mucho tiempo después del lenguaje oral. Pero otros lingüistas del Círculo de Praga, a diferencia de Saussure y a partir sobre todo del pensamiento de Artimovyc y Vachek (Luelsdorff, 1989), han argumentado que las normas del lenguaje escrito presentan otras funcionalidades, más allá de la mera transcripción del lenguaje hablado¹⁰. Vachek añade que suele ser un error deducir que la

¹⁰ Fundado en 1926, el Círculo Lingüístico de Praga, compuesto básicamente por lingüistas europeos, ha dedicado mayor atención a las cuestiones estructuralistas de la lengua, principalmente relacionadas con la

estructura del lenguaje escrito es, por su constitución, cualitativamente inferior a la estructura del lenguaje hablada: no es inferior sino, simplemente diferente.

El objetivo perseguido por Vachek es establecer el lenguaje escrito como un legítimo dominio de investigación lingüística. Para él, las dos formas de lenguaje difieren en medios (fonemas *versus* grafemas) y, principalmente, en función. El lenguaje oral provee medios para una reacción rápida e inmediata a la realidad extralingüística, mientras que el lenguaje escrito favorece una respuesta que es fácilmente conservable en el tiempo. Así se expresa Vachek:

Any sound linguistic theory must be based on concrete utterances of speech. It is often overlooked, however, that speech utterances are of two different kinds, i. e. spoken and written utterances. The latter cannot be simply regarded as optical projections of the former. To the difference of material existing between the two is added another difference, more profound and more essential, that is to say, a difference of functions. The function of spoken utterance is to respond to the given stimulus (which, as a rule, is urgent) in a dynamic way, i. e. quickly, readily, and with equal attention to both the intellectual and the emotional factors of the situation that gave rise to the stimulus. On the other hand, the function of the written utterance is to respond to the given stimulus (which, as a rule, is not urgent) in a static way, that is to say, the response should be permanent (i. e. preservable), affording full comprehension as well as clear survey of the situation concerned, and stressing the intellectual factors of the situation (Luelsdorff, 1989: 2).

Así, los resultados producidos por cada lenguaje, ya sea oral, ya sea escrito, serán diferentes, igual que los recursos lingüísticos disponibles para cada uno de ellos. Los recursos no están disponibles para ambos y, evidentemente, la estructura de uno no será inferior a la estructura del otro. En autores más actuales (Crystal, 1994; Halliday,

fonética. Pero estos dos lingüistas se han planteado un análisis de los textos escritos desde la perspectiva del estructuralismo.

1990; Nunberg, 1990) también se defiende que el habla y la escritura son manifestaciones del lenguaje a la vez diferentes e iguales.

De acuerdo con Crystal, son muchas las divergencias entre ambos lenguajes. El más evidente es la materialidad de los medios de expresión: el habla usa materia fónica, normalmente en forma de movimientos de la presión del aire, mientras que la escritura usa materia gráfica, especialmente en forma de marcas sobre una superficie. Las diferencias relacionadas con la función abordadas por Vichok también son citadas por Crystal:

Las diferencias en la estructura y en el uso entre lenguaje hablado y escrito son inevitables, porque son el producto de unos tipos de situación comunicativa radicalmente distintos. El habla está ligada al tiempo, es dinámica, pasajera; forma parte de una interacción en la que, normalmente, los dos participantes están presentes, y el hablante tiene en mente a un oyente (o grupo de oyentes) específico. La escritura está ligada al espacio, es estática, permanente; el emisor está lejos del receptor y, a menudo, puede que ni siquiera sepa quién es el receptor (como en la mayoría de las obras literarias) (Crystal, 1994: 179).

Esta ausencia de interlocutores presentes hizo con que, en el sentido más amplio, el sistema de escritura se desarrollara buscando ser capaz de representar ese contexto. Para esto, debía evitar algunos recursos orales: expresiones deícticas (por ejemplo, *este*, *allí arriba*), vaguedad y ambigüedad, etc. Y uno de los recursos que también se tuvo que modificar fue el de la puntuación. Aunque no todos los recursos utilizados en la puntuación se relacionan con pausas de planificación, sí que es cierto que se “crearon” estos recursos para representar aspectos del lenguaje oral: pausas breves representadas por comas, signos de exclamación o subrayado para indicar un aumento de intensidad, paréntesis representando una disminución del tempo, la intensidad y el tono, etc.

Crystal concluye sus observaciones diciendo que «puede que los sistemas de escritura deriven del habla, en sentido histórico, pero en la sociedad moderna la dependencia es mutua» (Crystal, 1994: 179).

No obstante, sí que es cierto que en ocasiones ambos sistemas de lenguaje intercambian de vez en cuando sus códigos: por ejemplo, una expresión de indignación puede ser caracterizada en la escritura con una sucesión de signos de exclamación al final de la oración; y hoy en día es muy común que una persona, refiriéndose de un modo especial a algo o alguien, modele sus dedos para poner comillas en la palabra recientemente hablada.

Halliday (1990), por su parte, añade otros aspectos del lenguaje hablado que no tienen equivalente en la escritura: el ritmo, la entonación, los grados de intensidad, la variación en la calidad de voz (“tono”), las pausas, así como las características contextuales mediante las cuales reconocemos que hay diferentes interlocutores, aparte de las características individuales de la voz de una persona en particular.

Pese a las diferencias descritas, es un hecho obvio que la escritura es capaz de representar todas las palabras posibles de un idioma: proporcionando expresiones ya conocidas para la mayoría de los sujetos y proporcionando los medios de creación (“codificación”) de expresiones para aquellos elementos que aun no están codificados. Esta capacidad de reconocimiento de elementos codificados y de codificaciones futuras tiende a llevar los sistemas de escritura a la uniformidad, para todas las formulaciones establecidas de la lengua, ya sea en los caracteres, ya sea en la ortografía silábica o alfabética, o en formas intermedias o compuestas. Y para cualquier cosa que no esté codificada, proporcionan los principios generales para que las expresiones se puedan crear.

Entonces la pregunta que en este trabajo se hace es la siguiente: y, con tanto esfuerzo histórico para hacer que el lenguaje escrito sea cada vez más uniforme, ¿por qué aún no es capaz de representar todos los rasgos propios de la oralidad?; ¿esta incapacidad es accidental, por lo difícil que es representarlos, o existe algún otro motivo? La respuesta para Halliday (cfr. *supra*) –y como es habitual en Lingüística– es que se trata un poco de ambos. Para él, otra vez, el motivo se encuentra en las diferentes funciones del habla y de la escritura.

¿Cuál es entonces la naturaleza de estos “ingredientes que faltan” y que la escritura sobrevive sin ellos? Esencialmente son las características locales del lenguaje, las cosas que la atan a este momento y contexto particular de hablar, el aquí y el ahora. El estado de la mente del hablante, las reservas y las dudas que él o ella puedan sentir,

las vacilaciones, el peso dado a las diferentes partes de un argumento... son aspectos que no tienen cabida en la mayoría de los usos del lenguaje escrito. Las condiciones particulares que prevalecen en el momento de escribir no van a estar presentes para el lector de ningún modo (Halliday, 1990).

Pero, como ya se ha dicho, no existe una dicotomía entre el lenguaje hablado y escrito. Hay mezclas entre los dos sistemas de expresión. Así que la omisión de los rasgos prosódicos en la lengua escrita es, en algunos aspectos y en ciertas circunstancias, una deficiencia genuina. Y como también se ha mencionado, existe un dispositivo que se utiliza con el fin de superar parcialmente esta deficiencia: el sistema de puntuación.

Para Nunberg (1990), así como para los otros autores citados, la puntuación, en principio, tenía el papel de representar los rasgos prosódicos del lenguaje oral en la escritura. Pero, con el paso del tiempo, fue adquiriendo otras funciones y otros modos de uso. Para Nunberg, las diferencias se intensificaron tanto que la puntuación pasó a constituirse en un subsistema lingüístico insertado en el sistema de la escritura:

I will be examining the graphical system that is most closely associated with the textual realization of written lexical expressions: the system of punctuation, or more accurately, the system of text-categories that punctuation, along with other graphical devices, is used to indicate. I will argue that punctuation is in fact a linguistic subsystem, and hence to be considered as part of the wider system of the written language, though its systematically is apt to be obscured if we try to analyze it contrastively, by reference to the set of spoken-language devices with which it has some functional overlap (Nunberg, 1990: 6).

Aunque ya ha quedado claro que existen diferencias entre la escritura y el lenguaje, es aún foco de discusión la naturaleza de algunas de esas diferencias. De hecho, límites o fronteras tan sutiles son normales en el ámbito de la Lingüística: qué separa completamente la fonética de la fonología y qué las une; cuál es el límite en la separación de lo lingüístico y lo paralingüístico y gestual en las discusiones sobre entonación (Nunberg, 1990).

Pero antes de pasar a la descripción de las características de la puntuación, quiero señalar tres temas para aclarar el papel real de la escritura. El primero es la falta de correspondencia total entre el lenguaje escrito y la transcripción de habla. Aunque los medios para expresar las dos cosas sean los mismos (marcas gráficas), y aunque utilicen las mismas unidades (palabras) para expresarse, se trata de dos producciones completamente diferentes. El lenguaje escrito posee normas específicas, como, por ejemplo, las que rigen la construcción de oraciones o el modo como las unidades lingüísticas se combinan. Estas normas difieren de las normas del lenguaje oral que, por su parte, poseen una serie de atributos peculiares a su forma de expresión. De este modo, el resultado de una transcripción tan solo es una transcripción. O sea es la expresión del lenguaje oral registrada con otros medios.

El segundo tiene que ver con el hecho de ambos lenguajes, el oral y el escrito, crean diferentes realidades. La escritura, por lo general, crea un mundo de cosas, mientras que el habla, también de forma general, crea un mundo de acontecimientos. Esto se refleja en la diversidad de verbos, más presente en el habla que en la escritura (Halliday, 1990).

El tercer tema va a exigir un poco más de explicación. Se trata de la existencia de otro tipo de lenguaje que difiere del lenguaje oral y del escrito, el denominado *lenguaje interno* por Vachek (1994), y que está asociado a su papel subjetivo en la conciencia humana. Este fenómeno ha recibido también otros nombres, dependiendo del énfasis que pretenden dar sus investigadores: habla implícita (para referirse a una conducta refleja que acompaña a determinadas actividades cognitivas, como la lectura), conducta oral encubierta (cuando lo relevante son sus componentes motrices de muy pequeña magnitud), lenguaje interior (para indicar la naturaleza del pensamiento verbal) o habla subvocal (cuando se consideran los pequeños movimientos en el aparato fonador del individuo) (Pueyo, 1987).

En las Tesis de 1929, el Círculo de Praga, en el apartado que trata de los problemas de las investigaciones sobre las diversas funciones de la lengua, aparece una referencia sobre este tema, que se transcribe a continuación:

Hay que distinguir el *lenguaje interno* y el *lenguaje manifiesto*¹¹. Este último no es, para la mayoría de los hablantes, sino un caso particular, pues estos emplean las formas lingüísticas pensando más a menudo que hablando; es también un error generalizar y sobrestimar la importancia, para la lengua, del aspecto fónico totalmente exterior, y es preciso tener en cuenta los hechos lingüísticos virtuales (Argente, 1980).

El *lenguaje manifiesto* puede ser el lenguaje oral o el lenguaje escrito, como ha señalado Vachek en un trabajo posterior de 1989 (apud Luelsdorff, 1989).

Vygotski, en *Pensamiento y Lenguaje* en el año de 1934 (Vygotsky, 1995) hace una evaluación de las características y de la función del *lenguaje interno*:

El habla interiorizada del adulto representa su *pensamiento para sí* más que la adaptación social, y tiene la misma función que el lenguaje egocéntrico en el niño. Fuera del contexto sería incomprendible porque omite *nombrar* lo que es obvio para el que habla. Estas similitudes nos conducen a suponer que cuando desaparece el habla egocéntrica no es que se atrofia simplemente, sino que permanece oculta, es decir, se convierte en lenguaje interiorizado.

El lenguaje interno, a diferencia del oral y del escrito, no tiene carácter comunicativo. Y por tener una función diferenciada, sus estructuras se distancian de la estructura de las otras dos. Vygotski afirma, incluso, que la sintaxis del lenguaje interno es directamente contradictoria con la sintaxis del lenguaje escrito, situándose la sintaxis del lenguaje oral en medio de ambas. Sería así, según el autor, porque el lenguaje interno sería altamente confuso, limitado y estenográfico, mientras que los enunciados producidos por el lenguaje escrito son simples, bien estructurados y desarrollados.

Otro autor citado por Vachek en el artículo que trata sobre el discurso interno y el lenguaje escrito (Luelsdorff, 1989) es Grodzinski, quien argumenta que el discurso

¹¹ El énfasis de las palabras en cursiva se puede leer en el texto original.

interno es reducido y fragmentario, hasta caótico a menudo, saltando de un objeto a otro.

A estas características del lenguaje escrito se debe añadir un rasgo que en su origen se encargaba, indiscutiblemente, de la reproducción de las características del lenguaje oral: la puntuación. Por su importancia para el presente trabajo, dedicaremos a ella un apartado.

2.1.3. Papel de la puntuación

Los signos de puntuación en el texto escrito en la cultura occidental se han desarrollado durante un largo período de tiempo. La escritura de los griegos antiguos comenzó como una cadena de letras, sin espacios y sin puntuación. La primera línea crecía en dirección izquierda-derecha y, al final de la línea, se seccionaba el texto poniendo las letras necesarias para rellenar todo el espacio de la línea para continuar en la siguiente, pero en dirección derecha-izquierda. Así lo describe (Tuson, 1996: 102):

La realitat és que els grecs, com els fenicis, van practicar tant l'escriptura de dreta a Esquerra, com l'anomenada "bustrofèdica": començaven en una direcció i, un cop acabada la ratlla, no tornaven al principi, sinó que continuaven a sota en direcció contrària.

En un principio, la puntuación tenía la función de indicar los espacios entre las palabras y las pausas que deberían producirse en la lectura en voz alta. Esta función se modificó cuando se implementó la lectura silenciosa. En ese momento, la puntuación se conectó más a la organización textual y pasó a ser tarea del propio escritor (Ferreiro y Pontecorvo, 1999). Los signos de interrogación y de exclamación fueron introducidos más tarde (Halliday, 1990).

Halliday (1990) enumera las innovaciones que se han ido incorporando a la escritura (durante siglos) y que han llevado a las normas actuales del uso de la puntuación. Se describen a continuación, en su orden de aparición:

1. Se estandarizó la dirección de la línea. Esto se hizo de forma diferente en sistemas diferentes: los griegos establecieron la dirección de izquierda a derecha, mientras que los alfabetos semíticos optaron por la dirección inversa.
2. Se introdujeron espacios entre las palabras.
3. Se introdujo el signo de puntuación “punto y aparte” para marcar los finales de de ciertas oraciones.
4. Se distinguieron sistemáticamente letras mayúsculas y minúsculas. Las mayúsculas asumieron funciones especiales como el iniciar frases.
5. Se crearon símbolos especiales para indicar vínculos, interpolaciones y omisiones (rayas, paréntesis y apóstrofes).
6. Se añadieron otros signos de puntuación más detallados: coma, dos puntos, punto y coma, guión.
7. Se añadieron algunos símbolos especiales: comillas (altas y bajas), signo de interrogación y signo de exclamación.

Un análisis de la puntuación en algunos escritos de la Edad Media indica la presencia de hábitos personales en la escritura, mostrando así rasgos personales del autor de la obra. Así lo asume, Parkes (2012) al afirmar que la puntuación es un aspecto únicamente personal. Según el mismo autor, este cambio se originó al producirse un aumento en la cantidad de signos de puntuación en esta época, a lo que debe añadirse el carácter novedoso de su uso y la inexistencia de reglas definidas. Entonces, cualquier autor podría apoyarse libremente en la puntuación para ayudar al lector a identificar más fácilmente la relación entre los elementos de la frase, determinando la función precisa de cada una de las estructuras utilizadas para indicar su sentido.

Según Crystal (1994), la puntuación posee dos funciones: permitir la lectura de textos escritos de un modo coherente e indicar el ritmo y el colorido del habla (aunque nunca se consigue del todo). Aunque su uso es el mismo, pues se utiliza para separar unas unidades gramaticales (como oraciones, sintagmas, palabras) de otras, existe una jerarquía entre los signos. Crystal cita los signos más utilizados, indicando con qué unidades gramaticales se relacionan en español –en relación con este trabajo, es

interesante notar que también se utilizan en inglés y portugués—. Los más importantes son los siguientes (cfr., también Marsá, 1986):

- *Espacio*: separa palabras e identifica párrafos.
- *Punto y aparte, signo de interrogación, signo de exclamación y puntos suspensivos*: identifican el final de una oración, aunque poseen otras funciones más específicas. El punto y aparte separa párrafos o conjuntos de frases con un tema común. Los signos (interrogación y exclamación) indican la presencia de una oración interrogativa (una pregunta) y de la expresión de un sentimiento o sorpresa (exclamación), de forma que existe una modulación prosódica. Los puntos suspensivos interrumpen una frase porque el conocimiento de la misma se da por conocido por parte del lector (por ejemplo, la primera parte de un refrán puede ir seguida por puntos suspensivos).
- *Punto y coma*: identifica las partes de un periodo complejo, o separa partes complejas dentro de una lista, relacionándolas al mismo tiempo.
- *Dos puntos*: indica que lo que sigue es una ampliación o explicación de lo que precede.
- *Coma*: marca la sucesión de unidades gramaticales, o el uso de una unidad dentro de otra. En todos los casos, el uso principal es una pausa breve y suele tener en muchos casos la función de evitar la repetición de conjunciones copulativas. Una pausa algo mayor se indica con el *punto y seguido*.
- *Paréntesis y corchetes; guión*: marcan la inclusión de una unidad gramatical al medio o al final de una oración. Cuando se ofrece una explicación complementaria integrada sintácticamente en el texto se usa una coma o un guión; cuando la explicación no se integra sintácticamente se usa el paréntesis. Además, los paréntesis se usan para incisos informativos, mientras que los corchetes (o paréntesis rectangulares) se emplean para incisos con un contenido informativo ajeno al texto.

- *Guión*: se usa como los paréntesis y corchetes, y también para separar un comentario o apreciación al final de una frase, y para expresar un enunciado incompleto.
- *Comillas sencillas y dobles*: identifican el principio y el final de un fragmento de habla, un título, una cita, o el uso especial de una palabra. Las comillas sencillas se utiliza para identificar un término o noción. Por su parte, las dobles pueden ser altas o bajas (de sargento): las altas suelen marcar la importancia de un contenido, mientras que las bajas se suelen utilizar para incluir una cita dentro del texto.

Por supuesto, en la descripción anterior solo se han indicado los usos principales de algunos componentes. Faltan otros signos (por ejemplo, el signo de guión largo, el signo de párrafo o §, el signo de un asterisco, etc.), mientras que se han obviado algunos usos: el uso de guión corto para marca separación de palabras al final de una línea, el uso del paréntesis para incluir el año de una edición, el valor enfático de los puntos suspensivos, el uso opcional de puntos en la escritura de siglas, etc.

El aprendizaje de la habilidad del uso de la puntuación aparece en los escritos de los niños solo cuando sus escritos se tornan más descriptivos, más elaborados. Los niños pre-alfabetizados saben que en los textos impresos hay marcas «que no son letras sino que van con las letras», dicen los niños cuando les preguntan y cuando todavía no conocen los nombres convencionales de estas marcas (Ferreiro y Pontecorvo, 1999). Cuando empiezan a construir frases sencillas usan los puntos, los signos de exclamación y los signos de interrogación.

La jerarquía comentada anteriormente se relaciona con la mayor o menor fuerza y mayor o menor pausa. Ya se ha explicado que la coma representa el signo con menor fuerza y que indica una pausa leve: sería como un marcador de claridad y como un desacelerador del flujo de la lectura (Nunberg, 1990). No obstante, el uso de estos signos es tremendamente personal, de forma que realizar una comparación entre el uso de signos de puntuación –como se propone en este trabajo– solo es posible si se compara la evolución de una misma persona a lo largo de los años.

Ahora bien, a pesar de las diferentes funciones y a pesar de la existencia de una jerarquía, todos los signos de puntuación pueden ser considerados como pausas

lingüísticas. De esta forma, la forma de puntuar se podría relacionar, al menos en algunos casos, con los procesos de planificación de la escritura, junto con la presencia de otros elementos sintácticos (Pontecorvo et al., 1996).

La falta de estandarización de los signos de puntuación, en estrecha conexión con la organización, estructura, y prosodia particular de un texto escrito (Chafe, 1988), obtiene su máxima representación en el uso actual del lenguaje escrito en sistemas informáticos de redes sociales. El lenguaje escrito tiene, en estos casos, una nueva funcionalidad, donde los sentimientos y estados mentales del escritor son importantes a la hora de construir el mensaje: esto lleva a reinventarse, añadiendo nuevos símbolos como los emoticonos.

Según Halliday (1990), esta falta de estándares está asociada a que existen dos modos de puntuar. Sí el escritor está componiendo el texto en el “modo de escritura”, la puntuación simplemente seguirá la gramática, pues no existe ninguna implicación para la vocalización y, por lo tanto, no hay prosodia, ningún modelo alternativo de organización. Pero si el escritor está componiendo “con su oído”, escribiendo lo que, para él, es esencialmente el discurso hablado, entonces sí existe una implicación de la expresión: el escritor está estructurando el texto prosódicamente a medida que avanza y, así, puede surgir un conflicto. Entonces, una pregunta que debemos hacernos es la siguiente: un escritor, ¿está puntuando gramaticalmente o está puntuando fonológicamente?; ¿cuál será, así, la unidad básica que está trabajando el escritor: la oración o el grupo tonal? Aquí no se entrará en esta discusión que se enlaza con la disputa entre si la unidad de planificación es semántica o sintáctica (para una revisión de este tema, remitimos al lector a Garrett, 1980 y 1992).

En resumen, los escritores son los que deciden si utilizarán uno u otro de estos dos modos, la puntuación gramatical o la puntuación fonológica. No obstante, también pueden usar una mezcla de los dos. Esto lleva a un uso personal cuando un escritor puntúa, lo que impide una valoración cuantitativa, salvo que se trate de estudios longitudinales. Además, lo que es innegable es que todos los usuarios de un idioma supuestamente aprenden las mismas reglas de uso de un lenguaje, y es este el punto que será abordado ahora.

2.1.4. Aprendizaje de la escritura y puntuación

Los conocimientos lingüísticos se van superponiendo. Se adquieren unos pocos sonidos antes de pronunciar las primeras palabras; estas palabras se adquieren conjuntamente con otros sonidos... y esta adquisición se extiende hasta el final de la vida del individuo. Y lo mismo sucede con el aprendizaje del lenguaje escrito: cuando el niño empieza a construir un conocimiento relacionado con la escritura (a los 5 años aproximadamente) aún está aprendiendo algunos conocimientos relacionados con el léxico y con la construcción de frases en el lenguaje hablado; y la puntuación acontece después, a los 7 años aproximadamente (Ferreiro y Pontecorvo, 1999).

Desde el punto de vista cognitivo, se trata de una sucesión de aprendizajes que, a pesar de que ocurran en paralelo, tiene un hito bien definido relativo al período de la vida en que empiezan. Así, por ejemplo, teóricamente, el desarrollo de un niño normal implica que aprenda a producir los sonidos, antes de aprender las letras, que sepa construir un número mínimo de palabras en el lenguaje hablado (conociendo, entonces, el mecanismo de su producción) antes de empezar a escribirlas; que ya haya producido verbalmente las frases simples, con una o dos palabras, antes de escribirlas.

Respecto a la escritura, según Crystal (1994), el aprendizaje de esta conducta exige desarrollar dos capacidades distintas: la motora y la lingüística. La destreza motora consiste en formar letras, elaborar de modo fluido y automático una secuencia con ellas y colocarlas en una página; y se relaciona también con el aprender cómo coger un lápiz, cómo sentarse y cómo disponer la hoja de papel sobre la mesa, entre otras conductas similares. Este aprendizaje suele durar unos tres años o más.

Ahora bien, la destreza lingüística no es similar y lleva a plantear un interrogante: ¿por qué la escritura se convierte en algo tan difícil para el escolar, hasta el punto que, en determinados periodos, se produce un retraso de 6 u 8 años entre su “edad lingüística” oral y escrita? Para Vygotski (1934), un niño de dos años utiliza pocas palabras y una sintaxis simple porque su vocabulario es reducido y no posee todavía el conocimiento de la estructura de oraciones más complejas, pero el escolar posee el vocabulario y las formas gramaticales necesarias para la escritura, puesto que son las mismas que para el lenguaje oral. El desarrollo de la escritura no repite la historia evolutiva del habla. El lenguaje escrito es una función lingüística separada, que difiere del lenguaje oral tanto en estructura como en su forma de funcionamiento.

Cuando aprende a escribir, el niño debe desembarazarse de los aspectos sensoriales del habla y reemplazar las palabras por imágenes de las mismas. La escritura representa un segundo nivel de simbolización y, por ello, naturalmente, debe resultar para el niño más difícil que el habla, así como el álgebra es más difícil que la cualidad abstracta del lenguaje escrito. Esta gradación también se produce por etapas.

La escritura, en los primeros años nace del habla, pero después se aleja de ella. El niño comienza a desarrollar la capacidad lingüística aprendiendo a emplear las estructuras de lenguaje de una manera apropiada. Hay pocos estudios longitudinales relacionados con este aprendizaje, pero Crystal lista unas etapas (citando a Kroll, 1981), que son las siguientes:

1. *Estadio preparatorio*: se desarrollan las destrezas motoras básicas y se adquieren los principios del sistema ortográfico.
2. *Estadio de consolidación*: ocurre en torno del séptimo año, cuando los niños empiezan a emplear el sistema de escritura para expresar lo que ya pueden decir mediante el habla. La escritura refleja de modo estrecho los patrones del lenguaje hablado.
3. *Estadio de diferenciación*: aproximadamente, en el noveno año la escritura comienza a separarse del habla y a desarrollar sus propios patrones y organización. Los niños aún cometen errores, pero estos se reducen gracias, sobre todo, a los ejemplos que adquieren con sus lecturas;
4. *Fase de integración*: es el momento en que la persona que escribe tiene un dominio elevado del lenguaje y puede variar sus elecciones estilísticas a voluntad y desarrollar un estilo personal. Es raro que esto suceda antes de la mitad de la adolescencia y, en cierto sentido, continúa desarrollándose durante toda la vida adulta.

La distinción en el modo de adquisición/aprendizaje del habla y de la escritura también pueden explicar el porqué los enfermos de Alzheimer pierden antes la lengua

escrita que la oral. Se pierde primero lo último que se aprende. De hecho, en esta misma línea, estaría seguramente la idea de la puntuación: es lo último que se aprende.

En el próximo apartado se presentará una breve exposición de los trastornos del lenguaje asociados al envejecimiento, con especial atención a la escritura y la EA.

2.2. Trastornos del lenguaje y el envejecimiento

Las primeras descripciones de una discapacidad en la escritura causada por un daño cerebral se registraron en la segunda mitad del siglo XIX. En esta época, Ogle (1867) sostuvo la idea de la independencia anatómica de la escritura y del habla, basándose en observaciones clínicas que indicaban una doble disociación entre la agrafia y la afasia (apud Rapcsak y Beeson, 2002). Además de esta disociación, propuso una clasificación de tipos de agrafia, distinguiendo entre daños lingüísticos de la ortografía y trastornos motores de la escritura.

Otros estudios realizados en la segunda mitad de ese mismo siglo, también detectaron trastornos del lenguaje entre diversos síntomas de lesiones cerebrales, pero en este caso se asociaban al lenguaje oral, como es el caso de la afasia de Broca (diagnosticada en 1861, con el nombre de “afemia”, tras una lesión en la parte posterior de la circunvolución frontal inferior izquierda, afectando también a algunas áreas adyacentes) y de la afasia de Wernicke (diagnosticada en 1874 tras una lesión en áreas de la mitad posterior de la circunvolución temporal superior, afectando también a una parte adyacente de la circunvolución temporal media) (Whitaker, 1998). Con el tiempo, estas dos afasias recibirían el nombre de sus descriptores, Broca y Wernicke, respectivamente, aunque la topografía y la sintomatología actual ha modificado bastante esos primeros estudios (Diéguez-Vide y Peña-Casanova, 2012).

Con el tiempo, se han observado trastornos del lenguaje en otras enfermedades, tanto neurológicas como psiquiátricas. Una de estas es la enfermedad de Alzheimer (EA), el objeto de estudio de este trabajo. También en esta enfermedad se han descrito alteraciones de la producción escrita. (Peña-Casanova, 1999; Murillo Ruiz, 1999; Murdoch, Chenery, Wilks y Boyle, 1987; Rapcsak, Arthur, Bliklen y Rubens, 1989).

2.2.1. Características clínicas de la EA

El primer diagnóstico de la EA lo realizó Alois Alzheimer, en el año 1906 (Peña-Casanova, 1999), en su paciente, Auguste Deter, una mujer de 51 años que había cambiado su comportamiento en corto tiempo (Gil, 2009). En los registros de Louis Alzheimer constaba una observación sobre el deterioro de la escritura de su paciente: «While writing, she repeated single syllables, omitting others and quickly become confused» (Maxim y Bryan, 1994: 173).

La EA es un tipo de demencia relacionado con la edad. Según la Organización Mundial de la Salud (OMS), en un informe publicado en este año (*Dementia - A public health priority*, 2012), estimase que en el 2010 habían 35,6 millones de personas en el mundo con demencia. Según datos estimados, este número se duplicará en el año 2030 y será más del triple en el año 2050. El número total de nuevos casos de demencia cada año en todo el mundo es de casi 7,7 millones, lo que implica un nuevo caso cada cuatro segundos. Además, es la única enfermedad en la que se produce un aumento del número de enfermos a consecuencia del incremento de la edad en la población de personas mayores.

Según la OMS (2012: 7), «la demencia es un síndrome debido a una enfermedad del cerebro –generalmente de naturaleza crónica o progresiva– en la que hay una alteración de múltiples funciones corticales superiores, incluyendo la memoria, el pensamiento, la orientación, la comprensión, el cálculo, la capacidad de aprendizaje, el lenguaje y el juicio. Por lo común, se acompañan estas alteraciones con otras de la función cognitiva y, a veces, vienen precedidas por el deterioro en el control emocional, en la conducta social o en la motivación». La EA es la más común de las demencias: entre el 50% y el 70% de los casos (Maxim y Bryan, 1994).

Los profesionales de la salud suelen clasificar a los pacientes con EA en uno de los siguientes tres estadios: fase inicial o leve, fase moderada y fase grave. Estas fases se corresponden con la *Escala Global del Deterioro (GDS)* de Reisberg y cols. (1982), de la cual también existe una versión en español (Peña-Casanova, 1999), recogida en el Anexo 1.

De acuerdo con esta Escala, existe un desacuerdo entre los investigadores por la débil y difusa frontera entre la EA y el Deterioro Cognitivo Leve (DCL) (Peña-Casanova, 1999; Petersen et al., 2001; Valls-Pedret, Molinuevo y Rami, 2010). La

detección de esta diferenciación se ha convertido en una importante área de estudio, pues ayudará en la identificación de síntomas en la fase prodrómica de la enfermedad. Actualmente, la caracterización y el diagnóstico del DCL implica la detección y cuantificación neuropsicológica de un deterioro cognitivo (que no es detectable en la entrevista clínica) y se trata de un paciente que no presenta un cuadro de demencia (Molinuevo, 2007). Aunque no es el objetivo de este trabajo, algunas de las evoluciones que se tratarán en los resultados podrían ser consecuencia, no de un paciente con EA, sino de un sujeto con DCL.

La dificultad en la delimitación del DCL provoca que se encuentren, en la literatura especializada, resultados discrepantes. Así, la prevalencia de sujetos con DCL sin demencia oscila entre el 3% y el 56% y estos resultados dependen de la metodología utilizada (uso de marcadores neuropsicológicos, entrevista clínica o junción de ambos) y de la población analizada. Aparte de estas discrepancias los estudios coinciden en que la prevalencia del DCL aumenta con la edad del grupo –en consonancia, en este sentido, con la EA– y es cuatro veces mayor que las demencias cuando comparados grupos por edad (Díaz, 2007).

Igual que en la EA, el DCL también presenta disfunciones en la memoria, pero los síntomas pueden estar asociados a otros dominios cognitivos como el lenguaje, la función ejecutiva, el ámbito visuo-perceptivo, la atención y la praxis; e, incluso puede asociar varios dominios. La idea de englobar observaciones clínicas de estos dominios podría facilitar el pronóstico, ya que «cualquier proceso neurodegenerativo demencial se inicia en un sujeto cognitivamente normal, [y] para alcanzar el cuadro clínico de enfermedad plenamente establecida ha de atravesar una fase intermedia entre el rendimiento normal y la demencia» (Díaz, 2007).

El diagnóstico precoz del deterioro cognitivo leve como también de la demencia maximiza los beneficios potenciales de un pronto acceso a la ayuda de especialistas y terapias a estos pacientes. Obstáculos para el diagnóstico precoz incluyen la presentación tardía de los síntomas más comunes, como el deterioro del lenguaje oral. Los cuidados y tratamientos se han centrado tradicionalmente en las personas *después de* un diagnóstico formal.

2.2.2. Características verbales de la EA

Como se acaba de comentar, el diagnóstico de las demencias se realiza a partir de un conjunto de síntomas cognitivos centrados en ámbitos como el de la memoria, el lenguaje, algunas actividades motoras y las funciones ejecutivas; unido a estas alteraciones también se aprecian dificultades en la convivencia social y familiar con los consecuentes desajustes en las actividades de la vida diaria (Peña-Casanova, 2007; Bayles, 1993; Cherrier, Mendez, Cummings y Benson, 1998). Uno de los síntomas importantes son las alteraciones en el lenguaje y en la comunicación y, como pueden presentarse también en el DCL o en otras demencias, en ocasiones provoca que el pronóstico no sea exacto y diferenciado. Las alteraciones del lenguaje suelen ser precoces y se agudizan con el progreso de la demencia.

Diversos estudios sobre la producción del lenguaje se centraron en los niveles fonológico, sintáctico y semántico, así como discursivo, y los primeros estudios parecían indicar que las habilidades fonológica y sintáctica se mantendrían preservadas, mientras que las habilidades semánticas se encontrarían extremadamente dañadas (Caramelli, Lessa Mansur, & Nitrini, 1998), constituyendo entonces la característica principal de los cambios en el lenguaje en la EA (Peña-Casanova, 2007; Bayles, 1993; Cherrier et al., 1998).

La presencia de anomias (problemas en la selección léxica, es decir, en el acceso a una palabra) y, en consecuencia, la presencia de circunloquios y palabras indefinidas en la fase inicial (como, por ejemplo, *esto, lo, cosa*, etc.) son indicativos de una pérdida de entradas léxicas detalladas; sin embargo, en estudios más actuales se ha observado que con el progreso de la enfermedad no solo aumentan los circunloquios, sino que los errores sintácticos son más frecuentes (Maxim y Bryan, 1994), sugiriendo una reducción en la especificidad de la información disponible. Así, los problemas léxico, sintáctico y semántico serían los componentes más alterados: de hecho, esta alteración es la que ha dado el nombre de *habla vacía* a la producción oral de los enfermos.

La capacidad de producción del discurso se ve reducida con el envejecimiento, y la recuperación de la palabra es el aspecto de la producción del lenguaje que está en el tope de las indicaciones de los ancianos (esto se denomina el fenómeno de la *punta de la lengua*, cuando, normalmente, el sujeto involucrado en esta situación suele decir: “sé que sé algo aunque no puedo recordarlo, aunque lo tengo en la punta de la lengua”).

Este estudio concluye que los procesos semánticos no son afectados con la edad, pero el déficit en la recuperación de las palabras ocurre en los niveles lexical y fonológico (Burke, 2002).

Algunos estudios indican que hay características lingüísticas que sufren alteraciones que se pueden detectar a través de métodos de investigación cuantitativos a partir de la producción textual, tanto para el DCL como para la EA. Estos estudios proponen evaluar la capacidad lingüística de los pacientes, considerando las dos dimensiones de la producción lingüística referenciada por Jakobson y Halle (1973) a partir de la propuesta de Saussure (1916): la selección y la combinación. La selección se refiere a los aspectos léxicos y la combinación se relaciona con la construcción de las frases (Snowdon et al., 1996; Croisile et al., 1996; Garrard y Hodges, 1999; Graham, Patterson, y Hodges, 2000; Garrard et al., 2005; Garrard et al., 2001; van Velzen y Garrard, 2008; Xu, 2008; Garrard, 2009 ; Pakhomov et al., 2010; Crane et al., 2010; Le et al., 2011).

Los métodos hasta ahora propuestos se basan en estudios estadísticos sencillos, principalmente en el recuento de palabras totales o específicas. Pero hay controversias sobre la importancia y precisión del recuento de palabras en el diagnóstico de la EA. Algunos autores (Tomoeda y Bayles, 1993 apud Murillo Ruiz, 1999; Kathrin A. Bayles, 1993; Ehrlich, 1997) defienden la importancia de las alteraciones lingüísticas más allá del recuento de palabras, incluso en el estadio inicial de la EA. Esta crítica se basa en que uno de los síntomas más precoces en relación con el lenguaje es la aparición de circunloquios, lo que implica hablar sobre las características y funciones de un objeto sin citar el nombre del objeto, o sea, expresar por un rodeo de palabras algo que hubiera podido decirse con menos. Este rodeo provoca, como es evidente, un uso mayor de vocabulario. Tomoeda y Bayles (cf. supra) hicieron un análisis de la producción del discurso a través de la descripción de imágenes y constataron que diversos pacientes con EA, en su estadio inicial, produjeron un número mayor de palabras que los pacientes de control, a consecuencia del uso de circunloquios. En otras palabras, para estos autores la contabilidad simple de palabras no es un buen parámetro para diferenciar entre una persona en fase inicial de Alzheimer y otra sin demencia.

Además, los circunloquios pueden ser un síntoma de otros cuadros clínicos. En un estudio sobre demencia semántica, Garrard y Hodges (2000) detectaron la presencia de circunloquios en el discurso de siete pacientes –de 32 casos recopilados de trabajos

anteriores de otros autores y de observaciones propias—. Los pacientes analizados tenían entre 42 y 73 años de edad y presentaban varios tipos de disfunción cognitiva. Los siete pacientes que presentaron los circunloquios como síntoma tenían entre 54 y 64 años de edad, siendo cuatro mujeres y tres hombres. De igual forma, Graham, Patterson y Hodges, (2000) evaluaron 14 pacientes con déficit de memoria semántica, que presentaban un habla fluente, sintácticamente correcta y sin errores fonológicos. No obstante, estos pacientes no tuvieron buena puntuación en las pruebas de denominación por un uso excesivo de circunloquios.

Rapcsak (1989 apud Ruiz, 1999) propuso y Rapsel (1993 apud Ruiz, 1999) reiteró la idea de que las alteraciones de escritura leve, relacionadas con la incapacidad total o parcial de expresar las ideas por escrito –agrafia–, que suelen ser de naturaleza léxico-semánticas, constituyen la principal característica de las alteraciones lingüísticas que se manifiestan al inicio de la EA. Incluso existe una posibilidad, según algunos autores (Ruiz, 1999), de que las agrafias sean evidentes antes de las dificultades orales. A partir de estos estudios, se ha constatado que la escritura se daña antes que el lenguaje oral, de forma que es factible asumir que lo primero que se ve afectado cuando se produce un deterioro cognitivo progresivo sea la puntuación, considerando el modelo de aprendizaje en capas expuesto al principio de este capítulo.

Esta característica secuencial de los principios de aprendizaje de las diversas características del lenguaje hablado y escrito es de fundamental importancia para el presente trabajo, ya que el fundamento del análisis realizado aquí se basa en que el deterioro cognitivo de las funciones del lenguaje, con el envejecimiento, evolucionará en sentido inverso. O sea, que el último aprendizaje adquirido será el primero que se va a deteriorar con el envejecimiento.

2.2.3. Trastornos de la escritura en la EA

Como se acaba de mencionar, para muchos investigadores, es en la escritura donde los daños suelen ser más tempranos (Appell, Kertesz y Fisman, 1982; McKhann et al., 1984 y Chédru y Geschwind 1972 – apud Glosser & Henderson, 2002; Croisile et al., 1996; Rapcsak et al., 1989). Esto es así por el carácter aprendido –y no adquirido– de la escritura: necesita poner en juego un amplio espectro de recursos cognitivos, desde los mecanismos de control motor a la praxis, pasando por la integración visual, espacial

y queirocinestésica (además del hecho de que la base de un sistema lingüístico es simbólica o abstracta).

Algunas investigaciones que evaluaron la descripción de imágenes (Neils, Boller, Gelderman y Cole (1989) apud Croisile et al., 1996) produjeron resultados que indicaban que la habilidad en la escritura correlacionaba significativamente con la severidad de la demencia, de forma que se produciría el mismo patrón que en las habilidades orales. Estas habilidades se evaluaban a partir de los siguientes criterios: evaluación de la organización, el vocabulario y el contenido informativo; gramática; ortografía; y mecánica de la escritura. Comparando grupos control de ancianos y pacientes con EA, estos escribieron párrafos más cortos y cometieron más errores. Ambos grupos mostraron constancia en los errores (palabras secuencialmente repetidas), pero solo el grupo de pacientes repitió contenidos de la historia.

No obstante, sí que existen otras características en la producción escrita de oraciones que se deteriora en pacientes con EA. Así, Kemper et al. en 1993 –apud Caplan y Waters, 2002– analizaron 350 pacientes con EA, solicitando que escribieran una oración sencilla. Los resultados de la producción de los pacientes se computaron para la longitud de las oraciones en número de palabras y frases, número de cláusulas producidas y producción de seis categorías gramaticales (pronombres, verbos principales, verbos secundarios, negaciones, conjunciones, preguntas). Los resultados indicaron que el 89% de la variancia en la clasificación clínica de la severidad de la demencia puede ser explicada por la longitud de las oraciones, y se expresaría por la cantidad de frases y por el contenido proposicional. Esto sugiere que la capacidad de producción de cláusulas en forma escrita se reduce con el incremento de la severidad en la EA.

Relativo al uso de la puntuación, no se ha encontrado ningún estudio sobre las características de su uso en la EA.

No obstante, se pueden presentar síntomas relacionados con problemas en el lenguaje en algunas demencias sin que cualquier deterioro cognitivo significativo esté presente. En estudios que evaluaron las capacidades cognitivas relacionadas con la comprensión de oraciones (Caplan y Waters, 1999) se ha intentado demostrar si existen bases neurales implicadas exclusivamente con el procesamiento sintáctico o si se trata de un deterioro cognitivo general y no exclusivo del lenguaje.

Según Kemper y Kemtes (2002), la memoria de trabajo sería la capacidad que se deteriora con el envejecimiento y se refleja, por lo tanto, en la producción y comprensión de textos. Su postura se basa en que las limitaciones en la memoria de trabajo afectan la capacidad de los adultos para procesar estructuras jerárquicas, como las construcciones sintácticas construidas con cláusulas incrustadas y subordinadas.

Capítulo 3

O objetivo de um escritor não é seguir as normas gramaticais, mas subvertê-las.

(Guimarães Rosa)

3.1. Material

En este apartado se especifican todos los recursos utilizados en esta investigación. Elegimos las novelas de la escritora inglesa Iris Murdoch para hacer una evaluación de los métodos aplicados en otros trabajos de investigación. Además, los escritores de lengua inglesa y portuguesa (junto con algunos de sus libros) se listarán ahora, con una breve descripción de sus vidas y de las características de sus obras. También se presentarán algunas consideraciones sobre el muestreo, situando el periodo de escritura de los libros y las características lingüísticas relevantes para esta investigación.

3.1.1. Autores

Se van a analizar los libros de diez escritores en lengua inglesa (incluyendo a Iris Murdoch) y siete escritores de habla portuguesa con el objetivo de realizar una comparación de los resultados obtenidos. El conjunto de libros elegidos para cada autor trata de representar su producción literaria a lo largo de sus vidas. En la mayoría de los casos se incluyen sus últimos trabajos para determinar la evolución en el curso del tiempo y, especialmente, si hay alguna evidencia de disfunciones en el período final. Solo se excluyen Arthur Conan Doyle y Harold Bindloss, de quienes no se incluyen sus últimas seis y quince novelas, respectivamente, porque no se encontraron en formato digital.

A continuación, se muestra una breve biografía de los autores elegidos, con algunas algunas informaciones relevantes acerca de sus obras y del período de la vida en que fueron escritas. Estas informaciones ayudarán a entender algunos de los resultados encontrados, proporcionando un contexto de la vida de los autores y de las circunstancias de sus muertes. Además, se pretende enfocar las características de las obras analizadas, incluso en lo que se refiere a la crítica literaria. En el



Anexo 2 se lista las informaciones bibliográficas.

3.1.1.1. Lengua Inglesa

Iris Murdoch (IM) (1919-1999). Nacida en Dublín (Irlanda), fue escritora y filósofa inglesa. A partir de 1954 comienza a escribir, produciendo 26 novelas en 40 años. Escribió sus primeras seis novelas sin dejar de enseñar a tiempo completo en tres escuelas. En 1977, a los 58 años, se retira de dos escuelas y dedica más tiempo a la producción de sus novelas. La primera novela de esta nueva fase gana el *Booker Prize*, convirtiéndose en una escritora conocida en todo el mundo. Existen indicios de que su última novela, *Jackson's Dilemma*, la escribió cuando ya sufría la enfermedad de Alzheimer (EA), aunque no presentaba los síntomas característicos de esta enfermedad. Sus dificultades en el momento de la escritura se atribuyeron a lo que suele llamarse *bloqueo del escritor*. Esta sospecha del deterioro cognitivo en esta etapa de su vida se basa en que, en menos de un año después de publicar su último libro, se le diagnosticó clínicamente la enfermedad de Alzheimer. Murió en 1999, a los 79 años y se confirmó su enfermedad post-mortem (Garrard et al., 2005).

Arthur Conan Doyle (ACD) (1859-1930). Sir Arthur Ignatius Conan Doyle nació el 22 de mayo de 1859 en Edimburgo. Fue un escritor escocés, un médico con poco éxito, que dedicó su tiempo libre en la clínica para escribir sus novelas, sobre todo de su personaje más famoso, el detective del mundo de ficción, Sherlock Holmes. Su libro *The Hound of the Baskervilles* fue considerado como una de las mejores aventuras de ficción detectivesca. Murió de un ataque al corazón a los 70 años el 7 de julio de 1930, escribiendo su último libro en el mismo año de su muerte. Nuestra colección incluye los libros escritos hasta 67 años, dejando de lado sus cinco últimos libros.

Charles Dickens (CD) (1812-1870). Charles John Huffam Dickens nació en Portsmouth (Inglaterra) el 7 de febrero de 1812. Fue un novelista inglés reconocido con apenas 20 años. Murió de muerte cerebral, a los 58 años en junio de 1870. La novela *David Copperfield* (1850) era su libro favorito y el más autobiográfico, siendo también una de sus obras más importantes. Este libro fue un hito en la carrera del autor, cuando sus novelas cambiaron a temas más serios y más planeados que las anteriores.

Frances Hodgson Burnett (FHB) (1849-1924). Frances Hodgson Burnett nació en Manchester (Inglaterra) en 1849. Fue una escritora inglesa que emigró a los Estados

Unidos a los 16 años, donde vivió la mayor parte de su vida. Tuvo una vida psicológicamente problemática intercalada con episodios constantes de depresión. *The Secret Garden* fue una de las novelas más populares de la escritora. Murió de paro cardíaco en octubre de 1924 a los 75 años de edad.

Harold Bindloss (HB) (1866-1945). Harold Edward Bindloss nació en Liverpool (Inglaterra) en diciembre de 1866. Viajó por el mundo antes de empezar a escribir, a los 32 años de edad. Ha publicado cerca de 100 novelas desde 1898 hasta 1943. Murió a los 79 años en diciembre de 1945. Sus principales novelas, producidas durante la época edwardiana, son *The Concession Hunters* (1902), *Alton of Somasko* (1905), y *The Gold Trail* (1910). Ha escrito también historias encantadoras sobre el noroeste de Canadá, de las cuales las más famosas son *Prescott of Saskatchewan* (1913) y *Winston of the Prairie* (1907).

Henry Collingwood (HC) (1851-1922). Bajo el seudónimo de William Joseph Cosens Lancaster, Henry Collingwood fue un ingeniero civil que se especializó en el transporte marítimo y en puertos. Nació en 1851 y murió en 1922. Henry Collingwood está entre los primeros autores a explorar el género de la ciencia ficción. No se han encontrado para este trabajo muchas referencias sobre este escritor, por lo que no es posible establecer cuáles son sus obras más importantes y/o populares.

Jacob Abbott (JAI¹²) (1803-1879). Jacob Abbott nació en Hallowell, Maine (EE.UU) en diciembre de 1803. Fue un escritor prolífico de ficción juvenil, cuentos, biografías, libros religiosos y algunas obras de divulgación científica, produciendo un total de unos 100 libros en sus 40 años de producción literaria. Su colección de libros *Rollo Books* son los más populares. Murió en 1879 a los 76 años.

James Fenimore Cooper (JFC) (1789-1851). James Fenimore Cooper nació en New Jersey (EE.UU) en septiembre de 1789. Fue un escritor prolífico y popular en América del Norte a principios del siglo XIX. Sus novelas históricas sobre la vida en la frontera y sobre el indio americano crearon una forma única de la literatura americana. Antes de iniciarse en su carrera como escritor sirvió en la Marina de los EE.UU. como un guardiamarina, lo que influyó mucho en sus novelas. Comenzó a escribir por un desafío de su esposa que decía que él podría escribir una novela mejor de la que estaba

¹² La “i” indica que es el autor de lengua inglesa Jacob Abbott.

leyendo. Su más famosa novela fue *The Last of Mohicans* (1826). Murió de un edema en septiembre de 1851, a los 62 años.

Robert William Chambers (RWC) (1865-1933). Robert William Chambers nació en Nueva York en 1865. Fue destacado como escritor de terror y misterio, pero pasó por diversos géneros literarios, como ficción científica, novelas históricas, relatos de detectives y novelas rosa ambientadas en la alta sociedad de Nueva York. Robert Chambers ha escrito más de 80 libros y acostumbraba a reciclar el material para volver a venderlo, por lo que no queda claro si las obras son completamente nuevas o una reedición o revisión. Fue un éxito de ventas con varios libros considerados *best sellers*, pero sin ningún que se pudiera destacar. Murió en diciembre de 1933, a los sesenta y ocho años, tras ser operado de una dolencia intestinal.

Wilkie Collins (WC) (1824-1889). Wilkie Collins nació en Londres en enero de 1824. Fue un novelista inglés, dramaturgo y autor de relatos cortos. Abogado de profesión, nunca ejerció pues se dedicó por entero a la literatura. Collins fue autor o co-autor de 27 novelas, más de 60 relatos cortos, al menos 14 obras de teatro y un centenar de piezas de no ficción. Tras la publicación de la novela *The Woman in White*, Collins disfrutó de diez años de gran éxito. La novela *Armadale* (1866) fue su primero fracaso comercial, y después de 1868, cuando escribió su última novela exitosa, *The Moonstone*, y tras la muerte de su mentor, Charles Dickens, se hizo adicto al opio. Comenzó a usar esta substancia como una medicina contra los fuertes dolores que tenía por un problema de salud, la gota reumática. Tras esto, su carrera literaria fue un fracaso que duró hasta el final de su carrera. Murió en 1889 de un ataque de parálisis.

3.1.1.2. Lengua portuguesa

Aluísio Azevedo (AA) (1857-1913) – Aluísio Tancredo Belo Gonçalves de Azevedo nació en São Luís do Maranhão (Brasil) en 14 de abril de 1857. Fue un novelista, cronista, diplomático, periodista y caricaturista brasileño. Se trasladó a Río de Janeiro en 1876, volviendo a su ciudad natal después de tres años para mantener a su familia donde, impulsado por las dificultades financieras, inició su actividad literaria, publicando su primera novela, *Lágrimas de uma Mulher*, a los 23 años de edad. Su vida literaria fue breve, pues escribió su última novela a los 38 años, cuando se convierte en diplomático y sirve en varios países (ya no volverá a Brasil). El último país en el que

trabajó como diplomático fue Argentina, donde murió el 21 de enero de 1913, a los 55 años, por razones desconocidas.

Camilo Castelo Branco (CCB) (1825-1890) – Camilo Ferreira Botelho Castelo Branco nació en Lisboa en 16 de marzo de 1825. Fue novelista, crítico, dramaturgo, historiador, poeta y traductor portugués. Prolífico escritor, con un carácter inestable, inquieto e irreverente, escribió en 40 años (1851-1890) más de 260 obras (lo que supone un promedio de más de 6 por año). *Amor de perdición* se considera su mejor obra, la más famosa y la que mejor representa el romanticismo portugués del siglo XIX. Aquejado de una ceguera progresiva que comenzó en 1865 como visión doble, escribió su última novela en 1886, a los 61 años, y desde entonces solo algunos poemas. Sin poder leer y trabajar, se suicida con un disparo el 1 de junio de 1890, a los 65 años de edad.

Eça de Queirós (EQ) (1845-1900) – José Maria Eça de Queirós nació en Póvoa de Varzim (Portugal) el 25 de noviembre de 1845. Fue diplomático, periodista y novelista portugués. Escribió muchos libros de gran éxito, como *O crime do Padre Amaro*, *O Primo Basílio* y *Os Maias*. Su última novela fue *A Ilustre Casa de Ramires*, escrita en 1900, el año de su muerte. Murió en París, cuando aún no tenía 55 años, el 16 de agosto de 1900, por causa desconocida.

Érico Veríssimo (EV) (1905-1975) – Érico Lopes Veríssimo nació en Cruz Alta, Río Grande do Sul (Brasil) en 17 de diciembre de 1905. Fue uno de los escritores más populares de Brasil del siglo XX. Escribió su primer trabajo a los 24 años: una colección de relatos cortos que fue publicada en 1932; su primera novela, *Clarissa*, la escribió a los 28 años. Tuvo una vida literaria de gran éxito, con varias obras como *best sellers*. Murió repentinamente de un ataque al corazón en noviembre de 1975, a los 70 años, en plena producción literaria.

Jorge Amado (JAp¹³) (1912-2001) – Jorge Leal Amado de Faria nació en Itabuna, Bahia (Brasil) el 10 de agosto de 1912. Fue diputado comunista, periodista y escritor, con una obra que ha sido traducida a 49 idiomas. Su primera novela la escribió a los 19 años. De sus novelas de éxito podemos mencionar *Jubiabá*, *Gabriela - Cravo e Canela*, *Dona Flor e seus dois maridos*, y *Tenda dos Milagres*. En 1987, a los 75 años,

¹³ La letra p identifica el autor de lengua portuguesa Jorge Amado, para diferenciar del autor de lengua inglesa Jacob Abbott que tiene las mismas letras iniciales.

comienza a tener dificultades para seguir produciendo debido a la pérdida progresiva de la visión. No obstante, escribió tres novelas más después de esta fecha, *O Sumiço da Santa* (1988), *Navegação de Cabotagem* (1992) y *O Descobrimento da América pelos turcos* (1994), obras que no se analizan en este trabajo. La pérdida de la visión lleva al autor a mostrar síntomas de depresión, más aguda en los últimos tres años de vida. Muere en Salvador de Bahía en el 6 de agosto de 2001, a los 89 años, por complicaciones cardio-respiratorias.

José Saramago (JS) (1922-2010) – José de Souza Saramago nació en Azinhaga (Portugal) el 16 de noviembre de 1922. A los 25 años publicó su primera novela, *Terra do pecado*; no publicó una nueva novela hasta 30 años después. José Saramago se hizo famoso por el uso de un estilo oral, contemporáneo de los cuentos populares de tradición oral, en que la vivacidad de la comunicación es más importante que la corrección del lenguaje escrito. Así, utiliza frases y oraciones largas, mediante el uso de una puntuación de un modo no convencional; los diálogos de los personajes se insertan en los propios párrafos precedentes, de modo que no hay rayas en los libros. Muchas de sus frases ocupan más de una página, con el uso de comas donde la mayoría de los escritores ponen puntos. Murió en junio de 2010 a los 87 años. El libro donde comenzó a utilizar este método de puntuación fue *Levantado do Chão*. Otros éxitos literarios de Saramago han sido *O ano da Morte de Ricardo Reis*, *O Evangelho segundo Jesus Cristo* y *Ensaio sobre a cegueira*.

Machado de Assis (MA) (1839-1908) – Joaquim Maria Machado de Assis nació en Río de Janeiro el 21 de junio de 1839. Fue poeta, novelista, columnista, dramaturgo, periodista y crítico literario. Su obra de mayor éxito es *Memórias Póstumas de Brás Cubas*, donde retrata la esclavitud, las clases sociales, el cientificismo y el positivismo de la época, en un Río de Janeiro pre modernista. Sufrió de una profunda depresión en los últimos cuatro años de su vida, debido a la muerte de su esposa, lo que también le llevó a la muerte en septiembre de 1908, a los 69 años.

3.1.2. Datos

Se han analizado 450 libros, 328 escritos en inglés y 122 en portugués. En su mayoría provienen del *Proyecto Gutenberg*, *Internet Archive Project* y de la biblioteca de dominio público del gobierno de Brasil. El resto de libros fueron escaneados y

procesados con un programa de reconocimiento óptico de caracteres (OCR): el *ABBYY FineReader 9.0 Professional Edition*.

En la Tabla 3.1 se ofrecen las características generales de la muestra de libros en portugués y en inglés. Se incluye la siguiente información: número de libros analizados para cada autor, año probable de la producción del último libro analizado, año de producción de la última obra del autor y año de su muerte. Como ya se ha comentado, en el

Anexo 2 se muestra la lista completa de los libros analizados, incluyendo información sobre el origen de los archivos digitales y la fecha de producción de las obras. Por su parte, la época en que los libros se escribieron se ha extraído de algunos sitios que contienen la bibliografía de los autores, incluyendo Wikipedia. En aquellos casos en que no se ha encontrado ninguna referencia, se consideró la fecha de la primera publicación. Los libros se han elegido para reflejar la carrera lo más larga posible de los autores.

Tabla 3.1 Muestreo de los libros en portugués e inglés.

Autor	N	Año de producción del último libro analizado (edad)	Año del último libro producido	Año de la muerte del autor (edad)
Inglés				
Arthur Conan Doyle	35	1926 (67)	1929 (70)	1929 (70)
Charles Dickens	23	1870 (58)	1870 (58)	1870 (58)
Frances Hodgson Burnett	31	1922 (73)	1922 (73)	1924 (75)
Harold Bindloss	44	1927 (61)	1936 (70)	1945 (79)
Harry Collingwood	32	1922 (71)	1922 (71)	1922 (71)
Iris Murdoch	24	1996 (76)	1996 (76)	1999 (79)
Jacob Abbott	45	1872 (68)	1872 (68)	1879 (76)
James Fenimore Cooper	37	1850 (61)	1850 (61)	1851 (62)
Robert William Chambers	28	1922 (57)	1933 (68)	1933 (68)
Wilkie Collins	29	1889 (65)	1889 (65)	1889 (65)
Portugués				
Aluísio Azevedo	11	1895 (38)	1895 (38)	1913 (55)
Camilo Castelo Branco	43	1886 (61)	1886 (61)	1890 (65)
Eça de Queirós	08	1900 (55)	1900 (55)	1900 (55)
Érico Veríssimo	10	1973 (68)	1973 (68)	1975 (70)
Jorge Amado	21	1984 (72)	1994 (82)	2001 (89)
José Saramago	17	2010 (87)	2010 (87)	2010 (87)
Machado de Assis	12	1908 (69)	1908 (69)	1908 (69)

Como nuestro objetivo es analizar la evolución de las métricas calculadas para cada autor y también para cada idioma, se manipularon dos conjuntos diferentes de datos en los análisis posteriores: uno compuesto por el idioma y otro formado por los datos de cada autor individual.

3.1.3. Consideraciones generales sobre el muestreo

Aparte de diferentes lenguas, la muestra de este trabajo se compone de libros de diversos orígenes, estilos y períodos: en concreto, consta de novelas, historias oficiales e historias cortas, escritas por autores de los siglos XIX y XX. Esto ofrece varios puntos de vista del análisis, ya sea por la comparación de la evolución de los escritos del autor, ya sea por la evolución de la lengua analizada, por los mecanismos de uso de la lengua o, incluso, por las diferencias entre las lenguas.

No obstante, es necesario hacer una advertencia. Para poder hacer comparaciones entre los escritores, en relación con el envejecimiento, hay que garantizar que los libros reflejen el texto original, o sea, sin haber sido sometidos a procesos posteriores de revisión por los editores. Esto no podemos asegurarlo completamente. Tampoco es posible valorar el estilo personal de cada escritor, pero se puede suplir este problema gracias a la comparativa longitudinal que representa toda o parte de su obra. A lo largo de la discusión de los resultados, se considerará este aspecto del muestreo y la importancia para los resultados encontrados.

3.1.4. Los signos de puntuación en inglés y portugués

El inglés es una lengua germánica occidental que se habló por primera vez en la actual Inglaterra, siendo ahora el idioma más utilizado en el mundo (y el segundo más hablado después del chino). Es la primera lengua en la mayoría de habitantes de varios países, entre ellos el Reino Unido, Estados Unidos, Canadá y Australia.

A pesar del amplio préstamo léxico de otros idiomas, la lengua inglesa posee rasgos claramente germánicos, como el uso de verbos modales. Desde aproximadamente el siglo IX, el inglés se escribe utilizando el alfabeto latino, en sustitución del sistema anterior, las runas anglosajonas, y también se adoptaron, junto

con las letras, los signos de puntuación. El alfabeto inglés moderno contiene 26 letras del alfabeto latino. Los signos de puntuación son un conjunto de símbolos con reglas específicas para el uso de cada uno.

Por su parte, el portugués es una lengua románica, oficial en Portugal, Brasil, Mozambique, Angola, Cabo Verde, Guinea-Bissau y Santo Tomé y Príncipe. La lengua portuguesa se ha desarrollado en la Península Ibérica Occidental a partir del latín y se expandió en la península por soldados y colonos romanos. En todos los aspectos (fonología, morfología, léxico y sintaxis) la lengua portuguesa es el resultado de una evolución del latín vulgar con algunas influencias de otras lenguas. Por lo tanto, los signos de puntuación usados para la escritura en portugués son, en la práctica, los mismos que para el idioma inglés.

Los principales signos de puntuación para ambos idiomas (y que han sido considerados en este estudio) están representados en la Tabla 3.2. En el Anexo 3 se presenta las principales reglas de utilización de los signos de puntuación para los idiomas inglés y portugués compilados de varios manuales de uso de la gramática inglesa y portuguesa.

Tabla 3.2 Signos de puntuación principales del portugués e inglés analizados.

Signos de Puntuación	Símbolo
Punto final	.
Coma	,
Dos puntos	:
Punto y coma	;
Signo de exclamación	!
Puntos suspensivos	...
Signo de interrogación	?
Guión	-

Las reglas de puntuación varían con el idioma, localización y tiempo, y están en constante evolución. Así, por ejemplo, los textos en portugués suelen ser más prolijos que los textos en inglés y las oraciones también son más largas. De esta forma, cuando se realizan traducciones del portugués hacia el inglés, lo recomendado es que la idea que se comunica en una oración única debe dividirse lo máximo posible en partes

menores: una oración demasiado larga, aparte de ampliar las posibilidades de error, siempre es más difícil de leer y entender que una serie de frases cortas. Los textos en inglés normalmente contienen más puntos y menos comas en comparación con los textos en portugués (dos Santos, 2007; Schütz, 2009).

Aunque las reglas de utilización de la puntuación son definidas y distintas para ambos idiomas, esto no significa que todos los escritores de lengua inglesa (o de lengua portuguesa) las utilicen de modo igual. Ciertos aspectos de la puntuación son estilísticos y, por lo tanto, son elegidas por el propio autor del texto, transformando este aspecto lingüístico en una característica individualizada. En el presente estudio, se estudiará cómo evoluciona, para cada individuo, el uso de la puntuación, conformando, entonces, un estudio intra-autor; así, por ejemplo, se puede describir cómo la autora Iris Murdoch ha generado sus oraciones, haciendo la medición de la longitud de sus oraciones y realizando una comparación con los resultados de ella misma en edad temprana y en la vejez. En la misma línea, se puede comparar, por ejemplo, cómo José Saramago ha evolucionado en el uso de la puntuación, otra vez mirando los datos desde la edad temprana hasta la vejez.

De este modo, no importa que, por razones del idioma, la autora Iris Murdoch utilice más puntos (y menos comas) que el autor José Saramago. Lo que se analizará es la trayectoria de la vida de cada autor y la forma como ha cambiado el modo en el uso de los signos de puntuación en sus novelas (reducido o incrementado). Una vez obtenidas estas puntuaciones, es posible comparar el comportamiento longitudinal de estos escritores sin que interfieran aspectos como el uso personal de cada autor.

Una comparativa inter-autores es también posible a partir de los datos de Iris Murdoch. Si se observan cambios en la escritura de esta autora, sería posible extrapolar estos resultados a otros autores en lengua inglesa para saber si podría haber existido algún trastorno cognitivo (un Deterioro Cognitivo Leve o una EA). Y lo mismo puede hacerse con los escritores en lengua portuguesa, lo que podría considerarse una métrica de detección precoz de la EA.

Por supuesto, esto es un ejemplo de cómo interpretar los datos, pero no refleja los resultados de este estudio que serán presentados en el capítulo 5.

3.2. Método

Para comparar y analizar los textos publicados por los autores elegidos para este estudio es necesario realizar una serie de tareas, desde limpiar a actualizar la ortografía de las palabras, pasando por el cuidado de los signos de puntuación para que representen las pausas lingüísticas, hacer la lematización de las palabras y calcular las métricas a analizar. A continuación, se describen todos los elementos considerados para la realización de estas tareas.

3.2.1. Tratamiento previo de los textos

Los textos han sido sometidos a una revisión con el fin de prepararlos para el correcto etiquetado de las clases gramaticales. La corrección de los textos se ha intentado igualar en los textos ingleses y portugués, de la forma que se describe a continuación. Las diferencias se muestran en la Tabla 3.3.

Las tareas para corregir los textos fueron las siguientes:

1. Eliminar todas las referencias al *Proyecto Gutenberg* que aparecen al principio y al final del texto, dejando solo el nombre del libro y el nombre del autor al comienzo de la obra, en su caso.
2. Eliminar listas de contenidos (índices).
3. Eliminar datos sobre la edición, publicación, autor y sobre las obras anteriores.
4. Eliminar prólogos, epílogos, notas, notas al pie de página, numeración de páginas, signos de numeración al principio de cada capítulo o de cada parte del libro.
5. Restaurar párrafos al formato original con la eliminación de saltos de líneas si el texto tiene el formato de esta manera.
6. Eliminar caracteres especiales como, por ejemplo, “, ‘, *, _ , «, », (,), [,] tanto al comienzo de párrafos como en la parte central y final.

7. Hacer la corrección de la ortografía para actualizar expresiones escritas en inglés/portugués antiguo o coloquial y que no serían reconocidas por el programa utilizado para etiquetar las palabras (ver ejemplos en la Tabla 3.3).
8. Eliminar todos los signos de puntuación (! ? ; : , . - ‘) duplicados.
9. Eliminar las abreviaturas que terminan con un punto: por ejemplo, EE.UU. o Dr.
10. Poner un punto al final de cada párrafo, incluyendo títulos de los capítulos.
11. Salvar el archivo como texto sin formato señalando la codificación de texto para ser compatible con el programa de etiquetado utilizado más adelante.

Lo importante es que todos los textos han sido sometidos al mismo tratamiento para lograr una comparación más fiable.

Tabla 3.3 Ejemplos de las correcciones para los textos en inglés y portugués.

Textos en Portugués

Eliminar las consonantes repetidas, como: *ll, nn, cc, tt, pp, mm, ff*.

Corregir pares de consonantes que no se usan en la actualidad, como: *th, sc, ph, cç, pt, xh, ct, hi, gn*.

Sustituir algunas letras, actualizando la escritura de ciertas palabras: *z > s, y > i, c > s, e > i*.

Corregir las enclisis y la mesoclisia como, por ejemplo, *vêl-o > vê-lo*.

Textos en Inglés

Sustituir las contracciones de palabras por sus correspondientes palabras como, por ejemplo,: *t'other > the other, Thankee > Thank you*.

Corregir la forma de algunas palabras: *ye > you, nat'ral > natural, calc'lated > calculated, 'em > them*.

Hay que precisar que algunas correcciones se hicieron de forma manual, mientras que otras fueron automatizadas. Los libros obtenidos en el proyecto *Internet Archive* y aquellos escaneados necesitaron una evaluación manual uno por uno, comparándolos con los originales y, en particular, en las digitalizaciones que se hicieron de los libros antiguos debido a problemas de reconocimiento de caracteres.

3.2.2. Lematización

Tras este tratamiento de los libros, se transformaron todas las palabras en minúsculas para evitar que ciertos grupos de palabras fueran tratados como expresiones idiomáticas, pues esto afectaría a las estadísticas de las palabras. Así, por ejemplo, una agrupación como *Bank of England* –con la primera y la tercera palabras escritas en letra mayúscula– sería considerada por el programa como una expresión única, y no se contabilizarían las tres palabras (siendo este nuestro interés).

Preparados los textos, utilizamos el programa *Freeling*¹⁴ para hacer la lematización. Este programa genera una lista de palabras y signos de puntuación donde cada entrada incluye la palabra o signo de puntuación tal como aparece en el texto (1), su lema o signo de puntuación correspondiente (2) y la etiqueta de clase gramatical (*Pos Tag*) o etiqueta de puntuación (*Tag*) a que pertenece la palabra o signo analizados (3). Así, por ejemplo, para la expresión *Were You can?* se obtendría lo siguiente:

were (1) *be* (2) *VBD* (3)
you (1) *you* (2) *PRP* (3)
can (1) *can* (2) *VB* (3)
? (1) *?* (2) *Fit* (3)

Las etiquetas de las clases gramaticales para las palabras en inglés se basan en el *Corpus The Penn Treebank tagset* (Marcus, Santorini, y Marcinkiewicz, 1993). En el Anexo 4 hay una tabla con las principales etiquetas utilizadas en este trabajo, con los siete *tags* genéricos para puntuación.

¹⁴ El programa *Freeling 2.2* se encuentra disponible, de forma gratuita, en la web de la Universidad Politécnica de Cataluña. Es un programa capaz de lematizar textos escritos en nueve idiomas. Para obtener más información, el lector puede visitar la página del proyecto en: <http://nlp.lsi.upc.edu/freeling/>

Para los textos en portugués utilizamos el estándar de etiquetas desarrollado por la Comunidad Europea en 1996, en el *Proyecto EAGLES*¹⁵. El proyecto ha contemplado todas las lenguas europeas, y el *Freeling* utiliza la versión 2.0 de este estándar¹⁶, para los textos en castellano y en portugués.

Para este trabajo, solamente se ha utilizado la segunda información: el lema de la palabra. Trabajar con el lema de las palabras –y no con las palabras como aparecen en el texto– es fundamental si se desea comparar el comportamiento de la escritura en dos idiomas como el inglés y el portugués. Esto es así por las diferencias en las estructuras lingüísticas de ambas lenguas: por ejemplo, la lengua portuguesa posee muchas más flexiones verbales que la lengua inglesa. Tras la lematización, los lemas en portugués corresponden a cerca del 50-65% de las palabras del texto original; en inglés, se corresponde solo con el 70-85%. De esta forma, la comparación es más fiable si se realiza a partir de la lematización.

Se enumeran, a continuación, las tareas del programa *Freeling* que se utilizan para realizar la lematización de las palabras.

Tokenizer Module

El primer módulo de la cadena de procesamiento es el *tokenizer*. Convierte un texto sin formato en un vector de palabras gracias a un conjunto de reglas de *tokenización*, es decir, de expresiones regulares con las cuales se comparan los componentes de cada línea de texto procesado. Cada componente, ya sea una palabra, ya sea un signo de puntuación, se extrae de la línea y el proceso se repite hasta que, o bien la línea se vacía, o bien el texto termina.

Number Detection Module

El módulo de detección de número depende del idioma: se reconoce una expresión numérica (por ejemplo, 1.220,54 o doscientos sesenta y cinco) y se le asigna un valor normalizado como lema. Básicamente, el módulo es un autómata de estados finitos (FSA de su sigla en inglés *Finite States Automata*) que reconoce expresiones numéricas válidas.

¹⁵ Para conocer más sobre el Proyecto EAGLES: <http://www.ilc.cnr.it/EAGLES96/home.html>

¹⁶ Para consultar la tabla con todas las etiquetas del castellano y portugués, se puede acceder a <http://www.lsi.upc.edu/~nlp/tools/parole-sp.html>.

El módulo de detección de puntuación. El constructor recibe contiene la lista de los *PoS tags* (del gramaticales o de desambiguación signos de puntuación. Este módulo anotará los componentes que han

Penn Treebank Tagset

15

~~individual, este módulo no tendrá~~
~~considere esta subcadena como una~~
~~este trabajo ver el Anexo 4~~

POS Tag	Description	Símbolo
Fat	Signo de exclamación	!
Fc	Coma	,
Fd	Dos puntos	:
Fg	Guión	-
Fit	Signo de Interrogación	?
Fp	Punto final	.
Fs	Puntos suspensivos	...
Fx	Punto y coma	;

Anexo 5).

Dates Detection Module

El módulo de detección de fechas, así como el módulo de detección de números, es un autómata de estados finitos que reconoce una colección específica de datos sobre fechas del lenguaje analizadas.

Quantity Recognition Module

La clase de cantidades es un autómata de estados finitos que reconoce relaciones, porcentajes y magnitudes físicas o monetarias (por ejemplo, un veinte por ciento, el 20%, uno de cada cinco, 1/5, a cien millas por hora, etc.).

Este módulo depende del módulo de detección de números. Si los números no están previamente detectados y anotados en la oración, las cantidades no serán reconocidas. Al igual que el reconocimiento de números, es un módulo que depende del idioma: es decir, un autómata de estados finitos tiene que ser programado para que coincida con los patrones de relación de expresiones en el idioma analizado. Así, por ejemplo, el valor 1.220,54, escrito en portugués es diferente del valor 1,220.54 escrito en inglés, pues se invierte el significado del punto y de la coma.

Dictionary Search Module

El módulo de búsqueda del diccionario posee dos funciones:

1. Buscar la forma de una palabra en el diccionario y encontrar su lema y su categoría gramatical.
2. Aplicar las reglas de afijación para encontrar la misma información en los casos en que la forma sea una forma derivada no incluida en el diccionario (por ejemplo, la palabra *quickly* puede no estar en el diccionario, pero la regla de sufijos puede eliminar el sufijo *-ly* y buscar por la forma *quick*).

Form Dictionary File

El diccionario es un texto plano que contiene una forma de la palabra por línea. Cada línea tiene el formato: `form [lemma1 PoS1] [lemma2 PoS2] [...]` (los corchetes se han puesto aquí para aclarar, pero no forman parte del formato).



Así, por ejemplo:

casa [*casa* NCF5000] [*casar* VMIP3S0] [*casar* VMM02S0]

En este ejemplo, la palabra *casa* puede ser etiquetada como un nombre, un verbo en el tiempo presente de indicativo o un verbo en el tiempo imperativo, según el estándar *EAGLES*.

Para el presente estudio no es importante etiquetar las palabras con la clase gramatical exacta, pues esta información no será utilizada en nuestro análisis. Lo que importa es cómo las palabras que son contracciones son tratadas por el programa: por ejemplo, el desmembramiento de una palabra contraída se presentará de esta forma: *do* (en portugués) -> *de+o*. Esto significa que cada vez que se encuentra la forma *do*, se sustituye por dos palabras: *de* y *o*. Las palabras serán divididas y esto provoca que la cantidad de palabras lematizadas sea un poco mayor que la cantidad de palabras del texto original.

Cuando ya se ha obtenido el fichero con la lista de lemas, se procede con los cálculos de las métricas elegidas para caracterizar el deterioro cognitivo de los escritores.

4.1.1. Cálculo de las métricas

Las métricas calculadas servirán para dos propósitos distintos: primero, para reproducir los trabajos realizados con los escritos de la escritora inglesa Iris Murdoch y, segundo, para evaluar la robustez de las nuevas medidas propuestas.

La finalidad de la reproducción de los resultados anteriores es identificar los puntos débiles de estos métodos en relación con el muestreo y también en relación con la interpretación de los resultados y con la influencia de otras características de la escritura de la autora que no fueron considerados en el análisis.

Para esta revisión de los resultados de trabajos anteriores relacionados con la utilización del vocabulario, se calcularán dos parámetros: la relación type-token ratio y la tasa de incremento de nuevas palabras a lo largo del texto. Igualmente, para la revisión de las alteraciones relacionadas con la sintaxis, se calculará la longitud

promedio de las oraciones. Aparte de la longitud promedio de las oraciones, la sintaxis también será evaluada a partir de dos nuevos parámetros: la longitud promedio de las pausas lingüísticas y la cantidad promedio de las pausas lingüísticas en las oraciones. Estos tres parámetros se relacionan con el uso de la puntuación.

En la Tabla 4.1 se muestran las abreviaturas de todas las métricas que serán utilizadas en este trabajo. Estas abreviaturas se relacionan con sus nombres en lengua inglesa.

Tabla 4.1: Código y descripción de las métricas analizadas.

código	descripción	código	descripción
WT	Cantidad de palabras lematizadas totales (<i>Words Total</i>)	WPS	Longitud promedio de las oraciones (medido en palabras lematizadas por oración) (<i>Words Per Sentences</i>)
WCD	Cantidad de palabras lematizadas diferentes (<i>Words Canonical Differents</i>)	TTR	Relación WCD/WT (<i>Type-Token Ratio</i>)
WPP	Longitud promedio de las pausas lingüísticas (medida en palabras lematizadas por pausa lingüística ¹⁷) (<i>Words Per Pauses</i>)	WTIR	Tasa de incremento de nuevas palabras lematizadas diferentes por intervalos de 10.000 palabras lematizadas del texto (<i>Word-Type Introduction Rates</i>)
PPS	Cantidad promedio de pausas lingüísticas por oración (<i>Pauses Per Sentences</i>)	TTR55	Relación WCD/WT en 55.000 palabras lematizadas (<i>Type-Token Ratio in 55,000 words</i>)
TPS	Cantidad de signos de puntuación de oraciones (<i>Total punctuation-marks of Sentence</i>)	TPP	Cantidad total de signos de puntuación (<i>Total punctuation-marks of Pauses</i>)

El *type-token ratio* (TTR) se calcula dividiendo la cantidad de palabras lematizadas diferentes (WCD) por la cantidad de palabras lematizadas totales utilizadas en el texto (WT).

¹⁷ Las pausas lingüísticas son delimitadas por cualquiera de los siete signos de puntuación analizados (ver Anexo 4).

$$TTR = \frac{WCD}{WT}$$

Para una mejor comparación, sin que la longitud diferenciada de los libros interfiera en los resultados, se efectuó un corte en los libros en 55.000 palabras lematizadas. Esto, supuestamente, permitirá que los resultados reflejen una igualdad de condiciones para las comparaciones pues, como ocurre casi siempre, los libros presentan longitud diferentes.

$$TTR55 = \frac{WCD_{en\ 55.000}}{55.000}$$

Le (2010) definió que el WTIR es la tasa de introducción de nuevas palabras lematizadas diferentes (WTIR), i.e. el numero acumulado de palabras lematizadas diferentes computados a cada intervalo de 10.000 palabras lematizadas, WCD (WT). La representación gráfica de esta función para tres libros de la escritora Iris Murdoch se puede observar en la Figura 4.1, donde los valores en los dos ejes están divididos por 1.000.

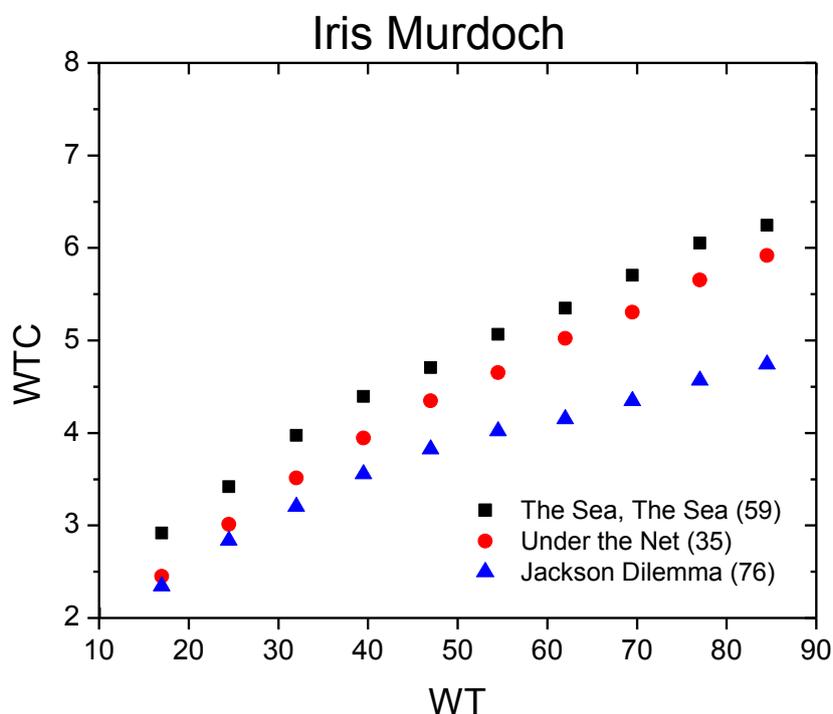


Figura 4.1 Reproducción del WTIR de Garrard et al. (2005).

Por su parte, como se ha visto en la Tabla 4.1, el WPS, la longitud promedio de las oraciones (medido en palabras lematizadas por oración), es el cociente entre WT, la

cantidad total de palabras lematizadas del texto, y el TPS, la cantidad total de signos de puntuación de finales de oraciones: punto (.), signo de exclamación (!), signo de interrogación (?) y puntos suspensivos (...). Los tres primeros signos fueron utilizados por Garrard (2005) y Le (2010 y 2011)¹⁸; los puntos suspensivos se han incluido en este estudio. Por tanto,

$$WPS = \frac{WT}{TPS}$$

Para las métricas WPP y PPS no se ha encontrado ninguna referencia sobre su aplicación en el tipo de investigación relacionado con el deterioro cognitivo. El WPP, longitud promedio de las pausas (medida en palabras lematizadas por pausa lingüística) es el cociente entre WT, el número total de palabras y el TPP, la cantidad total de signos de puntuación utilizados en el texto: punto (.), signo de exclamación (!), signo de interrogación (?), puntos suspensivos (...), coma (,), dos puntos (:), punto y coma (;) y guión (-).

$$WPP = \frac{WT}{TPP}$$

El PPS, la cantidad promedio de pausas lingüísticas por oración, se calcula dividiendo el WPS por el WPP.

$$PPS = \frac{WPS}{WPP}$$

Así, para los tres parámetros se calcula el valor promedio para cada libro. A modo de ejemplo, puede verse en el siguiente texto el cálculo de las métricas presentadas:

O perfume do Cravo, a cor de canela! (nueve palabras – dos signos)¹⁹

¹⁸ De aquí en adelante, dado que el número de apariciones u ocurrencias de los trabajos de Garrard et al. (2005) y Le et al. (2010, 2011) serán muy recurrentes, para agilizar la lectura y no “ensuciar” el texto, se citarán estos trabajos solo por el primer autor (Garrard y Le, respectivamente). Valga esta nota para reivindicar, además, el trabajo del resto de colaboradores de esas investigaciones.

¹⁹ Aunque en la oración aparecen ocho palabras, se cuentan nueve palabras pues la forma de la palabra *do* será transformada en *de+o* en el proceso de lematización.

Eu vim de longe, venho ver Gabriela – a deusa. (nueve palabras – tres signos)

A mulata... (dos palabras – un signo)

Después de la lematización, y de la extracción de los lemas, tenemos una lista de palabras lematizadas y de los signos de puntuación:

o

perfume

de

o

cravo

a

cor

de

canela

!

eu

vir

de

longe

,

vir

ver

gabriela

-

a

Deusa

.

a

mulata

...

El texto posee 20 palabras lematizadas totales ($WT = 20$) y 13 palabras lematizadas diferentes ($WCD = 13$). Las palabras *a*, *o*, *de*, *vir*, que se repiten, se cuentan una única vez. Relativo a los signos de puntuación, el texto tiene tres signos de final de oración ($TPS = 3$) y un total de seis signos de puntuación ($TPP = 6$).

Sobre los signos de puntuación es necesario explicar que para el cómputo del TPS fueron considerados los signos de exclamación (!), el punto (.) y los puntos suspensivos (...). Para el cómputo del TPP, aparte de los signos mencionados anteriormente, también se consideraron los otros signos en las partes intermedias de las oraciones, a saber, la coma y el guión. Junto a cada oración, entre paréntesis, se ha indicado la cantidad de palabras y de signos de puntuación utilizados. Así, tendremos los siguientes valores para las métricas:

$$TTR = \frac{13}{20} = 0,65$$

$$WPS = \frac{20}{3} = 6,67$$

$$WPP = \frac{20}{6} = 3,33$$

$$PPS = \frac{6,67}{3,33} = 2$$

Decimos que en el texto analizado, hay 3 oraciones con 6,67 palabras de promedio. Que, en las partes intermedias, las oraciones de este texto poseen 2 pausas lingüísticas de promedio y que estas pausas presentan una longitud promedio de 3,33 palabras. Como las métricas son valores promedios, son independientes de la longitud del texto y pueden ser comparados con muestras de longitudes diferentes.

• • •

4.1.2. Representación gráfica de las métricas

Además de los nuevos parámetros (WPP, PPS) en relación con el método de análisis, aquí se propone una forma diferente de analizar los datos. Como la intención de la investigación es la detección de puntos de alteración en la escritura de diversos autores –asociados con el envejecimiento–, la propuesta consiste en analizar el avance de estos parámetros con la edad. Y, para esto, se utilizará el método de la *trayectoria en espacios de fases o estados* (SPT).

El *espacio de fases* es un concepto introducido por primera vez en los estudios de física y matemática por Willard Gibbs en 1901 (Gibbs, 2010). Se trata de un espacio donde todos los posibles estados de un sistema son representados y se corresponden con un único punto en el espacio de fases. En el espacio de fase, cada grado de libertad o parámetro del sistema se representa como un eje del espacio multidimensional. La sucesión de puntos, normalmente, representa los estados del sistema a lo largo del tiempo. Al final, el diagrama de fases representa todo el que el sistema puede ser y su trazado (la trayectoria) puede elucidar calidades del sistema que no son detectables de otro modo.

En este trabajo, se relacionarán las métricas a analizar, proyectando la edad en relación con la vida del autor (las edades serán representadas por números escritos cerca de cada punto en los gráficos que corresponden a los libros). Así, por ejemplo, es posible mirar y detectar si ocurren y cuándo ocurren los cambios en el comportamiento lingüístico en relación con el uso de la puntuación.

Este método posee la ventaja de ampliar el potencial de caracterización de los textos, ya que amplía los grados de libertad del sistema²⁰. De este modo, consideramos las métricas WPP, WPS, TTR, PPS etc., como las variables de estado de los autores, representadas por sus libros, y su evolución en el tiempo como una trayectoria en el espacio de fase, donde cada índice define un eje en el sistema ortogonal (SPT) – ver ejemplo en la Figura 4.2.

²⁰ Cuando utilizamos dos variables representados en los ejes “x” e “y”, y además proyectamos la edad en el gráfico, trabajamos con un grado de libertad igual a 2.

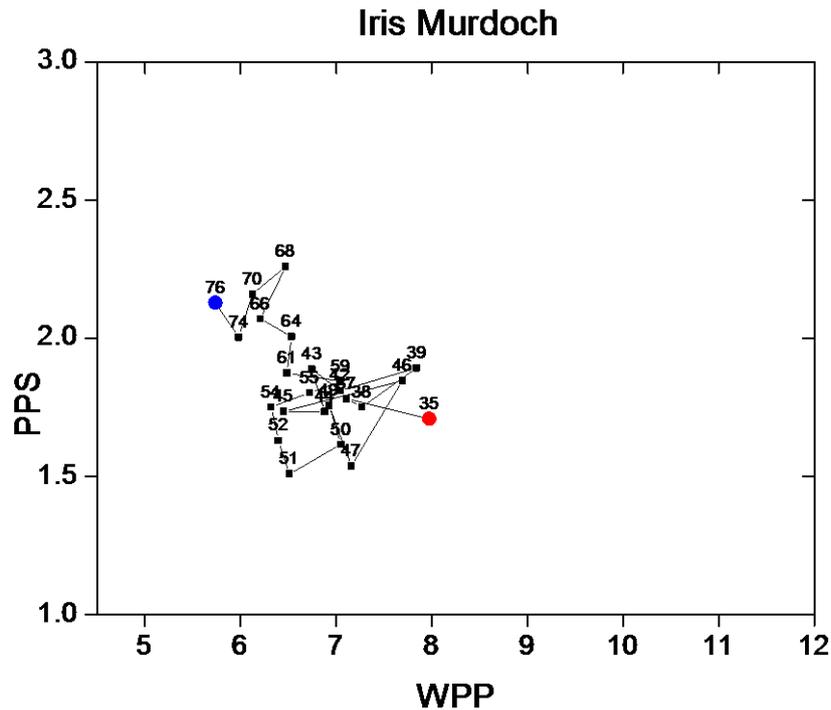


Figura 4.2 Ejemplo de un gráfico de trayectoria de estados.

De esta forma, es posible detectar si existe una tendencia a aumentar o disminuir el valor de las métricas analizadas y, lo más importante, detectar si hubo alguna alteración súbita asociada con el envejecimiento, ya que la edad es la tercera variable que se analizará.

Para hacer la comparación de algunas de las métricas se usará también los métodos de Regresión Lineal simple y múltiple (con uso de la ANOVA) y la correlación de Pearson.

4.1.3. Implementación

Para realizar el tratamiento de los textos y el cálculo de las métricas se ha construido un programa, el *palaBras*, que es el responsable de trabajar el tratamiento previo de los textos, la lematización y el cálculo de las métricas.

4.1.3.1. Tratamiento previo

Se ha utilizado un *script* de búsqueda y sustitución para reemplazar cadenas de palabras en varios archivos. En el Anexo 5 se muestra este *script*. El programa

•••

necesitaba un fichero con las reglas –o instrucciones simples– que se deben aplicar para que las sustituciones se lleven a cabo. Así, por ejemplo, si se quieren eliminar las repeticiones de signos de puntuación en que aparezca este signo seguido por otro signo cualquiera, las instrucciones serán:

```
\? \? ?  
\? ! ?  
\? ; ?  
\? : ?  
\? , ?  
\? \. \s ?  
\? \. \. \s ?  
\? \. \. \. \s ?  
\? - ?
```

En cada línea habrá una única instrucción escrita, utilizándose expresiones regulares (*regEx*). En este caso, si tenemos el signo de interrogación seguido de otros signos, estos serán eliminados y solo quedará el signo de interrogación.

Aparte de las sustituciones relacionadas en el Anexo 5 las sustituciones de la Tabla 3.3 también fueron implementadas en esta etapa.

4.1.3.2. Lematización

La operación del programa *Freeling* se implementa en el ambiente Linux. Para facilitar su ejecución por otro programa, se proporciona un *script* con el nombre de *analyze*. Este es un *script* capaz de localizar los archivos patrones de configuración y definir rutas de búsqueda de bibliotecas. A partir del *palaBras*, se hizo una llamada al programa *Freeling*, utilizando la línea de comando:

```
analyze -f <ficheroConfiguracion.cfg> --outf tagged <ficheroOrigen> <ficheroDestino>
```

Las opciones utilizadas fueron las siguientes:

- f para designar el nombre del fichero de configuración.
- outf para designar el formato del fichero de salida (*tagged*) y los nombres del fichero de entrada y de salida.

El fichero de Configuración fue diferente para cada idioma analizado. El Anexo 6 presenta un modelo de archivo de configuración utilizado por el *Freeling* para la lengua portuguesa. A continuación, se muestran las principales configuraciones a activar:

```
Lang=pt
InputFormat=plain
OutputFormat=tagged
AffixAnalysis=yes
MultiwordsDetection=no
NumbersDetection=yes
PunctuationDetection=yes
DatesDetection=yes
QuantitiesDetection=yes
DictionarySearch=yes
```

Esta funcionalidad del programa *palaBras* tiene como salida un fichero para cada libro analizado, que llamamos archivo TAG, y que contiene una lista de todas las palabras lematizadas.

4.1.3.3. Cálculo de las métricas

WTIR

El programa *palaBras* genera un fichero con las tasas de crecimiento de los textos a intervalos predefinidos (500 palabras). Calcula la cantidad de lemas diferentes y de palabras totales en cada etapa de 500 palabras, hasta el final de la novela. Estas tablas de datos sirven para construir los gráficos de WTIR.

TTR55, TTR, WPS, WPP e PPS

El programa *palaBras* permite elegir una lista de archivos para analizar y calcular las métricas, generando un fichero resumen que contiene el nombre del libro, la edad de producción del libro y los valores calculados para el texto total y para la etapa de 55.000 palabras.

Capítulo 4

*Realmente lo que caracteriza a la conducta científica
es un cierto escepticismo incluso en relación
a nuestras teorías más estimadas.
(Imre Lakatos)*

5.1. Trabajos anteriores

En este apartado presentaremos una revisión bibliográfica de los resultados de algunas investigaciones que han tenido como objeto de estudio el uso del lenguaje oral o escrito y su deterioro con el envejecimiento. Los estudios seleccionados aportan informaciones relevantes con la hipótesis presentada en este trabajo. Se trata de estudios longitudinales sobre el efecto del envejecimiento en la escritura y utilizan técnicas desde la descripción escrita de imágenes (que analiza pequeñas cantidades de frases producidas) a la producción de textos de mayor longitud (i.e., libros). Los resultados que se circunscriben a este estudio están relacionados con la cantidad de vocabulario utilizado y con todas las medidas que hacen referencia a la construcción de oraciones.

Kintsch y Glass (1974:57) –apud Bayles, Kaszniak y Tomoeda (1987)– han demostrado que es el número de proposiciones²¹, en lugar del número de palabras, lo que afecta al recuerdo del significado de una frase. En su investigación, comprobaron

²¹ Bayles y Kaszniak (1987) definen una *proposition* de la forma siguiente: “A proposition can be define as a relational expression that is grammatically analogous to a clause and comprises a relational term, such as a verb, and one or more nouns or noun phrases that function as subjects and objects of the relation” (p.56).

que el tiempo necesario para aprender el significado de un texto es dependiente del número de proposiciones y no de la cantidad de las palabras. Otro estudio de Kintsch y Keenan (1973) demostró que el tiempo necesario para la lectura de las oraciones estaba correlacionado, positivamente, con la cantidad de proposiciones presentadas y la tasa de recuerdo, corroborando los resultados de Kintsch y Glass.

Susan Kemper (1987a) –apud Kempler y Zelinski (1994) – ha apuntado unos cambios sutiles en la producción de construcciones sintácticas complejas. Se trata de un estudio longitudinal en que fueron analizados diarios, mantenidos durante siete o más décadas, de ocho personas adultas nacidas entre 1856 y 1876. Fueron analizadas las estructuras sintácticas usadas en estos diarios a lo largo de la vida de sus autores. Los resultados encontrados indican que, a los 70 años, las personas tienden a utilizar oraciones menos complejas, con un promedio de 1,4 cláusulas²², mientras que a los 20 años estas oraciones presentan un promedio de 3 cláusulas; sin embargo, la longitud promedio de las oraciones no cambia, siendo de 9,2 a los 70 años y de 9,8 a los 20 años. Estos pacientes fueron descritos por el autor del trabajo como “normales” (sin EA), pero haría falta constatar si desarrollaron o no la EA después de esa edad para sacar conclusiones más firmes sobre las causas de la reducción de la complejidad de las oraciones (medidas en número de cláusulas) y la conservación de las características sintácticas (como la longitud promedio de las oraciones) con el envejecimiento.

Steven Z. Rapcsak, Susan A. Arthur, David A. Bliklen y Alan B. Rubens (1989) constataron que, aunque los pacientes con la EA escribían correctamente palabras regulares y palabras sin significado (*pseudopalabras*), así como los controles, rindieron

²² Para el presente estudio usaremos la definición de cláusula y oración/frase de Dieguéz-Vide y Peña-Casanova (2012), a saber:

Cláusula: estructura sintáctica formada por la unión de un sujeto y un predicado que se encuentra inserida en otras estructuras (oraciones o frases).

Oración: estructura sintáctica formada por la unión de un sujeto y un predicado que no está inserida en otra estructura.

Frase: estructura mayor de descripción sintáctica.

No obstante, y como apuntan también los autores, los conceptos de “oración” y de “frase” son, en la práctica, sinónimos, por lo que aquí también se equiparán.

significativamente peor cuando escribían palabras irregulares²³. Además, la mayoría de los errores ortográficos en las palabras irregulares eran regularizaciones con un parecido fonológico (por ejemplo, *oner* por *honor*). Estos resultados indican una deficiencia selectiva de la ortografía léxica.

Rapcsak et al. (1989) han propuesto, y Platel (1993) –apud Ruiz, 1999– reiteró, la idea de que las alteraciones en la escritura, relacionadas con la incapacidad total o parcial de expresar las ideas por escrito –agrafia–, que suelen ser de naturaleza léxico-semánticas (al menos en la escritura espontánea), constituyen la principal característica de las alteraciones lingüísticas que se manifiestan en el principio de la EA. Incluso existe una gran posibilidad de que las agrafias sean evidentes antes de las dificultades orales (Ruiz, 1999). De acuerdo con estas investigaciones, que demuestran que la escritura se daña antes que el lenguaje oral, es factible asumir que lo primero que se afecta cuando se produce un deterioro cognitivo, como en el caso de la afasia, sea la puntuación escrita, considerando el modelo de aprendizaje en capas descrito en el capítulo 2.

Kelly Lyons, Susan Kempler, Emily LaBarge, F. Richard Ferraro, David Balota y Martha Storandt (1994) indican que los diversos trabajos presentados hasta la fecha sugieren que existe una disminución progresiva en la complejidad sintáctica en relación con la severidad de la demencia de Alzheimer, así como una pérdida de contenido informacional. También queda claro a partir de estos estudios que los adultos con demencia leve o moderada eran capaces de producir oraciones bien formadas gramaticalmente, evidenciando así una preservación de la habilidad sintáctica en estos pacientes. No obstante, también observaron una disminución en la longitud de las oraciones y en el uso de estructuras sintácticas complejas.

Snowdon (1996) intentó determinar si la capacidad lingüística en la vida temprana se asocia con la función cognitiva y con la EA en la vejez. Aunque no analizó la producción textual en el periodo de envejecimiento, estudió la relación entre las características de las producciones textuales en edad temprana de un conjunto de monjas (mediante la medida de complejidad lingüística y la densidad de ideas) con los síntomas

²³ Las palabras irregulares son léxicas (no es posible una conversión de fonemas en grafemas) y las regulares pueden ser léxicas o subléxicas. Las pseudopalabras solo pueden escribirse por medio de un proceso de conversión de fonemas-a-grafemas. Estas definiciones provienen de la escritura al dictado.

de deterioro cognitivo que esas monjas presentaron al final de sus vidas. El autor detectó que la baja densidad de ideas y la baja complejidad gramatical en las autobiografías escritas en edad temprana se asociaban a las bajas puntuaciones de los tests cognitivos administrados en la vejez.

La descripción de imágenes fue el método empleado por Bernard Croisile, Bernadette Ska, Marie-Josée Brabant, Annick Duchene, Yves Lepage, Gilbert Aimard y Andmarc Trillet (1996), y por Tomoeda y Bayles (1993). El análisis de deterioro relativo a la vejez en otros idiomas lo realizaron Xu (2008) para el idioma chino y Paul K. Crane, Jonathan C. Gruhl, Elena A. Erosheva, Laura E. Gibbons, Susan M. McCurry, Kristoffer Rhoads, Viet Nguyen, Keerthi Arani, Kamal Masaki y Lon White (2010) para el japonés.

Tomoeda y Bayles (1993 apud Murillo Ruiz, 1999) hicieron un análisis de la producción del discurso a través de la descripción de imágenes y constataron que diversos pacientes con la EA en su estadio inicial produjeron un número mayor de palabras que los pacientes control. Este mayor número de palabras apareció a consecuencia del uso de paráfrasis o circunloquios, es decir, a causa del uso de expresiones para denominar una palabra que no es accesible para el enfermo (anomia). Así, los pacientes deben expresarse con otras palabras para poder denominar un objeto o una acción. O sea, para esas autoras el sencillo conteo de palabras totales de un discurso, buscando la constatación de que el vocabulario disminuye, no es un buen parámetro para diferenciar entre una persona en fase inicial de Alzheimer y otra sin demencia. Para los enfermos que se encuentran en estadio más avanzados, este puede ser un parámetro válido para indicar su existencia, pero no sirve como parámetro para el diagnóstico precoz.

En otros estudios de Susan Kemper (Suzanne Norman, Susan Kemper, Donna Kynette, Hintat Cheung y Cheryl Anagnopoulos, 1991; Kemper, Herman y Lian, 2003 – apud Altmann y Kemper (2006)–) se indica que las personas mayores usan una sintaxis menos compleja en el habla espontánea cuando se comparan con personas más jóvenes. Kemper, Herman y Lian (2003) analizaron la producción de frases de 30 adultos jóvenes (18-28 años) y 30 adultos mayores (70-80 años) usando como *priming*²⁴

²⁴ El *priming* consiste en la presentación de un estímulo que facilite –no actúe o impida– la respuesta de un segundo estímulo, que es el realmente requerido: el estímulo facilitador puede mantener una relación semántica, fonológica, sintáctica... con el estímulo objetivo (*target*), y la presentación puede

conjuntos de 2, 3 y 4 palabras mostradas en una pantalla de ordenador. Para las oraciones con el *prime* de 4 palabras, los adultos mayores construyeron oraciones más fragmentadas.

Kemper, Thompson y Marquis (2001) han puesto de manifiesto cómo tanto la complejidad gramatical como el contenido proposicional están relacionados con los cambios al final de la vida en la cognición de adultos sanos, así como en personas con demencia. La EA acelera esta disminución, independientemente de la edad. Los autores sugieren que estos cambios en las habilidades lingüísticas de los adultos mayores se producen por limitaciones en la memoria de trabajo en la producción de construcciones sintácticas complejas.

Kemper et al. (2001) analizaron autobiografías escritas por monjas cuando ingresaron en el monasterio. Se trataba del mismo conjunto de escritos de monjas que el corpus de Snowdon y colaboradores (cfr. *supra*), La diferencia es que añadieron a la muestra nuevas versiones de estas autobiografías escritas a posteriori. Los investigadores solicitaron a las monjas que hicieran nuevas autobiografías en 1957-1959, en 1987-1989 y en 1995-1996. No se describió en el estudio de qué modo debían realizar la tarea. Solamente aquellas que presentaban lucidez suficiente para hacerla pudieron producir los escritos. El resultado de la investigación constató que aquellas monjas que presentaron síntomas de algún tipo de demencia, también presentaron valores menores en las medidas de complejidad gramatical y densidad de ideas, indicando un deterioro cognitivo relacionado con el cuadro de demencia. Las monjas que no presentaron cuadro característico de demencia no presentaron deterioro cognitivo relacionado.

5.2. Iris Murdoch

Las investigaciones mencionadas hasta ahora manipulaban pequeñas cantidades de textos producidas en descripciones de imágenes o métodos experimentales desarrollados en laboratorio, excepto el análisis de los escritos de las monjas. La

ser unimodal (ambos estímulos presentados en la misma modalidad) o transmodal (se presenta el *prime*, por ejemplo, en modalidad escrita y el estímulo objetivo en modalidad oral) (Dieguéz-Vide y Peña-Casanova, 2012).

mayoría de experimentos se administraron a grupos de personas “sospechosas” de tener un grado leve o moderado de demencia. Con el diagnóstico *post-mortem* de la EA para la escritora inglesa Iris Murdoch (1922-1999), surgió la oportunidad de investigar la evolución de las características del uso del lenguaje escrito realizando un estudio longitudinal con una cantidad de producción escrita hasta el momento no encontrado. Y aún más porque los textos fueron escritos con una cierta regularidad: casi un libro por cada dos años, de media, cubriendo todo el periodo de su vida adulta. La célebre novelista y filósofa inglesa fue diagnosticada con la enfermedad de Alzheimer en 1997, tras el deterioro de sus capacidades cognitivas, particularmente marcados en el dominio de la lengua. Un examen post-mortem de su cerebro confirmó la presencia de placas amiloides y de ovillos neurofibrilares, características patológicas dominantes en el diagnóstico de la enfermedad.

Son diversos los investigadores (Peter Garrard, Lisa M. Maloney, John R. Hodges y Karalyn Patterson (2005); Serguei Pakhomov, Dustin Chacon, Mark Wicklund y Jeanette Gundel (2010); Xuan Le, Ian Lancashire, Graeme Hirst y Regina Jokel (2011) que han trabajado con los libros de Iris Murdoch al ser la única escritora con un diagnóstico de EA confirmado. Estos trabajos tenían la finalidad de definir marcadores cognitivos de tipo lingüístico con el objetivo de identificar alguna clase de deterioro en las características de su escritura con el paso del tiempo.

A continuación analizaremos los tres estudios citados en el párrafo anterior, que son los únicos que han investigado la producción literaria de la autora con respecto a los efectos del Alzheimer sobre su obra escrita.

5.2.1. Aspectos Lingüísticos

En el trabajo de 2005, Garrard²⁵ encontró diferencias en las características léxicas en tres libros de la autora: el primero y el último libro de su carrera (*Under the Net* y *Jackson's Dilemma*), así como su libro de mayor éxito (*The Sea, The Sea*). Los datos indicaron que el último libro, *Jackson's Dilemma*, presentaba unos resultados

²⁵ Nuevamente reitero que los autores citados tuvieron la colaboración de otros colegas en la escritura de los artículos pero para que sea más fácil de leer vamos utilizar solamente el nombre del primeiro autor.

desfavorables, respecto a la variedad y frecuencia del vocabulario en comparación con los otros dos libros.

Este libro, *Jackson's Dilemma*, fue valorado de forma desfavorable por la crítica literaria. Estos resultados negativos fueron inmediatamente asociados a la mala calidad de la escritura detectada por Garrard que, a partir de entonces, trabajó con la hipótesis de que la autora lo escribió cuando ya sufría la enfermedad, pero sin presentar todavía síntomas. En el apartado 5.2.3 se mostrarán los datos y se discutirán los resultados encontrados.

Las pruebas neuropsicológicas repetidas dos años después del diagnóstico ofrecieron resultados más graves: empobrecimiento del vocabulario, deterioro semántico, anomia dependiente de la frecuencia y disgrafía²⁶ superficial.

Pakhomov (2010) analizó diversos parámetros referentes a la complejidad sintáctica. En relación con las observaciones realizadas sobre la medida de la longitud de oraciones y del número de frases por oraciones –parámetros que se relacionan con los objetivos de este trabajo–, se observó un patrón claro de disminución de estas dos medidas: los promedios de longitud de las oraciones y el número de frases por oración fueron menores. Ambas medidas presentan valores altos en el primer libro (*Under the Net*) y bajas en el último libro (*Jackson's Dilemma*). Además de los tres libros analizados por Garrard (2005), Pakhomov añadió el libro *The Green Knight* publicado en 1994, inmediatamente antes de su último libro.

Esta selección respondió al deseo de poner a prueba la sensibilidad de las herramientas del análisis automático sintáctico y léxico en las muestras de escritura. Igual que en el trabajo de Garrard, trabajaron con pequeñas partes del libro y no con la totalidad de los textos. Extrajeron 20 pasajes no contiguos aleatoriamente de cada uno de los cuatro libros, evitando la selección de diálogos (pues constituye un tipo diferente de discurso y no pertenecían al alcance de su estudio). Así, la mayoría de los pasajes resultantes consistieron en las descripciones de escenas y pensamientos atribuibles a los diversos personajes de los libros.

En un estudio posterior, Le (2011) utilizó 20 novelas de Iris Murdoch en su análisis. La riqueza de vocabulario se determinó al relacionarse el TTR²⁷ calculado para

²⁶ La grafía de esta palabra (sin acento en la “i”) sigue la indicación del libro de Dieguéz-vidé & Peña-Casanova, 2012.

la primera porción de las 55.000 palabras con la edad al escribir cada libro. Este procedimiento se denomina *normalización* y tiene la finalidad de eliminar la influencia de muestras de longitudes diferentes en la métrica medida. Para determinar lo que ocurre con el paso del tiempo los autores utilizaron el método de Regresión lineal y determinaron que no existen correlaciones estadísticamente significativas entre el parámetro TTR calculado a 55.000 palabras y con la edad para toda la muestra; sin embargo, dividieron la muestra en dos grupos formados por las primeras 15 novelas y las 5 últimas novelas, respectivamente, identificando un crecimiento en el primer grupo y un decrecimiento en el segundo. Según Xuan Le (2011), cuando se hace el cálculo del TTR en la primera porción de 55.000 palabras, esta métrica pasa a representar, en la comparación entre varias muestras, el vocabulario (número de palabras diferentes) utilizado por el autor en el libro.

En el apartado 4.1.1 se demuestra el cálculo del TTR, con una explicación más detallada sobre esta métrica, pero vale la pena observar un ejemplo para recordar los conceptos discutidos anteriormente y para aclarar el concepto de TTR55. Tomemos las oraciones 1, 2 y 3:

La caja de madera es roja con tiras de maderas envejecidas (oración 1).

La caja de madera es roja con tiras de maderas blancas // y telas blancas y rojas plegadas en la tapa (oración 2).

El bolso está lleno de papeles, documentos, fotografías, llaves y monedas (oración 3).

Presentamos 3 oraciones con los siguientes valores de longitud total (WT):

$$WT_{oración\ 1} = 11; \quad WT_{oración\ 2} = 20; \quad WT_{oración\ 3} = 11$$

²⁷ Como ya se ha comentado, el TTR (Type/token ratio) es la relación entre el número de palabras diferentes lematizadas (eliminando duplicados) utilizadas en la construcción de un texto, y la longitud total del texto analizado.

Para la cantidad de palabras diferentes (WCD) de cada oración analizada encontramos:

$$WCD_{oración\ 1} = 9; \quad WCD_{oración\ 2} = 15 \quad WCD_{oración\ 3} = 11$$

Debemos recordar que antes de hacer el conteo de palabras, el texto pasa por un procedimiento de lematización (descrito en el apartado 3.2.2) y las palabras que se repiten solo cuentan una única vez. Ahora hacemos el cálculo del TTR:

$$TTR_{oración\ 1} = \frac{WCD}{WT} = \frac{9}{11} = 0,82$$

$$TTR_{oración\ 2} = \frac{15}{20} = 0,75$$

$$TTR_{oración\ 3} = \frac{11}{11} = 1,00$$

Se miramos estos resultados, es posible decir que la oración 3 utiliza un mayor vocabulario que las otras 2. Pero esto no es verdad de todo. Podremos decir que la oración 3 posee más vocabulario que la oración 1, porque para la misma cantidad de palabras ($WT = 11$) la oración 3 tiene más palabras diferentes ($WCD = 11$), que la oración 1 ($WCD = 9$). La comparación con la oración 2 se ve perjudicada, pues el texto de la oración 2 posee más palabras que las otras 2. Para resolver esta cuestión los investigadores utilizan el procedimiento de contraer el texto analizado considerando solamente la longitud de interés. En nuestro ejemplo, en la oración 2 apenas se considerarán las primeras 11 palabras ($WT = 11$), lo que está indicado por doble barra en la oración referida. Haciendo los cálculos del $TTR_{contraído}$ de la oración 2, tenemos:

$$TTR_{contraído} = \frac{9}{11} = 0,82$$

Con los resultados actuales, las oraciones 1 y 2 ahora poseen la misma cantidad de palabras diferentes (WCD), pero aún por debajo del valor de la oración 3. De este modo, decimos que la oración 3 es más rica en vocabulario.

El punto de intereses a considerar sobre este procedimiento es pensar lo siguiente: ¿qué es lo que sucede con la producción del texto después que ultrapasar el límite impuesto por este procedimiento?

Es posible que puedan suceder dos cosas: el autor mantiene las tasas de incremento de nuevas palabras –y así el valor calculado refleja el comportamiento de la producción textual global del autor en el texto analizado– o el autor cambia la tasa de incremento de nuevas palabras –y, de este modo, el valor calculado reflejará el comportamiento del autor hasta este punto, pero no refleja el comportamiento global empleado en la producción del texto analizado–.

Lo que se conoce es que el incremento de nuevas palabras (este punto será discutido con más detalles en los apartados 5.2.2.5 y 5.2.3.2) no se mantiene igual en toda la longitud del texto (van Velzen y Garrard, 2008) porque las repeticiones de palabras son constantes a lo largo de un texto producido. Esto se puede observar en la oración 2. A medida que el texto crece, las repeticiones se hacen más constantes. De las 9 palabras añadidas después del límite impuesto por el método sugerido (WT = 11), hasta el final (WT = 20), 4 eran repetidas (*y*, *roja*, *blanca* y *la*).

Otro punto a considerar es si la dificultad de encontrar nuevas palabras para una persona que esté experimentando un deterioro cognitivo es una constante a lo largo de toda la producción de un texto (como una novela) o si, después de haber producido una cantidad de texto de esa novela (por ejemplo, después de la presentación inicial de los personajes, del ambiente y de los motivos, que se llevan en los primeros capítulos), el autor se ve con mayores dificultades para desarrollar la trama añadiendo elementos nuevos (que se reflejarían en el incremento de nuevas palabras) haciendo que la tasa de repeticiones sea mayor.

Estas son características de la producción del texto que, desde el punto de vista de este trabajo, no se cubre con el procedimiento de normalización. Una tentativa de responder a este cuestionamiento fue implementada por van Velzen y Garrard (2008). El estudio, basado en el análisis de los libros del escritor Gerard Reve (1923-2006), analizó la diversidad léxica en tres libros del novelista holandés quien escribió su última

novela poco antes de ser diagnosticado con la EA. Los datos muestran una disminución de la diversidad léxica, lo que coincide con el reportado “olvido” durante el proceso creativo en la escritura.

No obstante, el método difirió del empleado en los estudios anteriores para Iris Murdoch, pues cada libro fue dividido en bloques de textos consecutivos de misma longitud (1000 palabras), y se calculó el TTR para cada bloque y para los bloques agrupados en dos mitades: el inicio del libro y el final del libro, para cada uno de los tres libros analizados. El bloque inicial era de mismo tamaño, pero el bloque final estaba condicionado a la longitud final de cada libro. Según los investigadores, hubo una coincidencia entre el comienzo del deterioro cognitivo y la última novela escrita, pues se observó una marcada diferencia entre los valores de TTR de la primera respecto a la segunda mitad, lo que no pasó con los dos libros anteriores.

5.2.2. Aspectos metodológicos

En este apartado se ofrecerán algunas consideraciones sobre ciertas metodologías y métricas elegidas para la caracterización de los cambios en la producción textual asociados con el paso del tiempo y usados en los trabajos de investigación citados (aunque solo en relación con Iris Murdoch).

Referente a la metodología, se discutirán dos aspectos: el tratamiento dado a los textos y el método de análisis empleado para investigar el efecto de la edad sobre las métricas analizadas, la Regresión Lineal y las correlaciones.

Por su parte, sobre las métricas se discutirán el *type-token ratio* (TTR), la tasa de incremento de nuevas palabras (WTIR) y las alteraciones sintácticas relativas a la longitud promedio de las oraciones (WPS). Los dos primeros parámetros se centran en la selección de palabras, mientras que el último se relaciona con el ámbito de las combinaciones, según la especificación de Jakobson y Halle (1973).

Además, se discutirá una característica de la producción textual de la escritora: el aumento de la longitud de los textos producidos con el paso del tiempo (con excepción del último libro), lo que puede llevar a una interpretación equivocada de los resultados encontrados.

5.2.2.1. Tratamiento de los textos

Un aspecto a comentar sobre los trabajos previos es el tratamiento de los textos analizados. En su trabajo, Garrard (2005) eliminó los diálogos y los nombres propios, y resolvió las ambigüedades a nivel de categorías morfosintácticas de modo manual (por ejemplo, la palabra *hand* que puede ser utilizada como nombre o como verbo). Estos tratamientos no fueron aplicados al texto completo, sino a cinco muestras al azar de 100 palabras tomadas de cada uno de los tres libros analizados. El conteo de la cantidad de palabras por oraciones se obtuvo de modo manual (en un conjunto de tres porciones de 10 oraciones tomadas del principio, del medio y del final de cada uno de los libros) y la longitud promedia de las oraciones se calculó automáticamente para los textos completos.

El primer punto a discutir es la eliminación de algunas partes del texto: diálogos y nombres propios. La justificación de esta omisión es que el autor podría haber "simulado" una manera de hablar, que no es la propia, para caracterizar a su personaje. Ahora bien, en este trabajo se entiende que, en el caso de que un autor simule una manera de hablar que no es la propia, aún sigue siendo el autor quien produce lo que se escribe. Si bien difiere en su manera característica de hablar, lo que escribe es lo que sale de su mente. Serguei Pakhomov (2010) también eliminó los diálogos alegando que no forman parte del alcance de su trabajo, pero sin justificar los motivos.

Por su parte, Le (2010) discute la influencia que podrían causar los diálogos en los resultados. Confirma esta preocupación de la siguiente forma:

The fact that the target texts are novels, which inevitably contain dialogues, presents a complication that may affect the measures of syntactic complexity given above. Most fiction writers try to capture the essence of natural, real-life conversations in their dialogues; since spoken language tends to have lower complexity, with shorter sentences, fewer embedded clauses, less complex grammar, and more fragments, the proportion of dialogue in each novel partly determines its complexity scores. An optimal solution is to perform separate syntactic analysis on the dialogue portions and the narrative portions; however, this separation of dialogue from narrative cannot be accomplished, given the properties of our scanned

texts. Consequently, the results [...] measures might not reflect the absolute syntactic levels of the authors.

Una vez más se reitera la idea de que no existe otra persona produciendo el texto y si se crea un personaje que, por ejemplo, repite las palabras exhaustivamente, puede significar que se refleja en su escritura lo que está pasando en su cerebro.

Además, el autor puede utilizar la estrategia de construir diálogos cortos solo para no desarrollar una trama más descriptiva, más elaborada quizá por una dificultad cognitiva. Por lo tanto, en este trabajo se defiende que el texto debe ser analizado en su conjunto con todos los diálogos y los nombres propios, sin ninguna reserva, teniendo en cuenta incluso que estos aspectos también caracterizan la evolución de la escritura del autor. Los textos escaneados para esta tesis, también presentaron las mismas dificultades encontradas por Xuan Le en la separación de los diálogos del resto del texto. Por esto, queda pendiente verificar, en trabajos futuros, si los diálogos distorsionan o no la complejidad sintáctica.

El segundo punto a comentar se refiere al manejo de las ambigüedades y el cálculo de las métricas en muestras menores tomadas de los libros analizados de modo manual. El muestreo realizado por Garrard (2005) y Pakhomov (2010), seleccionando algunas partes del libro, ya fue exhaustivamente explicado y evaluado por Le (2010). Según lo que se plantea en este trabajo, lo mejor es que se use todo el libro en los análisis y no las 100 primeras páginas de cada libro extrapolando los resultados para el libro completo porque no hay garantías de la fiabilidad del análisis estructural. Así, por ejemplo, se mantienen los aspectos relativos a la longitud total del libro medido en palabras, el número de capítulos, el número de caracteres y la relación narrativa/diálogo cuando se comparan los resultados de libros completos con partes de otros libros.

También en este trabajo se considera difícil que medidas como las citadas en el párrafo anterior, realizadas para las primeras 100 páginas de los libros, puedan ser extrapoladas a todo el libro: en el caso del libro *The Sea, The Sea*, por ejemplo, haciendo la extrapolación de la cantidad total de palabras y la cantidad de palabras diferentes del libro, a partir del resultado de las primeras 100 páginas, apenas multiplicando el resultado calculado por el factor 5, supone una regularidad en el modo de escritura a lo largo del libro, que no es verdadera. Esta proyección se basa en el

supuesto cuestionable de que las restantes páginas de la novela eran idénticos en estructura a las primeras 100 páginas. Xuan Le va más allá y cuestiona la verdadera función de otras decisiones metodológicas utilizadas por Garrard:

Even if this problem with the data [calculados en porciones menores del texto] could be overlooked, we question the relevance of these measures, since little implication on the cognitive health of the author could be drawn from facts such as «the first book is subdivided into more chapters than either of the two later works, while the middle work is far and away the longest of the three» (p. 254). The variables measured in this structural analysis [como por ejemplo las relativas a la longitud total del libro medido en palabras, el número de capítulos, el número de caracteres, y la relación narrativa/diálogo, citadas en párrafo anterior] depend heavily on the topics, genres, settings, perspectives, plot development and other characteristics, which may vary among works by the same author regardless of age or cognitive health.

Basado en estas observaciones Xuan Le ha decidido realizar los análisis de modo automatizado para los libros completos. El mismo procedimiento será empleado en este estudio.

5.2.2.2. Métodos de Correlación

En relación con los métodos de análisis para las métricas de nuestro interés como, por ejemplo, el TTR o TTR55 y la longitud promedio de las oraciones, algunas de las investigaciones mencionadas evaluaban la relación de la edad con las métricas medidas por medio del método ANOVA (Pakhomov et al., 2010), mientras que otras lo hacían por medio de una Regresión Lineal simple usando la estadística ANOVA (Le et al., 2011) en busca de cambios en el tiempo.

Desde este trabajo se aboga que la Regresión Lineal asociada a cualquier prueba de correlación no es adecuada para esta clase de investigaciones porque es más apropiada para identificar tendencias a lo largo de la vida de un autor en lugar de

alteraciones o disfunciones en el comportamiento. En el presente estudio, lo que interesa es identificar un punto de cambio significativo en una edad avanzada, respecto al comportamiento previo y que permita detectar si esta disfunción ocurre al final de la carrera del escritor, así como advertir si existe una discapacidad cognitiva asociada al envejecimiento. Este es el objetivo de la tesis y se basa en el hecho de que la EA se caracteriza por alteraciones cognitivas que se manifiestan como una disfunción o alteración del comportamiento (en nuestro caso, de la escritura) que se producen a edades más avanzadas.

Supongamos, hipotéticamente, una situación en que una métrica cualquiera se compara con la edad de un autor mediante una regresión, y que la correlación entre ellos presenta un valor no significativo con la edad. En otras palabras, con el paso del tiempo los libros del autor no presentan ninguna tendencia de crecimiento o decrecimiento de este parámetro. Pero, cuando miramos el resultado de este parámetro para los últimos libros del autor, verificamos que hubo un incremento o una disminución drástica en comparación con los datos anteriores. Basándonos en el criterio de correlaciones estadísticamente significativas, este resultado podría ser considerado como no válido.

Pero lo que sucede con estas enfermedades asociadas a un deterioro cognitivo es que el paciente experimenta una disfunción, un cambio en su conducta, que se considera normal, cuando aparece la enfermedad. La existencia de datos que difieren de los anteriores en el periodo de envejecimiento del escritor es precisamente la situación que podría caracterizarse como una demencia senil o EA: el deterioro cognitivo, asociado a la edad, es un síntoma. Idealmente, necesitamos un método para identificar este punto de disfunción.

5.2.2.3. Longitud de los textos

La escritora Iris Murdoch presentó una tendencia a producir libros de longitud creciente con la edad, medido en cantidad total de palabras lematizadas, con una correlación lineal estadísticamente significativa [$F(1,22) = 18.93, p < 0,001$] y [Pearson = 0,69, $p < 0,001$], considerando todos los libros analizados de la autora (**¡Error! No se encuentra el origen de la referencia.a**). Si se elimina el último libro de la autora, *Jackson's Dilemma*, en el que existe la sospecha de un desvío en los datos estándares de escritura de la autora, las correlaciones son aún más significativas (**¡Error! No se**

encuentra el origen de la referencia.b), con valores de $[F(1,22) = 55.84, p < 0,0001]$ y $[\text{Pearson} = 0,85, p < 0,0001]$.

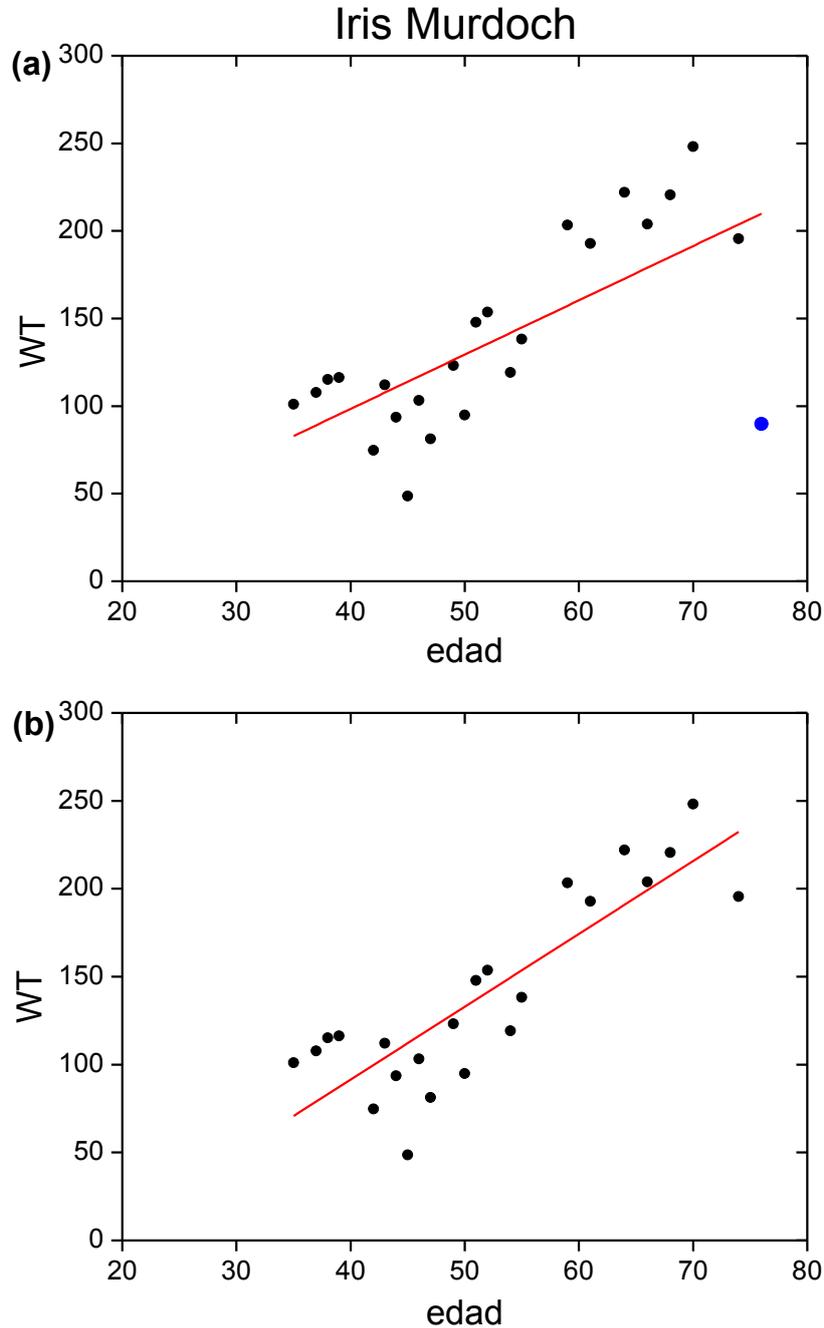


Figura 5.1 Crecimiento de la longitud del texto²⁸ en relación con la edad de todos los libros (a) y tras eliminar el libro *Jackson's Dilemma* (b).

²⁸ En todos los gráficos de aquí adelante los valores de WT y WCD serán presentados como múltiplos de 1.000.

Hay que tener en cuenta esta característica de la producción textual de Iris Murdoch en el análisis de los resultados, pues este crecimiento en la longitud de los textos puede llevar a interpretaciones erróneas. ¿En qué medida lo que vemos está relacionado con el envejecimiento y cómo se relaciona con el aumento de la longitud del texto? La respuesta a esta pregunta se presentará más adelante en el apartado 5.2.3.1.

5.2.2.4. Type/token ratio (TTR)

Type/token ratio (TTR) es la relación entre el número de palabras diferentes lematizadas (eliminando duplicados) utilizadas en la construcción de un texto, y la longitud total del texto analizado. Las medidas de la diversidad del vocabulario, utilizando el TTR como parámetro, se aplican en una amplia gama de investigaciones lingüísticas, incluidas las de desarrollo del lenguaje infantil (H. P. Chen y Irwin, 1946; Rondal y Turnure, 1976; C. Hess, Sefton y Landry, 1986; Owen y Leonard, 2002; Phillips, 2009; Heisler, Goffman y Younger, 2010), los trastornos del lenguaje (Harris Wright, Silverman y Newhoff, 2003; Gordon, 2008; Park, Rogalsky, Rodriguez y Zlatar, 2011), el aprendizaje de segundas lenguas (Warschauer, 1996; Richards y Malvern, 2000), los estudios sobre la esquizofrenia (Manschreck, Maher, & Ader, 1981; Manschreck, Maher, Hoover y Ames, 1984), los trastornos afectivos (Andreasen, 1975), y algunos estudios lingüísticos y psicológicos del lenguaje (Youmans, 1971; C. W. Hess, Ritchie y Landry, 1984; Biber, 1987; Y. S. Chen y Leimkuhler, 1989; Singh, 2001; Louwerse, McCarthy, Mcnamara y Graesser, 2004; Cheung y Kemper, 2008; McCarthy y Jarvis, 2010; Leedham, 2011).

De acuerdo con McKee y colaboradores (2000), existe una serie de programas informáticos para el cálculo del TTR, como el SALT –*Systematic Analysis of Language Transcripts*,²⁹ el *Oxford Concordance Program*³⁰ y el CLAN –*Computerized Language Analysis*³¹– (MacWhinney, 1995; MacWhinney y Snow, 1990). Aparte del análisis del TTR, existen también otras derivaciones matemáticas de este parámetro, como el *Root TTR* (Guiraud, 1960) o el *Corrected TTR* (Carroll, 1964) (apud McKee, Malvern y Richards, 2000). Estas derivaciones matemáticas intentan resolver un problema

²⁹ <http://www.saltsoftware.com/>

³⁰ <http://users.ox.ac.uk/~ctitext2/resguide/resources/o125.html>

³¹ <http://childes.psy.cmu.edu/clan/>

intrínseco al parámetro TTR: este parámetro es dependiente de la longitud final de la muestra de texto analizada, de forma que, cuanto mayor es la muestra, más pequeños son los valores del TTR. Según los mismos autores (McKee et al., 2000), este comportamiento, con frecuencia, ha distorsionado algunos resultados de ciertas investigaciones, pero, lamentablemente, se siguen publicando sin que se produzca ningún intento por controlar la variación en la longitud de las muestras (ver Tweedie & Baayen, 1998, para algunos ejemplos).

McKee y colaboradores siguen evaluando la aplicación de estos métodos y sostienen que existen varios intentos para superar el problema como, por ejemplo, mediante la estandarización de la cantidad de tokens (palabras) producidas por cada sujeto analizado. No obstante, este procedimiento no ha logrado garantizar que las medidas sean comparables con los resultados de otros investigadores que utilizan distintas cantidades de palabras.

La otra forma de controlar este problema es limitar la longitud de la muestra, haciendo recortes en tamaños iguales para lograr la comparación de los resultados de autores diversos y de muestras del mismo autor (como ya se ha explicado en el apartado 5.2.1). El problema es que cuando se reducen los textos a muestras más pequeñas, inevitablemente se pierden datos importantes sobre el texto completo. Además, el procedimiento de recorte de los textos no refleja lo que sucede con la totalidad del discurso (libro) producido por sus autores. Es decir, se pierde la referencia sobre la cantidad de vocabulario utilizado cuando el autor finaliza su obra, lo que significa que un corte en una parte inicial del texto implica la pérdida de informaciones si el autor, por ejemplo, ha experimentado alguna dificultad en encontrar nuevas palabras al final de la escritura de un libro. Esto fue comprobado por van Velzen y Garrard, 2008, y discutido en el apartado 5.2.1.

Y, por último, tampoco es posible conocer hasta qué punto las longitudes finales del texto ejercen una influencia sobre los valores calculados del TTR en recortes de 55.000 palabras. En otras palabras, cuál es la influencia, si existe, entre el modo como el autor construye un libro pequeño (con pocos personajes y pocas descripciones ambientales) y un libro mayor (en el cual se desarrollará una historia más larga) sobre el vocabulario usado en las primeras 55.000 palabras de ambos libros. Aunque no parezca un planteamiento lógico o trivial, cuando se han analizado los 23 libros de la autora Iris Murdoch se ha encontrado una correlación entre estas dos métricas (ver Figura 5.2).

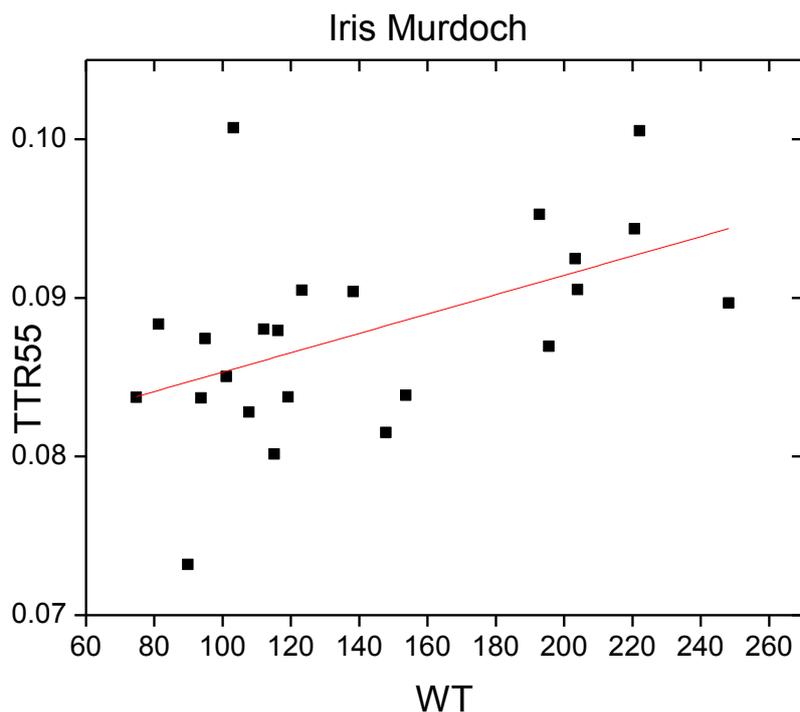


Figura 5.2 Relación entre las métricas TTR55 y WT.

La relación entre el TTR55 y la longitud final de los textos presentó un valor de correlación de Pearson estadísticamente significativo [Pearson = 0,50, $p < 0,05$], lo que indica que existe una correlación positiva entre las dos métricas, es decir, que cuanto mayor es el libro final mayor es el vocabulario utilizado en las primeras 55.000 palabras del libro producido por la autora.

Por su parte, la cantidad de vocabulario usado en las primeras 55.000 palabras de los libros de la escritora Iris Murdoch no se relacionó de modo estadísticamente significativo con la edad en que ellos fueron escritos [Pearson = 0,16, $p < 0,46$].

La influencia de la longitud del texto sobre el comportamiento de la métrica TTR55 está explicada más adelante en el apartado 5.2.3.1.

5.2.2.5. WTIR

Mediante el análisis de la construcción de un texto, en incrementos de pequeñas cantidades de 500 palabras (WTIR), podemos hacer un seguimiento de la evolución del comportamiento de la producción del texto en lo que se refiere a la cantidad de palabras nuevas añadidas a lo largo de su producción.

Garrard y Le utilizaron este procedimiento para identificar si existe una correlación entre la tasa de crecimiento del texto en cantidad de palabras diferentes, la longitud del texto y la EA diagnosticada cuando la autora escribió su último libro. Lo que se plantea es si existe una correlación entre un TTR bajo y una mala calidad literaria de un libro o, lo que es lo mismo, una dificultad en el uso del vocabulario asociado a la vejez.

Se presentan a continuación algunos resultados calculados solamente para corroborar las observaciones hechas sobre los trabajos anteriores de investigación de los escritos de la autora Iris Murdoch, que contribuirán a apoyar la crítica y que son relevantes para este estudio.

5.2.3. Análisis de los puntos considerados

En este apartado se discutirá la probable influencia de la longitud de un texto en los resultados y la adecuación de las métricas TTR55 y WTIR para identificar el deterioro cognitivo que los autores proponen.

5.2.3.1. Riqueza del vocabulario 1 (TTR55)

Para medir y comparar la riqueza de vocabulario utilizado en los tres libros (indicados en la leyenda del gráfico de la Figura 5.3a, con la edad en que fueron escritos entre paréntesis), se ha relacionado la cantidad de palabras diferentes con la cantidad total acumulada de palabras en cada serie de 10.000 palabras, y se ha trazado un gráfico que muestra el empobrecimiento del vocabulario en el último libro de la autora, *Jackson's Dilemma*.

En el estudio de Le (2011), la riqueza de vocabulario se determinó al relacionarse el TTR calculado sobre las 55.000 primeras palabras con la edad de la escritura de cada libro. El análisis se realizó sobre el muestreo completo del trabajo de Le, con las 20 novelas³², que vamos llamar de *Muestra Uno*. La *Muestra Uno* fue dividida en tres subconjuntos: el primero con el total de todas las novelas (*Muestra*

³² Aunque tenga analizado veinte novelas en el total, para el cálculo del TTR55 ha usado 19 novelas pues la novela *The Italian Girl* es muy pequeña y no alcanza esta cantidad de palabras totales.

Uno-T); el segundo con los primeros libros hasta *The Sea, the Sea* (*Muestra Uno-P*); y el tercero con los libros de la etapa final de producción de la autora (*Muestra Uno-F*).

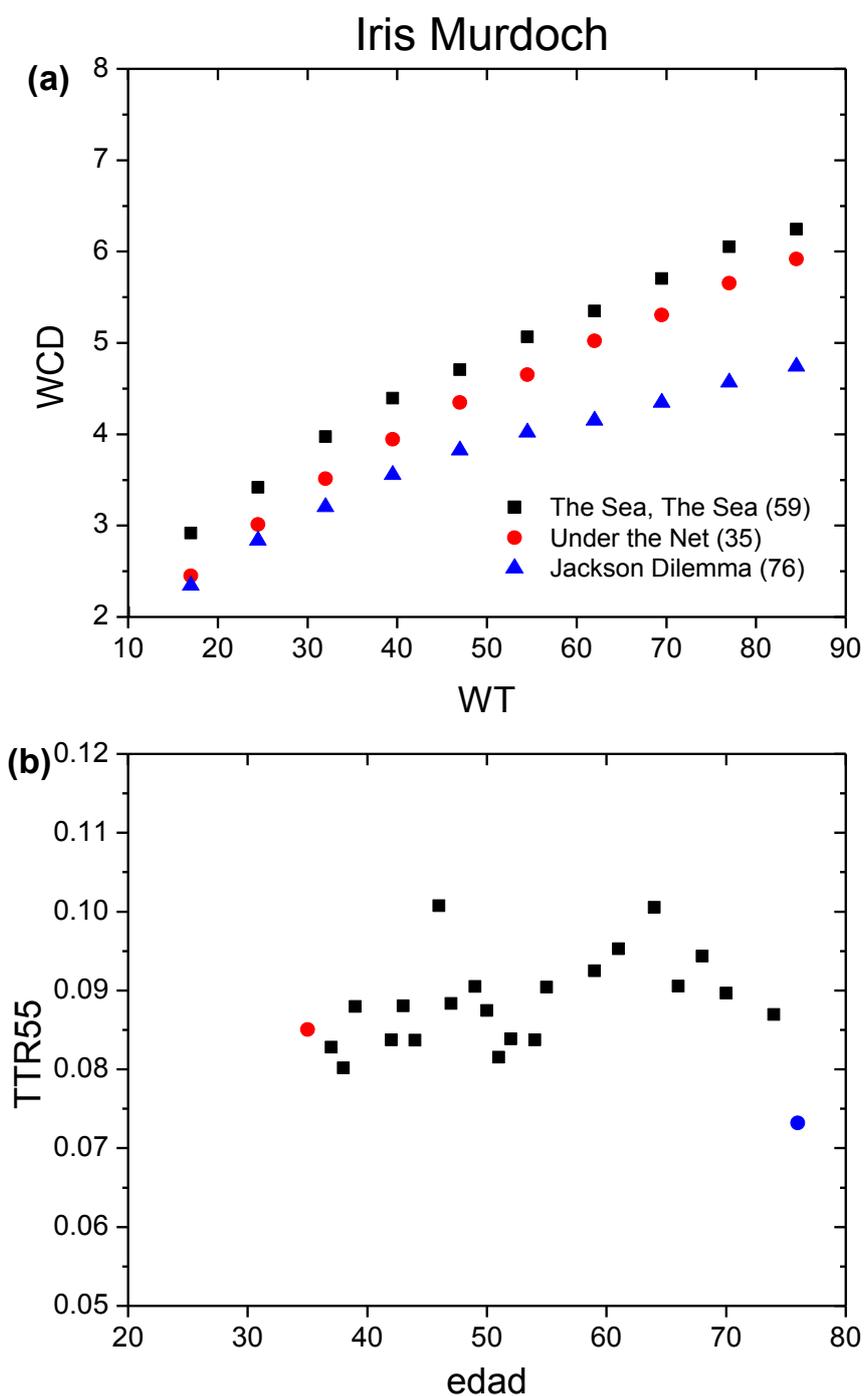


Figura 5.3 Reproducción del trabajo de Garrard (a) y de Xuan Le (b),

En la primera columna de la Tabla 5.1 están listadas las novelas analizadas por Le, de un total de 20 novelas, con la edad de la autora y el año en que fue escrita. He reproducido este mismo análisis, pero con otro conjunto de sus libros (24 novelas), que se encuentra en la segunda columna de la Tabla 5.1 para una comparación.



Tabla 5.1 El muestreo de Le (2011) y de este trabajo

Xuan Le et al			Este estudio		
Novela	Edad	Año	Novela	Edad	Año
Under the Net	35	1954	Under the Net	35	1954
The Flight from the Enchanter	36	1955	The Flight from the Enchanter	36	1955
-			The Sand Castle	38	1957
The Bell	39	1958	The Bell	39	1958
A Severed Head	42	1961	A Severed Head	42	1961
An Unofficial Rose	43	1962	An Unofficial Rose	43	1962
The Unicorn	44	1963	The Unicorn	44	1963
The Italian Girl	45	1964	The Italian Girl	45	1964
-			The Red and the Green	46	1965
The Time of the Angels	47	1966	The Time of the Angels	47	1966
The Nice and the Good	49	1968	The Nice and the Good	49	1968
Bruno's Dream	50	1969	Bruno's Dream	50	1969
A Fairly Honourable Defeat	51	1970	A Fairly Honourable Defeat	51	1970
-			An Accidental Man	52	1971
The Black Prince	54	1973	The Black Prince	54	1973
The Sacred and Profane	55	1974	The Sacred and Profane	55	1974
Love Machine			Love Machine		
Henry and Cato	57	1976	-		
The Sea, the Sea	59	1978	The Sea, the Sea	59	1978
-			Nuns and Soldiers	61	1980
The Philosopher's Pupil	64	1983	The Philosopher's Pupil	64	1983
The Good Apprentice	66	1985	Good Apprentice	66	1985
The Book and the Brotherhood	68	1987	The Book and the Brotherhood	68	1987
-			The Message to the Planet	70	1989
The Green Knight	74	1993	The Green Knight	74	1993
Jackson's Dilemma	76	1995	Jackson Dilemma	76	1995
Ntotal		20	Ntotal		24

La muestra del presente estudio (aquí referida como *Muestra Dos*) posee diferencias cuando se compara con la *Muestra Uno*. En el rango de 35 a 59 años se han incluido los libros *The Sand Castle* (38), *The Red and the Green* (46) y *An Accidental Man* (52), excluyendo el libro *Henry y Cato* (57). En el intervalo a partir de los 60 años, se han incluido los libros *Nuns and Soldiers* (61) y *The Message to the Planet* (70). El resto de los libros son los mismos para ambas investigaciones.

El procedimiento de los recortes del conjunto de textos analizados en 3 subconjuntos es el mismo que el utilizado en el trabajo de Le, o sea, el primero es compuesto por todos los libros que poseía para el análisis (24³³) – que llamaré de *Muestra Dos-T*; el segundo con los primeros libros hasta *The Sea, the Sea* (*Muestra Dos-P*); y el tercero con los libros de la etapa final de producción de la autora (*Muestra Dos-F*).

Como los libros analizados no son exactamente iguales para las *Muestras Uno* y *Muestra Dos*, el contenido de los subconjuntos generados no es igual para los dos estudios. Por esto, aquí se ha generado una *Muestra Modificada* intentado hacer un muestreo lo más similar posible al Muestreo de Le (*Muestra Uno*), o sea, excluyendo los libros *The Sand Castle*, *The Red and the Green*, *An Accidental Man*, *Nuns and Soldiers* y *The Message to the Planet*, para poder comparar los resultados de este estudio con los obtenidos por los autores. La diferencia reminiscente es que no se ha podido conseguir el libro *Henry and Cato* para añadir a esta muestra, pero aun así se analizarán los datos de esta *Muestra Modificada*. El procedimiento de generación de los tres subconjuntos usados para la muestra de Le (*Muestra Uno*) y para la muestra completa del presente estudio (*Muestra Dos*) fue aplicada a la *Muestra Modificada*, generando los subconjuntos *Muestra Modificada-T*, *Muestra Modificada-P* y *Muestra Modificada-F*.

Para disipar las dudas sobre si el muestreo ha influido en los resultados, aquí se ha realizado una cuarta muestra con los valores extraídos del artículo de Le (*Muestra Extrapolada*), escaneando el gráfico (ver Figura 5.4) y verificando qué valores de x e y correspondían a cada punto dibujado en el gráfico. Con estas informaciones, se ha

³³ Para el análisis del TTR55 fueron analizadas 23 novelas, pues la novela *The Italian Girl*, como ya se ha comentado, no atendía las condiciones de cantidad de palabras totales superior a 55.000 palabras.

calculado la Regresión Lineal simple entre TTR55 y la edad³⁴. Los recortes en 3 subconjuntos también se han producidos, generando las muestras *Muestra Extrapolada-T*, *Muestra Extrapolada-P* y *Muestra Extrapolada-F*.

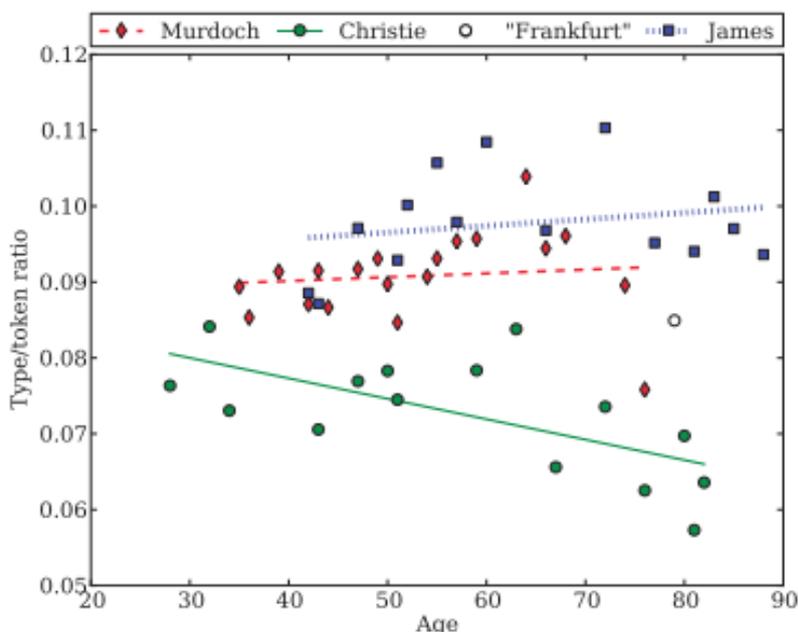


Figura 5.4 Reproducción del gráfico del artículo de Xuan Le (2011) escaneado.

El análisis de las cuatro muestras y sus subconjuntos se encuentran en la Tabla 5.2. En la primera línea de la tabla se muestra las letras T, P y F que corresponden a los tres subconjuntos: T (conjunto Total de los libros analizados), P (Primeros libros hasta *The, Sea, the Sea*, inclusive) y F (libros Finales de la carrera de la autora). Estos subconjuntos están asociados a las Muestras que se presentan en la primera columna de la tabla, *Muestra Uno*, *Muestra Dos*, *Muestra Modificada* y *Muestra Extrapolada*, explicadas en los párrafos anteriores.

Para el conjunto total de los libros (T), no existe ninguna correlación significativa entre la edad y el TTR55 en ninguna de las cuatro Muestras analizadas, como se puede observar en la primera columna de resultados de la Tabla 5.2.

³⁴ Se ha utilizado este procedimiento, pues no existía ninguna tabla con los datos en el artículo.

Tabla 5.2 Resultados de correlaciones entre la edad y el TTR55 para Iris Murdoch.

Muestreo	T	P	F
Muestra Uno	F(1,19) = 0.19 p=0.66	F(1,13) = 13.41 p=0.003	F(1,3)=14,17 p=0.03
Muestra Dos	F(1,22) = 0.56 p=0.46	F(1,15) = 1.51 p=0.24	F(1,6) = 12.07 p=0.02
Muestra Modificada	F(1,17) = 0.25 p=0.62	F(1,12) = 2.76 p=0.12	F(1,4) = 11.70 p=0.04
Muestra Extrapolada	F(1,18)=0.11 p=0.75	F(1,13)=7.59 p=0.02	F(1,4)=14.19 p=0.03

Para los primeros libros escritos por la autora hasta el libro *The Sea, the Sea* (P), Le (*Muestra Uno*) encontró una correlación positiva estadísticamente significativa entre los primeros libros y la edad. Este resultado no se puede observar en la *Muestra Dos* (los datos del presente estudio) ni tampoco en la *Muestra Modificada* (con los libros lo más similares posibles a la *Muestra Uno*, excepto el libro *Henry and Cato*), pues las correlaciones de estos datos no eran estadísticamente significativas. Aunque la correlación entre los datos adquiridos a partir del propio gráfico, incluido en el artículo de la autora, sea positiva y estadísticamente significativa, los valores no coinciden. (ver columna P, líneas 1 y 4 de la Tabla 5.2).

Para los libros finales de la carrera de la autora todas las muestras presentaron correlaciones positivas estadísticamente significativas con valores similares.

Las diferencias de valores encontrados para las correlaciones es una información pertinente, ya que la declaración de la autora (Le) sobre el mayor uso del vocabulario en edad temprana y la reducción del uso del vocabulario a partir de los 60 años depende de la muestra tomada para el estudio.

Aparte de las diferencias encontradas en los valores, como se ha identificado una relación entre la longitud de los textos (WT) con la edad, hace falta analizar sí esta correlación influye en las relaciones entre el TTR55 y la edad. Para realizar este

análisis, se ha utilizado la Regresión Lineal Múltiple (RLM) para verificar cuál es el grado de influencia entre la longitud y la edad en la métrica TTR55.

En el total de Libros (T)

Las estadísticas de Regresión Lineal individuales, utilizando ANOVA, para las dos métricas encontradas fueron TTR55 x Edad ($F = 0,56$ $p = 0,46$), lo que implica que la relación no es significativa, y TTR55 x WT ($7,11$ $p = 0,01$) que se considera significativa pues el $p \leq 0,05$.

El modelo propuesto, es decir, la explicación del comportamiento de la métrica TTR55 para dos variables (WT y Edad) es un modelo significativo, pues el análisis del RLM presenta el valor de $F = 4,66$ $p = 0,02$ y es, por lo tanto significativo, pues el $p \leq 0,05$.

Analizando el valor absoluto (sin considerar el señal negativo) de los Coeficientes Beta, se obtiene la contribución individual de cada métrica (WT y Edad) para el comportamiento del TTR55. Los valores encontrados son $Beta_{WT} = 0.745$ $p < 0,0011$ y $Beta_{Edad} = 0,352$ $p = 0,182$. Como el valor del Coeficiente Beta para el WT es mayor que el valor del coeficiente Beta para la edad, y como se trata de una relación significativa ($p \leq 0,05$), es posible afirmar que la longitud del texto presenta mayor contribución para explicar el comportamiento de la métrica TTR55 para el texto completo que la métrica Edad.

En los Primeros Libros (P)

Las estadísticas de Regresión Lineal individuales, utilizando ANOVA, para las dos métricas encontradas fueron TTR55 x Edad ($F = 1,51$ $p = 0,24$) y TTR55 x WT (0.25 $p = 0,62$), lo que implica que las dos relaciones son no significativas.

El modelo propuesto, es decir, la explicación del comportamiento de la métrica TTR55 para dos variables (WT y Edad) no es un modelo significativo, pues el análisis del RLM presenta el valor de $F = 7,56$ $p = 0,49$ y, por lo tanto, no es significativo, pues el $p \leq 0,05$.

Esto significa que el comportamiento del TTR55 no se relaciona con las dos métricas.

En los Últimos Libros (F)

Las estadísticas de Regresión Lineal individuales, utilizando ANOVA, para las dos métricas encontradas fueron TTR55 x Edad ($F = 12,07$ $p = 0,02$) y TTR55 x WT ($9,76$ $p = 0,03$), lo que se considera significativa, pues el $p \leq 0,05$.

El modelo propuesto, es decir, la explicación del comportamiento de la métrica TTR55 para dos variables (WT y Edad) es un modelo significativo. El análisis del RLM presenta el valor de la estadística $F = 18,05$ $p = 0,01$ y, por lo tanto, es significativo, pues el $p \leq 0,05$.

Analizando el valor absoluto (sin considerar el señal negativo) de los Coeficientes Beta, se obtiene la contribución individual de cada métrica (WT y Edad) para el comportamiento del TTR55. Los valores encontrados son $Beta_{WT} = 0,515$ $p = 0,05$ y $Beta_{Edad} = 0,573$ $p = 0,04$. Como el valor del Coeficiente Beta para las dos métricas es casi igual, con $p \leq 0,05$, es posible afirmar que la longitud del texto presenta una contribución igual a la contribución de la edad para explicar el comportamiento de la métrica TTR55 para los últimos libros escritos por la autora. Esto implica decir que no se puede concluir que el envejecimiento sea el factor que determina la caída del uso del vocabulario en los libros. Lo correcto es decir que en esta franja de edad, con el paso del tiempo, la autora ha escrito libros cada vez menores y, consecuentemente, con menor uso de vocabulario en las primeras 55.000 palabras.

Se requiere un análisis más detallado de estos parámetros para determinar si realmente las dos métricas son independientes. En caso negativo, lo que se puede hacer es eliminar una sobre la otra en la explicación del comportamiento de la métrica TTR55. Pero, en este caso, la pregunta es cuál sería eliminada, ¿la edad o la longitud de los textos?

Pero el objetivo de este análisis, realizado en este trabajo, es únicamente señalar que faltan elementos para las conclusiones de la Xuan Le cuando indica la relación entre el TTR55 y la edad, ya que no ha considerado la influencia de la longitud de los textos en esta relación.

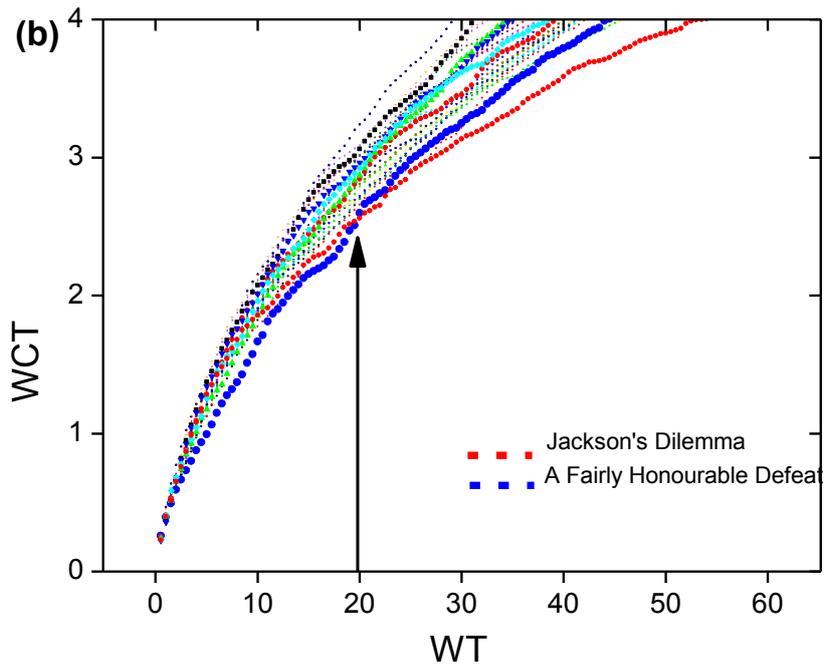
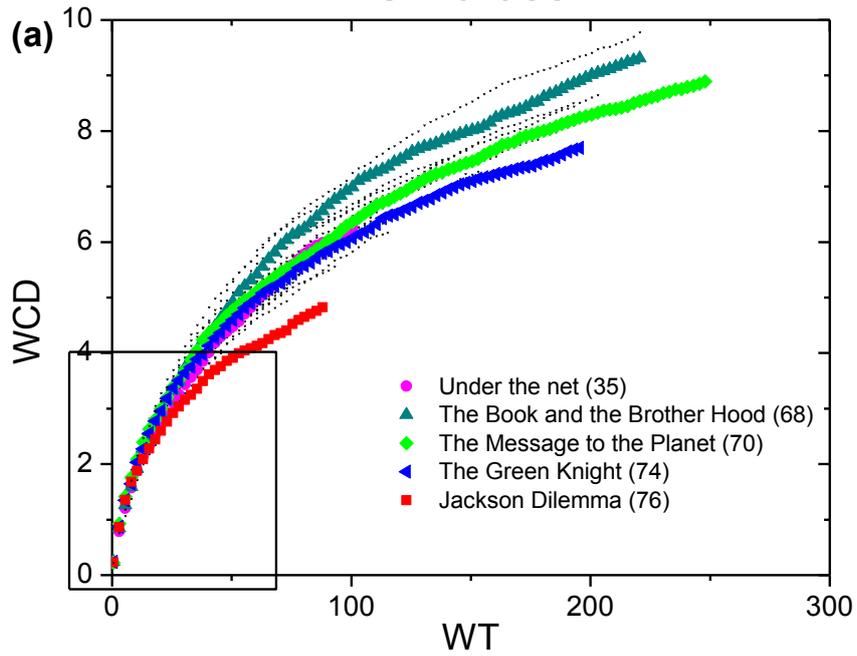
5.2.3.2. Riqueza del vocabulario 2 (WTIR)

En el mismo estudio de Le (2011) se han trazado las curvas de crecimiento de los textos, relacionando la cantidad de palabras diferentes (WCD) –añadidas en porciones crecientes del libro– con la longitud de estas porciones (WT) hasta el final del libro. El gráfico se ha hecho para todos los libros de la autora Iris Murdoch, y Le (2011) confirmó la disminución del vocabulario utilizado en la última novela de la autora a partir de un tiempo determinado de su escritura.

Aquí se ha reproducido la misma prueba, pero a intervalos de 500 palabras y aplicando la escala logarítmica en los dos ejes para visualizar mejor el comportamiento del gráfico (Figura 5.5c). Lo que se observa es la confirmación del comportamiento observado por los autores (Figura 5.5a); sin duda, se apreció un cambio repentino en el comportamiento de escritura de la autora en relación con la tasa de adición de nuevas palabras (WTIR³⁵) para su último libro. Pero este es un método no reproducible, ya que no existe ninguna seguridad de que el cambio de comportamiento de escritura se pueda detectar en un corte de 55.000 palabras para cualquier persona que tenga un deterioro cognitivo y que haya producido libros con estas longitudes.

³⁵ De su sigla en inglés: *word-type introduction rate*. En el capítulo 3 se definió detalladamente.

Iris Murdoch



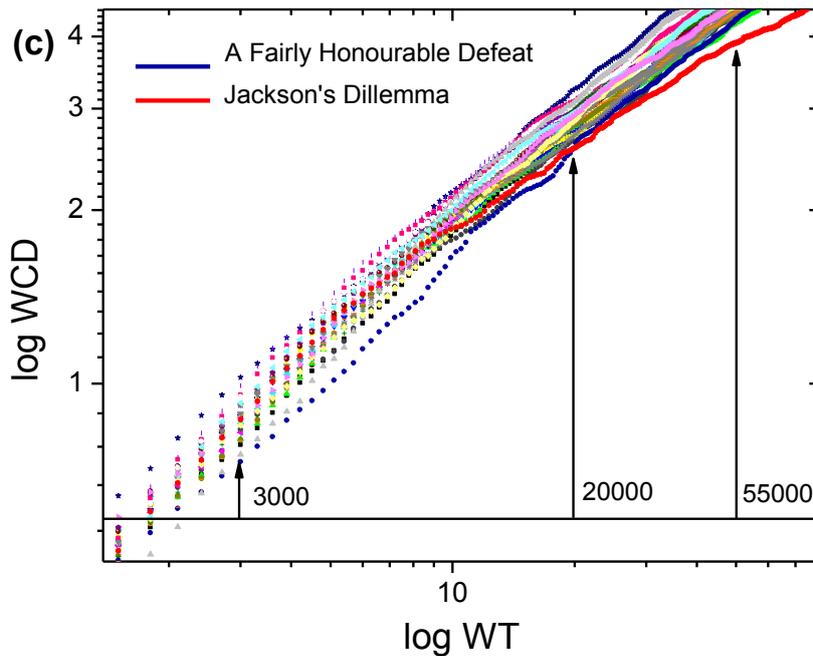


Figura 5.5 Crecimiento de los libros de Iris Murdoch³⁶ (a), Crecimiento en detalle (b); log-log de crecimiento en detalle (c).

Además, en el gráfico en detalle (Figura 5.5b y c) se puede ver que si se hubiera hecho el corte del texto, en el análisis anterior, en cualquier etapa situada entre 3.000 y 20.000 palabras, los resultados serían diferentes. Así, el libro *A Fairly Honourable Defeat*, escrito a los 51 años, presenta un comportamiento similar a lo que fue atribuido a un deterioro cognitivo para la escritora Iris Murdoch, de forma que este sería el libro apuntado como el “peor”.

En esta gama de corte se han detectado los valores más bajos de type/token ratio. Solo a partir de 20.000 palabras, el libro *Jackson's Dilemma* empieza a mostrar un comportamiento diferente, con valores más pequeños de este parámetro, a diferencia de lo que cabría esperar de un texto con su longitud. Esto sugiere que el recorte de 55.000 palabras podría no ser definitivo para identificar un deterioro cognitivo relacionado con el comportamiento en la escritura de un autor ni de varios autores.

En el próximo capítulo se presentará una propuesta, con dos nuevos parámetros y un nuevo modo de representar datos longitudinales para detectar el deterioro cognitivo en relación con el envejecimiento, incluso desde la perspectiva de intentar hacer un diagnóstico más precoz.

³⁶ Los valores de WT y WCD son expresos en múltiplos de 1000 en todos los gráficos.

Capítulo 5

*La ciencia de la matemática es
como un simple castillo de cristal,
donde adentro se ve todo,
pero de afuera no se ve nada.*
(Norma Banicevich)

6.1. Los resultados

Los resultados que se han obtenido en este trabajo se dividen en tres partes:

1. En la primera se mostrarán todos los resultados de los libros de la escritora Iris Murdoch, que se centrarán en las observaciones sobre el comportamiento de la escritura de la autora a lo largo de su vida, a partir de las trayectorias en espacio de fases (SPT³⁷) y se compararán estos resultados con los resultados encontrados utilizando el método de Regresión Lineal Simple (RLS). Esta comparación se realizará para los parámetros TTR55, WPS, WPP y PPS. En los gráficos, los valores del primer libro, *Under the Net*, están marcados en rojo, mientras que los valores del último, *Jackson's Dilemma*, se han resaltado en azul. El objetivo es explotar todas las posibilidades de comparación entre los parámetros y la edad. En los gráficos de la trayectoria se compararán dos métricas (por ejemplo, WT y TTR) y se indicará la evolución de las dos métricas analizadas teniendo en cuenta el paso del tiempo. En el gráfico, cerca de los

³⁷ De su sigla en inglés: *space phase trajectory*.

puntos que relacionan las variables, se podrá observar la evolución de las edades en que los libros fueron escritos.

2. En la segunda se describen todos los resultados de los escritores en lengua inglesa, aplicando la trayectoria (SPT) y sin hacer Regresión Lineal. Se realiza de esta forma por un doble objetivo: primero, porque funcionan como sujetos controles en relación con los escritos de Iris Murdoch; y, segundo, porque es necesario probar si el método de SPT y los parámetros propuestos son adecuados para identificar las alteraciones relacionadas con la producción escrita, aunque no exista ningún diagnóstico clínico de cualquier enfermedad cognitiva, para ninguno de ellos, y no se haya registrado ningún síntoma de deterioro cognitivo en sus biografías. En este sentido, se intentará buscar si es posible, a partir de estos análisis, una propuesta de detección precoz de la EA (o de síntomas más leves, como el DCL).

3. En la última parte se muestran los resultados de los libros para los escritores de lengua portuguesa, con los mismos objetivos que los descritos en el párrafo anterior, es decir, con el propósito de ver si la evolución de los escritos en una lengua “neutra” (respecto a la lengua de Iris Murdoch) permite probar la posible existencia de deterioro cognitivo solo a partir de la escritura.

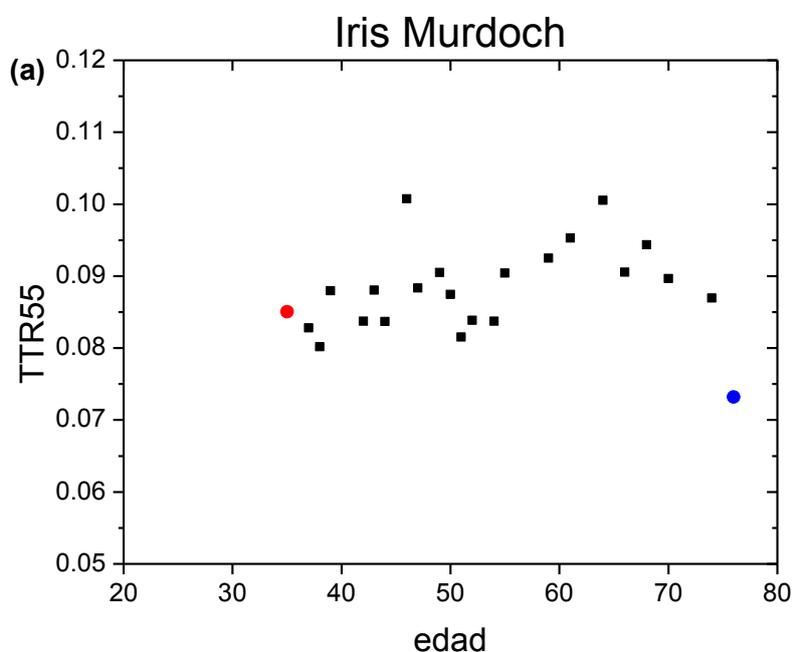
Aunque se trata de dos lenguas de orígenes diferentes con reglas de puntuación algo distintas (vid 3.1.4), la comparación relativa al uso de la puntuación entre los libros de Iris Murdoch, los libros de escritores de lengua inglesa y los libros escritos en lengua portuguesa es posible. Es posible porque la comparación se basa en los procesos de evolución individuales (estudio longitudinal) no entre los valores de las métricas medidas. Aunque cada lengua posee reglas diferentes de utilización de los signos de puntuación y aunque el autor pueda hacer uso de cantidades diferentes de las métricas analizadas (como, por ejemplo, José Saramago, quien utiliza la puntuación de un modo diferenciado), el proceso evolutivo individual, y no las cantidades, es lo que será comparado.

6.2. Iris Murdoch

Como Iris Murdoch es la única escritora con un diagnóstico de Alzheimer que pudo afectar, al menos, a sus últimas obras, los resultados del análisis de su obra son de fundamental importancia para este estudio, porque ofrece la posibilidad de entender cómo se ha desarrollado su escritura con el tiempo e, incluso, identificar si existen alteraciones en las características del uso de los elementos de puntuación, relacionados con su envejecimiento.

6.2.1. TTR55

El gráfico de la Figura 6.1a, donde se relaciona el TTR55 con la edad, indica que los libros *The Red and the Green* (46) y *The Philopher's Pupil* (64) presentaron los mayores valores de TTR55. Esto significa que en esos momentos de la vida de la autora su escritura tenía un vocabulario mayor, en comparación con sus otros libros. Pero estos dos libros poseen una diferencia en la longitud total, pues el primero (con 103.000 palabras) es casi la mitad del segundo (con cerca de 222.000 palabras).



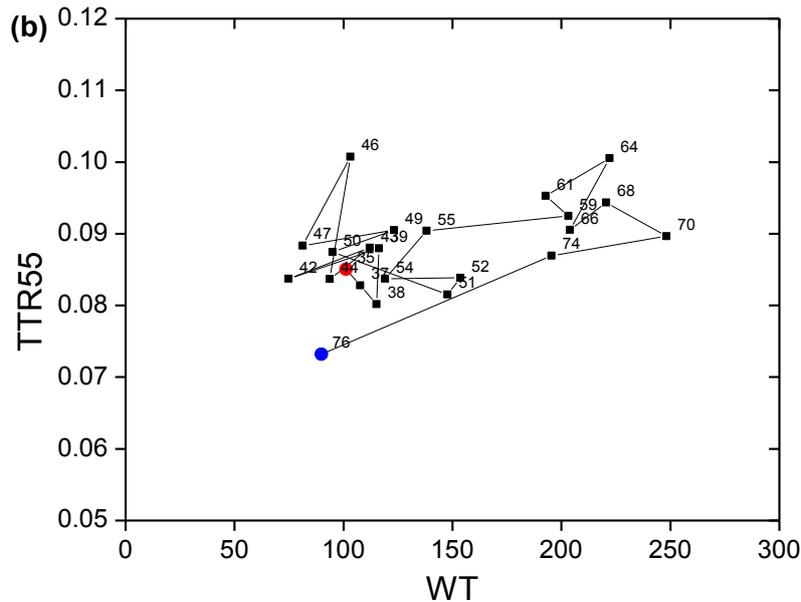


Figura 6.1 TTR55 representado por RLS (a) y por SPT (b) para Iris Murdoch.

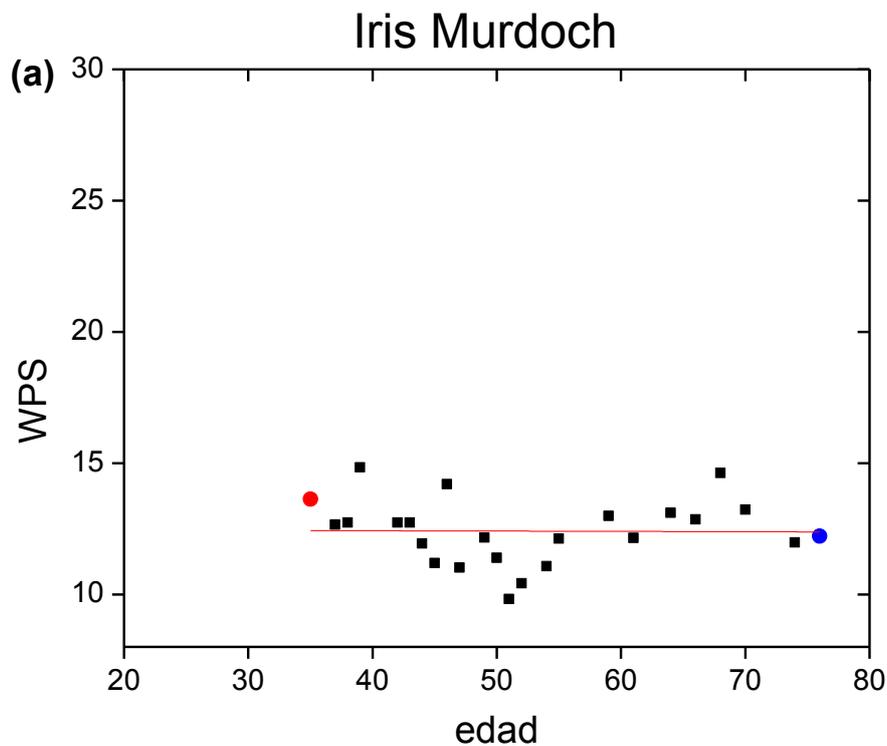
El corte se ha efectuado en un punto fijo de 55.000 palabras (Figura 6.1a), independiente de la longitud final del texto, por lo que se puede asumir que en una cuarta parte del segundo libro la autora ya había utilizado la misma cantidad de palabras diferentes que había utilizado en la mitad del primer libro. Este resultado es coherente con los trabajos que relacionan los circunloquios como uno de los síntomas más precoces de la EA respecto al uso del lenguaje. (Maxim y Bryan, 1994; Tomoeda y Bayles, apud Murillo Ruiz, 1999).

En la Figura 6.1b, donde se describe la trayectoria de la producción literaria a lo largo de la vida de la escritora con respecto a la longitud del texto (WT) –representado en el eje x – y el uso del vocabulario (TTR55) –representado en el eje y –, se puede verificar que el valor del parámetro, medido para el último libro, está influenciado por la longitud del texto porque el punto se ha dislocado hacia la izquierda (donde están representados los libros de longitud similar) y también por la edad, ya que libros con longitudes similares presentan valores del TTR55 mayores (el punto azul está situado por debajo de los demás). Es importante notar que estas últimas informaciones no son posibles de visualizar en el gráfico de la Figura 6.1a pues no considera la influencia de la longitud del texto en esta métrica (TTR55).

6.2.2. Longitud promedio de las oraciones (WPS)

La medida de la longitud promedio de las oraciones (WPS) no se relaciona de modo significativo con la edad (Edad) en que los libros fueron escritos [$F(1,24)^{38} = 0,003$, $p = 0,954$] (Figura 6.2a).

El hecho de que la escritora haya producido libros con longitudes cada vez mayores con el paso del tiempo no ha provocado ninguna influencia en la métrica WPS que mide la longitud promedio de las oraciones del texto. Esto se detecta cuando miramos el gráfico de la Figura 6.2b donde libros de la misma longitud presentan valores similares de la métrica WPS (el punto azul se mezcla con los demás puntos que representan libros de la misma longitud).



³⁸ La regresión fue calculada con todos los libros de la autora.

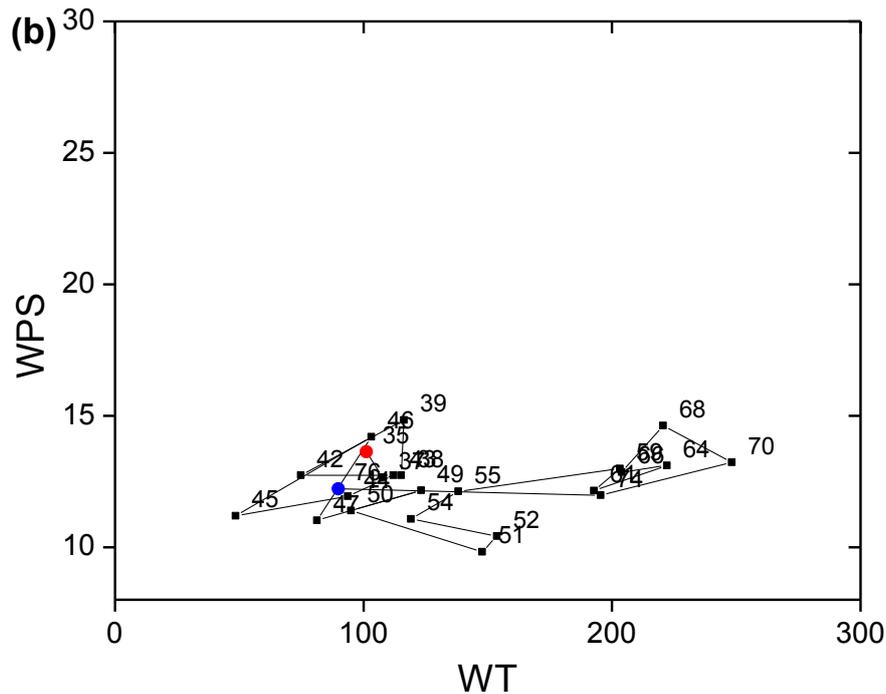


Figura 6.2 WPS representado por RLS (a) y por SPT (b) para Iris Murdoch.

Por lo tanto, el gráfico de la Figura 6.2b, que representa la trayectoria en el espacio de fase para el parámetro WPS, indica que, en realidad, el paso del tiempo no representa para la autora una pérdida en la capacidad cognitiva relacionada con la longitud de las oraciones ni tampoco se detecta una alteración al final de su vida, confirmando los resultados de la RLS. Estos resultados son consistentes con los encontrados en dos estudios previos (Garrard et al., 2005; Le et al., 2011).

6.2.3. Longitud promedio de las pausas lingüísticas (WPP)

En el gráfico de la Figura 6.3a puede verse que la autora, durante toda su vida, ha experimentado una reducción en la longitud de las pausas lingüísticas en una correlación negativa estadísticamente significativa [$F(1,24) = 40,67, p < 0,001$]. Una vez más, cuando se analiza la correlación no es posible detectar qué sucedió con los últimos libros de la autora. Esta disminución en la longitud de las pausas parece ser una tendencia en toda la vida de la autora y no representa un cambio de comportamiento cognitivo relacionado con la EA.

No obstante, en la Figura 6.3b se percibe que esta medida se relaciona con la longitud del texto, como se puede notar analizando el comportamiento del último libro con longitud menor que los inmediatamente anteriores (el punto azul) que se ha dislocado hacia la izquierda. Pero comparando este valor con los de libros de longitud similar, este presenta un valor más pequeño que los otros, escritos en épocas más tempranas.

Aquí es importante, igual que en el caso de la métrica TTR55 en el apartado 5.2.3.1, verificar cuál es la importancia de la longitud final de los libros (WT) y de la edad sobre el WPP. El cálculo de la Regresión Lineal Múltiple (RLM) indica que el modelo de explicación del comportamiento de la métrica WPP considerando las dos métricas es significativo [$F = 13,56$, $p < 0,001$]. Los valores de los Coeficientes Beta para la edad [$Beta_{edad} = 0,87$, $p < 0,001$] y para el WT [$Beta_{WT} = 0,18$, $p = 0,38$] señalan que la explicación del comportamiento de la métrica WPP está relacionada exclusivamente con la edad.

De este modo sería posible afirmar que esta es la única métrica, conocida hasta ahora, para el análisis del comportamiento lingüístico de la autora Iris Murdoch que describe un comportamiento de cambio relacionado con el paso del tiempo. Pero no se puede afirmar que este sea un comportamiento que se pueda considerar como un síntoma de la EA pues no se ha detectado ninguna alteración súbita del comportamiento al final de su vida para esta métrica a través del análisis de la trayectoria.

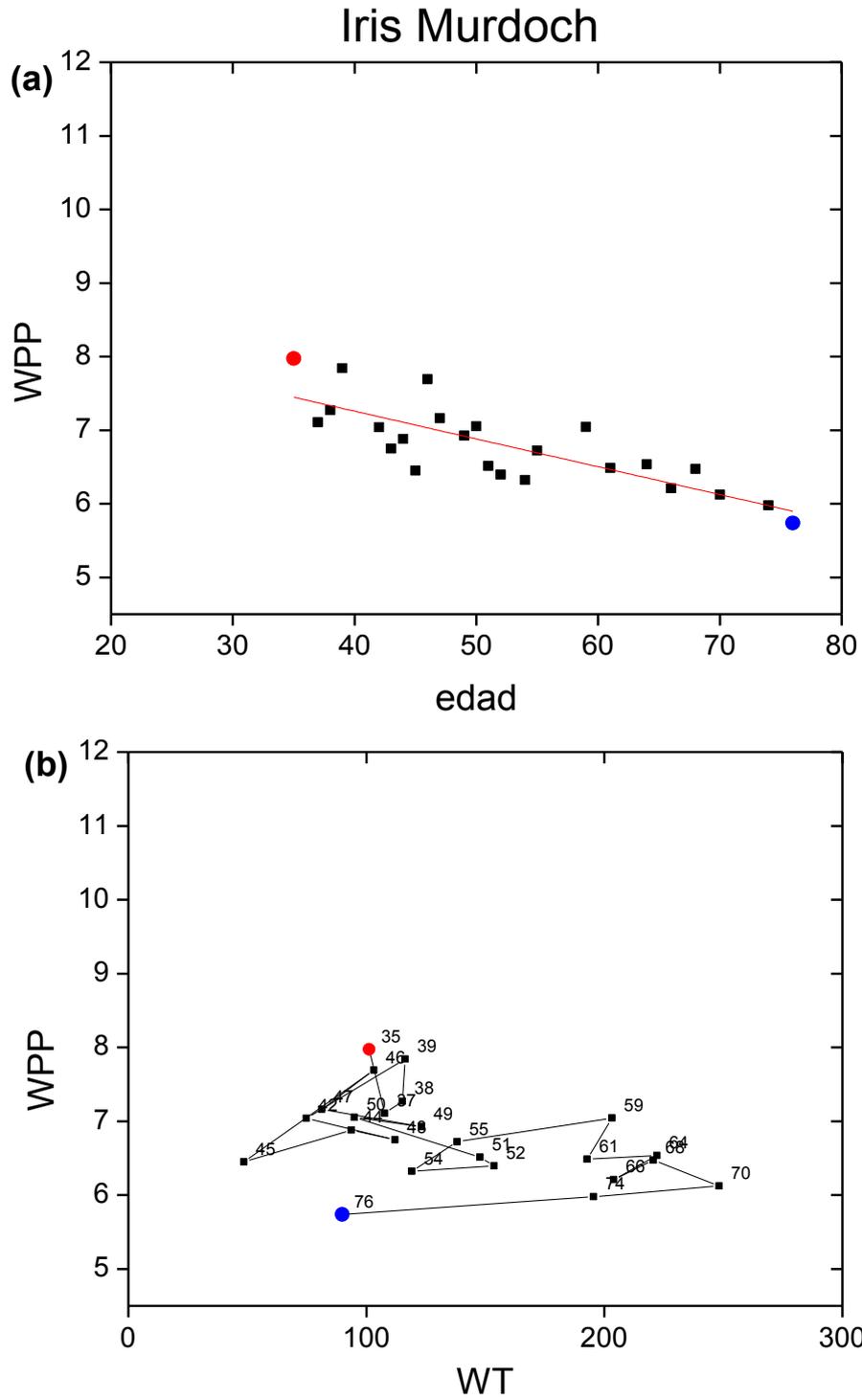


Figura 6.3 WPP representado por RLS (a) y por SPT (b) para Iris Murdoch.

6.2.4. Cantidad promedio de pausas lingüísticas por oración (PPS)

Este parámetro representa cómo el texto está fragmentado en pequeñas unidades relacionadas con las pausas. En la Regresión Lineal de la Figura 6.4a, la correlación entre los datos es estadísticamente significativa [$F(1,24) = 18.63, p=0,00$].

Pero hay que considerar cuidadosamente esta correlación, porque lo que se puede observar es una nube de puntos hasta la edad de 61 años y otra nube a partir de esta edad hasta el final de la producción de la autora a los 76 años, lo que sugiere dos etapas distintas de valores. Los cuadros azules en la Figura 6.4a representan estos rangos de valores.

Además de la comprobación visual, se ha realizado el cálculo de la Regresión Lineal Simple de los dos grupos de libros. La correlación entre los puntos de cada una de las dos nubes fue no significativas, con el primer grupo, compuesto por los libros escritos hasta la edad de 61 años, presentando [$F(1,28) = 0,02, p = 0,89$] y el segundo grupo, con el resto de los libros hasta el final de la producción de la autora, presentando [$F(1,6) = 0,02, p = 0,90$]. Estos resultados muestran que lo que ocurre es una correlación entre las dos nubes de puntos y no entre todos los puntos. Por tanto, la análisis de la Regresión Lineal entre la edad y el PPS no aporta informaciones relevantes sobre la evolución de la escritura de la autora Iris Murdoch con el paso del tiempo.

Seguendo la evolución longitudinal a través de la trayectoria en espacio de fases, en el gráfico de la Figura 6.4b, se observa que los libros escritos a los 59 y a los 61 años han cambiado de estado/fase, porque se ha producido un aumento en la longitud de los libros, dislocando los puntos hacia la derecha. No obstante, las características relacionadas con el parámetro PPS no han cambiado: los valores de PPS se mantienen en el intervalo del grupo anterior (entre 1,49 y 1,96).

A partir de los 64 años, a pesar de que la autora siga escribiendo libros de mayor longitud, el valor de PPS empieza a incrementar, alejándose de la gama de los valores anteriores y alcanzando un máximo a los 68 años para, después, comenzar a declinar; pero aun así se mantuvo en el nuevo rango de valores de los libros escritos en ese periodo (entre 2,00 y 2,30). Finalmente, en su último libro, a pesar de tener longitud

reducida, el PPS sigue dentro de la nueva gama de valores (con PPS = 2,19), alejándose de los valores para textos de longitud similar escritos en un periodo más temprano.

Iris Murdoch

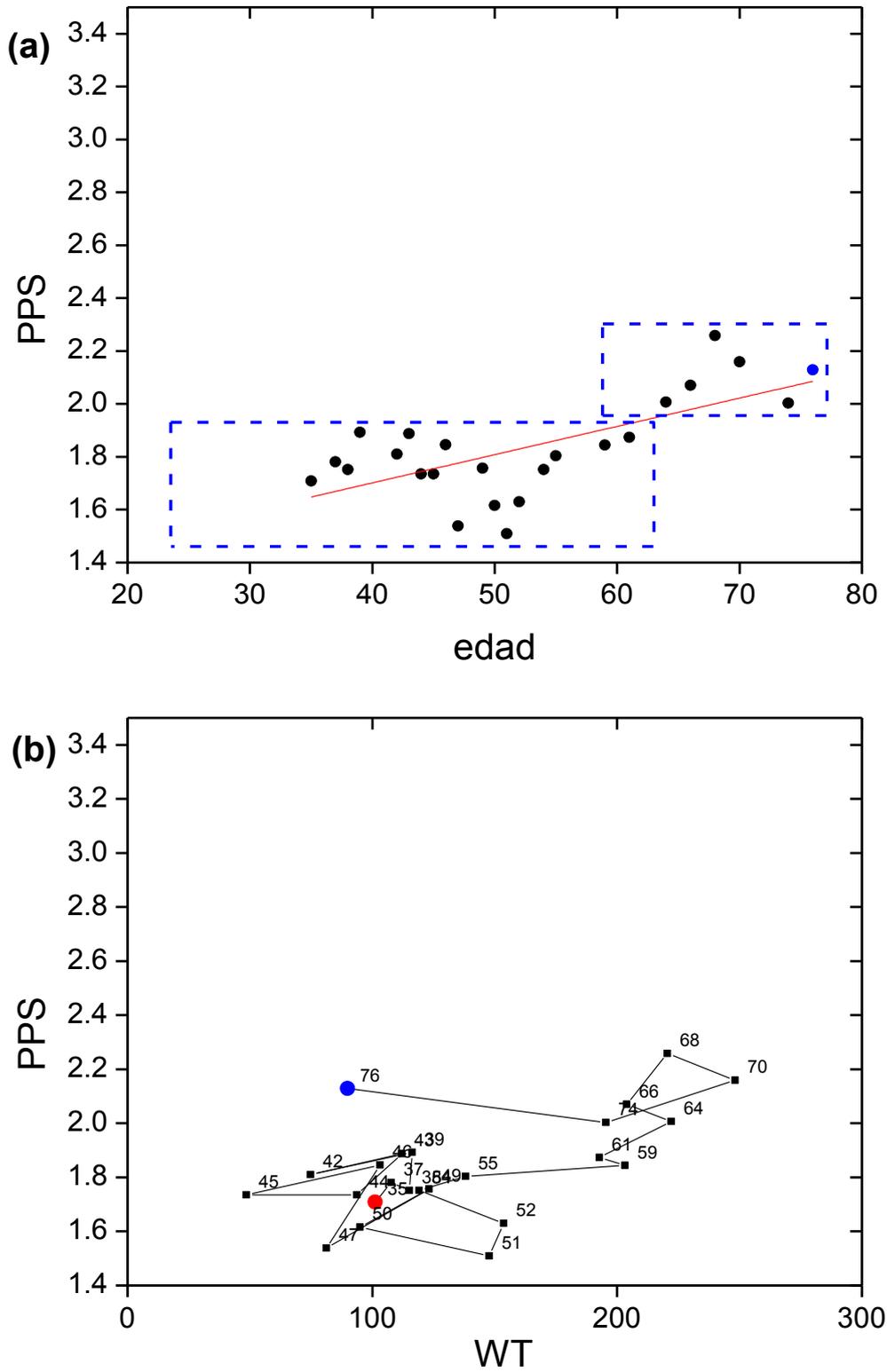


Figura 6.4 PPS representado por RLS (a) y por SPT (b) para Iris Murdoch.

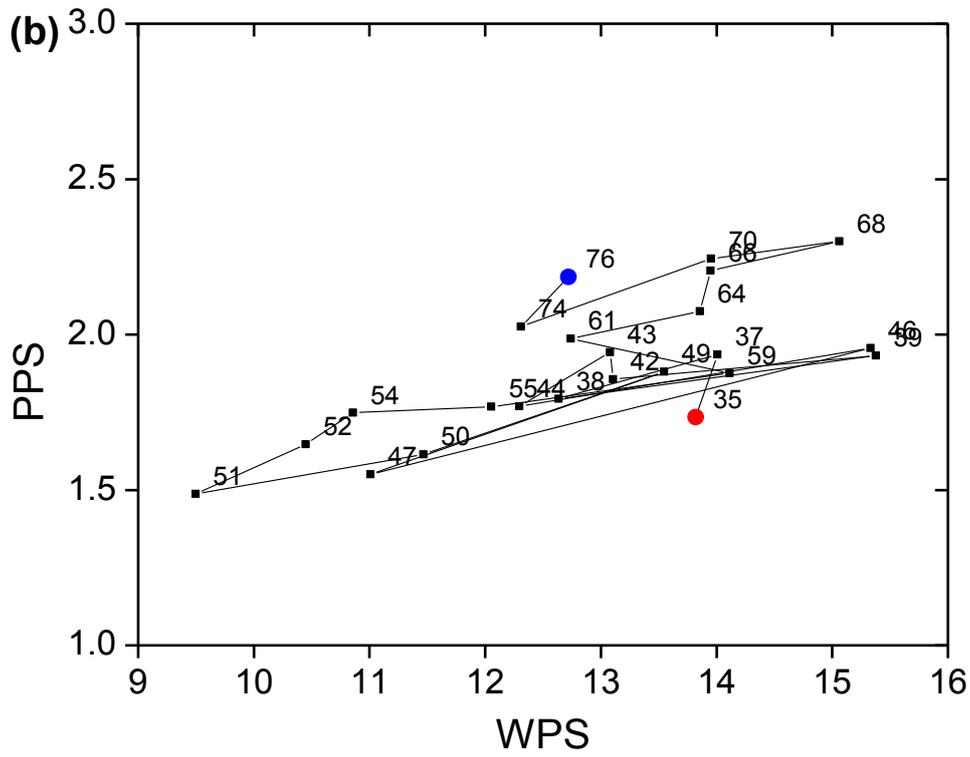
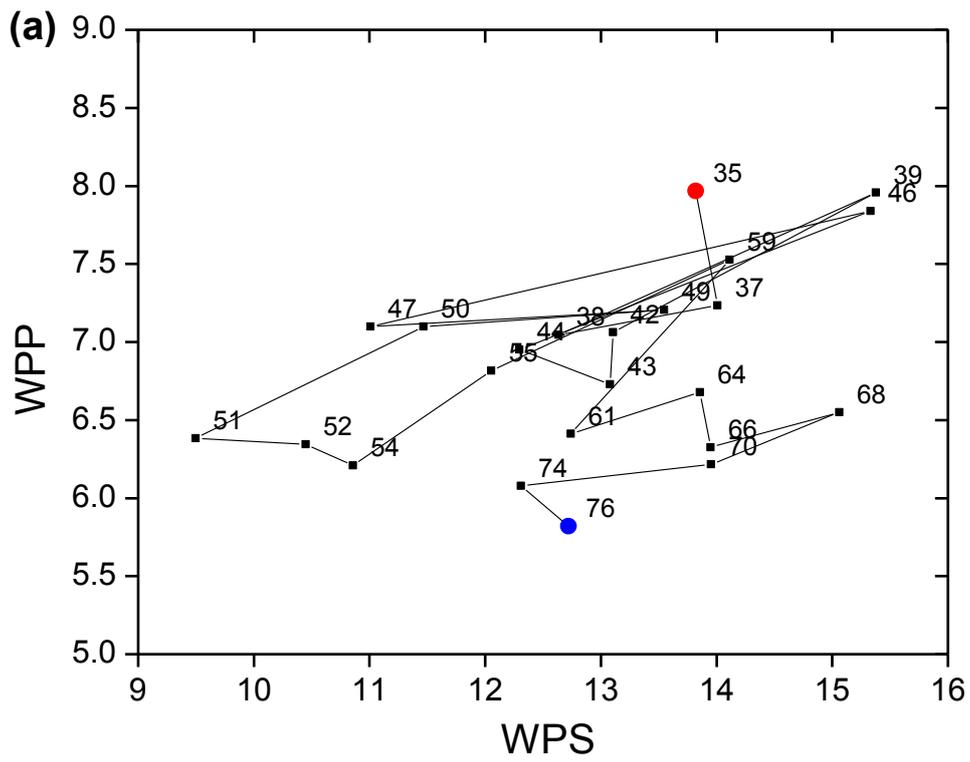
Reitero que el método de la regresión no fue capaz de identificar el cambio que se ha producido en las características de la escritura de la autora Iris Murdoch respecto a la cantidad de pausas lingüísticas utilizadas. En cambio, el análisis de la trayectoria identifica la existencia de dos estados distintos en relación con la utilización de pausas lingüísticas: uno hasta la edad de 61 años y el otro a partir de los 64 años. Es una evidencia el hecho de que la autora ha fragmentado sus textos con el paso del tiempo; sin embargo, este comportamiento quedará más claro cuando se realice la comparación de las métricas en el siguiente análisis.

6.2.5. Comparando WPS, WPP e PPS

Hay que recordar, a partir del cálculo de la Regresión Lineal descrito en el apartado 6.2.2, que la longitud promedio de las oraciones (WPS) no se relaciona de modo significativo con la edad [$F(1,24) = 0,003$, $p = 0,95$], lo que se pudo comprobar visualmente analizando el gráfico de la Figura 6.5a. Por consiguiente, los valores del WPP y PPS, cuando se relacionan con el WPS (Figura 6.5a y b) no presentan ninguna relación con la edad, pues los valores se mezclan sin presentar una diferenciación entre los calculados en edad temprana y en edad más avanzada.

No obstante, sí puede apreciarse, relacionando el WPS y el PPS, que a partir de los 64 años, dentro de las oraciones, se produce un aumento en la cantidad de pausas lingüísticas y una reducción de sus longitudes, lo que lleva a un proceso de segmentación del texto como un todo. (Figura 6.5c).

Iris Murdoch



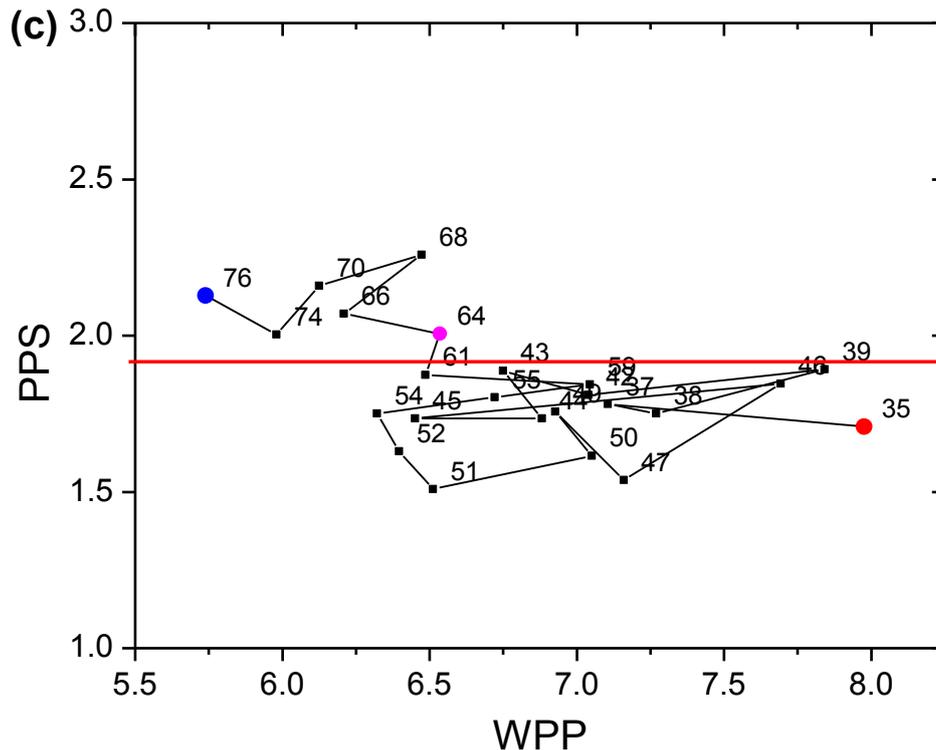


Figura 6.5 Comparación de SPT del WPS, WPP y PPS para Iris Murdoch.

La Regresión Lineal entre el WPP y WPS es positiva y estadísticamente significativa [$F = 15,17$, $p = 0.00$], lo que indica una dependencia entre las dos. Es un factor no trivial el crecimiento del PPS cuando se reduce el WPP, como se observa en el gráfico de la Figura 6.5c.

Este comportamiento permite que estos parámetros, cuando se combinan en el análisis de los libros de Iris Murdoch, se hayan transformado en un mecanismo idóneo para identificar los cambios sintácticos producidos bajo el efecto del envejecimiento. Esto se puede afirmar sí se tiene en cuenta que la edad donde los cambios de fragmentación descritos en este apartado empezaron fue la misma edad en que se produjo el incremento del vocabulario (que podría ser un síntoma precoz de la EA) a los 64 años. La identificación de esta alteración del comportamiento en la escritura de la autora Iris Murdoch ha sido posible con la representación de los datos como una trayectoria en el espacio de fases, que ha permitido observar tres variables al mismo tiempo. Por los resultados encontrados se podría afirmar que las alteraciones en el uso de la puntuación puede ser utilizado como un marcador para la detección precoz de la Enfermedad de Alzheimer.

6.3. Escritores en lengua inglesa

El objetivo de este apartado es doble: por un lado, comparar los datos de otros autores en relación con Iris Murdoch en la misma lengua de la autora; por otro, comprobar si otros escritores en la lista de elegidos para la comparación presentan comportamientos similares a los de la escritora Iris Murdoch. A continuación, se presentarán los resultados para WTIR, TTR55, WPS, WPP y PPS. Los gráficos que se representan aquí son los de los escritores en lengua inglesa cuyos comportamientos de sus libros demostraron alguna alteración con la edad. El resto de gráficos –aquellos que no mostraron diferencias significativas– se encuentran en el Anexo 7.

Aparte de este objetivo, también se comprobará si la tendencia de escribir libros de mayor longitud con el paso del tiempo es una característica de la escritura de Iris Murdoch o si existen otros escritores con un comportamiento similar al suyo.

Se han elegido nueve autores que escriben en inglés para realizar un análisis comparativo de los resultados. Son autores que no tuvieron ninguna sospecha o diagnóstico comprobado de cualquier deterioro cognitivo al final de sus vidas.

Por otra parte, no todos los conjuntos de novelas analizados aquí representan la producción del autor en la vejez, cuando los síntomas de una demencia suelen manifestarse; sin embargo, el análisis de este material es útil para verificar si el uso del lenguaje (en el caso de los escritores de lengua inglesa) es la causa de algunos de estos comportamientos identificados para la autora Iris Murdoch, lo que se podrá comprobar si se encuentran cambios similares del comportamiento en otros autores independientemente de la edad en que ocurren.

6.3.1. Longitud del texto

Como se ha mostrado, la escritora Iris Murdoch presentó una tendencia hacia la escritura de libros cada vez mayores con el paso del tiempo. De los nueve conjuntos de libros probados, se han encontrado cinco correlaciones estadísticamente significativas entre la edad y la longitud del texto utilizando la Regresión Lineal simple. En la **¡Error! No se encuentra el origen de la referencia.¡Error! No se encuentra el origen de la referencia.** se muestra la estadística ANOVA (F) y se indica también la edad en que se escribió el último libro que forma parte del conjunto analizado (*EdadI*), así como la

edad en que fallecieron (*Edad2*). A partir de este punto se mantendrán los valores de la escritora Iris Murdoch en todas las tablas (en la primera línea señaladas en rojo) para facilitar la comparación de los resultados.

Tabla 6.1 Correlación entre la edad y la longitud final de los textos en inglés.

Autor	Edad1	Edad2	F
IM	76	79	+19.63***
ACD	67	70	-
CD	57	58	-
FHB	73	75	+4.39*
HB	61	79	-72.73***
HC	72	72	-9.65**
JAi	69	76	+22.62***
JFC	61	62	-
RWC	57	68	-
WC	65	65	-4.50*

* $p < 0.05$; ** $p < 0.01$; *** $p < 0.001$

- no presenta correlaciones estadísticamente significativas.

Las novelas de los escritores Frances Hodgson Burnett y Jacob Abbott tuvieron una correlación positiva estadísticamente significativa de la longitud final del texto con la edad de la escritura, mostrando una tendencia a escribir más con el paso del tiempo.

Se observó una correlación negativa estadísticamente significativa para todas las novelas de los autores Harold Bindloss, Harry Collingwood y Wilkie Collins, lo que indica una tendencia de estos escritores hacia una reducción de la longitud de sus libros con la edad. Los escritores Arthur Conan Doyle, Charles Dickens, James Fenimore Cooper y Robert William Chambers no presentan ninguna alteración respecto a la longitud de sus libros con la edad. Esta diversidad de resultados indicaría que la tendencia detectada en la escritura de Iris Murdoch sería más un comportamiento individual y no una tendencia en el uso de la lengua, como era de esperar.

6.3.2. TTR55

En la **¡Error! No se encuentra el origen de la referencia.** se presenta un resumen estadístico del parámetro TTR55 para los autores de lengua inglesa. El número de libros de cada autor en este resumen es menor que lo indicado en el listado del muestreo (Tabla 3.1), porque para este análisis se descartan aquellos libros que tienen longitud inferior a 55.000 palabras (*vid.* capítulo 3).

Tabla 6.2 Resumen estadístico del TTR55 de los libros de los diez autores en lengua inglesa.

Autor	N	Mean	SD	Min	Max
IM	23	0,0866	0,0066	0,0730	0,1005
ACD	24	0,0915	0,0092	0,0763	0,1066
CD	16	0,0932	0,0068	0,0787	0,1097
FHB	15	0,0728	0,0106	0,0483	0,0876
HB	44	0,0721	0,0085	0,0518	0,0938
HC	32	0,0861	0,0055	0,0758	0,0960
JAi	14	0,0699	0,0054	0,0605	0,0784
JFC	33	0,0859	0,0082	0,0712	0,1027
RWC	21	0,0966	0,0051	0,0883	0,1053
WC	25	0,0772	0,0106	0,0630	0,1051

El valor del TTR55 para todos los autores varía en el rango de 0,0483 a 0,1097 (señalados en azul en la **¡Error! No se encuentra el origen de la referencia.**). Los autores que presentaron libros con menor riqueza de vocabulario fueron Frances Hodgson Burnett, Harold Bindloss, Jacob Abbott y Wilkie Collins, cuyos valores están por debajo de la media global calculada mediante el análisis de todos los textos, que es de 0,0830. Hay que hacer notar que los valores son menores que los encontrados para la autora Iris Murdoch.

Aplicando el método de trayectoria en el espacio de fase (SPT) para la relación entre el TTR55 y la longitud de los textos para los nueve autores de lengua inglesa,

observamos que solamente el autor Wilkie Collins (Figura 6.6) presentó una trayectoria con un cambio de comportamiento en el parámetro TTR55 respecto a la edad.

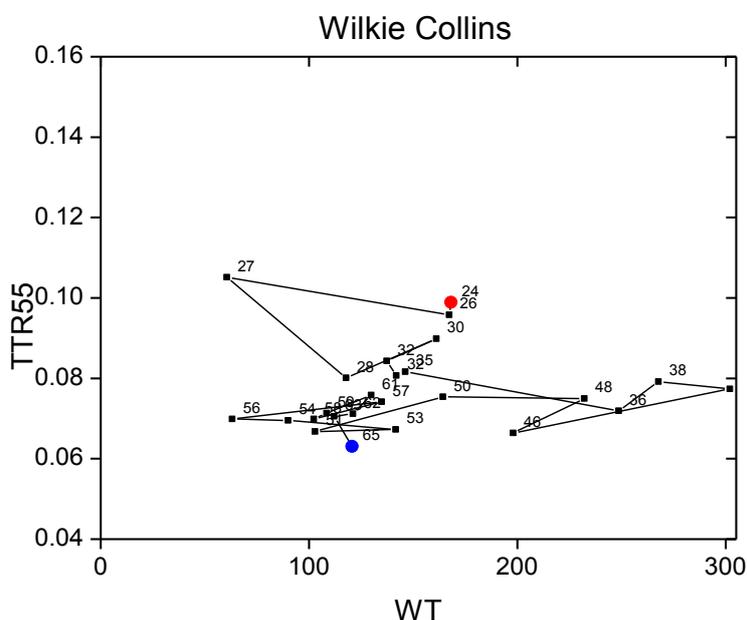


Figura 6.6 SPT en TTR55 y WT para Wilkie Collins.

Independientemente de la longitud de los textos, el TTR55 presenta un cambio de estado que indica una reducción del vocabulario utilizado por el autor a lo largo de su vida, aunque esta reducción puede ser muy sutil. El punto rojo (primer libro escrito) con el mayor valor del TTR55 está en la parte de arriba del gráfico, mientras que el resto de puntos se van reduciendo con la edad hasta el punto azul (último libro escrito).

6.3.3. WTIR

El análisis WTIR, como se recordará, se refiere a la tasa de adición de nuevas palabras por cada 10.000 palabras. En su primer estudio sobre la escritora Iris Murdoch, Garrard y colaboradores (2005) identificaron visualmente un cambio en el último libro de la autora Iris Murdoch, cambio que se reproduce en la Figura 4.1 del capítulo 3 (página 77). La reducción en la cantidad de palabras nuevas la asociaron los autores al deterioro cognitivo de la escritora. Esto fue, precisamente, lo que, sucedió en su último libro, *Jackson's Dilemma*, pues recibió críticas negativas de los expertos, aunque se sospechaba ya que sufría la EA (Garrard, 2005). Le y colaboradores (2011) confirmaron

esta teoría al afirmar la evidencia de cambio en el comportamiento de la escritura en el último libro.

La última novela de Murdoch, *Jackson's Dilemma*, posee una tasa inusualmente baja de crecimiento del vocabulario en comparación con sus obras anteriores, pues todas ellas (con la excepción de *The Philosopher's Pupil*) se aglomeran en un mismo rango de crecimiento. Esto se confirma con los resultados de TTR de los trabajos anteriores (Garrard et al., 2005; Le et al., 2011) según los cuales el descenso en el vocabulario de Murdoch se produjo abruptamente.

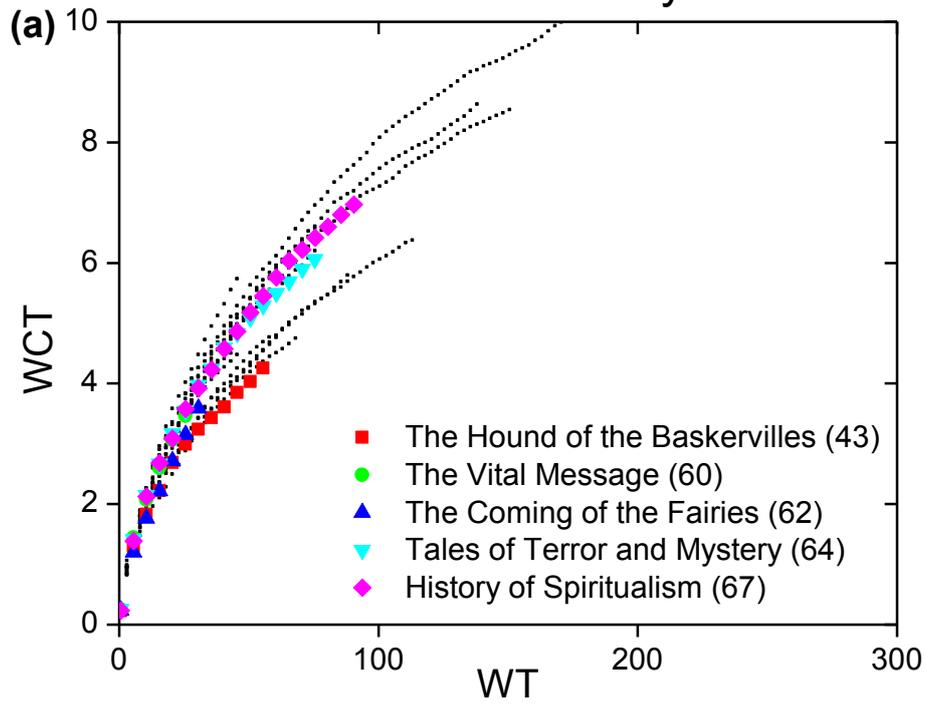
En cuanto a los libros de otros autores en lengua inglesa, se representa en la Figura 6.7 la tasa de incorporación de nuevas palabras.

Cabe señalar que en los gráficos de otros autores también se ha detectado algún libro con un comportamiento en el parámetro WTIR que indica una disminución abrupta en la tasa de adición de nuevas palabras, similar al último libro escrito por Iris Murdoch. Pero, a diferencia de este, los libros eran de gran aceptación de público y crítica. Es el caso, por ejemplo, de los libros *The Hound of the Baskervilles* de Arthur Conan Doyle (b), *Secret Garden* de Frances Hodgson Burnett (d) y *Rollo's Philosophy* de Jacob Abbott (g). En estos casos, se trata de las series de mayor éxito de algunos autores.

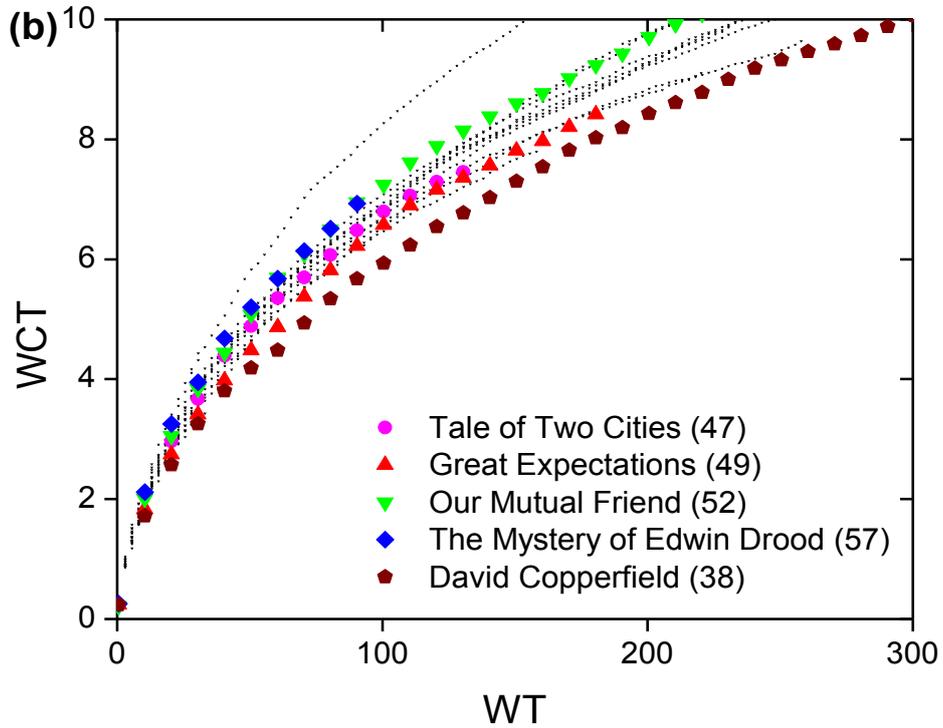
La tasa más baja de adición de nuevas palabras en los libros de Charles Dickens se obtuvo en la novela *David Copperfield*, su mayor éxito literario (c). Aunque sea uno de los más pobres en vocabulario, no demostró un comportamiento anormal, de tipo alteración abrupta, en la evolución de su escritura.

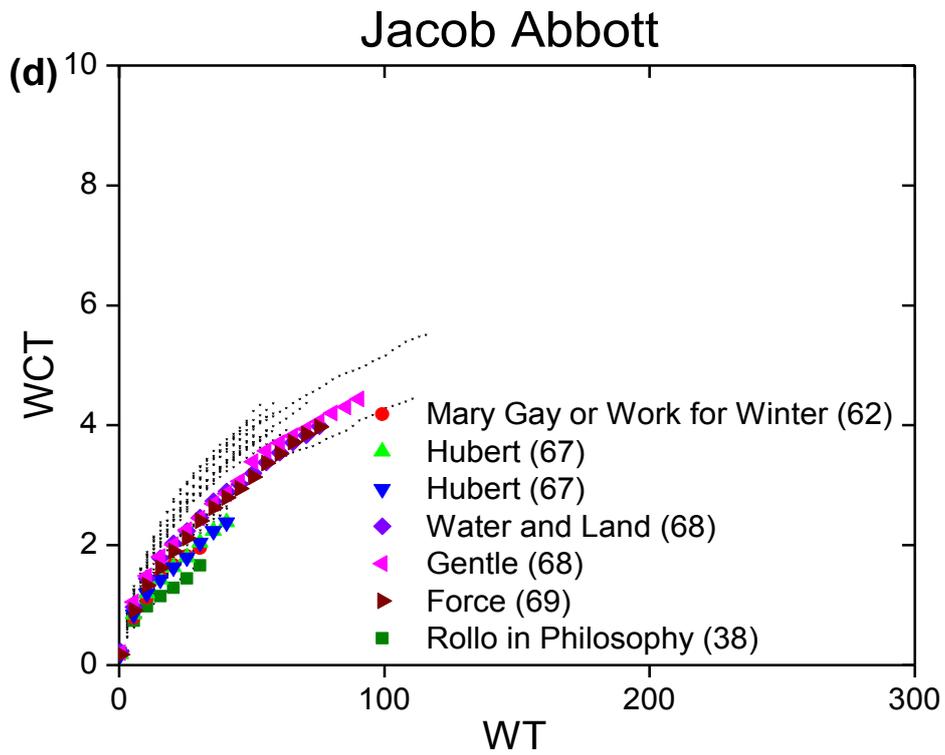
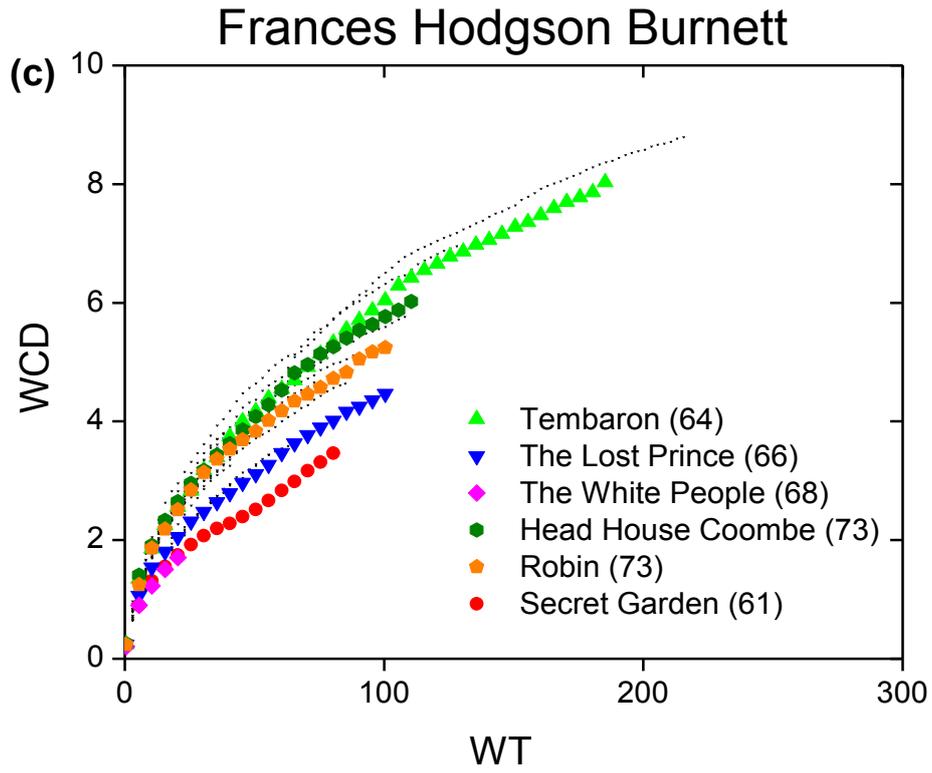
El único autor en el que se ha asociado un comportamiento diferente en la tasa de incorporación de nuevas palabras en su última novela fue Wilkie Collins. No obstante, los resultados demuestran que fue de manera muy tenue y sin evidencia de alteración importante.

Arthur Conan Doyle



Charles Dickens





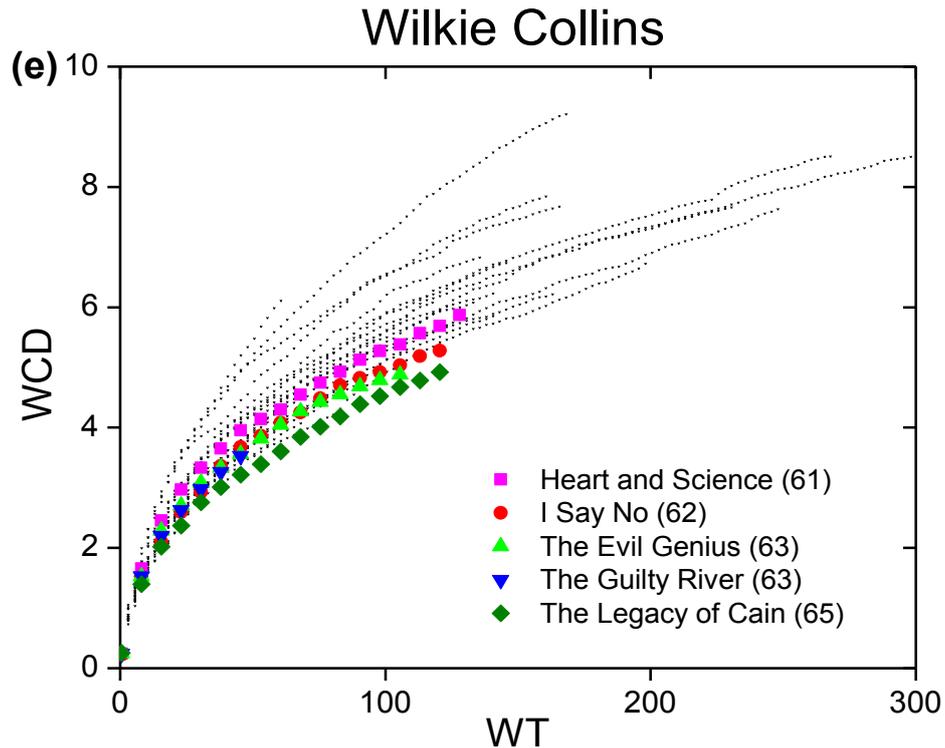


Figura 6.7 WTIR para cinco autores en lengua inglesa.

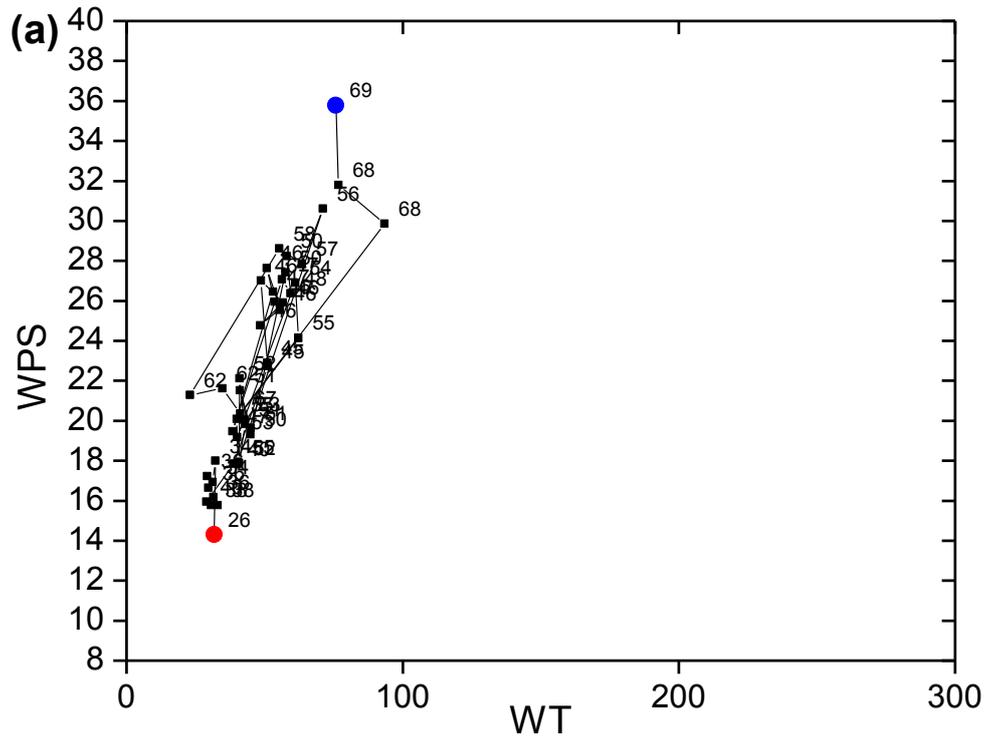
A partir de los resultados presentados, decir que la baja utilización del vocabulario en el último libro de la autora Iris Murdoch, junto con una aceptación negativa por la crítica literaria, se podría considerar como un síntoma de la enfermedad de Alzheimer. Pero esto no es del todo acertado, puesto que hay otros autores con libros que presentan una menor tasa de utilización del vocabulario y, en cambio, aparecen como bien conceptualizados por la crítica (además, se trata de obras que fueron escritas en edades tempranas).

6.3.4. Longitud promedio de las oraciones (WPS)

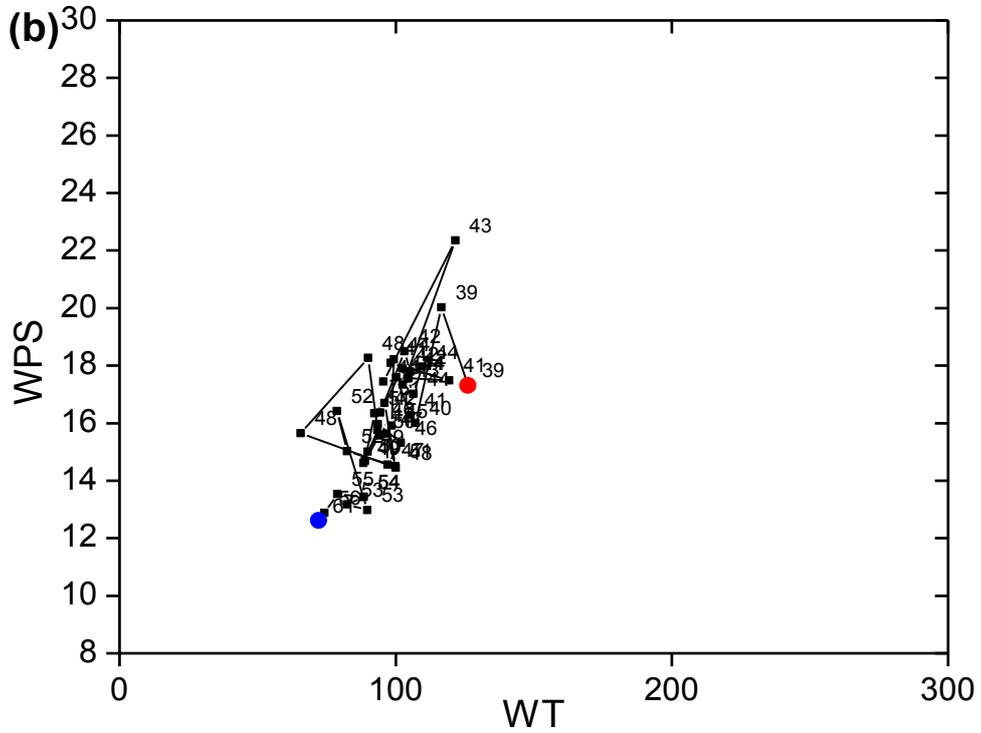
El único escritor que ha aumentado la complejidad de sus oraciones con el paso del tiempo ha sido Jacob Abbott (Figura 6.8a), pues aumentó la longitud promedio de las oraciones de aproximadamente 15 palabras a los 26 años a algo más de 35 a los 69 años. Además, es relevante tener presente que la longitud de sus textos no difiere mucho de un libro a otro.

Se produjo un cambio en la tendencia, en relación con la edad, para los autores Harold Bindloss (Figura 6.8b) y Wilkie Collins (Figura 6.8c), en los que se observa una disminución con la edad de la longitud promedio de las oraciones.

Jacob Abbott



Harold Bindloss



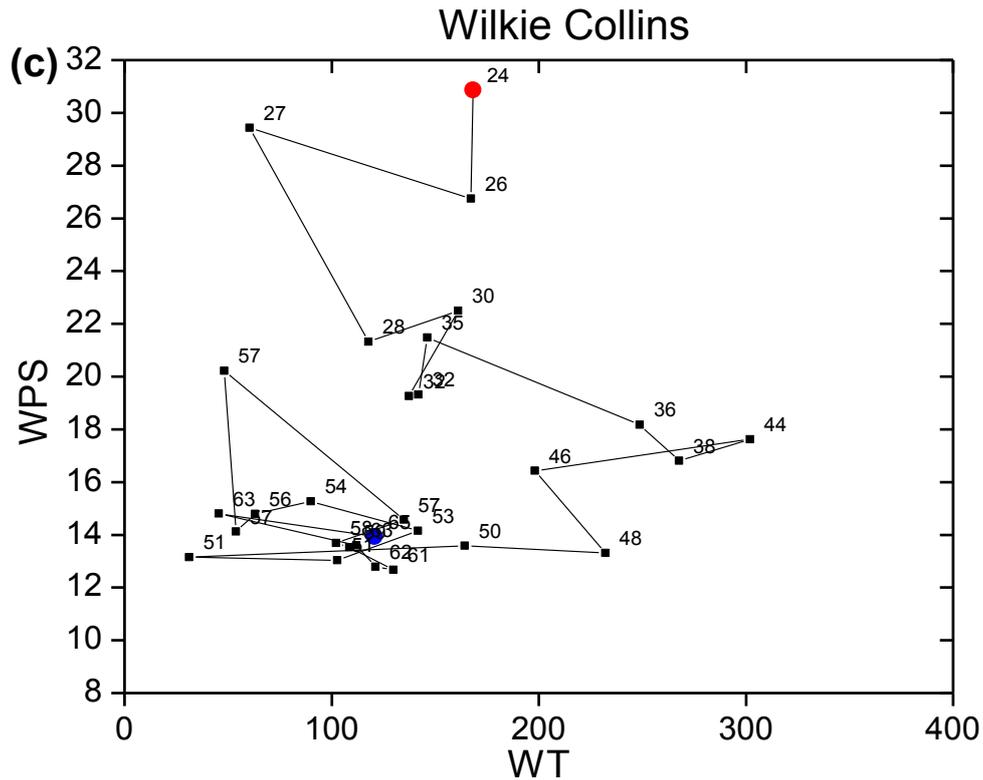


Figura 6.8 SPT del WPS para tres escritores en lengua inglesa.

En la Tabla 6.3 se muestra el resumen estadístico del WPS para los libros de los autores en lengua inglesa. El WPS ha variado desde el rango de 9,83 a 39,77 palabras por oración, tomando la media para todas las obras.

La escritora Iris Murdoch presentó los menores valores de longitud de oraciones, sugiriendo que era una característica presente en su producción literaria durante toda su vida.

Tabla 6.3. Resumen estadístico del WPS de los libros de los autores en lengua inglesa.

Autor	N	Mean	SD	Min	Max
IM	24	12,41	1,24	9,83	14,84
ACD	35	18,21	3,88	12,22	26,79
CD	23	18,92	3,59	14,44	27,85
FHB	31	16,22	2,56	12,46	23,15
HB	44	16,16	2,02	12,61	22,36

Autor	N	Mean	SD	Min	Max
HC	33	33,44	3,56	26,64	39,77
JAi	45	22,64	5,11	14,31	35,79
JFC	37	24,99	2,46	19,69	30,81
RWC	28	14,02	1,48	11,67	17,02
WC	29	17,28	5,01	12,67	30,87

Otro escritor que ha presentado un comportamiento diferente es Harry Collingwood (señalado en azul en la Tabla 6.3), ya que sus escritos tienen las mayores oraciones, con un promedio de 33,44 palabras. Para los demás escritores los valores de WPS fluctuaron en torno a la media global, que es de 19,93. El desarrollo del comportamiento en relación con esta métrica para seis de los autores analizados es compatible con los resultados encontrados para la escritora Iris Murdoch que no presentó cambios significativos en el estudio longitudinal.

6.3.5. Longitud promedio de las pausas lingüísticas (WPP)

La escritora Iris Murdoch presentó una clara tendencia a disminuir la longitud de las pausas lingüísticas a lo largo de su vida, utilizando más signos de puntuación dentro de las oraciones, fragmentándolas (ver 6.2.3).

Analizando los otros escritores de lengua inglesa, notamos que la escritora Frances Hodgson Burnett (Figura 6.9a), a diferencia de Iris Murdoch, presenta un aumento de la cantidad de palabras utilizadas en las pausas lingüísticas, pero no se aprecia ningún momento de cambio repentino en el comportamiento. Además de ella, el autor Jacob Abbott (Figura 6.9b) también aumentó la longitud de las pausas con la edad, pero es de notar que extendió la longitud de las oraciones, lo que significa que no existe una fragmentación en el texto. De este modo la escritora Iris Murdoch sigue siendo la única escritora de lengua inglesa de este estudio que presentó una relación de decrecimiento estadísticamente significativa de la longitud de las pausas lingüísticas (WPP) con la edad.

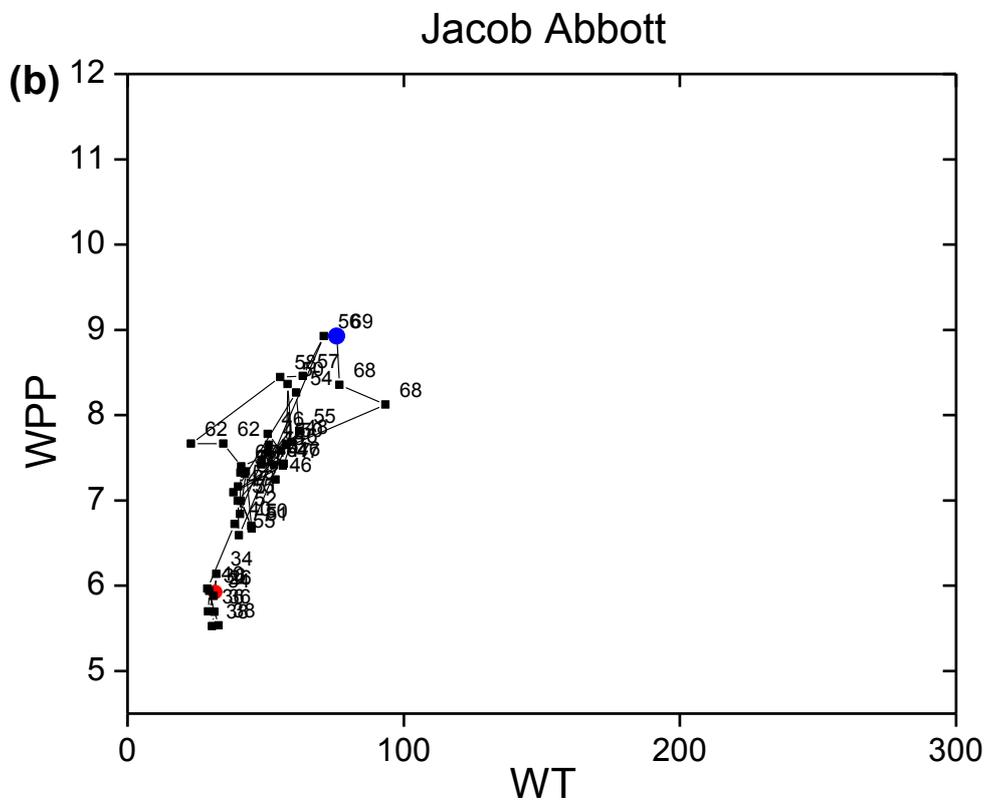
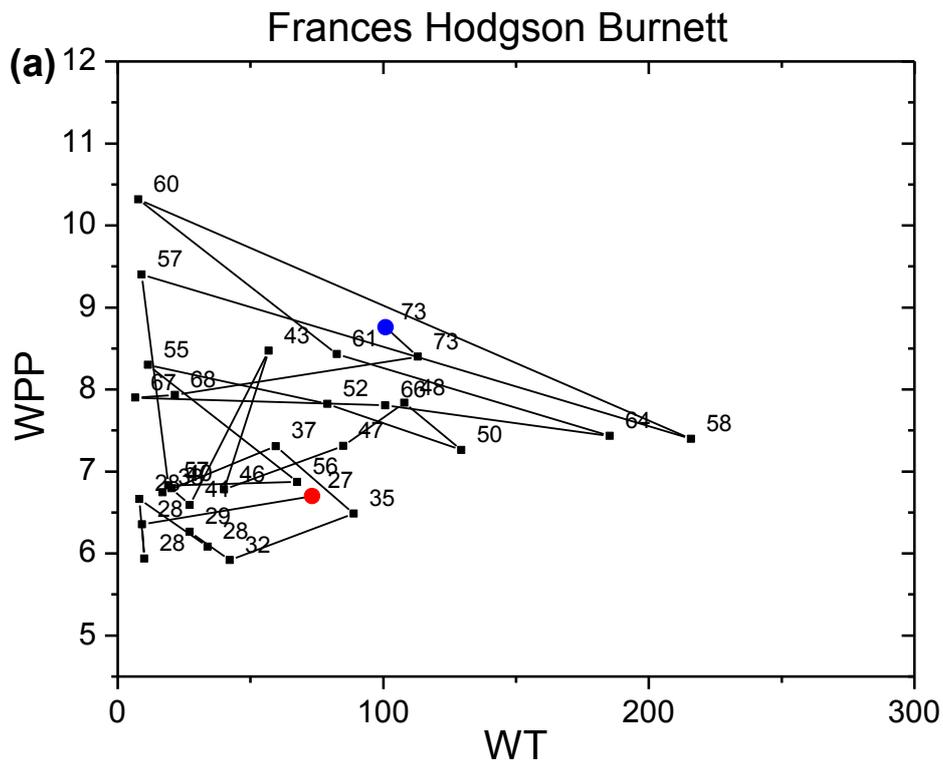


Figura 6.9 SPT del WPP para escritores en lengua inglesa.

En la Tabla 6.4 se muestra el resumen estadístico del WPP para los libros de los autores en lengua inglesa. Los valores del WPP oscilaron entre 5,07 y 10,32 palabras por pausa lingüística como promedio, teniendo en cuenta todos los autores.

Tabla 6.4. Resumen estadístico del WPP para los libros de los autores en lengua inglesa.

Autor	N	Mean	SD	Min	Max
IM	24	6,78	0,57	5,74	7,98
ACD	35	7,68	0,78	6,18	9,33
CD	23	5,93	0,61	5,07	7,43
FHB	31	7,39	1,04	5,92	10,32
HB	44	7,25	0,38	6,00	8,60
HC	33	7,95	0,28	7,13	8,48
JAi	45	7,22	0,90	5,52	8,93
JFC	37	6,93	0,46	6,07	8,29
RWC	28	5,86	0,33	5,36	6,52
WC	29	6,85	0,42	6,17	7,79

El escritor Charles Dickens presenta el menor promedio de la longitud de las pausas lingüísticas en sus libros, mientras que la escritora Frances Hodgson Burnett mostraba el mayor. El libro con el mayor longitud de pausas lingüísticas fue *The Land of the Blue Flower*, escrito a los 60 años por la escritora Frances Hodgson Burnett. El promedio de este parámetro para todos los escritores es $7,06 \pm 0,87$.

6.3.6. Cantidad promedio de pausas lingüísticas por oración (PPS)

Como vimos, para la escritora Iris Murdoch, el comportamiento relacionado con la cantidad de pausas por oraciones cambió a partir de los 64 años, con valores más altos que los presentados en el período anterior. Un comportamiento similar se observó en los libros de Jacob Abbott (a).

En la PPS

A pesar de que el valor de la cantidad de las pausas en el primer libro de Arthur Conan Doyle (Figura A7.5a) fue inferior al valor del último libro, no se puede decir que exista una evolución, pues los dos valores se enmarcan dentro del rango general de valores. En esta misma perspectiva se incluyen los resultados de los escritores Charles Dickens (Figura A7.5b), Frances Hodgson Burnett (Figura A7.5c), Harold Bindloss (Figura A7.5d), Harry Collingwood (Figura A7.5e), James Fenimore Cooper (Figura A7.5f) y Robert William Chambers (Figura A7.5g).

los gráficos muestran la trayectoria en espacio de fases del PPS y de la longitud de los textos en el tiempo para los escritores Jacob Abbott y Wilkie Collins.

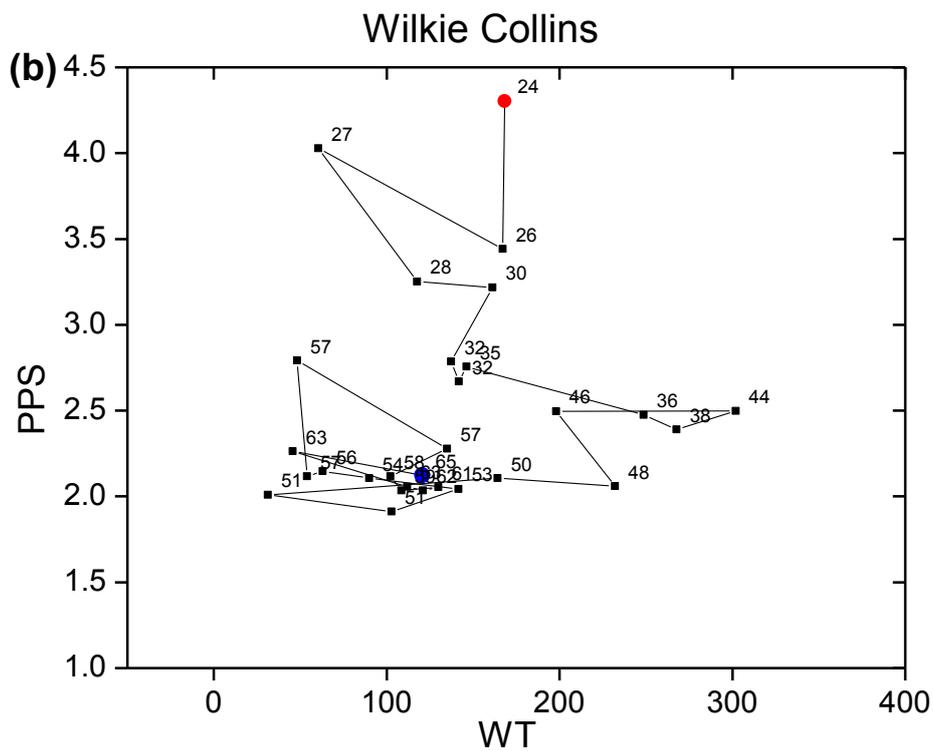
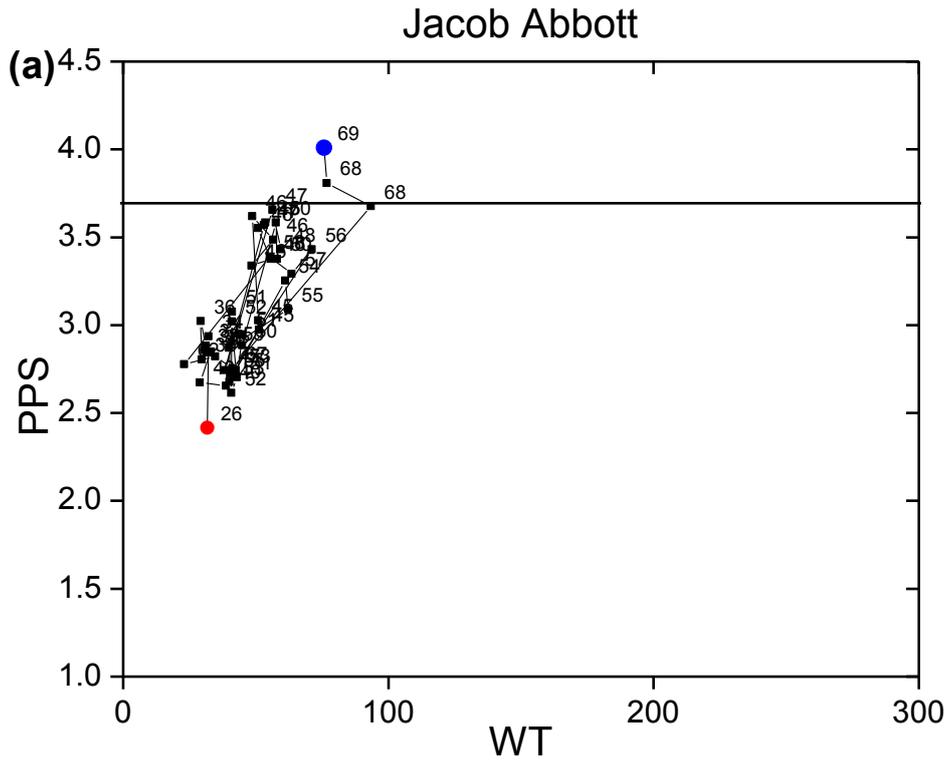


Figura 6.10 SPT del PPS para los escritores en lengua inglesa.

En los libros del escritor Wilkie Collins se observó un comportamiento inverso: la cantidad de pausas lingüísticas disminuye con el paso del tiempo, pero solo hasta los

44 años. Comenzó con 4,5 pausas por oración, decayendo hasta cerca de 2,5 a los 44 años y, posteriormente, fluctuaba entre dos y tres pausas en la mayoría de sus libros.

En la Tabla 6.5 se muestra el resumen estadístico del PPS para los libros en los escritores ingleses. Los valores del PPS oscilaron entre 1,51 y 4,91 pausas por oraciones, teniendo en cuenta todos los autores. El promedio fue de $2,80 \pm 0,78$.

Tabla 6.5. Resumen estadístico del PPS para los libros de los autores en lengua inglesa.

Autor	N	Mean	SD	Min	Max
IM	24	1,84	0,19	1,51	2,26
ACD	35	2,35	0,28	1,94	3,00
CD	23	3,18	0,36	2,64	3,96
FHB	31	2,22	0,40	1,62	3,27
HB	44	2,23	0,25	1,86	3,00
HC	33	4,20	0,36	3,56	4,91
JAi	45	3,11	0,38	2,42	4,01
JFC	37	3,61	0,29	2,76	4,09
RWC	28	2,40	0,25	1,96	2,98
WC	29	2,50	0,62	1,91	4,30

La escritura de Iris Murdoch se caracteriza por una menor cantidad promedio de pausas lingüísticas por oración y que, a lo largo de su vida se ha reducido aun más. Harry Collingwood mostró los valores más altos.

6.3.7. Comparando parámetros (WPS, WPP e PPS)

Los tres parámetros que se han utilizado en este estudio pretenden aclarar la evolución de las características cognitivas en relación con el lenguaje (precisamente sobre el uso de la puntuación) y el envejecimiento. Un aspecto particular tiene que ver con las preferencias de los escritores respecto al uso de los signos de puntuación dentro de las oraciones. Una visión general de las correlaciones entre estas métricas está en la



Tabla 6.6. Para esta comparación fueron utilizados todos los libros de los autores (*vid.* Tabla 3.1 en la página 65) y los valores no significativos están señalados en azul.

Tabla 6.6 Correlaciones entre las tres métricas: WPS, WPP y PPS.

Autor	WPS x WPP ^{*1}		WPS x PPS ^{*1}		WPP x PPS	
	F	p-value	F	p-value	F	p-value
IM	4,65	0,04	15,17	0,00	4,99 (-)	0,04
ACD	216,08	0,00	292,93	0,00	48,99 (+)	0,00
CD	42,95	0,00	57,81	0,00	4,28(+)	0,05
FHB	2,37	0,14	24,53	0,00	10,39 (-)	0,00
HB	10,99	0,00	208,05	0,00	0,11	0,74
HC	16,38	0,00	241,40	0,00	2,67	0,11
JAi	181,79	0,00	213,54	0,00	30,89 (+)	0,00
JFC	17,76	0,00	42,46	0,00	0,48	0,49
RWC	0,99	0,33	70,42	0,00	2,72	0,11
WC	24,70	0,00	837,04	0,00	12,11(+)	0,00

^{*1} Todas las correlaciones son positivas

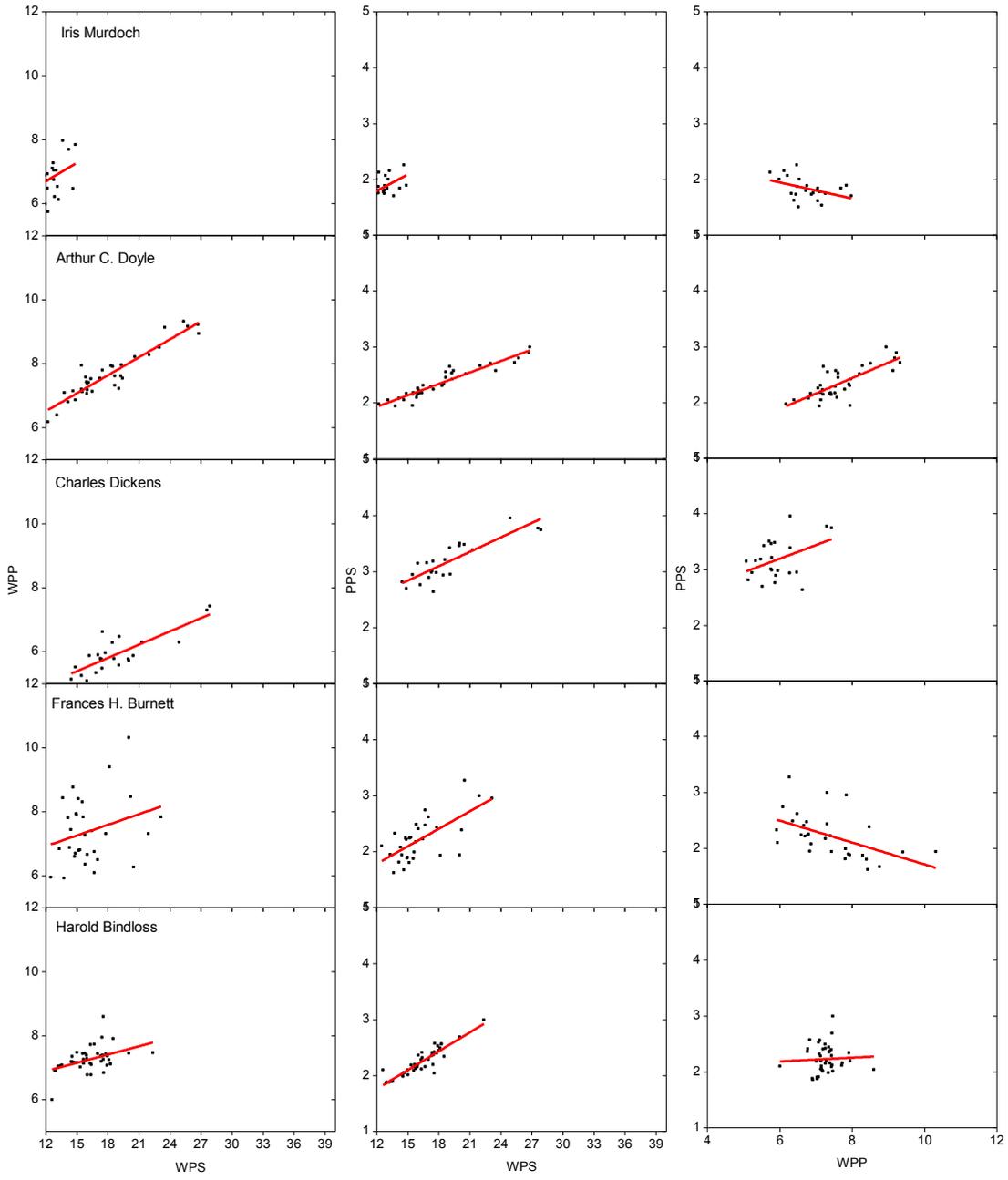
Como se ha visto anteriormente, la correlación entre la longitud promedio de las oraciones (WPS) y la longitud promedio de las pausas lingüísticas (WPP) para los libros de Iris Murdoch es positiva. Lo mismo encontramos para los libros de la mayoría de los autores, excepto para Frances Hodgson Burnett y Robert William Chambers. La interpretación para estas correlaciones es la siguiente: cuando el autor aumenta el tamaño de las oraciones (en promedio), lo hace ampliando la longitud de las pausas lingüísticas (en promedio).

Ahora bien, si se observa la correlación entre la longitud promedio de las oraciones (WPS) y la cantidad de pausas lingüísticas (PPS), se constata que el aumento de la longitud promedio de las oraciones es, por unanimidad (todas las correlaciones son fuertemente significativas), fruto del aumento en la cantidad de las pausas lingüísticas utilizadas en estas oraciones. Entonces, la interpretación conjunta de estas dos primeras estadísticas de correlación, analizadas hasta ahora, podría ser la siguiente: cuando el autor escribe oraciones más grandes, lo hace aumentando la cantidad de las pausas y estas pausas adicionales son mayores (en promedio).

Pero esto no es verdad de todo. Las dos variables (WPP y PPS) influyen en el valor de la variable WPS, pero cuando se incrementa la cantidad de pausas no siempre

estas pausas son de tamaño mayor. Es posible constatar esta afirmación mirando los valores de la última columna de la Tabla 6.6. De los diez escritores analizados los datos se distribuyen de la siguiente forma: cuatro no presentaron correlaciones estadísticamente significativas; cuatro corroboran la explicación de que cuanto más pausas tiene el libro, mayores son estas pausas; y dos escritoras no confirman esta explicación (Iris Murdoch y Frances Hodgson Burnett).

En el gráfico de la Figura 6.11 se puede mirar las inclinaciones de las rectas de las regresiones lineales entre WPS x WPP (la primera columna), entre WPS x PPS (la segunda columna) y entre el WPP x PPS. En cada línea se representa los gráficos de un escritor. Las inclinaciones de las rectas se interpretan del siguiente modo: si de izquierda-a-derecha es ascendente, significa que el incremento de la variable representada en el eje x se relaciona con el incremento de la variable representada en el eje y ; si de izquierda-a-derecha es descendente, significa que cuando se reduce el valor de la variable en el eje x , los valores de la variable en el eje y se incrementan; cuando no se observa ascendencia ni descendencia en la recta, hay que mirar los valores de la estadística F y su respectivo p -value en la Tabla 6.6 para concluir que no existe ninguna correlación entre las variables representadas en el eje x e y .



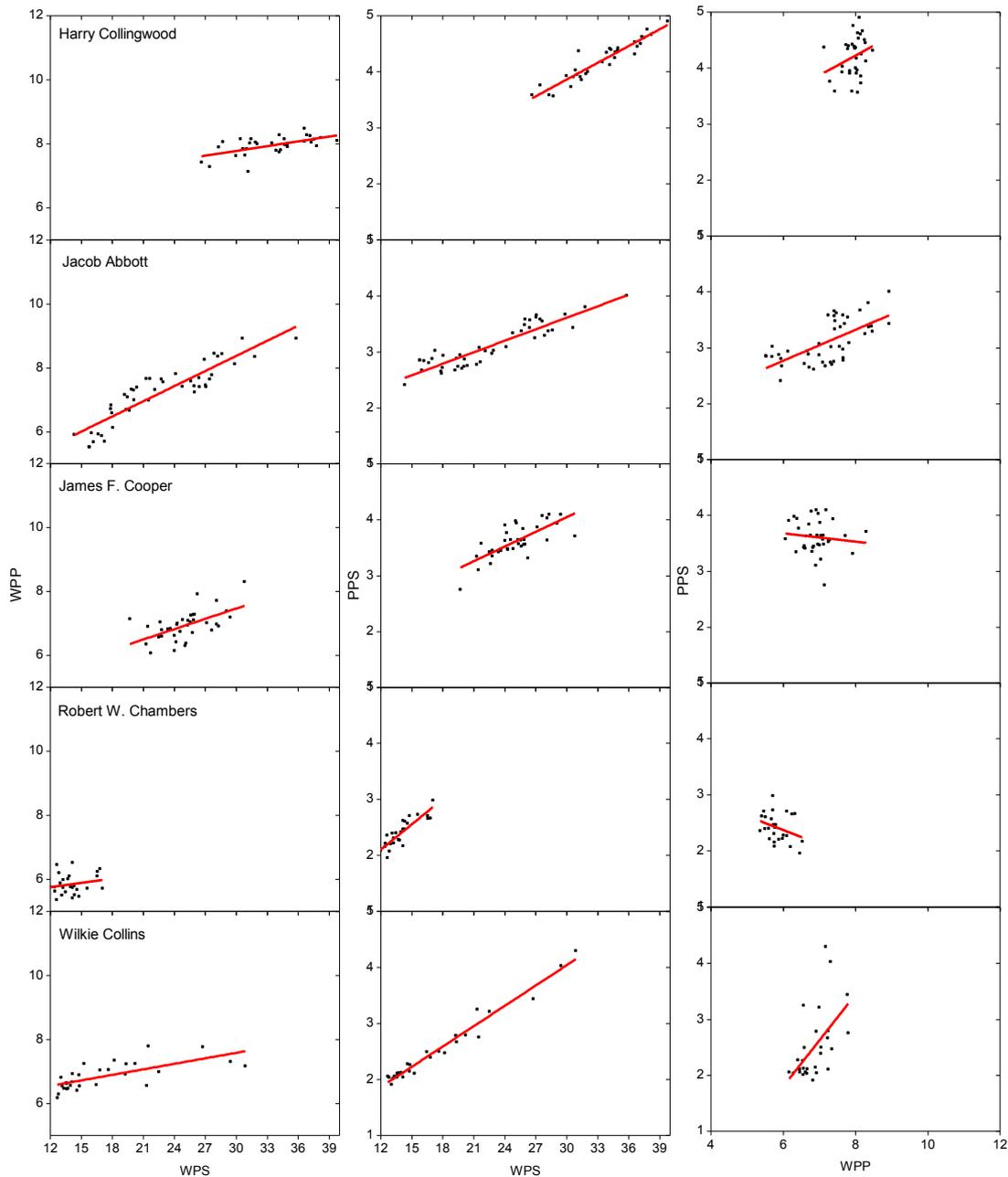


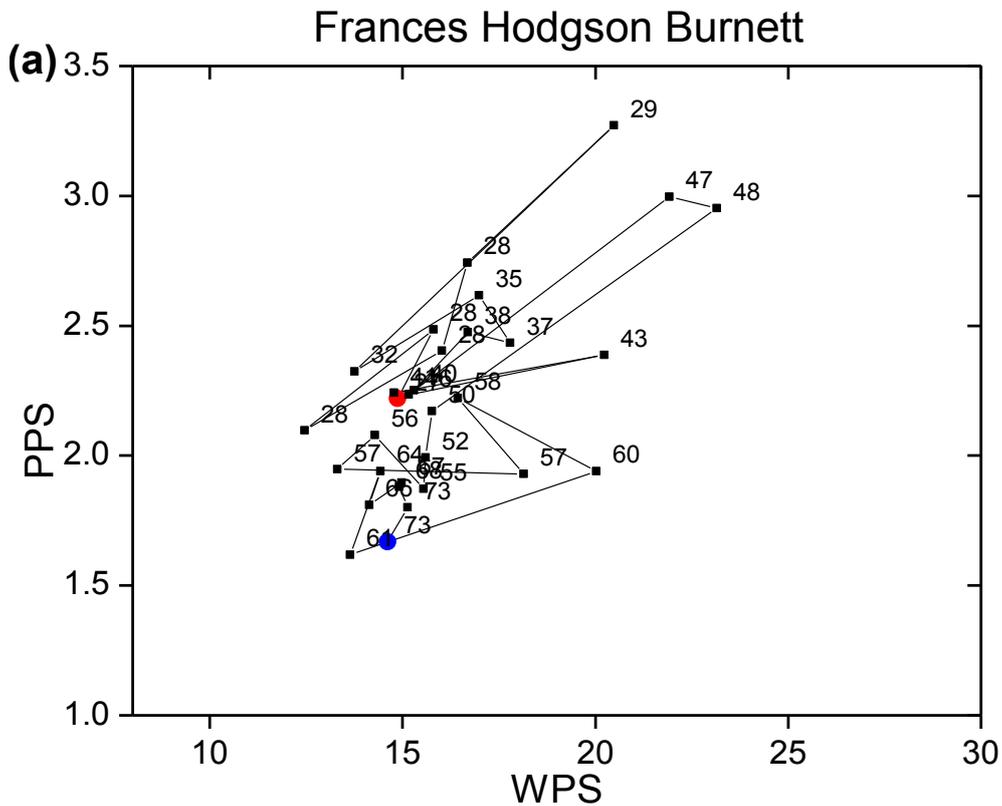
Figura 6.11 Regresión lineal de WPS x WPP (primera columna), WPS x PPS (segunda columna) y WPP x PPS (tercera columna) para todos los autores de lengua inglesa.

Estos datos, relativos a la Regresión Lineal y a la correlación, presentados hasta ahora es muy útil para entender los mecanismos de la escritura en lo que se refiere al uso de la puntuación en escritores de lengua inglesa. Pero no permite concluir mucho sobre el comportamiento de los libros de los escritores (respecto a estas métricas) y sobre el envejecimiento. La edad no está representada en ninguna de estas correlaciones. Para este análisis consideramos que la representación de la trayectoria en espacio de fases/estados es más adecuada. A continuación, se mostrarán en detalle el



comportamiento de estas métricas propuestas en este trabajo para la autora Frances Hodgson Burnett, pues fue la única que presentó una similitud con los resultados de Iris Murdoch. Los gráficos de los demás autores están en el Anexo 7, juntamente con una corta explicación de las métricas analizadas.

El análisis de los dos espacios de estados/fases podría proporcionar información sobre los cambios en la escritura de los autores. El primer espacio, como se ha comentado anteriormente, relaciona la longitud media de las oraciones (WPS) con el promedio de las pausas lingüísticas (PPS) que el autor utiliza en esas oraciones. El segundo muestra la relación entre la longitud promedio de las pausas lingüísticas (WPP) y el número medio de pausas que utilizadas en las oraciones (PPS). A continuación, se realiza la descripción de lo que sucede con estos parámetros para cada autor.



6.4. Escritores en lengua portuguesa.

A continuación, se analizarán los resultados de WTIR, TTR, WPS, WPP y PPS para los escritores en lengua portuguesa con el objetivo de comprobar si existen diferencias en estos parámetros relacionados con el idioma.

Se han elegido siete autores en lengua portuguesa para un análisis comparativo de los resultados. Son autores que no tenían un diagnóstico –o no se tenía constancia de que tuvieran un diagnóstico– relacionado con cualquier deterioro cognitivo al final de sus vidas. Por otra parte, no todas las obras analizadas aquí representan la producción del autor en la vejez (momento en que por lo general, aparecen los síntomas de enfermedades neurodegenerativas).

6.4.1. Longitud del Texto

Como ya se ha comentado, Iris Murdoch presentó una tendencia hacia una escritura de libros cada vez más grandes con el tiempo. En este trabajo, se ha comprobado si este es un comportamiento que se reproduce también para autores en lengua portuguesa.

Para esta verificación se calculó la Regresión Lineal simple usando la estadística ANOVA (F). La Tabla 6.7 indica también la edad en que escribieron el último libro que forma parte del conjunto analizado (Edad1) y la edad de su fallecimiento (Edad2).

Tabla 6.7. Correlaciones entre la edad y la longitud final de los textos en portugués.

Autor	Edad1	Edad2	F
IM	76	79	+19.63***
AA	38	55	-
CCB	61	65	7.06**
EQ	55	55	-
EV	68	70	-
JAp	72	89	17.25***
JS	87	87	-
MA	69	69	-

* p < 0.05; ** p < 0.01; *** p < 0.001

- no presenta correlaciones estadísticamente significativas.

Solo los escritores Camilo Castelo Branco y Jorge Amado presentaron una correlación estadísticamente significativa entre la edad y la longitud del texto. Mientras que en el primero se observa una disminución de la longitud de sus novelas con la edad, en el segundo se aprecia un aumento del número de palabras que utilizó en la construcción de sus novelas, un comportamiento que coincide con el de la escritora Iris Murdoch.

El escritor Jorge Amado escribió más tres novelas después del último libro incluido en este estudio para análisis: *O Sumiço da Santa*, con 438 páginas (aproximadamente 82.000 palabras), *Navegação de Cabotagem*, con 512 páginas (aproximadamente 95.000 palabras) y *A descoberta da América pelos Turcos*, con 171 páginas (aproximadamente 32.000 palabras).

Otra observación importante es que los escritores Érico Veríssimo y José Saramago, aunque no muestran una correlación significativa en el conjunto de los textos, muestran una tendencia hacia la reducción de la longitud de sus libros al final de sus carreras literarias (Figura 6.13).

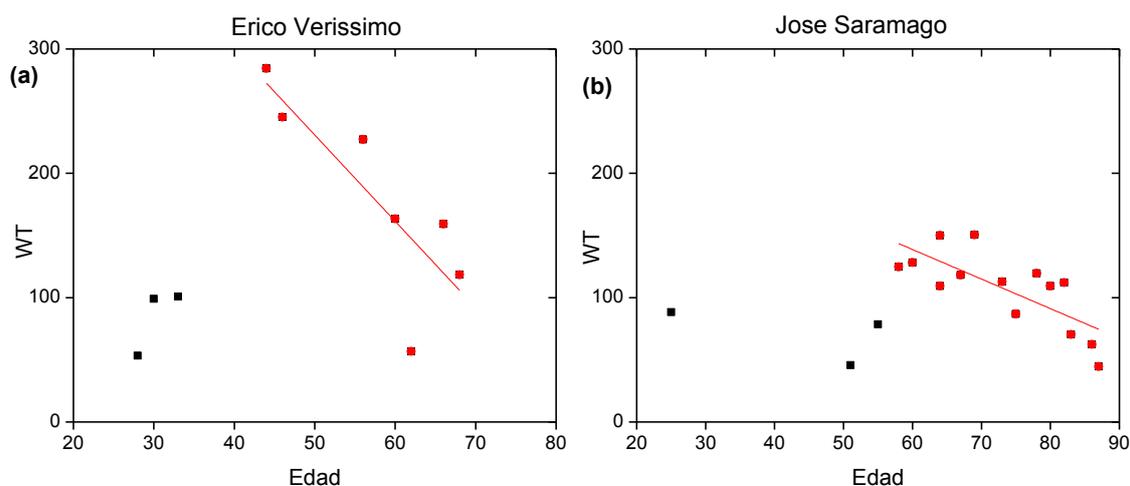


Figura 6.13 Correlación de la edad con la longitud de los últimos libros para EV (a) y JS (b).

A título ilustrativo, se ha realizado un análisis de esta métrica considerando apenas los libros finales de dos autores: Érico Veríssimo y José Saramago. En el conjunto de los libros de Érico Veríssimo, si se excluyen los tres primeros, y se realiza una correlación con el resto de los libros, se puede ver una clara tendencia a disminuir la longitud de los libros respecto a la edad (Figura 6.13a) con $[F = 10,79, p > 0,05]$. Para los libros de José Saramago, si también se excluyen los tres primeros, también aparece

una correlación negativa estadísticamente significativa (Figura 6.13b), con $[F = 9,90, p > 0,05]$.

6.4.2. Riqueza del vocabulario 1 (TTR55)

En la Tabla 6.8 se presenta un resumen estadístico de la métrica TTR55 para los libros de los siete autores de lengua portuguesa. El número de libros de cada autor analizado es menor que toda la muestra (Tabla 3.1), pues en este análisis se descartan todos los libros que poseen una longitud inferior a 55.000 palabras.

Tabla 6.8. Resumen estadístico del TTR55 de los libros de los siete autores en lengua portuguesa.

Autor	N	Mean	SD	Min	Max
IM	23	0,0866	0,0066	0,0730	0,1005
AA	9	0,0915	0,0079	0,0812	0,1052
CCB	13	0,1028	0,0112	0,0824	0,1257
EQ	7	0,1051	0,0102	0,0934	0,1175
EV	9	0,1029	0,0125	0,0858	0,1241
JAp	18	0,0910	0,0215	0,0561	0,1262
JS	15	0,0909	0,0086	0,0769	0,1026
MA	16	0,0829	0,0085	0,0770	0,0994

El valor de TTR55 para todos los autores varía desde 0,0561 a 0,1262. Jorge Amado obtuvo el valor más bajo de TTR en el recorte de 55.000 palabras estimado para el libro *Mar Morto*, escrito a los 24 años. El valor más alto se asocia también con una obra de Jorge Amado, el libro *A Tenda dos Milagres*, escrito a los 57 años. Los otros escritores mantienen el TTR55 cerca del valor promedio general de 0,0951, calculado para los 77 libros.

Los resultados de todos los escritores en lengua portuguesa difieren de los resultados para la escritora Iris Murdoch, que presentó el valor promedio por debajo de todos los valores promedio de los otros autores: su mínimo está por encima solamente del menor libro de Jorge Amado y su máximo está debajo de todos los otros.

Aunque se constate que las características de escritura de los textos en portugués son diferentes de la escritura en inglés (por lo menos, respecto a Iris Murdoch), es necesario reiterar que las comparaciones del comportamiento de las métricas sugeridas entre los textos escritos en lengua portuguesa y los textos escritos en lengua inglesa por Iris Murdoch es posible y no presenta ningún prejuicio, pues se trata de una comparación basada en las trayectorias dentro de un estudio longitudinal (no se basa, entonces, en la comparación entre los valores de las variables). Se han realizado algunas indicaciones, como la que se ha efectuado en el párrafo anterior, sobre el comportamiento de los valores de los libros en portugués en comparación con los valores de los libros de Murdoch, pero solo como una descripción de lo que podría haber sucedido, es decir, sin ninguna intención de buscar diferenciar las lenguas a través del análisis de los valores. Las comparaciones se han efectuado utilizando la trayectoria en espacio de fases/estados.

Calculando la correlación entre el TTR55 y la edad para todos los escritores de lengua portuguesa, se han encontrado correlaciones estadísticamente significativas para los libros de los autores Camilo Castelo Branco, Eça de Queirós, Érico Veríssimo y Jorge Amado. Pero, tras aplicar el método de trayectoria en el espacio de fases (SPT), se ha verificado que solamente el escritor Jorge Amado presenta un comportamiento relevante para el presente estudio, pues su comportamiento fue similar al de Iris Murdoch, con el TTR55 relacionado positivamente y de modo significativo con la longitud del texto [$F(1,18) = 5,56$ $p = 0,03$] y con la edad [$F(1,18) = 76,42$ $p < 0,001$, (Figura 6.14). Las demás trayectorias, no interesantes respecto a este trabajo, se muestran en el Anexo 7.

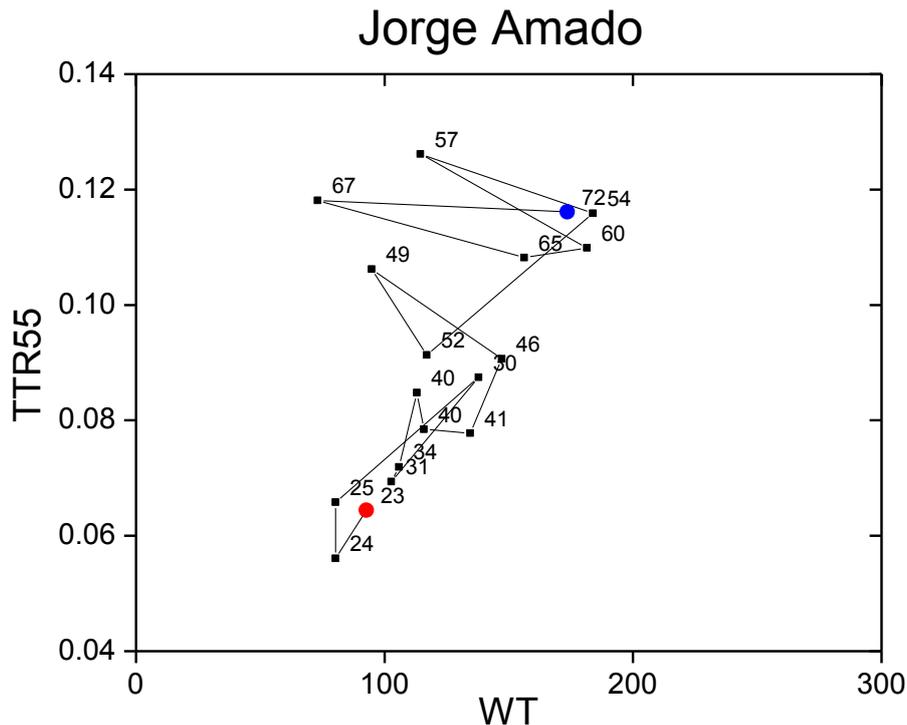


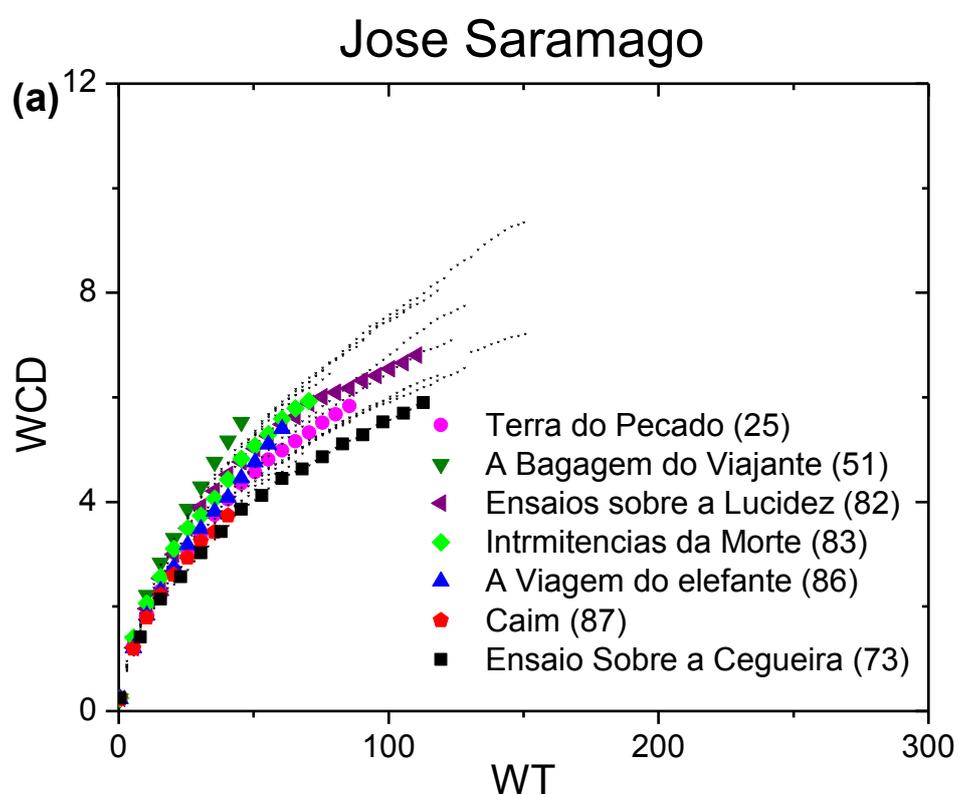
Figura 6.14 SPT en el espacio de TTR55 x WT para Jorge Amado.

Para los libros de Jorge Amado, a pesar de no contar con sus tres últimos, es evidente que presenta un gran avance en los valores de la TTR55, como si a lo largo de su carrera literaria hubiera experimentado una evolución en su escritura, haciéndose cada vez más rica en vocabulario. Cabe destacar que este autor tuvo el valor del TTR55 más bajo para todo el grupo y que este valor se asocia con una edad más temprana.

Como el TTR55 se relaciona significativamente con la edad y con la longitud del texto (WT), y entre ellos también existe una relación significativa [$F(1,18) = 6,22$ $p = 0,02$], se hace necesario evaluar cuál es la participación de cada una de las métricas sobre el TTR55 a través de la RLM. El modelo de explicación considerando las dos variables es significativo [$F = 36,07$ $p < 0,001$], pero, a diferencia del comportamiento de las obras de la escritora Iris Murdoch, la RLM ha constatado que la contribución de la edad en relación con el comportamiento de la métrica TTR55 es casi exclusiva, como se puede verificar analizando los valores del Coeficiente Beta y sus respectivos p-value, $Beta_{edad} = 0,890$ $p < 0,001$ y $Beta_{WT} = 0,037$ $p = 0,724$. Aunque el modelo sea significativo, lo mejor es considerar la contribución de la edad y despreciar la contribución de la longitud de los textos sobre el comportamiento del TTR55.

6.4.3. WTIR

Para el análisis de la tasa de incorporación de nuevas palabras en los libros de los autores en lengua portuguesa, se ha generado la Figura 6.15 de la misma forma que se hizo para los escritores en lengua inglesa. Apenas los gráficos de los libros de los escritores José Saramago y Machado de Assis están aquí representados, pues el resto no presentaron informaciones relevantes para la corroboración –o rechazo– de la hipótesis **propuesta**. Pero, como anteriormente, se reproducen todos los gráficos en el Anexo 7 para su consulta.



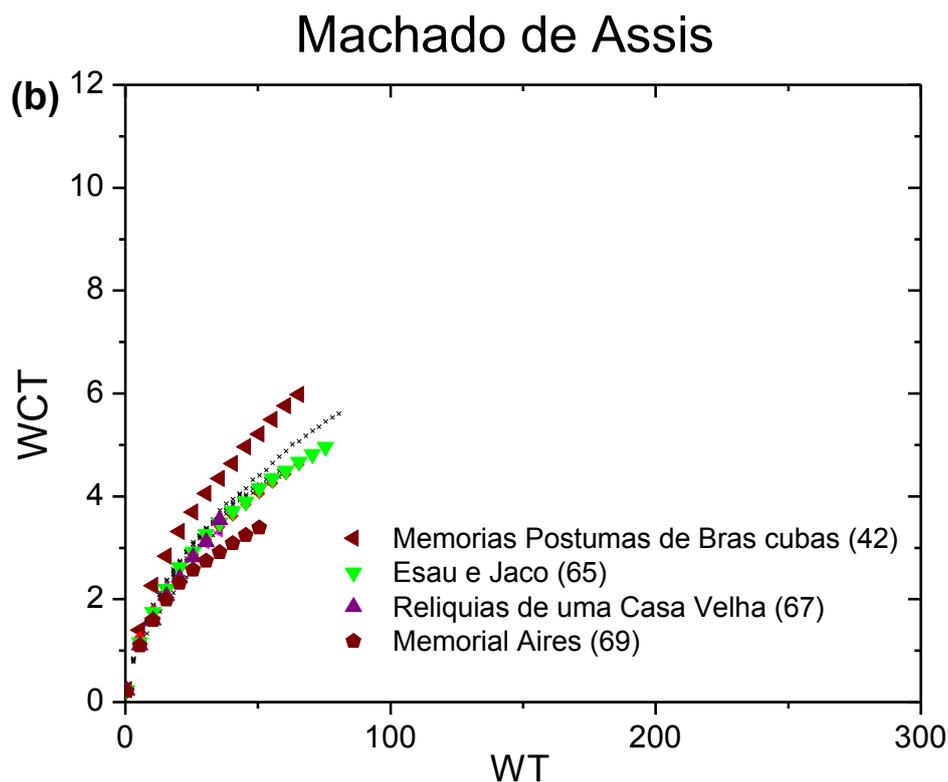


Figura 6.15 WTIR para José Saramago y Machado de Assis.

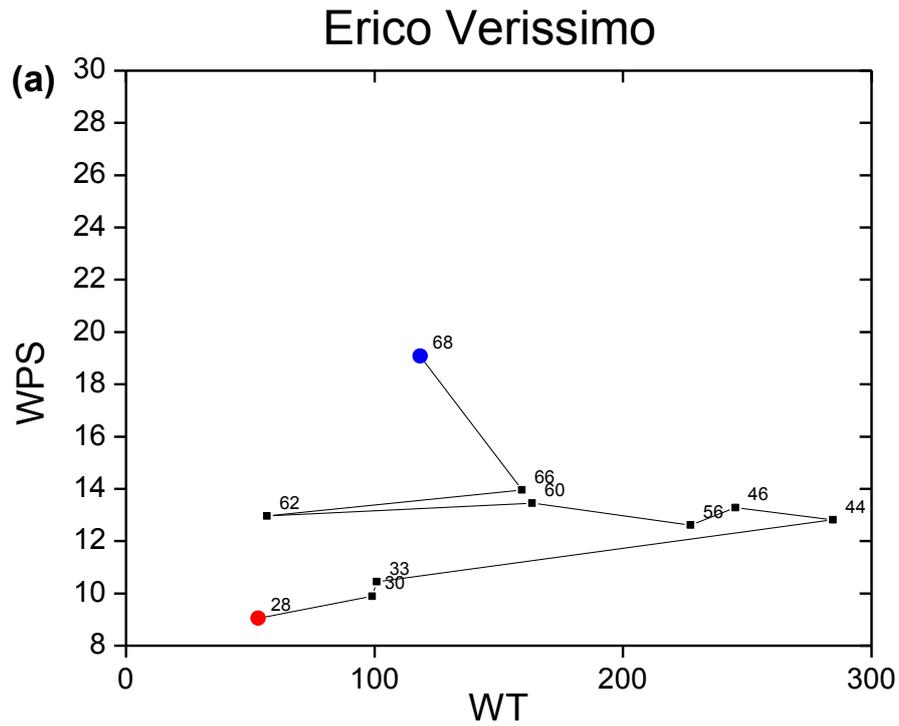
Ya se ha comentado que la escritora Iris Murdoch obtuvo la menor tasa de crecimiento de nuevas palabras en su último libro, *Jackson's Dilemma* y fue una tasa atribuida a su deterioro cognitivo asociado con el envejecimiento. Para los escritores en lengua portuguesa, solo el escritor Machado de Assis (b) demostró un comportamiento similar. Su último libro, *Memorial Aires*, se encuentra en la parte inferior de la gráfica y muestra un trazado irregular diferente de los libros anteriores. Se trata de un libro escrito a los 69 años, la edad de su fallecimiento a consecuencia de una profunda depresión por la muerte de su esposa. En este caso, no se puede concluir hablando de la existencia de un deterioro relacionado con el envejecimiento, pues el comportamiento detectado podría ser consecuencia de la depresión sufrida por el autor.

El libro más popular del escritor José Saramago (a), *Ensaio sobre a Cegueira*, se sitúa en la parte inferior de la gráfica, lo que indica que es el más pobre en vocabulario de todos los libros escritos y analizados; sin embargo, no se aprecia ninguna irregularidad o caída importante en su trazado. Otra cosa a destacar es que los libros del autor escritos al final de su carrera se distribuyen en el rango de todos los libros, y esto indica que la riqueza de su vocabulario se mantuvo regular durante todo el período de su vida. Otra vez se reafirma la inconclusión en atribuir una relación entre la mala

referencia de críticos literarios y la pobreza de vocabulario relacionados con un deterioro cognitivo.

6.4.4. Longitud promedio de las oraciones (WPS)

En la Figura 6.16 se presentan las gráficas de SPT para los autores en lengua portuguesa Érico Veríssimo, Jorge Amado y José Saramago.



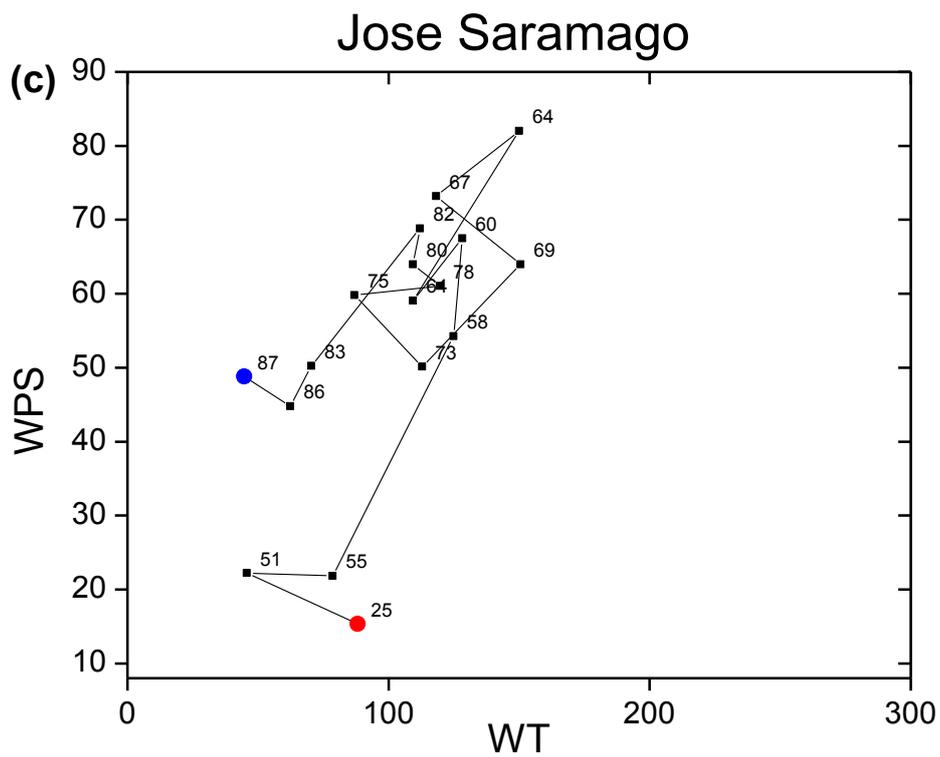
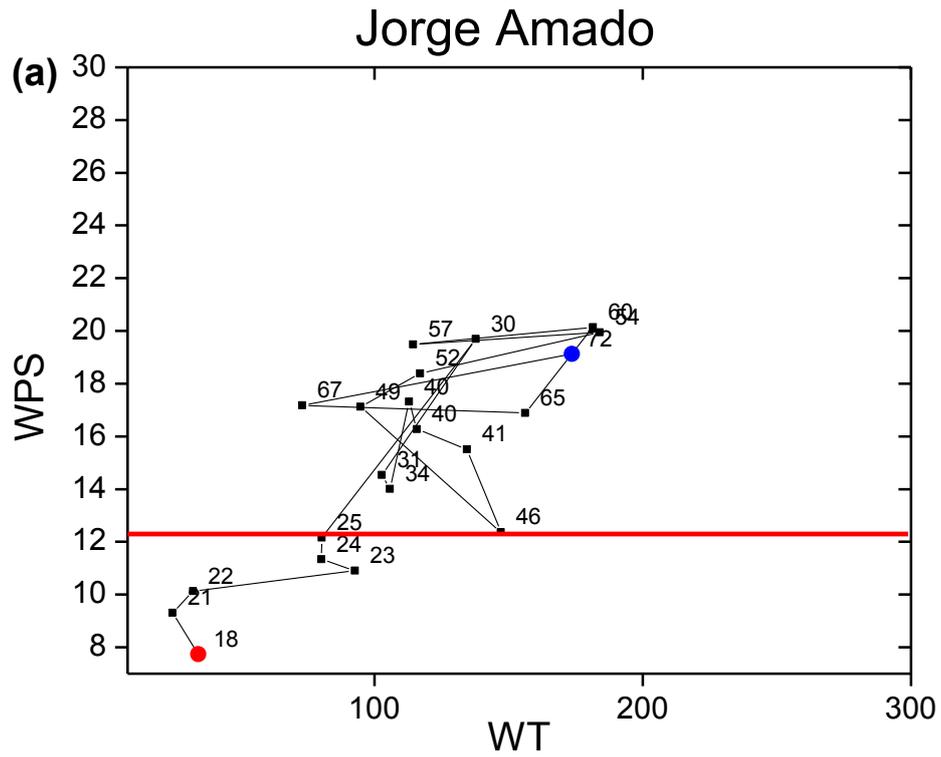


Figura 6.16 SPT de WPS para los escritores en lengua portuguesa.

Los tres autores presentaron valores de correlación significativos entre la longitud promedio de las oraciones (WPS) y la edad, $[F = 18,30 p = 0.00]$, $[F = 30,50, p$

$< 0,001]$ y $[F = 5,84 \text{ p} = 0,03]$, respectivamente³⁹. Los tres libros poseen una escala ascendente de la métrica WPS.

Para el escritor Érico Veríssimo se puede notar que la ascendencia de los valores del WPS es independiente de la longitud de los textos. Aunque escriba libros mayores o menores (ver desplazamientos de los puntos hacia la derecha y la izquierda del gráfico de la Figura 6.16a, que representan cambios de estado asociados a la longitud de los textos), el valor de la métrica WPS crece (desplazamiento ascendente de los puntos) con la edad (esto se observa en la secuencia de los números registrados cerca de cada punto, que representan las edades en que los libros fueron escritos).

Aun sobre el escritor Érico Veríssimo (Figura 6.16a), es fácil apreciar que en el último libro el escritor experimenta una subida brusca en el valor del WPS (pasa de 14 a 20 como promedio). Debido a que es un libro autobiográfico, tal vez utilice una narrativa más cercana a la oralidad y, por lo tanto, posea oraciones mayores: podría ser que el autor tenga pensamientos más espontáneos sobre su narrativa.

En los primeros libros del escritor Jorge Amado (Figura 6.16b), escritos hasta los 25 años de edad, se ha encontrado un mayor número de oraciones cortas (no excedieron el promedio de 14 palabras), pero con una clara tendencia de crecimiento. Después de los 25 años, no se observa una tendencia relacionada con la edad. Esta es una información importante para la caracterización de la escritura del escritor Jorge Amado, información que el método de correlación no puede detectar. Cuando se observa el valor de la correlación es posible entender la tendencia de crecimiento o decrecimiento de la métrica, pero no se puede concluir si esto suceda de modo regular en toda la vida del autor, como tampoco si existe una tendencia mayor o menor en una edad temprana o más avanzada. La representación de la información en un gráfico de trayectoria de fases/estados, en cambio, sí permite un análisis más detallado de, como el nombre ya indica, la trayectoria de vida del autor en relación con su escritura.

Es necesario reiterar, por la importancia que posee para la interpretación, que los tres últimos libros del autor no están incluidos en el análisis y, por lo tanto, no es posible extraer conclusiones sobre el comportamiento de su escritura asociado con el envejecimiento.

³⁹ En los cálculos que implican el WPS, WPP y el PPS, se consideran todos los libros.

José Saramago, al haber desarrollado un método propio de escritura, en relación con el uso de la puntuación, merece cierta consideración especial. Desde la novela *Levantado do Chão*, escrito a los 58 años, el escritor empieza a usar una puntuación diferente de la puntuación convencional: por ejemplo, elimina el signo de interrogación y el guión de su escritura: pasa de 463 signos de interrogación en el libro *Manual de Pintura e Caligrafía* a 11 signos en *Levantados do Chão* y, desde esta obra, avanza la disminución hasta llegar a cero en otros libros. Según el investigador Carlos Reis (*Levantado do Chão (documental)*, 2008), José Saramago reinventó la puntuación y la utilizó para crear un flujo continuo en el discurso. No obstante, es erróneo decir que Saramago no utiliza puntuación en sus libros: lo único que hace es transformar la forma de utilizarlos. Según sus propias palabras:

Era como si tuviera que contar la historia que me habían dicho. Y como usted sabe, cuando hablamos, no usamos puntuación. Tenemos pausas [de respiración, de vibración en consonantes oclusivas, de desambiguación] e, incluso, como digo en mis libros, los dos únicos signos de puntuación, el punto y la coma, no son signos de puntuación, son pausas, una pausa breve y una larga pausa. En el fondo, como suelo decir, hablar es hacer música (Saramago, entrevista al *Semanário Expresso*, 2004).

Analizando el gráfico de la Figura 6.16 (c) se identifican estos dos momentos en la escritura de José Saramago, con cambios de estado relacionados con la edad de producción de sus libros. El primer momento comienza a los 58 años, cuando empieza a escribir oraciones más largas de entre 50 a 82 palabras; el segundo se produce después del libro escrito a los 83 años, *Intermitências da Morte*, cuando se observa una caída repentina en la longitud promedio de las oraciones. No obstante, es necesario anotar dos aspectos. El primero es que son libros más pequeños que los escritos anteriormente. El segundo es que, aunque sea una caída brusca, los valores caen todavía dentro del rango de valores de los libros anteriores, con 50 palabras por oración, coincidiendo con el libro *Ensaio sobre a Cegueira*. Es esperable que se pueda obtener una claridad mayor sobre este asunto con el uso de los otros dos parámetros relativos a la puntuación, el WPP y el PPS.

Para los otros autores, el comportamiento de la métrica WPS no presentó ninguna relación significativa con la edad, al igual que lo acontecido con la autora Iris Murdoch (ver gráficos en el Anexo 7).

En la Tabla 6.9 se presenta el resumen estadístico del WPS para los libros de los autores en lengua portuguesa. Los valores del WPS oscilaron entre 7,74 y 82 palabras por oración, tomando como referencia todos los autores.

Tabla 6.9. Resumen estadístico de WPS de los libros de los autores en lengua portuguesa.

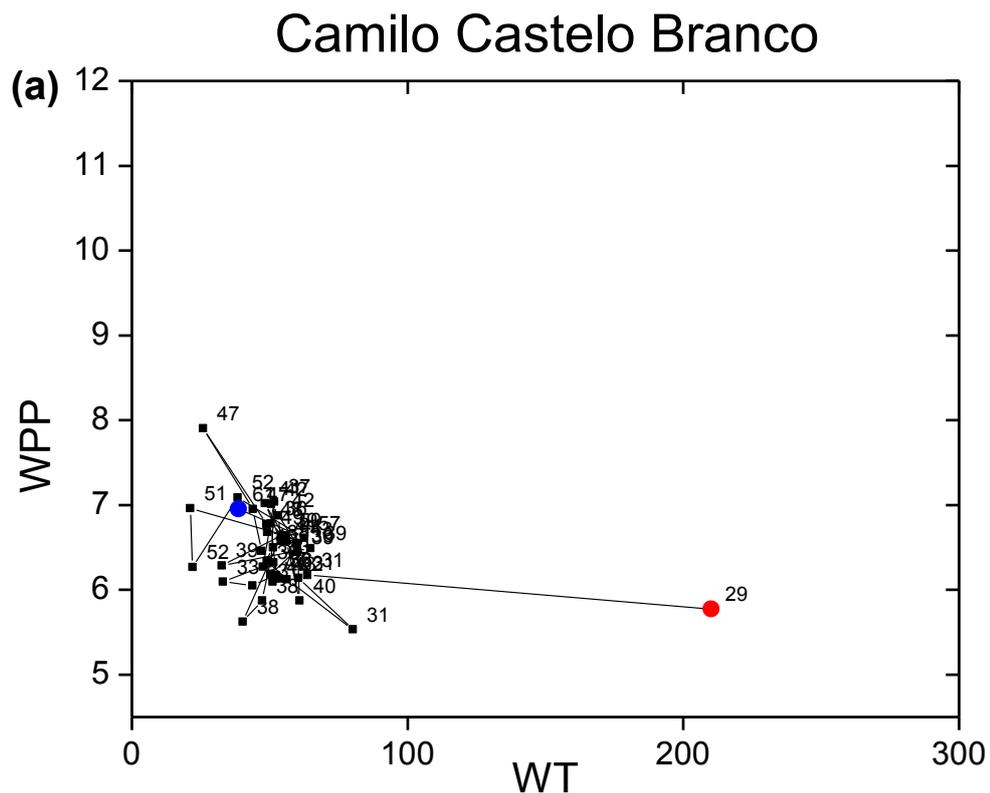
Autor	N	Mean	SD	Min	Max
IM	24	12,41	1,24	9,83	14,84
AA	11	15,47	3,14	12,26	22,62
CCB	43	16,41	2,85	12,13	24,60
EQ	8	16,12	3,36	11,32	22,58
EV	10	12,75	2,78	9,05	19,08
JAp	21	15,22	3,85	7,74	20,14
JS	17	53,35	18,60	15,32	82,00
MA	12	16,02	1,68	13,82	19,33

Salvo Érico Veríssimo y José Saramago, no existe desacuerdo entre los valores promedios de la longitud promedio de las oraciones de los otros escritores. Y estos valores están por encima de los valores presentados por la autora Iris Murdoch, tanto el valor medio como el mínimo y el máximo. Así, comparando con los escritos de Iris Murdoch, los valores medio, mínimo y máximo están por encima de los encontrados para sus libros (excepto para el valor mínimo de los libros de Jorge Amado).

6.4.5. Longitud promedio de las pausas lingüísticas (WPP)

La escritora Iris Murdoch mostró (vid. 6.2.3) una clara tendencia hacia la disminución de la longitud de las pausas lingüísticas a lo largo de su vida, con más signos de puntuación dentro de las oraciones, fragmentándolas.

El análisis del comportamiento de la métrica WPP para los libros de Camilo Castelo Branco (Figura 6.17a) es un caso típico en que el método de la trayectoria aclara los resultados. La métrica WPP presenta una correlación estadísticamente significativa con la edad [$F = 18,96$ $p < 0,001$], pero por la trayectoria se puede notar que no existe ningún marcador de alteraciones bruscas (como las que se ha producido en la trayectoria de José Saramago para el WPS en edad temprana y edad más avanzada (vid. gráfico en la página 158)).



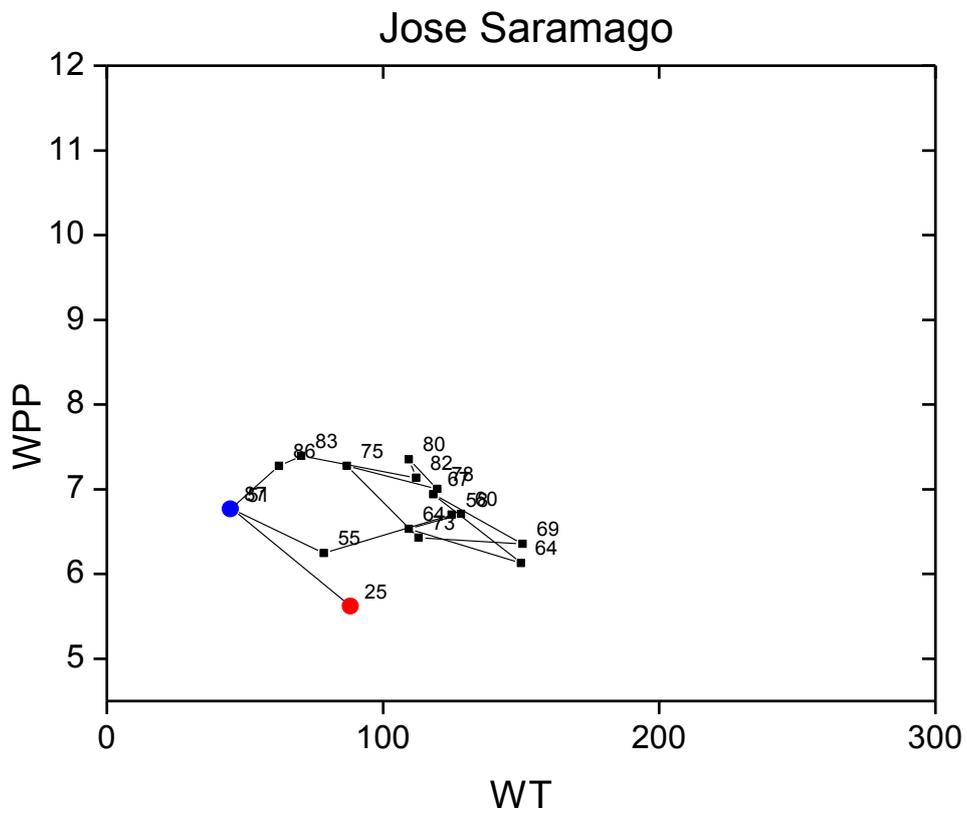
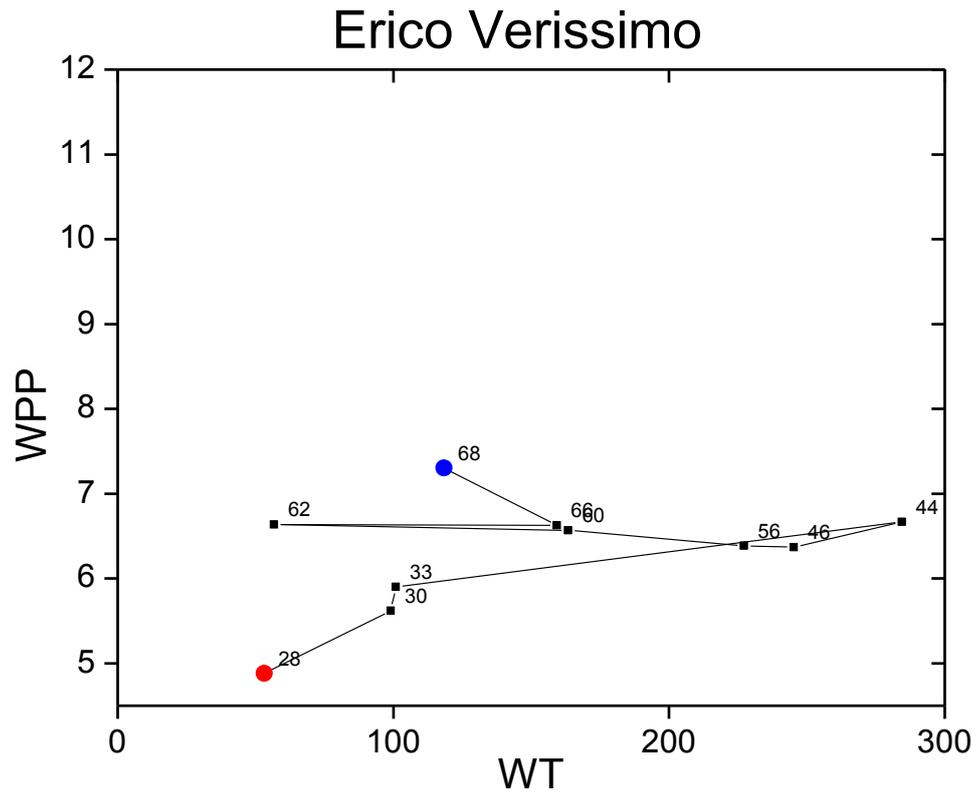


Figura 6.17 SPT del WPP para escritores en lengua portuguesa.

Para el escritor Érico Veríssimo, la correlación entre la edad y la longitud de las pausas lingüísticas (WPP) es significativa [$F = 23,63$, $p < 0,001$] y la trayectoria no sugiere ninguna alteración en el comportamiento en ninguna fase de la vida del autor. Apenas se puede notar que utilizó en su último libro pausas lingüísticas algo mayores que en sus anteriores trabajos. Pero, como también mostraba un aumento en la longitud promedio de las oraciones, es plausible decir que este aumento se asocia más con la propia escritura que con el envejecimiento.

Para el escritor José Saramago, que también presentó una correlación estadísticamente significativa entre la edad y el WPP con [$F = 23,20$, $p < 0,001$], la trayectoria no aporta informaciones relevantes.

El resto de autores no presenta correlaciones significativas entre la edad y el WPP y sus trayectorias tampoco muestran ningún cambio significativo respecto a la longitud de pausas lingüísticas y a la longitud del texto final, en relación con la edad, y que pudiera indicar una disfunción en el comportamiento. Sus gráficos se encuentran en el Anexo 7.

En la Tabla 6.10 se muestra el resumen estadístico del WPP para los libros de los autores en portugués. Los valores del WPP oscilaron entre 4,63 y 7,91 palabras por pausas lingüísticas, teniendo en cuenta todos los autores.

Tabla 6.10. Resumen estadístico del WPP de los libros de los autores en lengua portuguesa.

Autor	N	Mean	SD	Min	Max
IM	24	6.78	0.57	5.74	7.98
AA	11	6.05	0.43	5.30	6.78
CCB	43	6.45	0.45	5.54	7.91
EQ	8	5.41	0.40	4.63	5.87
EV	10	6.30	0.67	4.88	7.30
JAp	21	6.09	0.59	4.70	6.98
JS	17	6.74	0.49	5.62	7.39
MA	12	5.79	0.66	4.73	6.86

La longitud promedio de las oraciones calculada para los libros de José Saramago (vid. Tabla 6.9) es, notablemente, más elevada, ya que cambió, como ya se ha comentado, la manera de utilizar los signos de puntuación, reemplazando los puntos finales, los signos de interrogación y los signos de exclamación por comas con la intención de fusionar el estilo directo e indirecto del discurso. No obstante, el análisis de la longitud de las pausas no sugiere que hubiera cambios significativos en el comportamiento con el envejecimiento.

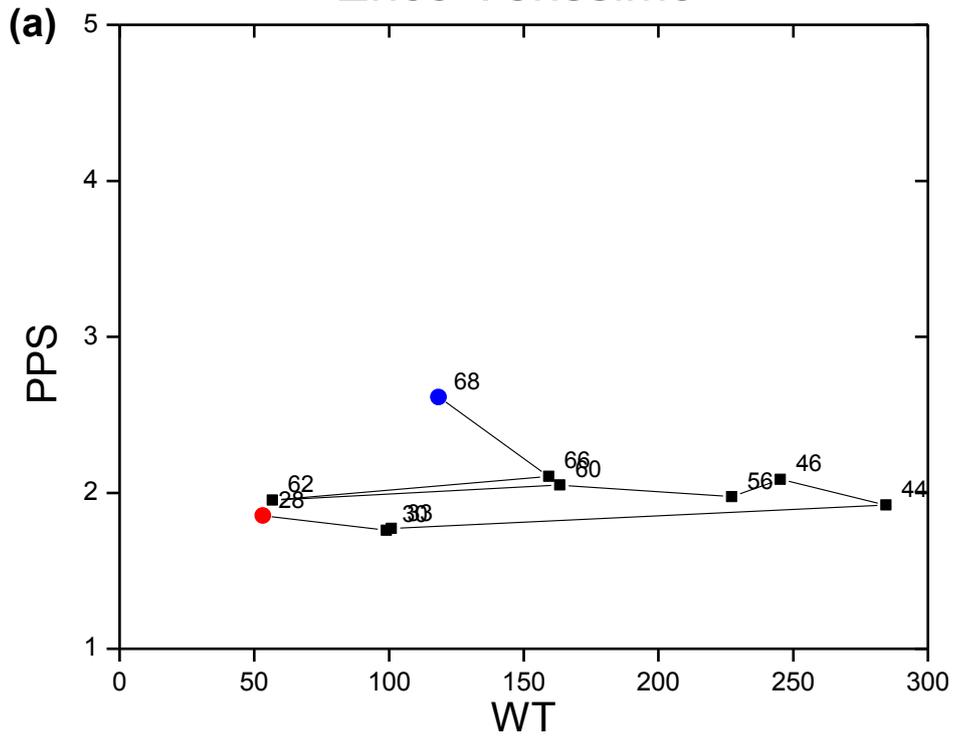
Para el resto de escritores, existen similitudes en los valores medios de este parámetro, pero todos se encuentran por debajo del valor promedio calculado para la escritora Iris Murdoch.

6.4.6. Pausas lingüísticas por Oraciones (PPS)

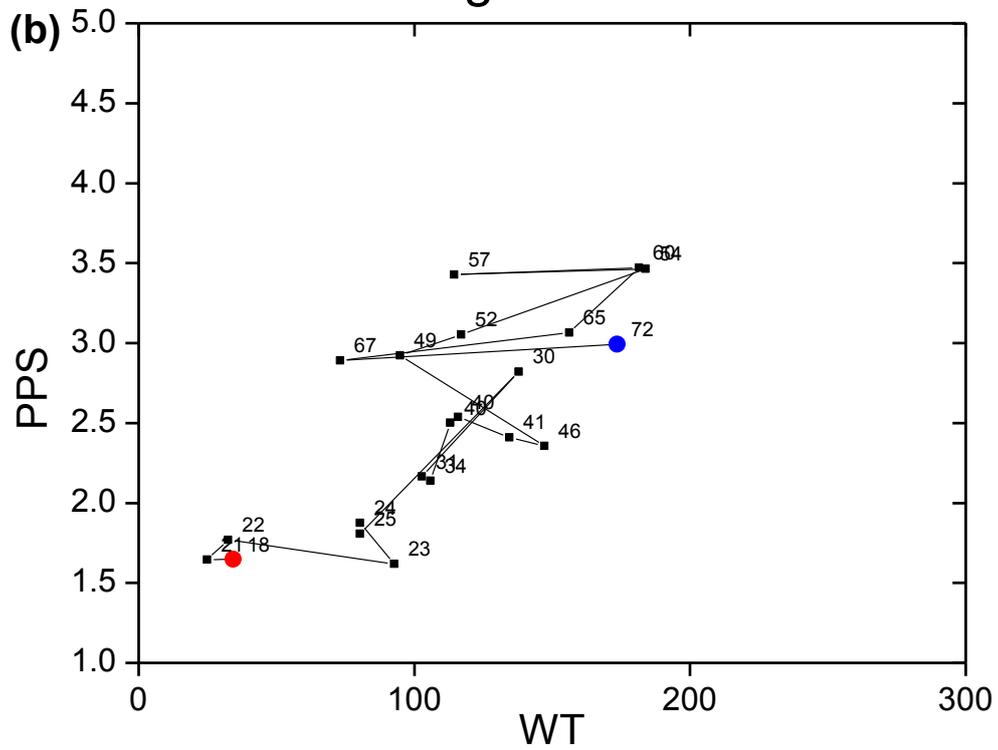
En la Figura 6.18 se presentan los gráficos de la trayectoria en espacio de fase del PPS y de la longitud del texto para los escritores Érico Veríssimo, Jorge Amado y José Saramago. El comportamiento de la escritura de la escritora Iris Murdoch (a) en relación con la cantidad promedio de pausas lingüísticas por oraciones comenzó a cambiar a partir de los 64 años, con valores más altos que los presentados en el período anterior (vid. página 124).

¿Cómo se comporta esta métrica para los escritores en lengua portuguesa en cuanto a la correlación y la trayectoria? Los libros de Érico Veríssimo y Jorge Amado presentaron una correlación positiva estadísticamente significativa entre la edad y el PPS, con $[F = 9,47 \text{ p} = 0,02]$ e $[F = 58,90, \text{ p} < 0,001]$, respectivamente. Sus trayectorias representan bien esta correlación, como se puede ver en los gráficos (a) y (b) de la Figura 6.18.

Erico Verissimo



Jorge Amado



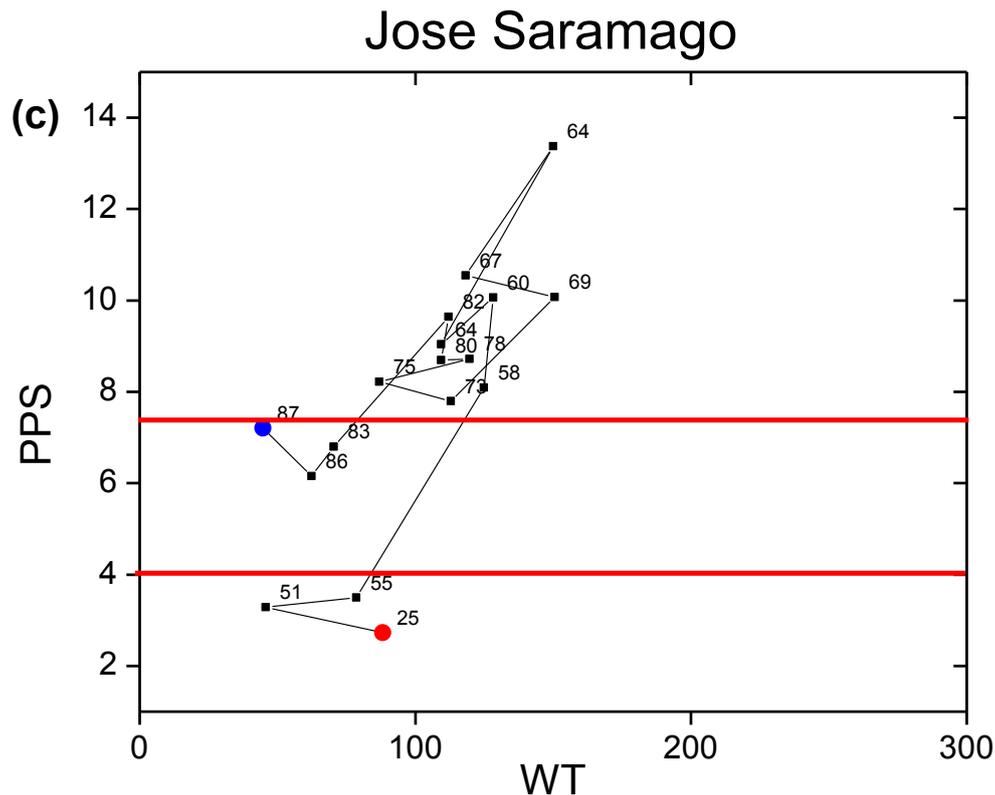


Figura 6.18 SPT del PPS para escritores en lengua portuguesa.

El gráfico de la trayectoria del PPS para los libros del escritor José Saramago presenta un trazado similar al gráfico de la trayectoria del WPS (vid. Figura 6.16). Esto era lo esperable, ya que el aumento de la longitud de las oraciones es consecuencia de la sustitución de diversos signos por comas. Sí aumentan las longitudes de las oraciones gracias a este mecanismo, entonces, consecuentemente, aumentan las cantidades de pausas lingüísticas (como promedio). La correlación del PPS con la edad no es significativa.

También en José Saramago se detecta un cambio en sus últimos tres libros. No obstante, este cambio se produce en la dirección opuesta, es decir, en lugar de fragmentar el texto con el aumento del número de interrupciones, se observa una disminución en la cantidad de pausas.

En la Tabla 6.11 se muestra el resumen estadístico del PPS para los libros de los autores en portugués. Los valores del PPS oscilaron entre 1,62 y 13,38 pausas por oraciones, tomando como referencia todos los autores. El promedio fue de $3,30 \pm 2,16$.

Todas las medias están por encima de la media calculada para la autora Iris Murdoch, caracterizando que, para este conjunto, los escritores de lengua portuguesa aquí representados usan más pausas lingüísticas que la autora.

Tabla 6.11. Resumen estadístico del PPS de los libros de los autores en lengua portuguesa.

Autor	N	Mean	SD	Min	Max
IM	24	1.84	0.19	1.51	2.26
AA	11	2.54	0.38	2.16	3.36
CCB	43	2.54	0.36	1.97	3.91
EQ	8	2.96	0.49	2.45	3.85
EV	10	2.01	0.24	1.76	2.61
JAp	21	2.50	0.63	1.62	3.47
JS	17	7.88	2.79	2.73	13.38
MA	12	2.81	0.49	2.16	3.86

6.4.7. Comparando parámetros (WPS, WPP y PPS)

Una vez más se analizarán las tres métricas relativas al uso de la puntuación, pero ahora para los autores en lengua portuguesa. Una visión general de las correlaciones entre estas métricas se presenta en la Tabla 6.12. Para esta comparación se han utilizados todos los libros de los autores en lengua portuguesa (*vid.* Tabla 3.1 en la página 65); los valores no significativos se han señalados en azul.

Tabla 6.12 Correlaciones entre WPS, WPP y PPS para autores en lengua portuguesa..

Autor	WPS x WPP^{*1}		WPS x PPS^{*1}		WPP x PPS	
	F	p-value	F	p-value	F	p-value
IM	4,65	0,04	15,17	0,00	4,99 (-)	0,04
AA	11,58	0,01	95,18	0,00	3,43	0,10
CCB	28,48	0,00	205,72	0,00	3,41	0,07
EQ	5,15	0,06	68,24	0,00	1,50	0,27

Autor	WPS x WPP ^{*1}		WPS x PPS ^{*1}		WPP x PPS	
	F	p-value	F	p-value	F	p-value
EV	34,21	0,00	82,24	0,00	9,26(+)	0,02
JAp	1,53	0,23	140,62	0,00	0,08	0,77
JS	2,07	0,17	398,16	0,00	0,48	0,50
MA	0,39	0,55	10,99	0,01	18,54 (-)	0,00

^{*1} Todas las correlaciones son positivas

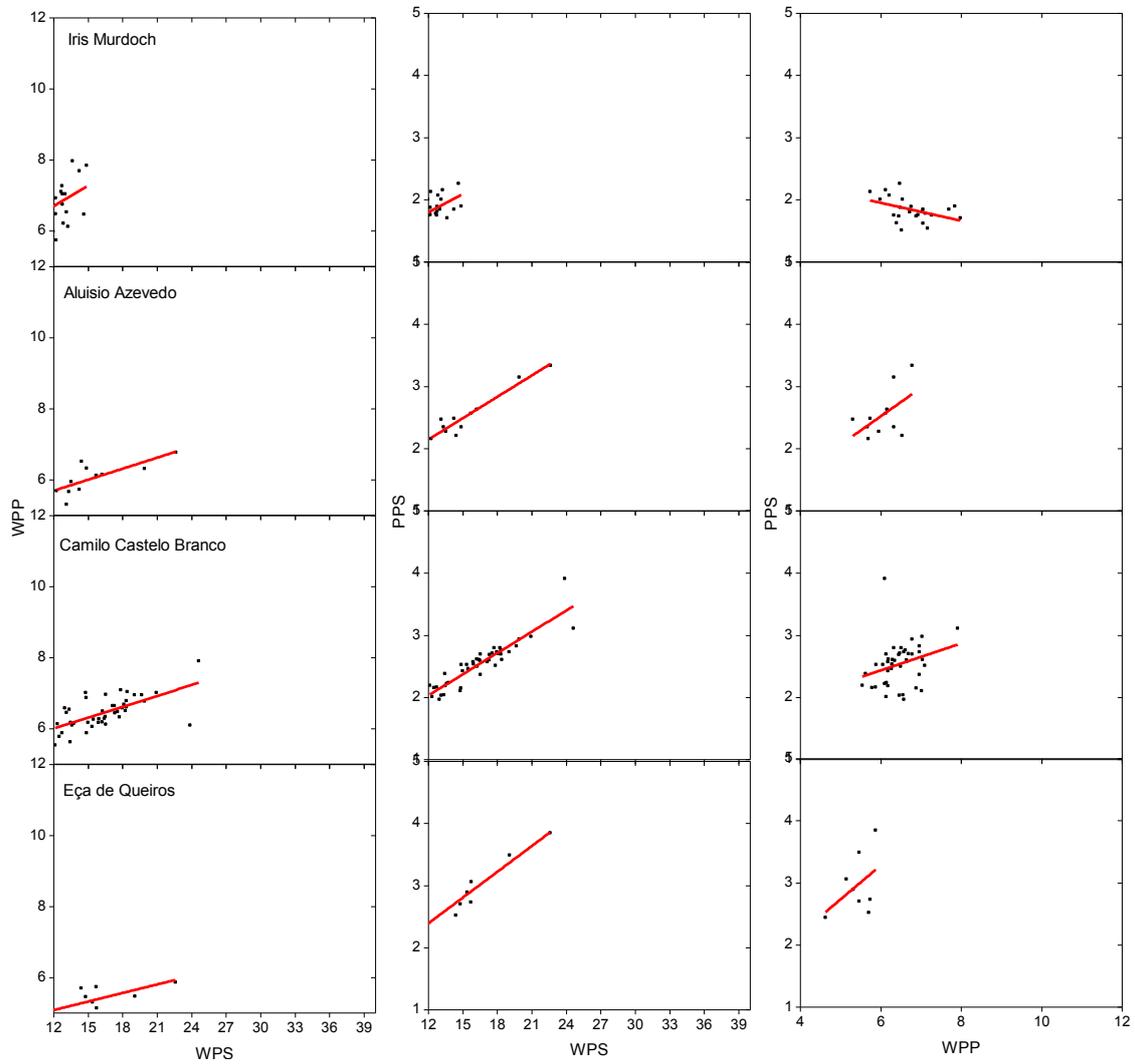
Como se ha descrito anteriormente, la correlación entre la longitud promedio de las oraciones (WPS) y la longitud promedio de las pausas lingüísticas (WPP) para los libros de Iris Murdoch es positiva. Este mismo comportamiento se ha encontrado para los libros de los autores Aluísio Azevedo, Camilo Castelo Branco y Érico Veríssimo. Para los demás, el aumento en la longitud de las oraciones no implica un aumento en la longitud de las pausas lingüísticas.

Sí se observa la correlación entre la longitud promedio de las oraciones (WPS) y la cantidad de pausas lingüísticas (PPS), el resultado con autores de lengua portuguesa se repite igual que los libros en lengua inglesa. El aumento de la longitud promedio de las oraciones es, por unanimidad (todas las correlaciones son fuertemente significativas), fruto del aumento en la cantidad de las pausas lingüísticas usadas en estas oraciones.

Para la mayoría de los autores (que no presentaron correlaciones significativas entre el WPP y el PPS) el aumento de la longitud de las oraciones ocurre básicamente por adición de nuevas pausas, pero estas no son de tamaño mayor. La excepción son los autores Érico Veríssimo (para quien las pausas añadidas son mayores) y Machado de Assis (para quien las pausas añadidas son menores), lo que coincide con el comportamiento de esta métrica para Iris Murdoch.

En el gráfico de la Figura 6.19 se muestran las inclinaciones de las rectas de las regresiones lineales entre WPS x WPP (la primera columna), entre WPS x PPS (la segunda columna) y entre el WPP x PPS. Las inclinaciones de las rectas se interpretan del siguiente modo: sí la dirección de izquierda-a-derecha es ascendente, significa que el incremento de la variable representada en el eje *x* está relacionada con el incremento de la variable representada en el eje *y*; si la dirección de izquierda-a-derecha es descendiente, significa que cuando se reduce el valor de la variable en el eje *x*, los valores de la variable en el eje *y* se incrementan; cuando no existe ni ascendencia ni

descendencia en la recta, es necesario mirar los valores de la estadística F y su respectivo p-value en la Tabla 6.12 para concluir que no exista ninguna correlación entre las variables representadas en el eje x e y .



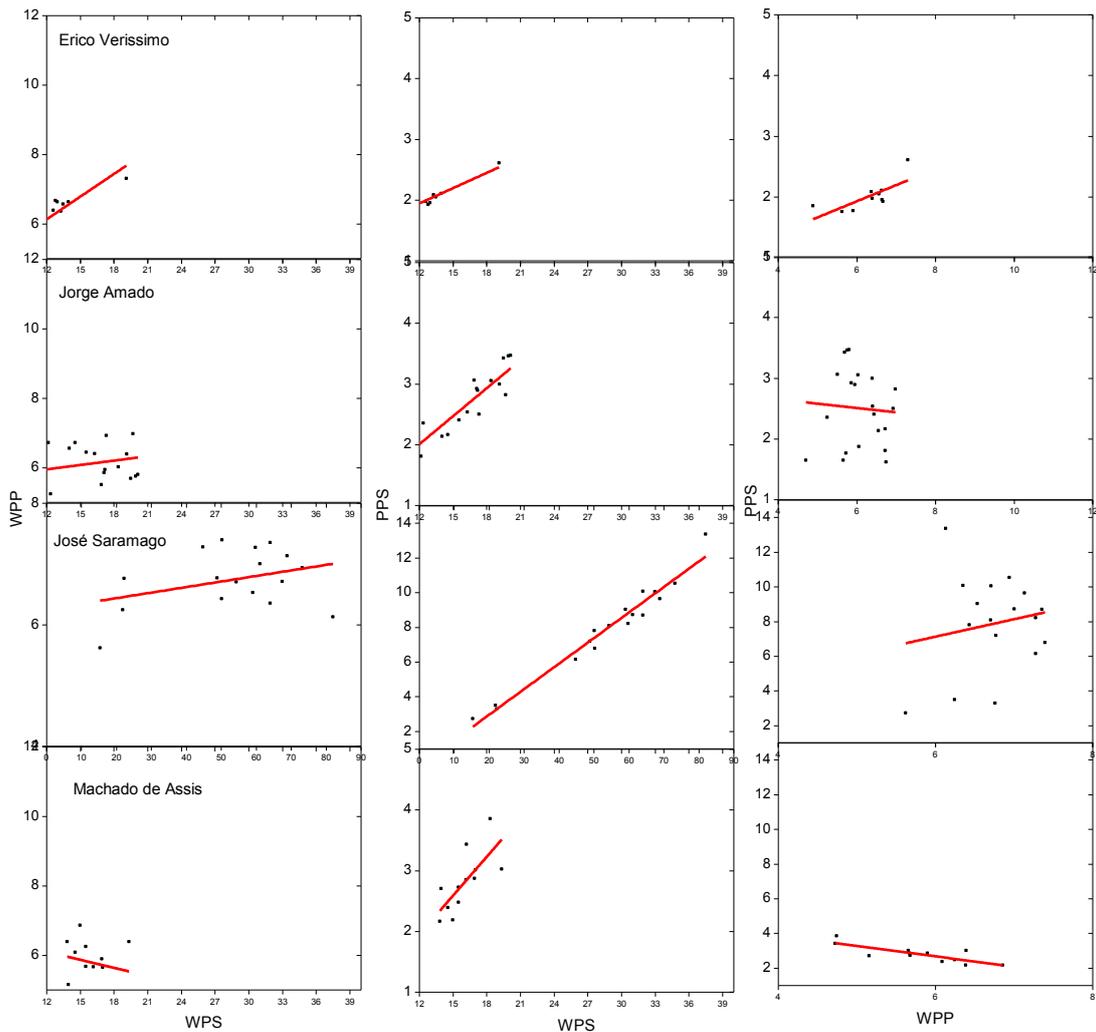


Figura 6.19 Regresión lineal de WPS x WPP (a), WPS x PPS (b) y WPP x PPS (c) para todos los autores de lengua portuguesa.

Una vez más se resalta que los datos presentados no son concluyentes respecto al envejecimiento. Para este análisis consideramos que la representación de la trayectoria en espacio de fases/estados es más adecuada. A continuación, se describirá con detalle el comportamiento de las métricas propuestas en este trabajo para el escritor Machado de Assis, pues fue el único que presentó una similitud en los resultados respecto a la autora inglesa. Los gráficos de los demás autores se muestran en el Anexo 7, junto a una breve explicación de los resultados obtenidos con las métricas analizadas.

Como ya se ha visto anteriormente, el primer espacio (Figura 6.20) se relaciona con la longitud media de las oraciones (WPS) y con el promedio del número de pausas lingüísticas (PPS) que el autor utiliza en las oraciones. El segundo relaciona la longitud

promedio de las pausas lingüísticas (WPP) y el número medio de pausas que los autores utilizan en las oraciones (PPS).

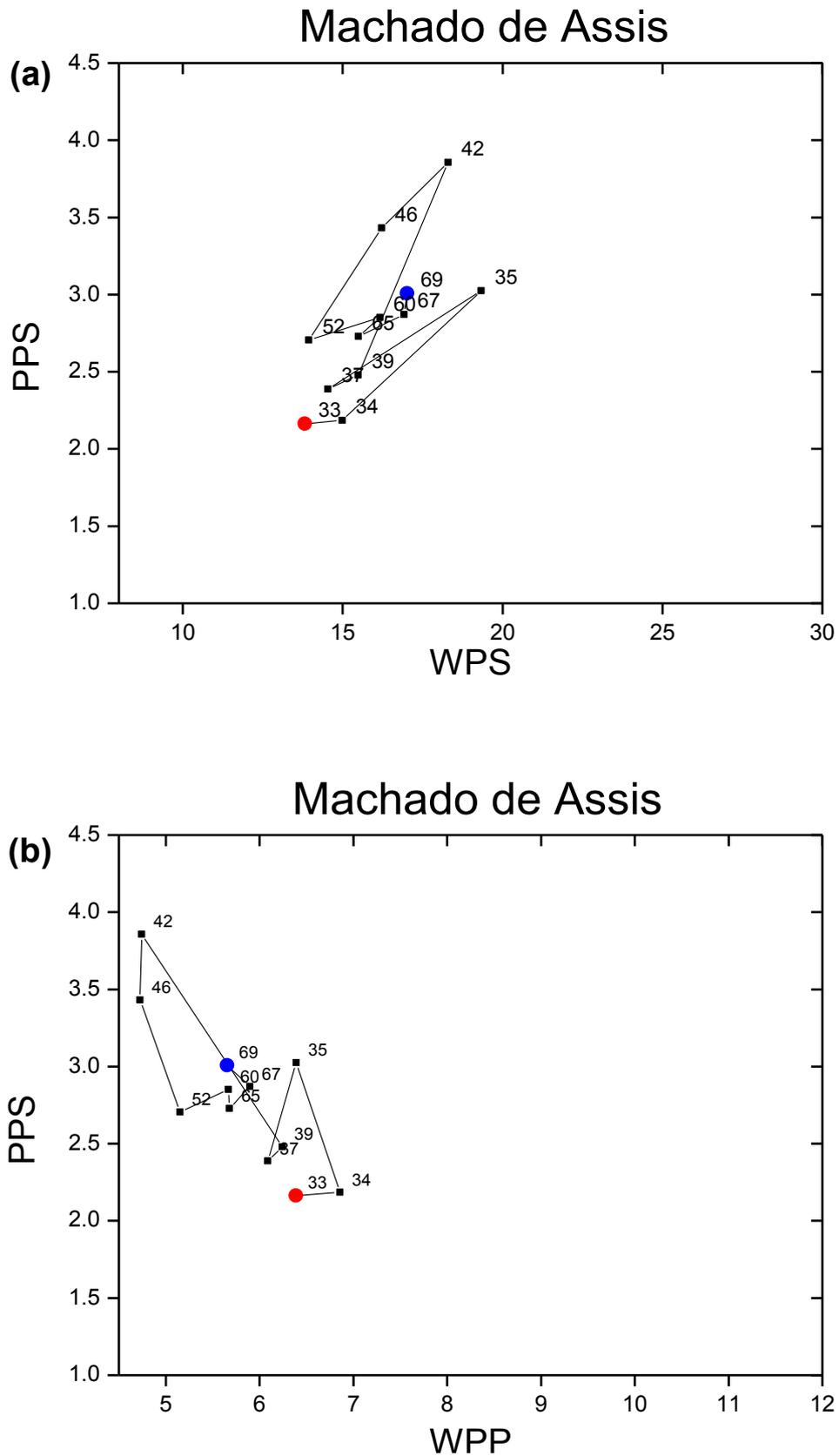
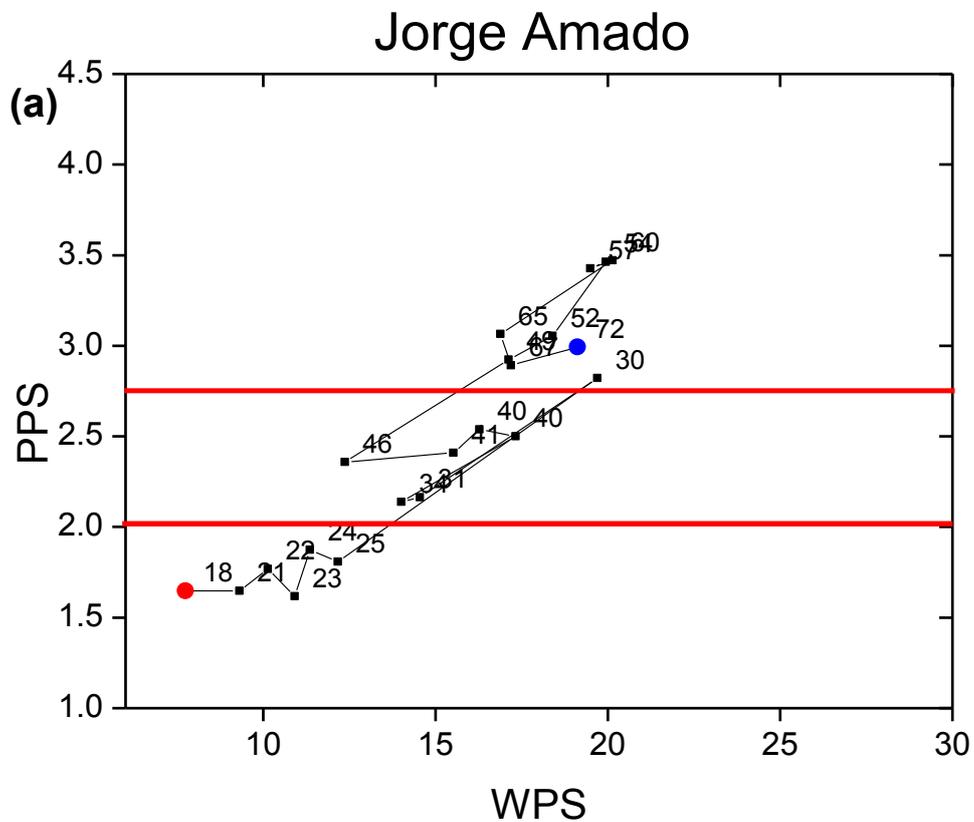


Figura 6.20 Comparación de WPS y WPP (a) y WPP y PPS (b) para Machado de Assis.



Machado de Assis es un ejemplo de que la simple observación de la correlación de las métricas no es suficiente para sacar conclusiones sobre lo que ocurre con el paso del tiempo. Aunque el comportamiento de las métricas WPP y PPS fueron similares si se compara Iris Murdoch y Machado de Assis, cuando se observan los gráficos de las trayectorias se percibe que, a diferencia de la autora inglesa (Figura 6.5 en la página 127) (para quien la caída en la longitud de las pausas lingüísticas se asocia a la caída de la longitud de estas pausas, lo que estaría directamente vinculado con el envejecimiento), para el autor Machado de Assis esto no es igualmente cierto. Mirando el gráfico (b) de la Figura 6.20 es fácil apreciar que los últimos libros no son los que presentan menores pausas lingüísticas.



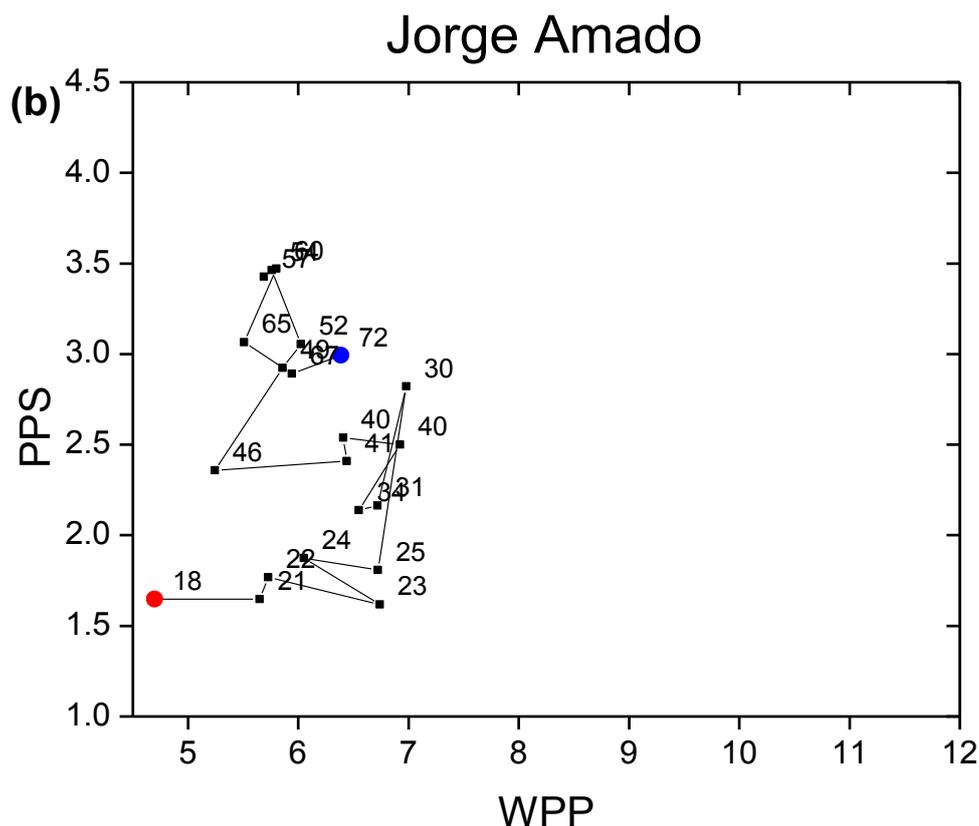
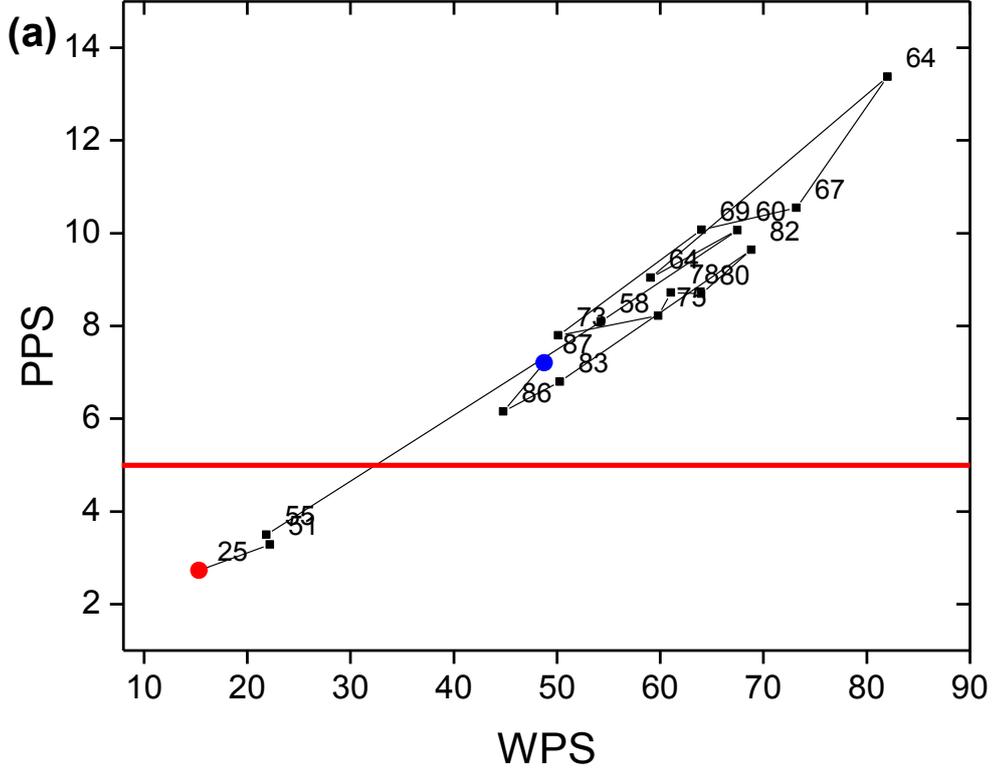


Figura 6.21 Comparación de WPS y WPP (a) y WPP y PPS (b) para Jorge Amado.

Las oraciones de los libros del escritor Jorge Amado (en promedio) incrementaron la longitud de modo significativo [$F = 30,50$, $p < 0,001$] con el paso del tiempo, cuando consideramos todos los puntos. Pero cuando analizamos la trayectoria se ha podido notar tres distintas fases: la primera entre 18 y 25 años, la segunda entre 30 y 46 años y la tercera de los 49 a los 76 años (indicadas por las líneas rojas en la Figura 6.21a). La existencia de las tres fases se confirma con la Regresión Lineal para cada una de las fases individualmente, donde encontramos los valores de correlación respectivos: [$527,9$, $p < 0,001$], [$1,45$, $p = 0,28$] y [$0,03$, $p = 0,86$]. Otra vez podemos decir que la correlación encontrada para todas las medidas de la métrica es entre la nube de puntos y no entre todos los puntos, dado que las dos últimas fases no son correlacionadas significativamente. En resumen, la tendencia de crecimiento de la longitud de las oraciones por medio del incremento de la cantidad de pausas no es un comportamiento que ocurre con el envejecimiento del autor.

Jose Saramago



Jose Saramago

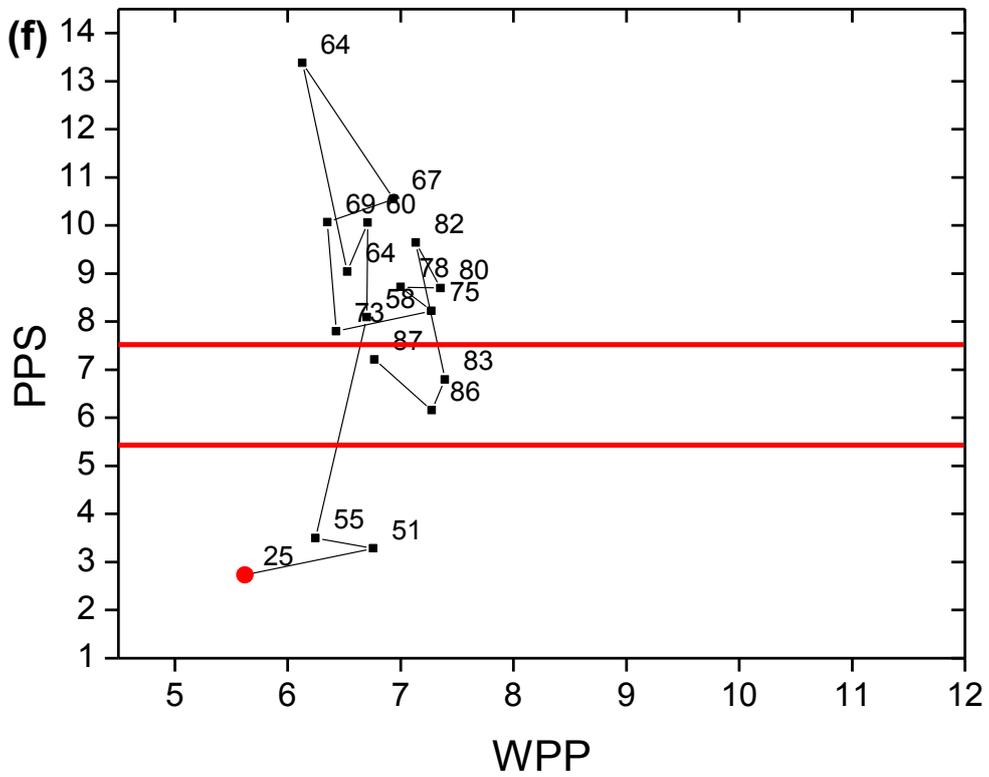


Figura 6.22 Comparación de WPS y WPP (a) y WPP y PPS (b) para José Saramago.

Para el autor José Saramago la trayectoria del WPP y PPS con el paso del tiempo (Figura 6.22a) caracteriza la existencia de dos fases distintas igual que ha pasado con la autora Iris Murdoch (Figura 6.5) para estas mismas métricas, pero esto no ha ocurrido en la etapa final de su producción literaria y sí en el principio de su carrera literaria. El análisis de la trayectoria del WPP y PPS no indica ninguna tendencia ni cambio para estas métricas.

En resumen, ninguno de los conjuntos de libros de los autores de lengua portuguesa presentó trayectorias similares a las observadas en los libros de Iris Murdoch, por lo que solo esta autora sería la única que presentó ese comportamiento.

Capítulo 6

Enquanto eu tiver perguntas e não houver respostas, continuarei a escrever.

(Clarice Lispector)

7.1. Conclusiones

El conjunto de datos presentados en el apartado anterior nos permite realizar algunas conclusiones generales.

1. *Inadecuación de la relación entre baja utilización del vocabulario con mala calidad de escritura y deterioro cognitivo*

Peter Garrard, Lisa M. Maloney, John R. Hodges y Karalyn Patterson (2005), los primeros a investigar los libros de Iris Murdoch, han hecho la asociación entre la baja utilización del vocabulario en su último libro, *Jackson's Dilemma*, con la mala evaluación de la crítica literaria y, conseqüentemente, con el deterioro cognitivo de la escritora indicando que esto podría haber ocurrido porque la escritora ya estaría con la Enfermedad de Alzheimer sin presentar síntomas aparentes.

En esta investigación se puede constatar que asociar el reducido vocabulario utilizado en un libro con la mala crítica literaria y el deterioro cognitivo no es de todo correcto. Primero porque en función del punto de corte de los libros para análisis, el libro de la autora que presenta menor vocabulario no es el último, y sí el libro *A Fairly Honourable Defeat*, escrito a los 51 años. Por lo tanto, aunque el corte en las 55.000 primeras palabras para analizar el vocabulario haya detectado el libro *Jackson's Dilemma* con el que tenía el menor vocabulario, es un resultado dependiente de la decisión del investigador, por consiguiente, no reproducible.

En segundo, analizando los libros de otros escritores hemos encontrado que no siempre la baja cantidad de vocabulario se puede asociar a libros de baja calidad literaria. Así, autores como Arthur Conan Doyle, Charles Dickens, Frances Hodgson Burnett (lengua inglesa) y José Saramago (lengua portuguesa) presentan sus libros de menor uso de vocabulario asociados a éxitos de público.

Además si la intención de la investigación es definir un marcador cognitivo capaz de convertirse en un marcador de Deterioro Cognitivo Leve, de detección precoz, hay que considerar que en la primera fase de la Enfermedad de Alzheimer lo que ocurre no es una baja cantidad de vocabulario y si un incremento, basado en la evidencia del uso de las circunlocuciones.

De este modo hace falta, para trabajos futuros, investigar con respecto a las características léxicas no el último libro de la autora y si el libro *The Philosopher's Pupil*, escrito a los 64 años, por presentar mayor vocabulario en las primeras 55.000 palabras.

2. *Insuficiencia de los métodos de correlación*

Los métodos de regresión lineal e correlación, utilizados por los investigadores que hicieron estudios longitudinales de producción del lenguaje (sea oral o escrito) de un modo general y en particular los estudios de la producción literaria de escritores como Iris Murdoch no son suficientes para sacar conclusiones sobre la evolución del comportamiento lingüístico de personas mayores a lo largo de sus vidas relativos a cambios abruptos.

Esto porque este método, cuando tiene la edad como una de las variables, está diseñado para investigaciones de tendencias de comportamiento a lo largo de la vida. Y, como es sabido, la Enfermedad de Alzheimer se caracteriza por disfunción o alteración del comportamiento sea social, cognitivo o emocional a partir de una determinada edad.

En el caso de este estudio evaluamos las alteraciones relacionadas con el comportamiento cognitivo del uso del lenguaje escrito, más precisamente el uso de la puntuación, pero se puede decir que este método no es capaz de identificar cualquier cambio de comportamiento cuando se compara la edad temprana y la vejez.

Este aspecto de la inadecuación del método se ha podido perceber cuando relacionamos la métrica PPS con la edad (página 124). Por la Regresión Lineal había una correlación positiva entre los puntos, pero lo que realmente ocurrió fue una correlación entre dos nubes de puntos que representaban los dos estados detectados por el análisis de la trayectoria del uso de la puntuación.

El método de Regresión Lineal también se presentó como insuficiente cuando hemos analizado y comparado los datos de Machado de Assis y Iris Murdoch (página 172). En este caso, los datos de la correlación entre WPP y PPS fueron similares para los dos autores, pero para la escritora Iris Murdoch la relación entre las dos métricas estaba relacionada con el envejecimiento lo que no ocurría con el escritor brasileño Machado de Assis.

A parte de la insuficiencia general del método para investigaciones longitudinales con la finalidad de detectar cambios de comportamiento en la vejez, hay que apuntar que, en el caso específico de las investigaciones relacionadas a los libros de Iris Murdoch, no han considerado la influencia de la longitud de los libros sobre la métrica que define el uso del vocabulario en las primeras 55.000 palabras (TTR55) para esta autora para sacar conclusiones referentes a cambios asociados a la edad. Esta fue una característica compartida con algunos otros escritores (de lengua inglesa y portuguesa) pero que no es una regla general. Esto ha causado un oscurecimiento de los resultados, pues asignaron al envejecimiento los cambios detectados en la cantidad de vocabulario utilizado en los libros cuando en verdad estos cambios también estaban asociados a la longitud de los libros en la franja de edad de 61 a 76 años (página 109).

3. Validez de la Trayectoria en espacio de fases/estados

El método de la Trayectoria en espacio de fases/estados es más adecuado para identificar las alteraciones cognitivas relacionadas con el uso de la puntuación con el paso del tiempo.

Esta adecuación del método se ha hecho presente en las evaluaciones de los resultados de Iris Murdoch cuando hemos comparado las métricas PPS y WT (página 122) y, principalmente, WPP y PPS (página 127). En esta última, como se ha explicado anteriormente, por los métodos de Regresión Lineal y Correlación tendríamos una

correlación entre los puntos sin considerar que estos hacían parte de dos estados distintos de comportamiento.

Cuando comparamos los resultados de autores de lengua inglesa y portuguesa con los resultados de la escritora Iris Murdoch el análisis del WPS y de las nuevas métricas propuestas, WPP y PPS, también se ha beneficiado de la trayectoria para dilucidar el comportamiento de uso de la puntuación.

Por ejemplo, los autores Harold Bindloss, Wilkie Collins y Jacob Abbott presentaron correlaciones significativas entre el WPS y WT pero por la correlación no se ha podido verificar si la edad influyó en el resultado (página 138). Por la trayectoria observamos que en cuanto en los dos primeros la tendencia fue de decrecimiento con la edad el último ha incrementado la longitud promedio de las oraciones cuando los libros se tornaban mayores.

Para el escritor brasileño Jorge Amado la trayectoria fue importante pues se observó que no hubo alteraciones bruscas en edades más avanzadas (página 158), aunque el WPS ha incrementado con la edad, este crecimiento se ha mantenido constante (al menos hasta la edad de 72 años cuando escribió el último libro aquí analizado, sin considerar los tres últimos). Sin embargo, se nota dos fases distintas, una que abarca los primeros libros del autor escritos entre 18 y 25 años, y la segunda los demás libros escritos entre 30 y 72 años.

Comparando los resultados de WPS, WPP y PPS para Jorge Amado (página **¡Error! Marcador no definido.**) y José Saramago (página **¡Error! Marcador no definido.**) y recurriendo a las trayectorias podemos concluir que los dos escritores difieren de modo igual con los resultados de Iris Murdoch precisamente por que no hubo ninguna alteración significativa negativa de la longitud de las pausas lingüísticas relacionada con el paso del tiempo.

En el caso particular de José Saramago la trayectoria ha demostrado los cambios importantes en dos etapas de la vida del autor. El primer cuando empezó a utilizar la puntuación de un modo especial a los 58 años, con el libro *Levantado do Chão*, y al final de su vida, en los tres últimos libros. En este último periodo de su escritura, aunque tenga presentado un cambio este fue en sentido contrario a del Iris Murdoch, o sea, Saramago ha disminuido la fragmentación de sus textos (en promedio) la lucidez de José Saramago es un hecho inquestionable y que se puede comprobar en videos de

entrevistas concedidas a los 83 años (“Entrevista 1,” 2005), cuando fue acometido de una grave enfermedad respiratoria pero en este mismo año ha lanzado el libro *Intermitências da Morte* y a los 87 años (“Entrevista 2,” 2009) cuando ha lanzado su último libro *Caim*, precisamente seis meses antes de su muerte.

Así como el método de la trayectoria sirvió para detectar la presencia y ausencia de alteraciones en el uso de la puntuación ocurridas en edades avanzadas, podremos afirmar que el método es adecuado para evaluar cualquier otro aspecto de deterioro cognitivo relativo al lenguaje, sea oral o escrita. Lo importante para el buen funcionamiento de este método es tener datos longitudinales que cubran la mayor parte de la vida de los individuos en estudio y que mantengan una regularidad periódica en la medición.

3. *Iris Murdoch: ¿detección de la enfermedad?*

El análisis del comportamiento lingüístico de la escritora Iris Murdoch por el método de la trayectoria en espacio de fases/estados, utilizando la longitud promedio de las oraciones (WPS) y las nuevas métricas, longitud promedio de las pausas lingüísticas (WPP) y cantidad promedio de pausas por oración (PPS) indicaron que la autora, a lo largo de su vida, ha utilizado pausas cada vez menores y ha incrementado la cantidad de pausas, pero sin cambiar de modo significativo la longitud de las oraciones.

Los resultados encontrados indican que hubo una alteración de comportamiento relacionado al uso de la puntuación, y este cambio ha empezado en una edad más temprana que otros síntomas que caracterizan la Enfermedad de Alzheimer, como por ejemplo los resultados positivos de pruebas clínicas (Peter Garrard et al., 2005), como también de lo propuesto cambio en la cantidad de vocabulario, asociado a la mala crítica literaria en que se han basado las otras investigaciones (Peter Garrard et al., 2005); (Le, 2010); (Le et al., 2011) de los libros de la escritora inglesa Iris Murdoch.

Para mejor entender lo que ha pasado, reproduzo el gráfico (Figura 7.1) que resume los cambios en la puntuación que hemos relacionado con la Enfermedad de Alzheimer en este trabajo. Analizando el gráfico vemos que a los 64 años la autora empieza a elevar los valores de cantidades de pausas lingüísticas (PPS) en relación a los valores anteriores y continúa a reducir el tamaño de estas pausas. Esta combinación de

las dos métricas en este punto caracteriza la aparición de una nueva fase o estado, que se mantendrá hasta el final de la vida de la autora.

Hay que notar el hecho de que no es solamente la tendencia a disminuir la longitud de las pausas lingüísticas que caracteriza el cuadro asociado con el diagnóstico de la Enfermedad de Alzheimer propuesto. Esto ya se había detectado con los métodos de Regresión Lineal y la correlación, indicando que a lo largo de la vida la autora ha reducido significativamente el valor del WPP con $[F(1,24) = 40,67, p < 0,001]$. Pero la constatación de que hay un cambio a los 64 años solo fue posible utilizando las trayectorias combinando tres variables en el análisis.

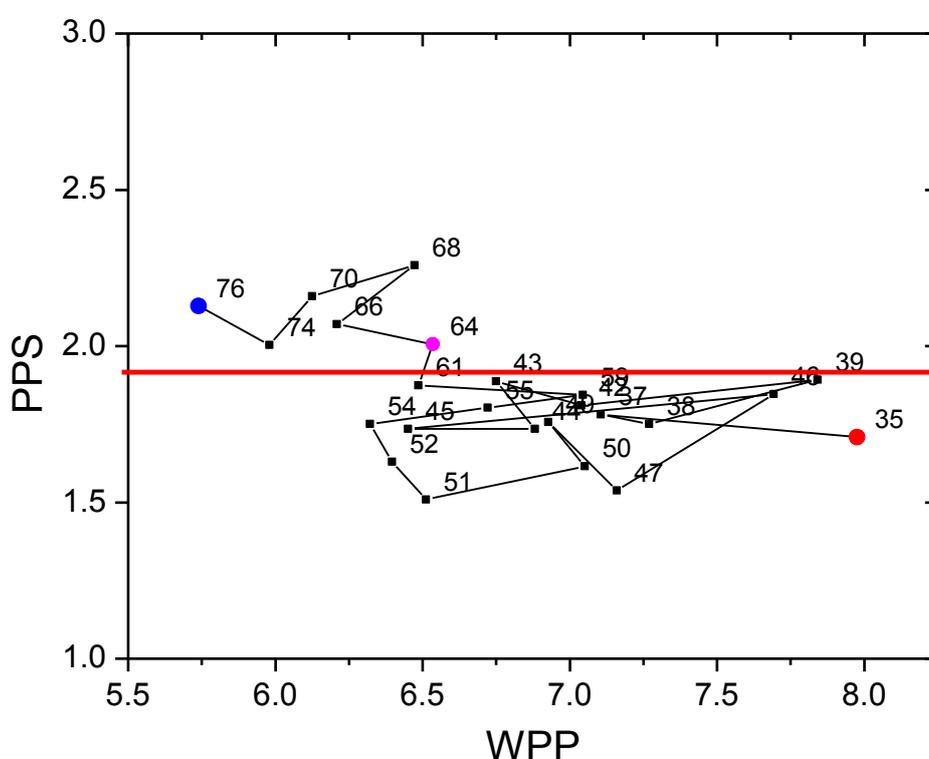


Figura 7.1 Trayectoria en espacio de fases (SPT) del WPP y PPS para Iris Murdoch

La importancia de estos resultados como un síntoma de la Enfermedad de Alzheimer se basa en el hecho de que con estas alteraciones de comportamiento lo que está indicado es que la autora ha fragmentado su discurso dividiendo las oraciones (en promedio) en trozos más pequeños. Esto puede estar asociado a una pérdida cognitiva relacionada al hecho de que el aprendizaje de la puntuación es la última etapa de conocimiento adquirido relativo al lenguaje, considerando que el aprendizaje se da en capas. En este caso, la estructura de las oraciones pasa a presentar una semejanza mayor

con el *lenguaje interno* (fragmentado) do que con la estructura característica del lenguaje escrito.

La autora Iris Murdoch presentó dos libros con una mayor cantidad de palabras diferentes (WCD) cuando analizado el corte de 55.000 palabras. Fueron los libros *The Red and Green*, escrito a los 46 años con WCD = 5540 y el libro *The Philosopher's Pupil* con WCD = 5529, escrito a los 64 años. Consideramos el último como el libro de mayor vocabulario, pues la diferencia entre el tamaño final de ellos es grande (más del doble). El primer con longitud igual a 103.186 palabras y el último con 222.065 palabras. Por esto, cuando la autora escribió las primeras 55.000 palabras del primero libro estaba en la mitad del tamaño final (53,4%), y las primeras 55.000 palabras del segundo representaban a penas 24,8% del total del libro. Significa decir que aunque ella repita las palabras como suele ocurrir en la escritura (vid. explicación en las páginas 90-92), nuevas palabras también irán surgir hasta que finalice el texto, y para el segundo texto esto irá ocurrir más veces.

El hecho de que la alteración en el patrón de utilización de la puntuación empieza a la edad de 64 años, por lo tanto mucho antes de que la autora empiece a tener dificultades en el uso del vocabulario (12 años antes), coincide con el libro de mayor vocabulario de la autora de toda su vida, el libro *The Philosopher's Pupil*.

Este incremento del vocabulario utilizado por la autora sería compatible con las argumentaciones de Tomoeda y Bayles (1995 apud Murillo Ruiz, 1999) y Bayles (Kathrin A. Bayles, 1993) que se refieren a las circunlocuciones (donde el vocabulario es más amplio por que la persona no consigue nominar directamente los objetos y voltea para describirlo) como un síntoma precoz de un deterioro cognitivo relativo a la edad.

Estudios actuales (Hydén y Örvulv, 2009; Ogar et al., 2011; Costa, Calabria, Marne, y Hernández, 2012; Gómez-ruiz, Aguilar-alonso, y Antònia, 2012) identifican la aparición de las circunlocuciones como un síntoma precoz de demencia de un modo general cuando se analiza el lenguaje oral. No hay registros de utilización de circunlocuciones en el lenguaje escrito en pacientes de Enfermedad de Alzheimer o de otros tipos de demencia.

Es factible decir que las dos alteraciones en el modo de escritura de la escritora Iris Murdoch, ocurridos a los 64 años (el uso de circunlocuciones y la fragmentación del texto definido por el uso de cantidad mayor de puntuación de pausas breves dentro de

las oraciones), están relacionados y pueden ser considerados como síntomas precoces de la Enfermedad de Alzheimer.

4. *¿Es posible una detección precoz de la EA a partir de la escritura?*

El diagnóstico precoz de la enfermedad de Alzheimer es el objetivo principal de los investigadores involucrados en la detección de esta enfermedad, sea a través de marcadores biológicos o cognitivos (Peña-Casanova, 1999; Molinuevo, 2007; Cuetos-Vega, Menéndez-González y Calatayud-Noguera, 2007; Valls-pedret, Molinuevo y Rami, 2010).

Dentre los marcadores cognitivos, la memoria y el lenguaje son las capacidades que más se evalúan, teniendo el lenguaje oral recibido mayor destaque por parte de los investigadores. Los diferentes estudios realizados del lenguaje oral hasta ahora, buscan detectar las alteraciones en la producción del lenguaje oral, sea por su contenido (estudios léxicos) o por su estructura (estudios sobre la sintaxis). No hay un consenso sobre el orden en que ocurren estos daños. Podemos encontrar innumerables investigadores que defienden que lo que se daña primero es la sintaxis y una cantidad similar de otros que preconizan el inverso, o sea, que el léxico se ve deteriorado antes de que las estructuras sintácticas sufran alteraciones.

El presente trabajo se concentra en la estructura de las oraciones, pero no del modo como se llevan las investigaciones sobre el tema, a través de la longitud de las oraciones (WPS) y la secuencia en que las clases gramaticales se presentan en ellas (por ejemplo SVO –Sujeto-Verbo-Objeto–. Precisamente las estructuras se analizan a partir del uso de la puntuación y de como representan la fragmentación del discurso escrito, añadido más dos métricas (WPP y PPS) al WPS para definir este comportamiento.

Los 16 autores de lengua inglesa (9) y de lengua portuguesa (7) fueron analizados para las métricas WPS, WPP y PPS y ninguno de ellos presentó el comportamiento que ha caracterizado la fragmentación del discurso de la escritora Iris Murdoch. Como se puede notar en las Figuras Figura 6.11 y Figura 6.19 en algunos incluso se puede notar relaciones similares, como para Frances Hodgson Burnett, James Fenimore Cooper o Robert William Chambers, pero en ninguno de ellos este comportamiento estuvo asociado a la edad más avanzada. Para autores en lengua portuguesa, el único que suele comportarse de modo similar a la escritora es el autor

Jorge Amado, pero hace falta analizar los últimos 3 libros para concluir algo sobre este tema.

Por ello, la única escritora que presento indicios de la fragmentación de los textos fue la inglesa Iris Murdoch que tubo el diagnostico de la Enfermedad de Alzheimer comprobado en un diagnostico posmortem y los otros escritores han servido como control para esta afirmación.

5. Sobre la hipótesis de partida

El lenguaje es un marcador cognitivo utilizado en diversas investigaciones y muchos de los cuales fueron citados en este trabajo. Pero el lenguaje escrito no recibió la misma cantidad de atención por parte de los investigadores que intentan decifrar “el misterio” sobre en que momento de la vida empiezan a ocurrir alteraciones cognitivas que irán transformarse en la EA en un futuro. Quizá por el lenguaje escrito, por mucho tiempo, ter sido considerada como un processo cognitivo inferior frente a la grandeza del language hablado. O solamente por cuestiones pragmáticas relativas a la facilidad de se obserbar y evaluar la producción oral de individuos, asociados al hecho de que la escritura es una capacidade cognitiva que se ha popularizado en periodo histórico muy reciente en el mundo (que aun cuenta con cerca de 800 millones de personas analfabetas). Pero hay una cantidad significativa de personas en el mundo que usan el lenguaje escrito, de orden de miles de millones.

La escritura puede se transformar en un potente fuente de datos sobre las capacidades cognitivas que se deterioran con el envejecimiento, principalmente cuando consideramos la posibilidad de que la adquisición/aprendizaje del lenguaje se realiza en una sequencia cronológica (en capas) y que la pérdida cognitiva ocurre en sentido inverso (lo último que se adquiere/aprende es el primero que se daña). Así, consideramos que el diagnostico precoz de la Enfermedad de Alzheimer o de otros tipos de demencia que se caracterizan por deterioro cognitivo relativo a las capacidades lingüísticas pueden encontrar en el lenguaje escrito (y en especial en la puntuación) los primeros señales de daños, incluso antes de que el lenguaje oral presente alteraciones.

Nuestra conclusión es que el comportamiento lingüístico asociado al uso de la puntuación es un marcador cognitivo para la enfermedad de Alzheimer en edades tempranas, corroborando la hipótesis de este trabajo de que el deterioro cognitivo se da

en sentido inverso a la adquisición del lenguaje, cuando lo primero que se pierde es el último que se ha adquirido.

6. Límites de este trabajo

El principio de las alteraciones en el uso de la puntuación por la escritora Iris Murdoch se ha detectado a la edad de 64 años, justamente cuando ella ha escrito su libro de mayor uso de vocabulario en la primera porción de 55.000 palabras lo que parece un indicio de alteración cognitiva. Hay que aclarar las características del vocabulario empleado por la autora en este libro para constatar si el uso excesivo de palabras diferentes (vocabulario) está relacionada a una dificultad incipiente en nominar los objetos.

La escritura también puede estar asociada con otros disturbios cognitivos como la depresión. En el caso de este estudio, autores como Frances Hodgson Burnett, Wilkie Collins, Jorge Amado y Machado de Assis, que tuvieron episodios de depresión asociados a su histórico de vida presentaron resultados que sugieren alteraciones en el uso de la puntuación, pero que no estaban asociados a la Enfermedad de Alzheimer, y que merecen la pena una investigación más profundizada en el futuro para verificar si hay una relación entre los cambios detectados y la depresión. Incluso en los últimos tres libros del escritor José Saramago, aunque no presentó ningún síntoma de esta enfermedad al final de su vida, hay una alteración en el uso de la puntuación relativo a los libros anteriores que sugiere un cambio cognitivo.

Otro punto a destacar es que los marcadores cognitivos relativos al uso del lenguaje que asocian la caída del vocabulario en edades avanzadas a la EA no fueron corroborados por este estudio, ya que otros autores presentaron libros de baja tasa de inclusión de nuevas palabras pero esto no estaba asociado al último libro del autor ni a libros que no tuvieron aceptación de crítica o de público.

7.2. Consideraciones Finales

En trabajo anterior (Caldeira, Lobao, Andrade, Neme y Miranda, 2006) he analizado una diversidad de textos utilizando la metodología de redes complejas y siempre tuve la idea de que la escritura es como una red o como una tela cosida. En

este trabajo tuve la oportunidad de mirar en el interior de la costura y ver que los hilos que unen las diversas partes que componen la tela son conectados por puntos, pero no puntos cualesquiera, y si de exclamación, interrogación, comas, guiones, dos puntos, punto y coma, punto y aparte y punto final.

Referencias

- Albert, R., & Barabasi, A. (2002). Statistical mechanics of complex networks. *Reviews of Modern Physics*, 74(47).
- Altmann, G. (1978). Towards a Theory of Language. *Glottometrika*, 1.
- Altmann, L. J. P., & Kemper, S. (2006). Effects of age, animacy and activation order on sentence production. *Language and Cognitive Processes*, 21(1-3), 322–354. doi:10.1080/0169096054400006
- Andreasen, N. J. C. (1975). Linguistic Analysis of Speech in Affective Disorders. *Archives of general psychiatry*, 33.
- Appell, J., Kertesz, a, & Fisman, M. (1982). A study of language functioning in Alzheimer patients. *Brain and language*, 17(1), 73–91. Retrieved from <http://www.ncbi.nlm.nih.gov/pubmed/7139272>
- Argente, J. (1980). *El Círculo de Praga*. (J. A. Argente, Ed.) (p. 129). Barcelona: Editorial Anagrama.
- Bayles, Kathrin A. (1993). Pathology of Language Behavior in Dementia. In G. Blanken (Ed.), *Linguistic Disorders and Pathologies: An International Handbook* (pp. 388–408). New York: Walter de Gruyter.
- Bayles, Kathryn A., Kaszniak, A. W., & Tomoeda, C. K. (1987). *Communication and cognition in normal aging and dementia* (p. 400). Little, Brown. Retrieved from <http://books.google.com/books?id=RnZxz38BORgC&pgis=1>
- Biber, D. (1987). A Textual Comparison of British and American Writing. *American Speech*. doi:10.2307/455273
- Burke, D. M. (2002). Language Production and Aging. In Susan Kemper & R. Kliegl (Eds.), *Constraints on language: Aging, grammar, and memory* (pp. 3–28). Boston: Kluwer Academic Publishers.

- Caldeira, S. M. G., Lob, T. C. P., Neme, A., & Miranda, J. G. V. (2006). The network of concepts in written texts. *Eur. Phys. J. B*, 49, 523–529. doi:10.1140/epjb/e2006-00091-3
- Caplan, D., & Waters, G. (1999). Age effects on the Functional Neuroanatomy of Syntactic Processing in Sentence. In S. Kemper & R. Kliegl (Eds.), *Constraints on language: Aging, grammar, and memory* (pp. 283–297). Boston: Kluwer Academic Publishers.
- Caplan, D., & Waters, G. (2002). Sentence Comprehension in Alzheimer's Disease - Studies of Normal Aging and Brain Damage. In L. T. Connor & L. K. Obler (Eds.), *Neurobehavior of Language and Cognition* (pp. 61–76). New York: Kluwer Academic Publishers.
- Caramelli, P., Lessa Mansur, L., & Nitrini, R. (1998). Language and communication disorders in dementia of the Alzheimer type. In B. Stemmer & H. A. Whitaker (Eds.), *Handbook of Neurolinguistics* (pp. 463–474). London: Academic Press. (pp. 463–474). London: Academic Press.
- Chafe, W. (1988). Punctuation and the Prosody of Written Language. *Written Communication*, 5(4), 395–426. doi:10.1177/0741088388005004001
- Chen, H. P., & Irwin, O. C. (1946). The type-token ratio applied to infant speech sounds. *The Journal of speech disorders*, 11, 126–130.
- Chen, Y. S., & Leimkuhler, F. F. (1989). A type-token identity in the Simon-Yule model of text. *Journal of the American Society for Information Science*, 40(1), 45–53. Retrieved from <http://www3.interscience.wiley.com/cgi-bin/jtoc/27981/>
- Cherrier, M. M., Mendez, M. F., Cummings, J. L., & Benson, F. (1998). Language and Communication in Non-Alzheimer's Dementias. In B. Stemmer & H. A. Whitaker (Eds.), *Handbook of Neurolinguistics* (pp. 447–461). California: Academic Press.
- Cheung, H., & Kemper, S. (2008). Competing complexity metrics and adults' production of complex sentences. *Applied Psycholinguistics*, 13(01), 53. doi:10.1017/S0142716400005427
- Costa, A., Calabria, M., Marne, P., & Hernández, M. (2012). On the parallel deterioration of lexico-semantic processes in the bilinguals' two languages : Evidence from Alzheimer's disease. *Neuropsychologia*, 50, 740–753. doi:10.1016/j.neuropsychologia.2012.01.008
- Crane, P. K., Gruhl, J. C., Erosheva, E. A., Gibbons, L. E., Mccurry, S. M., Rhoads, K., Nguyen, V., et al. (2010). Use of Spoken and Written Japanese Did Not Protect Japanese-American Men From Cognitive Decline in Late Life. *Society*, 654–666. doi:10.1093/geronb/gbq046.
- Croisile, B., Ska, B., Brabant, M. J., Duchene, a, Lepage, Y., Aimard, G., & Trillet, M. (1996). Comparative study of oral and written picture description in patients with

Alzheimer's disease. *Brain and language*, 53(1), 1–19.
doi:10.1006/brln.1996.0033

Crystal, D. (1994). *Enciclopedia del lenguaje de la Universidad de Cambridge* (p. 476).
Madrid: Taurus Ediciones.

Cuetos-Vega, F., Menéndez-González, M., & Calatayud-Noguera, T. (2007).
Descripción de un nuevo test para la detección precoz de la enfermedad de
Alzheimer. *Revista de Neurología*, 44(8), 469–474.

Daniels, H. (2005). *An Introduction to Vygotsky* (2th ed.). New York.

De Villiers, J. G. (1978). Fourteen grammatical morphemes in acquisition and aphasia. In
A. Caramazza & E. B. Zurif (Eds.), *Language acquisition and language
breakdown* (pp. 121–144). Baltimore: The John Hopkins University Press.

Dieguéz-vide, F., & Peña-Casanova, J. (2012). *Cerebro y Lenguaje Sintomatología
Neurolingüística* (1a. ed., p. 364). Barcelona: Editorial Médica Panamericana.

Dorogovstev, S. N., & Mendes, J. F. . F. (2001). Language as an evolving word web.
Proceedings of the Royal of London B, 268, 2603–2606.

dos Santos, A. S. (2007). *Guia Prático da Tradução Inglesa*. São Paulo: Campus.

Díaz, R. S. del V. (2007). Clínica y subtipos de deterioro cognitivo leve. In J. L.
Molinuevo (Ed.), *Deterioro cognitivo leve* (pp. 25–46). Barcelona: Editorial Glosa.

Ehrlich, J. S. (1997). IDEATIONAL AND SEMANTIC CONTRIBUTIONS TO
NARRATIVE PRODUCTION IN ADULTS WITH DEMENTIA OF THE
ALZHEIMER ' S TYPE, 30, 79–99.

Entrevista 1. (2005). Retrieved from
<http://www.youtube.com/watch?v=i2vARBxzjaE&feature=related>

Entrevista 2. (2009). Retrieved from
<http://www.youtube.com/watch?v=lSk4IyDZEho&feature=relmfu>

Ferreiro, E., & Pontecorvo, C. (1999). Managing the written text : the beginning of
punctuation in children ' s writing, 9, 543–564.

Ferrer i Cancho, R. (2003). *Language: universals, principles and origins*. Universitat
Politécnica de Catalunya (Barcelona).

Ferrer-i-Cancho, R., & Solé, R. (2001). The small world of human language.
Proceedings of Royal Society of London B, 268, 2261.

Ferrer-i-Cancho, R., Solé, R., & Köhler, R. (2004). Patterns in Syntactic Dependency
Networks. *Physical Review E*, 69, 051915.

- Fisher, R. A. (1935). The logic of inductive inference. *Journal of the Royal Statistical Society*, A(98), 39–54.
- Gardner, H. (1988). *La nueva ciencia de la mente: Historia de la revolución cognitiva* (p. 449). Barcelona: Paidós.
- Garrard, P, & Hodges, J. R. (2000). Semantic dementia: clinical, radiological and pathological perspectives. *Journal of neurology*, 247(6), 409–22. Retrieved from <http://www.ncbi.nlm.nih.gov/pubmed/10929269>
- Garrard, P, Lambon Ralph, M. a, Watson, P. C., Powis, J., Patterson, K., & Hodges, J. R. (2001). Longitudinal profiles of semantic impairment for living and nonliving concepts in dementia of Alzheimer’s type. *Journal of cognitive neuroscience*, 13(7), 892–909. doi:10.1162/089892901753165818
- Garrard, Peter. (2009). Cognitive archaeology: Uses, methods, and results. *Journal of Neurolinguistics*, 22(3), 250–265. doi:10.1016/j.jneuroling.2008.07.006
- Garrard, Peter, & Hodges, J. R. (1999, August). Semantic dementia: implications for the neural basis of language and meaning. *Aphasiology*. doi:10.1080/026870399401966
- Garrard, Peter, Maloney, L. M., Hodges, J. R., & Patterson, K. (2005). The effects of very early Alzheimer’s disease on the characteristics of writing by a renowned author. *Brain : a journal of neurology*, 128(Pt 2), 250–60. doi:10.1093/brain/awh341
- Garrett, M. (1980). Levels of processing in sentence production. In B. Butterworth (Ed.), *Language production*. London: Academic Press.
- Garrett, M. (1992). *Niveles de procesamiento en la producción de oraciones*. (F. Valle, F. Cuetos, J. M. Igoa, & S. del Viso, Eds.) *Lecturas de psicolingüística* (p. vol. 1). Madrid: Alianza.
- Gibbs, J. W. (2010). *Elementary Principles in Statistical Mechanics: Developed with Especial Reference to the Rational Foundation of Thermodynamics*. Retrieved from <http://books.google.com/books?hl=es&lr=&id=7VbC-15f0SkC&pgis=1>
- Gil, A. M. (2009). *El Alzheimer* (p. 126). CSIC-Dpto. de Publicaciones. Retrieved from <http://books.google.com/books?id=A9S-DNxuX-wC&pgis=1>
- Glosser, G., & Henderson, V. W. (2002). Writing Impairments in Alzheimer’s Disease. In L. T. Connor & L. K. Olber (Eds.), *Neurobehavior of Language and Cognition - Studies of Normal Aging and Brain damage* (pp. 77–91). New York: Kluwer Academic Publishers.
- Gordon, J. K. (2008). Measuring the lexical semantics of picture description in aphasia. *Aphasiology*, 22(7-8), 839–852. doi:10.1080/02687030701820063.Measuring

- Graham, N. L., Patterson, K., & Hodges, J. R. (2000). The impact of semantic memory impairment on spelling: evidence from semantic dementia. *Neuropsychologia*, 38(2), 143–63. Retrieved from <http://www.ncbi.nlm.nih.gov/pubmed/10660226>
- Gómez-ruiz, I., Aguilar-alonso, Á., & Antònia, M. (2012). Language impairment in Catalan-Spanish bilinguals with Alzheimer ' s disease. *Journal of Neurolinguistics*, 25(6), 552–566. doi:10.1016/j.jneuroling.2011.06.003
- Halliday, M. A. K. (1990). *Spoken and written language*. (F. Christie, Ed.). Oxford: Oxford University Press.
- Harris, J. R. (1883). Stichometry - part II. *The American Journal of Philology*, 4(3), 309–331.
- Harris Wright, H., Silverman, S., & Newhoff, M. (2003). Measures of lexical diversity in aphasia. *Aphasiology*, 17(5), 443–452. doi:10.1080/02687030344000166
- Heisler, L., Goffman, L., & Younger, B. (2010). Lexical and articulatory interactions in children's language production. *Dev Sci*, 13(5), 722–730. doi:10.1111/j.1467-7687.2009.00930.x.Lexical
- Hess, C., Sefton, K., & Landry, R. (1986). Sample size and type-token ratios for oral language of preschool children. *Journal of speech and*. Retrieved from <http://jslhr.highwire.org/cgi/content/abstract/29/1/129>
- Hess, C. W., Ritchie, K. P., & Landry, R. G. (1984). Type-token ratio and vocabulary performance. *Psychological Reports*, 55, 51–57 ST – Type–token ratio and vocabulary perfor.
- Hydén, L., & Örvulv, L. (2009). Narrative and identity in Alzheimer ' s disease : A case study. *Journal of Aging Studies*, 23(4), 205–214. doi:10.1016/j.jaging.2008.01.001
- J. Rendel Harris. (1883). Stichometry. *The American Journal of Philology*, 4(2), 133–157.
- Jakobson, R., & Halle, M. (1973). *Fundamentos del Lenguaje* (2a. ed.). Madrid: Editorial Ayuso.
- Johnson, K. (2008). *Quantitative methods in linguistics*. Australia: Blackwell Publishing Ltd.
- Jones, B. (1996). *What's The Point ? A (Computational) Theory of Punctuation*. University of Edinburgh.
- Kemper, S, Greiner, L. H., Marquis, J. G., Prenovost, K., & Mitzner, T. L. (2001). Language decline across the life span: findings from the Nun Study. *Psychology and aging*, 16(2), 227–39. Retrieved from <http://www.ncbi.nlm.nih.gov/pubmed/11405311>

- Kemper, Susan, Herman, R., & Lian, C. (2003). Age Difference in Sentence Production. *Journal of Gerontology*, 58B(5), 260–268.
- Kemper, Susan, & Kemtes, K. A. (2002). Limitations on Syntactic Processing. In Susan Kemper & R. Kliegl (Eds.), *Constraints on language: Aging, grammar, and memory* (pp. 79–106). Boston: Kluwer Academic Publishers.
- Kemper, Susan, Thompson, M., & Marquis, J. (2001). Longitudinal change in language production: Effects of aging and dementia on grammatical complexity and propositional content. *Psychology and Aging*, 16(4), 600–614. doi:10.1037//0882-7974.16.4.600
- Kempler, D., & Zelinski, E. M. (1994). Language in Dementia and normal aging. *Dementia and Normal Aging* (1a. ed., pp. 331–365). Great Britain: Cambridge University Press.
- Kennedy, J. B. (2010). Plato 's Forms , Pythagorean Mathematics , and Stichometry. *APEIRON*, (Academic Printing and Publishing), 1–32.
- Kintsch, W., & Keenan, J. (1973). Reading Rate and of Propositions Retention as a Function of the Number in the Base Structure of Sentences '. *Cognitive Psychology*, 274, 257–274.
- Kohler, R., & Rieger, B. B. (1993). *Contributions to Quantitative Linguistics*. (R. Köhler & B. B. Rieger, Eds.). Dordrecht: Springer Netherlands. doi:10.1007/978-94-011-1769-2
- Köhler, R. (1987). System Theoretical Linguistics. *Theoretical Linguistics*, 14, 241–257.
- Le, X. (University of T. (2010). Longitudinal Detection of Dementia Through Lexical and Syntactic Changes in Writing. *Science*, (January).
- Le, X., Lancashire, I., Hirst, G., & Jokel, R. (2011). Longitudinal detection of dementia through lexical and syntactic changes in writing: a case study of three British novelists. *Literary and Linguistic Computing*, 26(4), 435–461. doi:10.1093/lc/fqr013
- Leedham, M. (2011). *A corpus-driven study of features of Chinese students ' undergraduate writing in UK universities Thesis*. The Open University.
- Levantado do Chão (documental)*. (2008). Rede de Televisão Portuguesa - RTP. Retrieved from http://www.youtube.com/watch?v=yz_Ht0roxQw
- Louwerse, M. M., McCarthy, P. M., Mcnamara, D. S., & Graesser, A. C. (2004). Variation in Language and Cohesion across Written and Spoken Registers. *26th Annual Conference of the Cognitive Science Society* (pp. 1035–1040). Hillsdale, NJ: LEA.

- Luelsdorff, P. A. (1989). *Josef Vachek - Written Language Revisited*. (P. A. (University R. Luelsdorff, Ed.). Philadelphia: John Benjamins Publishing Company.
- Lust, B. C., & Foley, C. (Eds.). (2004). *First Language Acquisition: the Essential Readings. Studies in Second Language Acquisition* (Vol. 27, p. 442). Oxford: Blackwell Publishing Ltd. doi:10.1017/S0272263105230280
- Lyons, K., Kemper, S., LaBarge, E., Ferraro, F. R., Balota, D., & Storandt, M. (1994). Oral language and Alzheimer's Disease - A reduction in syntactic complexity. *Aging and Cognition*, 1(4), 271–281.
- Manschreck, T. C., Maher, B. A., & Ader, D. N. (1981). Formal thought disorder, the type-token ratio and disturbed voluntary motor movement in schizophrenia. *The British Journal of Psychiatry*, 139(1), 7–15. Retrieved from <http://bjp.rcpsych.org/content/139/1/7.abstract>
- Manschreck, T. C., Maher, B. A., Hoover, T. M., & Ames, D. (1984). The type--token ratio in schizophrenic disorders: clinical and research value. *Psychological Medicine*, 14(1), 151–157.
- Marcus, M. P. (University of P., Santorini, B. (Northwestern U., & Marcinkiewicz, M. A. (University of P. (1993). Building a Large Annotated Corpus of English : The Penn Treebank. *Computational Linguistics*, 19(2), 313–330.
- Maroco, J. (2007). *Análise estatística - com utilização do SPSS*. (M. Robalo, Ed.) (3rd ed., p. 822). Lisboa: Sílabo.
- Marsá, F. (1986). *Diccionario normativo y guía práctica de la lengua española*. Barcelona: Ariel.
- Maxim, J., & Bryan, K. (1994). *Language of the Elderly*. (D. Crystal, R. Lesser, & M. Snowling, Eds.). London: Whurr Publishers Ltda.
- McCarthy, P. M., & Jarvis, S. (2010). MTL-D, vocd-D, and HD-D: a validation study of sophisticated approaches to lexical diversity assessment. *Behavior research methods*, 42(2), 381–92. doi:10.3758/BRM.42.2.381
- McCune, L. (2008). *How Children Learn to Learn Language*. New York: Oxford University Press.
- McKee, G., Malvern, D., & Richards, B. (2000a). vocd. *Literary and Linguistic Computing*.
- McKee, G., Malvern, D., & Richards, B. (2000b). Measuring vocabulary diversity using dedicated software. *Literary and Linguistic Computing*, 15(3), 323–337.
- McKhann, G., Drachman, D., Folstein, M., Katzman, R., Price, D., & Stadlan, E. M. (1984). Clinical diagnosis of Alzheimer's disease: Report of the NINCDS-ADRDA Work Group* under the auspices of Department of Health and Human

- Services Task Force on Alzheimer's Disease. *Neurology*, 34(7), 939–939.
doi:10.1212/WNL.34.7.939
- Molinuevo, J. L. (2007). Deterioro cognitivo leve: antecedentes históricos y concepto. In J. L. Molinuevo (Ed.), *Deterioro cognitivo leve* (pp. 9–23). Barcelona: Editorial Glosa.
- Montemurro, M. A., & Pury, P. A. (2002). Long-range fractal correlations in literary corpora. *ArXiv:cond-mat*, 1, 0201139.
- Motter, A. E., Moura, A. P. S. de, Lai, Y. C., & Dasgupta, P. (2002). Topology of the conceptual network of language. *Physical Review E*, 65, 065102.
- Murdoch, B. E., Chenery, H. J., Wilks, V., & Boyle, R. S. (1987). Language disorders in dementia of the Alzheimer type. *Brain and language*, 31(1), 122–37. Retrieved from <http://www.ncbi.nlm.nih.gov/pubmed/2437993>
- Murillo Ruiz, B. (1999). *Estudio de la evolución del lenguaje en la demencia tipo Alzheimer*. Barcelona: Isep.
- Nowak, M. A., Plotkin, J. B., & Jansen, V. A. A. (2000). The evolution of syntactic communication. *Nature*, 404, 495 – 498.
- Nunberg, G. (1990). *The Linguistics of Punctuation* (p. 160). Center for the Study of Language (CSLI). Retrieved from <http://books.google.com/books?id=Sh-sruuKjJwC&pgis=1>
- Ogar, J. M., Baldo, J. V., Wilson, S. M., Brambati, S. M., Miller, B. L., Dronkers, N. F., & Gorno-tempini, M. L. (2011). Semantic dementia and persisting Wernicke ' s aphasia : Linguistic and anatomical profiles. *Brain and Language*, 117(1), 28–33. doi:10.1016/j.bandl.2010.11.004
- OMS. (2012). Dementia - a public health priority. United Kingdom: World Health Organization.
- Owen, A. J., & Leonard, L. B. (2002). Lexical diversity in the spontaneous speech of children with specific language impairment: application of D. *Journal of Speech, Language and Hearing Research*, 45(5), 927–937. Retrieved from <http://www.ncbi.nlm.nih.gov/pubmed/12381050>
- Owens Jr., R. E. (2000). *Language development an introduction* (p. 542). Boston, USA: A Pearson Education Company.
- O'Grady, W. D. (2005). *How Children Learn Language*. Cambridge: Cambridge University Press. doi:10.1017/CBO9780511791192
- Pakhomov, S., Chacon, D., Wicklund, M., & Gundel, J. (2010). Computerized assessment of syntactic complexity in Alzheimer's disease: a case study of Iris Murdoch's writing. *Behavior Research Methods*, 43(1), 136–144. doi:10.3758/s13428-010-0037-9

- Park, H., Rogalsky, Y., Rodriguez, A. D., & Zlatar, Z. (2011). Perceptual cues used by listeners to discriminate fluent from nonfluent narrative discourse. *Aphasiology*, 25(9), 998–1015. doi:10.1080/02687038.2011.570770.Perceptual
- Parkes, M. B. (2012). *Pages from the past : medieval writing skills and manuscript books*. (P. R. Robinson & R. Zim, Eds.). Farnham : Ashgate.
- Petersen, R. C., Doody, R., Kurz, a, Mohs, R. C., Morris, J. C., Rabins, P. V, Ritchie, K., et al. (2001). Current concepts in mild cognitive impairment. *Archives of neurology*, 58(12), 1985–92. Retrieved from <http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=2887763&tool=pmcentrez&rendertype=abstract>
- Peña-Casanova, J. (1999). *Enfermedad de Alzheimer* (p. 116). Barcelona: Fundació La Caixa.
- Peña-Casanova, J. (2007). *Neurología de la conducta y neuropsicología*. Madrid: Editorial Medica Panamericana.
- Phillips, J. R. (2009). Syntax and Vocabulary of Mothers' Speech to Young Children: Age and Sex Comparisons. *Child Development*, 44(1), 182–185.
- Pontecorvo, C., Orsolini, M., Burge, B., Resnick, L. B., Simone, R., & Marques Abaurre, M. B. (1996). *Children's Early Text Construction - Google Libros*. New Jersey: Lawrence Erlbaum Associates.
- Pueyo, A. A. (1987). *Lenguaje Interno ó Habla Subvocal* (p. 309). Promociones y Publicaciones Universitarias. Retrieved from <http://books.google.com/books?id=0FTOAAAACAAJ&pgis=1>
- Rapcsak, S. Z., Arthur, S. A., Bliklen, D. A., & Rubens, A. B. (1989). Lexical Agraphia in Alzheimer ' s Disease. *Arch Neurol.*, 46(1), 65–68.
- Rapcsak, S. Z., & Beeson, P. M. (2002). Neuroanatomical Correlates of Spelling and Writing. In A. Hillis (Ed.), *The Handbook of Adult Language Disorders* (pp. 71–99). New York: Psychology Press.
- Reisberg, B., Ferris, S. H., de Leon, M. D., & Crook, T. (1982). The Global Deterioration Scale for Assessment of Primary Degenerative Dementia. *American Journal of Psychiatry*, 139, 1136–1139.
- Richards, B., & Malvern, D. (2000). Measuring vocabulary richness in teenage learners of French. *British Educational Research Association Conference*. Cardiff, Wales, UK.
- Rondal, J., & Turnure, J. E. (1976). Maternal speech to normal and Down's syndrome children matched for mean length of utterance. *Annual Meeting of the American Association on Mental Deficiency*. Retrieved from <http://www.ncbi.nlm.nih.gov/pubmed/155777>

- Sausurre, F. de. (2006). *Course in General Linguistics*. USA: Open Courts Publisher.
- Schütz, R. (2009). English made in Brazil. <http://www.sk.com.br/sk-inst.html>. Retrieved October 23, 2012, from <http://www.sk.com.br/sk-inst.html>
- Singh, S. (2001). A pilot study on gender differences in conversational speech on lexical in richness measures. *Literary and Linguistic Computing*, 16(3), 251–264. doi:10.1093/lc/16.3.251
- Snowdon, D. A., Kemper, S. J., Mortimer, J. A., Wekstein, D. R., & Markesbery, W. R. (1996). Linguistic Ability in Early Life and Cognitive Function From the Nun and Alzheimer ' s Disease in Late Life Study. *JAMA : the journal of the American Medical Association*, 275(7), 528–532.
- Talman, R. (2008). *Geometric Mechanics (Google eBook)* (p. 582). John Wiley & Sons. Retrieved from <http://books.google.com/books?id=ESOsKbCIVmIC&pgis=1>
- Tuson, J. (1996). *L'escriptura*. Barcelona: Empúries.
- Tweedie, F. J., & Baayen, R. H. (1998). How variable may a constant be? Measures of Lexical Richness in Perspective. *Computers and Humanities*, 32(5), 323–352.
- Valls-pedret, C., Molinuevo, J. L., & Rami, L. (2010). Diagnóstico precoz de la enfermedad de Alzheimer : fase prodrómica y preclínica. *Revista de Neurologia*, 51(8), 471–480.
- van Velzen, M., & Garrard, P. (2008). From hindsight to insight – retrospective analysis of language written by a renowned Alzheimer's patient. *Interdisciplinary Science Reviews*, 33(4), 278–286. doi:10.1179/174327908X392852
- Vygotsky, L. (1995). *Pensamiento Y Lenguaje* (p. 237). Paidós. Retrieved from http://books.google.com/books?id=kwm_AAAACAAJ&pgis=1
- Warschauer, M. (1996). Comparing Face-to-Face and Electronic Discussion in the Second Language Classroom. *CALICO Journal*, 13(2), 7–26.
- Whitaker, H. A. (1998). Neurolinguistics from the Middle Ages to the Pre-Modern Era " Historical Vignettes. In B. Stemmer & H. A. Whitaker (Eds.), *Handbook of Neurolinguistics* (pp. 27–54). Academic Press.
- Wilcox, R. (University of S. C. (2002). Least squares regression and pearson's correlation. *Applying Contemporary Statistical Techniques* (pp. 173–206). Academic Press.
- Xu, X. (2008). *EXPLORING THE IMPACT OF AGING THROUGH TRANSITIVITY : A SYSTEMIC FUNCTIONAL APPROACH TO CHINESE ELDERS ' NARRATIVES*. The Pennsylvania State University.
- Youmans, G. (1971). Measuring Lexical Style and Competence: The type-token Vocabulary Curve. *Vocabulary*, (10), 1–13.

Zipf, G. K. (1968). *The Psycho-Biology of Language An Introduction to Dynamic Philology*. Cambridge: MIT Press.

Zipf, G. K. (1972). *Human behavior and the principle of least effort. An introduction to human ecology*. *Psychological Bulletin* (Vol. 47). New York: Hafner Publishing Company. doi:10.1037/h0052442



Anexos

Anexo 1

Escala de Deterioro Global de Reisberg

Estadio	Déficit cognitivo	Características Clínicas
GDS-1	Ausencia de Alteración Cognitiva	Ausencia de quejas subjetivas. Ausencia de trastornos evidentes de memoria en la entrevista clínica. Ausencia de trastornos evidentes de la memoria en la entrevista clínica.
GDS-2	Disminución cognitiva muy leve	Quejas subjetivas de defectos de déficit de memoria, sobre todo en las áreas siguientes: <ul style="list-style-type: none">– El paciente olvida los lugares donde ha dejado objetos familiares.– El paciente olvida nombres previamente muy conocidos. No hay evidencia objetiva de defectos de memoria en la entrevista clínica. No hay evidencia de defectos objetivos en el trabajo o en situaciones sociales. Preocupación apropiada respecto a la sintomatología.
GDS-3	Defecto Cognitivo Leve	Aparecen los primeros defectos claros. Manifestaciones en más de una de las siguientes áreas: <ul style="list-style-type: none">– El paciente puede haberse perdido yendo a un lugar no familiar.– Sus compañeros de trabajo son conscientes de su escaso rendimiento laboral.– Su dificultad para evocar palabras y nombres se hace evidente para las personas más próximas.– El paciente puede leer un pasaje de un libro y recordar relativamente poco material.– El paciente puede mostrar una capacidad disminuida para recordar el nombre de las personas que ha conocido recientemente.– El paciente puede haber perdido un objeto de valor o haberlo colocado en un lugar equivocado.– En la exploración clínica puede hacerse evidente un defecto de concentración. Solo en una entrevista intensiva se observa evidencia objetiva de defectos de memoria.

		<p>Los rendimientos del paciente en actividades laborales y sociales exigentes han disminuido.</p> <p>El paciente niega estos defectos o manifiesta desconocerlos.</p> <p>Los síntomas son acompañados por una ansiedad de discreta a moderada.</p>
GDS-4	Defecto Cognitivo Moderado	<p>Defectos claramente definidos en una entrevista clínica cuidadosa.</p> <p>Déficits manifiestos en las áreas siguientes:</p> <ul style="list-style-type: none"> – Disminución del conocimiento de los acontecimientos actuales y recientes. – Cierta déficit en el recuerdo de su historia personal. – Defecto de concentración, que se manifiesta en la sustracción seriada. – Disminución de la capacidad para viajar, controlar la propia economía... <p>Con frecuencia, no se da defecto alguno en las siguientes áreas:</p> <ul style="list-style-type: none"> – Orientación en el tiempo y la persona. – Reconocimiento de las personas y caras familiares. – Capacidad para viajar a lugares familiares. <p>El paciente es incapaz de realizar tareas complejas.</p> <p>La negación es el mecanismo de defensa dominante.</p> <p>Se observa una disminución del afecto y se da un abandono en las situaciones más exigentes.</p>
GDS-5	Defecto Cognitivo Moderado Grave	<p>El paciente no puede sobrevivir mucho tiempo sin algún tipo de asistencia.</p> <p>Durante la entrevista, el paciente es incapaz de recordar aspectos importantes y relevantes de su vida actual:</p> <ul style="list-style-type: none"> – Una dirección o un número de teléfono que no han cambiado desde hace muchos años. – Nombres de familiares próximos (como los nietos). – El nombre de la escuela o el instituto donde estudió. <p>A menudo, el paciente presenta cierta desorientación con respecto al tiempo (fecha, día de la semana, estación del año...) o al lugar.</p> <p>Una persona con educación formal puede tener dificultad para contar hacia atrás desde 40 de 4 en 4 o desde 20 de 2 en 2.</p> <p>En este estadio, el paciente mantiene el conocimiento sobre muchos de los hechos de más interés que le afectan a sí mismo y a otros.</p> <p>Invariablemente, conoce su nombre y, en general, el nombre de su cónyuge e hijos.</p> <p>No requiere asistencia ni en el aseo ni para comer, pero puede tener alguna dificultad en la elección de la indumentaria adecuada.</p>
GDS-6	Defecto Cognitivo Grave	<p>Ocasionalmente, el paciente puede olvidar el nombre del cónyuge, de quien, por otra parte, depende totalmente para sobrevivir.</p>



GDS-7 Defecto
Cognitivo
Muy Grave

Además, desconoce gran parte de los acontecimientos y experiencias recientes de su vida.

Mantiene cierto conocimiento de su vida pasada, pero de forma muy fragmentaria.

Por lo general, no reconoce su entorno, el año, la estación...

Puede ser incapaz de contar desde 10 hacia atrás, y a veces hacia adelante.

El paciente requiere cierta asistencia en las actividades cotidianas:

- Puede presentar incontinencia.
- Necesita asistencia para viajar, pero en ocasiones será capaz de viajar a lugares familiares.

Con frecuencia, su ritmo diurno está alterado.

Casi siempre recuerda su nombre.

A menudo sigue siendo capaz de distinguir entre las personas familiares y las no familiares de su entorno.

Tienen lugar cambios emocionales y de personalidad que son bastante variables, entre los que se incluyen:

- Conducta delirante, por ejemplo, acusar a su cónyuge de ser un impostor, hablar con personas imaginarias o con su imagen reflejada en el espejo.
- Síntomas obsesivos, por ejemplo, repetir continuamente actividades de limpieza.
- Síntomas de ansiedad, agitación e incluso una conducta violenta previamente inexistente.
- Abulia cognitiva, por ejemplo, pérdida de deseos por falta del desarrollo suficiente de un pensamiento para determinar una acción propositiva.

A lo largo de esta fase, el paciente va perdiendo todas las capacidades verbales.

Al principio, puede articular palabras y frases, pero su lenguaje es muy limitado.

Al final, no se expresa por medio del lenguaje; solo emite sonidos inarticulados.

Se van perdiendo también las habilidades psicomotoras básicas (por ejemplo, caminar).

Aparece la incontinencia urinaria. El paciente requiere asistencia en el aseo y también para ingerir alimentos.

El cerebro parece incapaz de indicarle al cuerpo lo que tiene que hacer.

A menudo se detectan signos y síntomas neurológicos generalizados y corticales.

Anexo 2

Código	Idioma	Autor	Libro	WT ⁴⁰	Edad
PAA23	Portugués	AA	Uma Lagrima de Mulher	36054	23
PAA24	Portugués	AA	O Mulato	102275	24
PAA25_1	Portugués	AA	A Condessa de Vesper	114316	25
PAA25_2	Portugués	AA	Girandola de Amores	101756	25
PAA27_1	Portugués	AA	A Casa de Pensao	101603	27
PAA27_2	Portugués	AA	Filomena Borges	57053	27
PAA30	Portugués	AA	O Homem	57121	30
PAA33_1	Portugués	AA	O Cortico	87095	33
PAA33_2	Portugués	AA	A Coruja	116613	33
PAA37_2	Portugués	AA	A Mortalha de Alzira	50975	37
PAA38	Portugués	AA	Livro de uma Sogra	57395	38
PCCB29	Portugués	CCB	Misterios de Lisboa	210102	29
PCCB31_1	Portugués	CCB	Homem de Brios	63584	31
PCCB31_2	Portugués	CCB	A Neta do Arcediago	60414	31
PCCB31_3	Portugués	CCB	Onde está a Felicidade?	80134	31
PCCB32_1	Portugués	CCB	Cenas da Foz	56160	32
PCCB32_1	Portugués	CCB	Lágrimas Abençoadas	49032	32
PCCB33_1	Portugués	CCB	Carlota Angela	51312	33
PCCB33_2	Portugués	CCB	Eusébio Macário	32950	33
PCCB33_3	Portugués	CCB	Que Fazem as Mulheres?	43677	33
PCCB33_4	Portugués	CCB	A Vinganca	52103	33
PCCB36_1	Portugués	CCB	Amor de Perdição	52304	36
PCCB36_2	Portugués	CCB	Doze Casamentos Felizes	50419	36
PCCB36_3	Portugués	CCB	O Romance de um Homem Rico	59984	36
PCCB37_1	Portugués	CCB	Coisas Espantosas	50250	37
PCCB37_2	Portugués	CCB	Coração, Cabeça e Estomago	51652	37
PCCB38_1	Portugués	CCB	Agulha em Palheiro	47528	38
PCCB38_2	Portugués	CCB	Anos de Prosa	51223	38
PCCB38_3	Portugués	CCB	Aventuras de Basílio Fernandes	47288	38
PCCB38_4	Portugués	CCB	Memórias de Guilherme Amaral	40270	38
PCCB39_1	Portugués	CCB	Amor de Salvação	54216	39
PCCB39_2	Portugués	CCB	Bom Jesus do Monte	32532	39

⁴⁰ Cantidad total de palabras lematizadas, mejor explicada en el capítulo 3.

PCCB39_3	Portugués	CCB	A Filha do Doutor Negro	64819	39
PCCB39_4	Portugués	CCB	Vinte Horas de Liteira	60165	39
PCCB40	Portugués	CCB	O Esqueleto	60853	40
PCCB41_1	Portugués	CCB	A Enjeitada	56016	41
PCCB41_2	Portugués	CCB	O Olho de Vidro	48320	41
PCCB41_3	Portugués	CCB	A Queda de um Anjo	54065	41
PCCB42_1	Portugués	CCB	A bruxa do Monte Córdova	53082	42
PCCB42_2	Portugués	CCB	A Doida do Candal	50582	42
PCCB43_1	Portugués	CCB	Os Mistérios de Fafe	59621	43
PCCB43_2	Portugués	CCB	O Retrato de Ricardina	53678	43
PCCB44	Portugués	CCB	Os Brilhantes do Brasileiro	51068	44
PCCB45	Portugués	CCB	A Mulher Fatal	47024	45
PCCB47_1	Portugués	CCB	O Carrasco de Victor Hugo José Alves	43854	47
PCCB47_2	Portugués	CCB	A Infanta Capelista	25856	47
PCCB48	Portugués	CCB	O dêmónio do Ouro	48849	48
PCCB49	Portugués	CCB	O Regicida	49166	49
PCCB50	Portugués	CCB	A Filha do Regicida	55557	50
PCCB51	Portugués	CCB	Novelas do Minho VVI - O Filho Natural	21235	51
PCCB52_1	Portugués	CCB	Novelas do Minho VII e VIII - Maria Moisés	22009	52
PCCB52_2	Portugués	CCB	Novelas do Minho - Aviúva do Enforcado	38432	52
PCCB57	Portugués	CCB	A Brasileira do Prazins	62565	57
PCCB61	Portugués	CCB	Os vulcões de Lama	38635	61
PEQ25	Portugués	EQ	O Mistério da Estrada de Sintra	71335	25
PEQ30	Portugués	EQ	O Crime do Padre Amaro	152549	30
PEQ33	Portugués	EQ	O Primo Basílio	128553	33
PEQ35	Portugués	EQ	O Mandarim	21829	35
PEQ42	Portugués	EQ	A Relíquia	93096	42
PEQ43	Portugués	EQ	Os Maias	232828	43
PEQ45	Portugués	EQ	A Campanha Alegre	63643	45
PEQ55	Portugués	EQ	A Ilustre Casa de Ramires	114733	55
PEV28	Portugués	EV	Clarissa	53236	28
PEV30	Portugués	EV	Caminhos Cruzados	99106	30
PEV33	Portugués	EV	Olhai os Lírios do Campo	100825	33
PEV44	Portugués	EV	O Tempo e o Vento - OContinente	284416	44
PEV46	Portugués	EV	O Tempo e o Vento - O Retrato	245199	46
PEV56	Portugués	EV	O Tempo e o Vento - Arquipélago	227222	56
PEV60	Portugués	EV	Senhor Embaixador	163356	60
PEV62	Portugués	EV	O Prisioneiro	56792	62
PEV66	Portugués	EV	Incidente em Antares	159335	66

PEV68	Portugués	EV	Solo de Clarineta 01	118407	68
PJA18	Portugués	JA	O País do Carnaval	34304	18
PJA21	Portugués	JA	Cacau	24811	21
PJA22	Portugués	JA	Suor	32512	22
PJA23	Portugués	JA	Jubiabá	92746	23
PJA24	Portugués	JA	Mar Morto	80257	24
PJA25	Portugués	JA	Capitães de Areia	80339	25
PJA30	Portugués	JA	O Cavaleiro da Esperança	137807	30
PJA31	Portugués	JA	Terras do Sem Fim	102732	31
PJA34	Portugués	JA	Seara Vermelha	105791	34
PJA40_1	Portugués	JA	Subterrâneos 1 - Ásperos Tempos	112968	40
PJA40_2	Portugués	JA	Subterrâneos 2 - Agonia da Noite	115863	40
PJA41	Portugués	JA	Subterrâneos 3 - Luz no Fim do Tunel	134484	41
PJA46	Portugués	JA	Gabriela	147199	46
PJA49	Portugués	JA	Velhos Marinheiros	94831	49
PJA52	Portugués	JA	Pastores da Noite	117017	52
PJA54	Portugués	JA	Dona Flor e Seus Dois Maridos	183928	54
PJA57	Portugués	JA	Tenda dos Milagres	114420	57
PJA60	Portugués	JA	Tereza Batista Cansada de Guerra	181496	60
PJA65	Portugués	JA	Tieta do Agreste	156158	65
PJA67	Portugués	JA	Farda, Fardão, Camisola de Dormir	73133	67
PJA72	Portugués	JA	Tocaia Grande	173595	72
PJS25	Portugués	JS	Terra do Pecado	88248	25
PJS51	Portugués	JS	A Bagagem do Viajante	45685	51
PJS55	Portugués	JS	Manual de Pintura e Caligrafia	78557	55
PJS58	Portugués	JS	Levantado do Chão	124820	58
PJS60	Portugués	JS	Memorial do Convento	128280	60
PJS64_1	Portugués	JS	Jangada de Pedra	109422	64
PJS64_2	Portugués	JS	O Ano da Morte de Ricardo Reis	150055	64
PJS67	Portugués	JS	História do Cerco de Lisboa	118201	67
PJS69	Portugués	JS	O Evangelho Segundo Jesus Cristo	150556	69
PJS73	Portugués	JS	Ensaio Sobre a Cegueira	112744	73
PJS75	Portugués	JS	Todos os Nomes	86963	75
PJS78	Portugués	JS	A Caverna	119608	78
PJS80	Portugués	JS	O Homem Duplicado	109369	80
PJS82	Portugués	JS	Ensaio Sobre a Lucidez	112083	82
PJS83	Portugués	JS	Intermitências da Morte	70456	83
PJS86	Portugués	JS	A Viagem do Elefante	62441	86
PJS87	Portugués	JS	Caim	44689	87

PMA33	Portugués	MA	Ressureição	38663	33
PMA34	Portugués	MA	Histórias da Meia Noite	44181	34
PMA35	Portugués	MA	A Mão e a Luva	37383	35
PMA37	Portugués	MA	Helena	59995	37
PMA39	Portugués	MA	Iaiá Garcia	60481	39
PMA42	Portugués	MA	Memórias Póstumas de Brás Cubas	65847	42
PMA46	Portugués	MA	A Casa Velha	23980	46
PMA52	Portugués	MA	Quincas Borba	81979	52
PMA60	Portugués	MA	Dom Casmurro	69866	60
PMA65	Portugués	MA	Esaú e Jacó	76268	65
PMA67	Portugués	MA	Relíquias da Casa Velha	39493	67
PMA69	Portugués	MA	Memorial Aires	54536	69
EACD28_1	Inglés	ACD	Beyond the City	39783	28
EACD28_2	Inglés	ACD	Study in Scarlet	43628	28
EACD29	Inglés	ACD	Micah Clarke	174913	29
EACD30	Inglés	ACD	The Mystery of Cloomber	49137	30
EACD31_1	Inglés	ACD	The Captain of the Polestar	84533	31
EACD31_2	Inglés	ACD	The Firm of Girdlestone	138701	31
EACD31_3	Inglés	ACD	The Sign of the Four	43271	31
EACD31_4	Inglés	ACD	The White Company	151120	31
EACD32_1	Inglés	ACD	The Doings of RafflesHaw	38147	32
EACD32_2	Inglés	ACD	The Great Shadow	49982	32
EACD32_3	Inglés	ACD	The Refugees	123399	32
EACD35_1	Inglés	ACD	Rounded the Red Lamp	68882	35
EACD35_2	Inglés	ACD	The Memoirs of Sherlock Holmes	88032	35
EACD36	Inglés	ACD	Rodney Stone	92063	36
EACD37	Inglés	ACD	Uncle Bernac	57797	37
EACD39	Inglés	ACD	The Tragedy of the Korosko	47005	39
EACD40	Inglés	ACD	Duet	68086	40
EACD41_1	Inglés	ACD	The Great Boer War	221563	41
EACD41_2	Inglés	ACD	The Green Flag	76070	41
EACD43_1	Inglés	ACD	The Hound of the Baskervilles	59523	43
EACD43_2	Inglés	ACD	The War in South Africa	59142	43
EACD44	Inglés	ACD	The Exploits Brigadier Gerard	74747	44
EACD46_1	Inglés	ACD	Sir Nigel	130902	46
EACD46_2	Inglés	ACD	The Return of Sherlock Holmes	113306	46
EACD48	Inglés	ACD	Through the Magic Door	47127	48
EACD49	Inglés	ACD	Round Fire Stories	101484	49
EACD53	Inglés	ACD	The Lost World	76174	53

EACD54	Inglés	ACD	The Poison Belt	29491	54
EACD56	Inglés	ACD	The Valley of Fear	58556	56
EACD57	Inglés	ACD	The British Campaign France and Flanders	68466	57
EACD59	Inglés	ACD	Danger	58051	59
EACD60	Inglés	ACD	The Vital Message	28826	60
EACD62	Inglés	ACD	The Coming of the Fairies	33189	62
EACD64	Inglés	ACD	Tales of Terror and Mystery	75422	64
EACD67	Inglés	ACD	The History of Spiritualism	91543	67
ECD24_1	Inglés	CD	Sketches by Boz	256353	24
ECD24_2	Inglés	CD	The Pickwick Papers	304406	24
ECD25	Inglés	CD	Mudfog	30753	25
ECD27_1	Inglés	CD	Nicholas Nickeby	326170	27
ECD27_2	Inglés	CD	Oliver Twist	159866	27
ECD29_1	Inglés	CD	Barnaby Rudge	257851	29
ECD29_2	Inglés	CD	Master Humphreys Clock	46165	29
ECD29_3	Inglés	CD	The Old Curiosity Shop	220090	29
ECD31	Inglés	CD	The Christmas Carol	28872	31
ECD32_1	Inglés	CD	Life and Adventures Martin Chuzzlewit	340984	32
ECD32_2	Inglés	CD	The Chimes	31249	32
ECD33	Inglés	CD	The Cricketon the Hearth	32141	33
ECD34	Inglés	CD	The Battle of Life	30105	34
ECD36	Inglés	CD	theHauntedManandGhostsBargain	34386	36
ECD37	Inglés	CD	dombeyandSon	360238	37
ECD38	Inglés	CD	davidCopperfield	359653	38
ECD41	Inglés	CD	bleakHouse	358920	41
ECD42	Inglés	CD	hardTimes	104493	42
ECD45	Inglés	CD	littleDorrit	341335	45
ECD47	Inglés	CD	taleofTwoCities	136965	47
ECD49	Inglés	CD	greatExpectations	187248	49
ECD52	Inglés	CD	ourMutualFriend	330877	52
ECD57	Inglés	CD	theMysteryofEdwinDrood	95001	57
EFHB27	Inglés	FHB	thatLassoLowries	73231	27
EFHB28_1	Inglés	FHB	Esmeralda	9229	28
EFHB28_2	Inglés	FHB	leMonsieurdeLaPetitDame	9966	28
EFHB28_3	Inglés	FHB	Seth	8203	28
EFHB28_4	Inglés	FHB	Theo	33922	28
EFHB29	Inglés	FHB	lindsaysLuck	27159	29
EFHB32	Inglés	FHB	fairBarbarian	42129	32
EFHB35	Inglés	FHB	Vagabondia	88895	35

EFHB37	Inglés	FHB	littleLordFauntleroy	59503	37
EFHB38	Inglés	FHB	SaraCrewe	16911	38
EFHB40	Inglés	FHB	thePrettySisterofJose	20348	40
EFHB41	Inglés	FHB	LittleSaintElizabeth	27108	41
EFHB43	Inglés	FHB	giovanni	56871	43
EFHB46	Inglés	FHB	twoLittlePilgrims	40140	46
EFHB47	Inglés	FHB	aLadyofQuality	85012	47
EFHB48	Inglés	FHB	hisGraceOsmonde	107998	48
EFHB50	Inglés	FHB	inconnectionwithDeWilloughbyClaim	129397	50
EFHB52	Inglés	FHB	makingOfMarchionessMethodsLadyWalderhurst	78994	52
EFHB55	Inglés	FHB	InTheClosedRoom	11453	55
EFHB56	Inglés	FHB	littleprincess	67600	56
EFHB57_1	Inglés	FHB	theDawnofaTomorrow	19093	57
EFHB57_1	Inglés	FHB	theRackettyPackettyHouse	9016	57
EFHB58	Inglés	FHB	theShuttle	215921	58
EFHB60	Inglés	FHB	theLandoftheBlueFlower	7768	60
EFHB61	Inglés	FHB	secretGarden	82497	61
EFHB64	Inglés	FHB	tembaron	185317	64
EFHB66	Inglés	FHB	theLostPrince	100749	66
EFHB67	Inglés	FHB	theLittleHunchbackZia	6604	67
EFHB68	Inglés	FHB	theWhitePeople	21523	68
EFHB73_1	Inglés	FHB	headHouseCoombe	112959	73
EFHB73_2	Inglés	FHB	robin	100827	73
EHB39_1	Inglés	HB	altonSomasco	126177	39
EHB39_2	Inglés	HB	mistressBonaventure	116613	39
EHB40	Inglés	HB	cattleBaronsDaughter	107138	40
EHB41_1	Inglés	HB	DelilahoftheSnows	105343	41
EHB41_2	Inglés	HB	forJacinta	100181	41
EHB41_3	Inglés	HB	theDustofConflict	119435	41
EHB41_4	Inglés	HB	winstonPrairie	98276	41
EHB42_1	Inglés	HB	aDamagedReputation	102442	42
EHB42_2	Inglés	HB	byRightofPurchase	94465	42
EHB42_3	Inglés	HB	LongOdds	103104	42
EHB42_4	Inglés	HB	ThriceArmed	105258	42
EHB43_1	Inglés	HB	greaterPower	102600	43
EHB43_2	Inglés	HB	lorimerNorthwest	121606	43
EHB44_1	Inglés	HB	aPrairieCourtship	95496	44
EHB44_2	Inglés	HB	goldTrail	104754	44
EHB44_3	Inglés	HB	hawtreysDeputy	109645	44

EHB44_4	Inglés	HB	mastersWheatlands	106460	44
EHB44_5	Inglés	HB	theBoyRanchersofPugetSound	104259	44
EHB44_6	Inglés	HB	thurstonOrchardValley	99275	44
EHB45	Inglés	HB	vaneTimberlands	98432	45
EHB46_1	Inglés	HB	longPortage	93497	46
EHB46_2	Inglés	HB	ranchingSylvia	101890	46
EHB47_1	Inglés	HB	forTheAllisonHonor	97117	47
EHB47_2	Inglés	HB	prescottSaskatchewan	96679	47
EHB48_1	Inglés	HB	buccaneerFarmer	99995	48
EHB48_2	Inglés	HB	intriguers	65657	48
EHB48_3	Inglés	HB	theLeagueoftheLeopard	90016	48
EHB48_4	Inglés	HB	theSecretoftheReef	93393	48
EHB49_1	Inglés	HB	HardingofAllenwood	88336	49
EHB49_2	Inglés	HB	theCoastofAdventure	89915	49
EHB50_1	Inglés	HB	brandonEngineers	93963	50
EHB50_2	Inglés	HB	JohnstoneoftheBorder	88905	50
EHB51_1	Inglés	HB	blakesBurden	92255	51
EHB51_2	Inglés	HB	carmensMessenger	95973	51
EHB51_3	Inglés	HB	girlKellers	99840	51
EHB52_1	Inglés	HB	lureNorth	82375	52
EHB52_2	Inglés	HB	theProtector	78801	52
EHB53_1	Inglés	HB	partnersOutTrail	89782	53
EHB53_2	Inglés	HB	WyndhamsPal	82421	53
EHB54_1	Inglés	HB	theWildernessMine	88389	54
EHB54_2	Inglés	HB	wildernessMine	88452	54
EHB55	Inglés	HB	listersGreatAdventure	78857	55
EHB56	Inglés	HB	northwest	74277	56
EHB61	Inglés	HB	theDarkRoad	72118	61
EHC33	Inglés	HC	undertheMeteorFlag	142107	33
EHC34_1	Inglés	HC	thePirateIsland	114501	34
EHC34_2	Inglés	HC	theVoyageoftheAurora	87787	34
EHC35	Inglés	HC	theCongoRovers	101887	35
EHC36	Inglés	HC	theLogFlyinffish	105464	36
EHC37	Inglés	HC	theRoversSecret	122569	37
EHC38	Inglés	HC	theMissingMerchantman	107400	38
EHC44	Inglés	HC	thePirateSlaver	103501	44
EHC46_1	Inglés	HC	forTreasureBound	104236	46
EHC46_2	Inglés	HC	theLogPrivateersman	102654	46
EHC47	Inglés	HC	pirateoftheCaribbees	85448	47

EHC48	Inglés	HC	theCastaways	81167	48
EHC55_1	Inglés	HC	acrosstheSpanishMain	94446	55
EHC55_2	Inglés	HC	dickLesliesLuck	118398	55
EHC57	Inglés	HC	withAirshipSubmarine	101657	57
EHC58_1	Inglés	HC	middyinCommand	104835	58
EHC58_2	Inglés	HC	undertheChilianFlag	84642	58
EHC59_1	Inglés	HC	harryEscombe	91806	59
EHC59_2	Inglés	HC	theCruisioftheThetis	104276	59
EHC60_1	Inglés	HC	middyoftheSlaveSquadron	101593	60
EHC60_2	Inglés	HC	overdue	83457	60
EHC61_1	Inglés	HC	middyoftheKing	105716	61
EHC61_1	Inglés	HC	theAdventuresofDickMaitland	84755	61
EHC62_1	Inglés	HC	turnedAdrift	88788	62
EHC62_2	Inglés	HC	twoGallantSonsDevon	108794	62
EHC63_1	Inglés	HC	theCruiseNonsuchBuccaneer	103409	63
EHC63_2	Inglés	HC	theFirstMate	78978	63
EHC63_3	Inglés	HC	throughVeldForest	104444	63
EHC64_1	Inglés	HC	chineseCommand	103257	64
EHC64_2	Inglés	HC	pairAdventuresSearchElDorado	92241	64
EHC65	Inglés	HC	undertheEnsignRisingSun	106585	65
EHC71	Inglés	HC	theStrangeAdventuresEricBlackburn	75866	71
EIM35	Inglés	IM	undertheNet	101113	35
EIM37	Inglés	IM	FlightfromtheEnchanter	107723	37
EIM38	Inglés	IM	theSandcastle	115139	38
EIM39	Inglés	IM	theBell	116223	39
EIM42	Inglés	IM	ASeveredHead	74760	42
EIM43	Inglés	IM	AnUnofficialRose	112039	43
EIM44	Inglés	IM	theUnicorn	93664	44
EIM45	Inglés	IM	TheItalianGirl	48520	45
EIM46	Inglés	IM	theRedeandtheGreen	103186	46
EIM47	Inglés	IM	theTimeoftheAngels	81257	47
EIM49	Inglés	IM	theNiceandtheGood	123240	49
EIM50	Inglés	IM	BrunosDream	94902	50
EIM51	Inglés	IM	FairlyHonourableDefeat	147784	51
EIM52	Inglés	IM	AnAccidentalMan	153605	52
EIM54	Inglés	IM	theBlackPrince	119151	54
EIM55	Inglés	IM	theSacredandProfaneLoveMachine	138215	55
EIM59	Inglés	IM	theSeatheSea	203242	59
EIM61	Inglés	IM	NunsandSoldiers	192750	61

EIM66	Inglés	IM	goodApprentice	203960	66
EIM68	Inglés	IM	theBookAndTheBrotherHood	220582	68
EIM70	Inglés	IM	theMessagetoThePlanet	248210	70
EIM74	Inglés	IM	theGreenKnightfinal	195502	74
EIM76	Inglés	IM	JacksonDilemma	89835	76
EJA26	Inglés	JA	conversationsonBible	31713	26
EJA34_1	Inglés	JA	rolloatPlay	32150	34
EJA34_2	Inglés	JA	rolloatWork	31222	34
EJA36_1	Inglés	JA	CalebCountry	29154	36
EJA36_2	Inglés	JA	rolloExperiments	29640	36
EJA36_3	Inglés	JA	rolloMuseum	31520	36
EJA38_1	Inglés	JA	jonasFarm	30647	38
EJA38_2	Inglés	JA	rolloPhilosophy	32982	38
EJA40_1	Inglés	JA	forests	28974	40
EJA40_2	Inglés	JA	marcoPolo-vermont	38742	40
EJA45_1	Inglés	JA	CharlesI	51255	45
EJA45_2	Inglés	JA	maryQueenScots	50977	45
EJA46_1	Inglés	JA	alfred	48697	46
EJA46_2	Inglés	JA	charlesII	56559	46
EJA46_3	Inglés	JA	elizabeth	48402	46
EJA46_4	Inglés	JA	hannibal	55543	46
EJA46_5	Inglés	JA	julius	50799	46
EJA46_6	Inglés	JA	williamConqueror	53613	46
EJA47_1	Inglés	JA	darius	53044	47
EJA47_2	Inglés	JA	maryErskin	38403	47
EJA47_3	Inglés	JA	xerxes	56271	47
EJA48	Inglés	JA	cleopatra	59393	48
EJA50_1	Inglés	JA	nero	58030	50
EJA50_2	Inglés	JA	pyrrhus	57632	50
EJA50_3	Inglés	JA	rolloonAtlantic	44897	50
EJA51_1	Inglés	JA	rolloinLondon	41015	51
EJA51_2	Inglés	JA	rolloinParis	45020	51
EJA51_3	Inglés	JA	rolloinSwitzerland	42927	51
EJA52_1	Inglés	JA	rolloinScotland	40819	52
EJA52_2	Inglés	JA	rolloonRhine	40924	52
EJA53_1	Inglés	JA	rolloinGeneva	42462	53
EJA53_2	Inglés	JA	rolloinHolland	40117	53
EJA54	Inglés	JA	richardI	61104	54
EJA55_1	Inglés	JA	richardII	62269	55

EJA55_2	Inglés	JA	rolloinNaples	39899	55
EJA55_3	Inglés	JA	rolloinRome	40407	55
EJA56	Inglés	JA	peterGreat	71062	56
EJA57	Inglés	JA	genghis	63506	57
EJA58	Inglés	JA	margaretOfAnjou	55293	58
EJA62_1	Inglés	JA	maryGayorWorkforSummer	22985	62
EJA62_2	Inglés	JA	maryGayorWorkforWinter	34761	62
EJA67	Inglés	JA	hubert	41179	67
EJA68_1	Inglés	JA	gentle	93362	68
EJA68_2	Inglés	JA	waterandLand	76826	68
EJA69	Inglés	JA	force	75837	69
EJFC31	Inglés	JFC	precaution	140086	31
EJFC32	Inglés	JFC	theSpy	149237	32
EJFC34_1	Inglés	JFC	talesforFifteen	39912	34
EJFC34_2	Inglés	JFC	thePilot	157387	34
EJFC34_3	Inglés	JFC	thePioneers	170499	34
EJFC36_1	Inglés	JFC	lionelLincoln	151011	36
EJFC36_2	Inglés	JFC	theLAsstoftheMohicans	145794	36
EJFC37	Inglés	JFC	thePrairie	166082	37
EJFC39_1	Inglés	JFC	redRover	165312	39
EJFC39_2	Inglés	JFC	theWeptofWish	147614	39
EJFC40	Inglés	JFC	theWaterWitch	156168	40
EJFC42	Inglés	JFC	theBravo	140257	42
EJFC43_1	Inglés	JFC	heidenmauer	144312	43
EJFC43_2	Inglés	JFC	theHeadsman	157300	43
EJFC45	Inglés	JFC	lettertoHisCountrymen	33119	45
EJFC46	Inglés	JFC	theMonikins	141123	46
EJFC47	Inglés	JFC	sketchesSwitzerland02	55905	47
EJFC49_1	Inglés	JFC	homeasFound	155730	49
EJFC49_2	Inglés	JFC	homewardBound	177600	49
EJFC50	Inglés	JFC	pathfinder	178429	50
EJFC51	Inglés	JFC	theDeerslayer	214875	51
EJFC53_1	Inglés	JFC	theTwoAdmirals	180057	53
EJFC53_2	Inglés	JFC	theWingandWing	169322	53
EJFC54_1	Inglés	JFC	autobiographyPocket-Handkerchief	50819	54
EJFC54_2	Inglés	JFC	battleofLakeErie	42189	54
EJFC54_3	Inglés	JFC	nedMyers	94064	54
EJFC54_4	Inglés	JFC	wyandotte	164139	54
EJFC55_1	Inglés	JFC	afloatandAshore	200468	55

EJFC55_2	Inglés	JFC	milesWallingford	168772	55
EJFC56_1	Inglés	JFC	chainbearer	173697	56
EJFC56_2	Inglés	JFC	satanstoe	174929	56
EJFC57_1	Inglés	JFC	jackTier	181034	57
EJFC57_2	Inglés	JFC	livesDistinguishedAmericanNavalOfficers	147214	57
EJFC58	Inglés	JFC	theCrater	179320	58
EJFC59	Inglés	JFC	oakOpenings	174261	59
EJFC60	Inglés	JFC	theSeaLions	169008	60
EJFC61	Inglés	JFC	theWaysoftheHour	164159	61
ERWC31	Inglés	RWC	makerofMoons	81346	31
ERWC32	Inglés	RWC	lorraine	84883	32
ERWC36	Inglés	RWC	cardigan	183013	36
ERWC37	Inglés	RWC	theMaidatArms	97890	37
ERWC39_1	Inglés	RWC	inSearchoftheUnknown	71461	39
ERWC39_2	Inglés	RWC	youngManHurry	65623	39
ERWC40	Inglés	RWC	theReckoning	101446	40
ERWC41	Inglés	RWC	theTracerofLostPersons	54417	41
ERWC42_1	Inglés	RWC	theGreenMouse	47537	42
ERWC42_2	Inglés	RWC	theYoungerSet	140465	42
ERWC43	Inglés	RWC	theFiringLine	128879	43
ERWC44	Inglés	RWC	theDangerMark	131025	44
ERWC45	Inglés	RWC	ailsaPaige	117302	45
ERWC46	Inglés	RWC	theCommonLaw	122766	46
ERWC47_1	Inglés	RWC	BlueBirdWeather	13891	47
ERWC47_2	Inglés	RWC	theStreetsofAscalon	111172	47
ERWC48	Inglés	RWC	theGayRebellion	46608	48
ERWC49	Inglés	RWC	hiddenChildren	160157	49
ERWC50_1	Inglés	RWC	athalie	94116	50
ERWC50_2	Inglés	RWC	police	46424	50
ERWC52_1	Inglés	RWC	barbarians	56344	52
ERWC52_2	Inglés	RWC	theDarkStar	108257	52
ERWC54_1	Inglés	RWC	crimsonTide	93762	54
ERWC54_2	Inglés	RWC	inSecret	75633	54
ERWC55_1	Inglés	RWC	slayerofSouls	62710	55
ERWC55_2	Inglés	RWC	theLittleRedFoot	130156	55
ERWC57	Inglés	RWC	theFlamingJewel	66828	57
EWC24	Inglés	WC	memoirslifeWilliamCollins	168166	24
EWC26	Inglés	WC	antonina	167246	26
EWC27	Inglés	WC	ramblesBeyondRailways	60486	27

EWC28	Inglés	WC	basil	117740	28
EWC30	Inglés	WC	hideandSeek	161142	30
EWC32_1	Inglés	WC	afterDark	137257	32
EWC32_2	Inglés	WC	deadSecret	141868	32
EWC35	Inglés	WC	theQueenofHearts	146227	35
EWC36	Inglés	WC	theWomaninWhite	248673	36
EWC38	Inglés	WC	noName	267663	38
EWC44	Inglés	WC	armadale	302021	44
EWC46	Inglés	WC	theMoonstone	198031	46
EWC48	Inglés	WC	theManandWife	232138	48
EWC50	Inglés	WC	poorMissFinch	164206	50
EWC51_1	Inglés	WC	missorMrs	31204	51
EWC51_2	Inglés	WC	theNewMagdalen	102791	51
EWC53	Inglés	WC	theLawtheLady	141603	53
EWC54	Inglés	WC	theTwoDestinies	89916	54
EWC56	Inglés	WC	theHauntedHotel	63086	56
EWC57_1	Inglés	WC	myLadysMoney	53902	57
EWC57_2	Inglés	WC	theFallenLeaves	134922	57
EWC58	Inglés	WC	jezebelsDaughter	102245	58
EWC59	Inglés	WC	theBlackRobe	108525	59
EWC61	Inglés	WC	heartandScience	129845	61
EWC62	Inglés	WC	iSayNo	121128	62
EWC63_1	Inglés	WC	theEvilGenius	111975	63
EWC63_2	Inglés	WC	theGuiltyRiver	45592	63
EWC65	Inglés	WC	theLegacyofCain	120677	65

Autor	Origen de los libros	Origen de las informaciones
AA	X	A, C, K
CCB	V, W, X	A, C, G
EQ	X	A, C
EV	Z	A, C
JAp	Z	A,C
JS	Z	A, C, Q
MA	X	A, C
ACD	V, W	A, B
CD	V, W	A, B, H
FHB	V, W	A, B, L, P
HB	V	A, B, H
HC	V, W	A, B, F
IM	Y, Z	A, B, O

JAp	V, W	A, B, D
JFC	V, W	A, B, E, J, L, N
RWC	V, W	A, B, I
WC	V, W	A, B, M

Origen de los libros	
V	www.gutenberg.org
W	www.archive.org
X	www.dominiopublico.gov.br
Y	Digitalização
Z	www.esnips.com

Origen de las informaciones de los libros	
A	on book
B	http://en.wikipedia.org
C	http://pt.wikipedia.org
D	http://www.merrycoz.org/bib/ABBOTT.HTM
E	http://external.oneonta.edu/cooper/bibliography/works.html
F	http://www.athelstane.co.uk/collngwd/hc_books.htm
G	http://www.citi.pt/cultura/literatura/romance/c_castelo_branco/bibliografia.html
H	http://www.fantasticfiction.co.uk
I	http://www.isfdb.org/
J	http://www.online_literature.com/cooperj
K	http://www.portalsaofrancisco.com.br/alfa/biografia-aluizio-azevedo/aluisio-azevedo.php
L	http://www.tickledorange.com/
M	http://www.web40571.clarahost.co.uk/wilkie/bibliog/books.htm
N	http://xroads.virginia.edu/~ug02/COOPER/timeworks.html
O	http://fass.kingston.ac.uk/research/iris-murdoch/biography/
P	http://www.booksellerworld.com/frances-hodgson-burnett.htm
Q	http://www.josesaramago.org/

Anexo 3

Cuadro comparativo de las reglas de puntuacion para el inglés y el portugués.

Descripción	Símbolo	Inglés	Portugués
Punto final	.	Se utiliza para marcar el final de una oración que expresa una declaración.	Se utiliza para marcar el final de una oración que expresa una declaración.
Coma	,	<p>En oraciones compuestas:</p> <ul style="list-style-type: none"> - para separar las partes de una oración compuesta, poniendo la coma antes de la conjunción. - Frases con dos verbos o cláusulas verbales unidas por la palabra “and” no suelen incluir una coma antes de la “and”. - para indicar una cláusula no restrictiva o dependiente (por lo general introducida por la “which”). - No utilice la coma con una cláusula restrictiva (por lo general introducida por “that”, y por lo general el tipo de cláusula “needed–which” em que a menudo la coma es usada en exceso y mal empleada). 	<p>Hace una pequeña pausa en el texto escrito, no siempre corresponde a las interrupciones (más arbitrarias) del texto hablado.</p> <p>Se utiliza como una marca de separación para el vocativo, atributos, los elementos de una frase no conectadas por las conjunciones “e”, o, “nem”, oraciones asindéticas coordinadas (no conectadas por conjunciones), las oraciones relativas, oraciones intercaladas, y las cláusulas subordinadas y adversativas introducidas por “mas”, “contudo”, “todavía”, “entretanto” y “porém”.</p>

		<p>Para separar los componentes de una fecha;</p> <p>Para separar ítems de una serie;</p> <p>Antes y después de un nombre de estado cuando se utiliza con un pueblo o ciudad en el texto;</p> <p>al final de una dirección de calle en el texto, cuando esté en el medio de una oración;</p>	
Dos puntos	:	<p>Para introducir una oración completa, más de una frase, una declaración formal, cotización, o habla en un diálogo, una serie o lista.</p> <p>Cuando en una frase puede entrar casi como un punto final;</p> <p>No utilizar los dos puntos entre un verbo y una preposición.</p> <p>Usar entre el lugar de publicación y el nombre del editor en las referencias bibliográficas.</p>	Señala una pausa para anunciar una citación, un dialogo, una enumeración, una aclaración o un resumen.
Punto y coma	;	<p>Las siguientes palabras se consideran adverbios en lugar de conjunciones y deben estar precedida por un punto y coma cuando se usa entre las cláusulas de una oración compuesta: “then”, “however”, “thus”, “hence”, “indeed”, “yet”, “so”;</p> <p>También se utilizan para unir oraciones completas cuando un período único crearía demasiada cantidad de pausas en la línea de pensamiento;</p>	Señal intermedio entre el punto y la coma, que indica que el significado de la frase se complementará. Representa una pausa más larga que la coma y más corta que el punto. Se utiliza en frases con varias oraciones, algunas de las cuales que ya contienen una o más comas; se usa también para separar clausulas subordinadas; y sustituyendo la coma en la separación de la clausula coordinada adversativa de la oración principal.
Signo de exclamación	!	Se utiliza al final de una oración o una frase breve que	Se utiliza al final de cualquier frase que expresa sentimientos, emociones, dolor, asombro, sorpresa, ironía y



		<p>expresa sentimiento muy fuerte;</p> <p>Son muy normales en el tipo de escritura que intenta representar el lenguaje hablada - por ejemplo, en las novelas. No se utiliza en escritas formales;</p> <p>Es habitual después de un exclamación que empiece con “What” o “How”;</p> <p>Para mostrar que una afirmación es muy sorprendente;</p> <p>También es admisible utilizar un signo de exclamación para llamar la atención a una interrupción;</p>	humor.
Puntos suspensivos	...	<p>Evite el uso de los puntos suspensivos, si es posible. Ellos hacen la frase más difícil de leer y entender.</p> <p>Los puntos suspensivos se utilizan adecuadamente para indicar la omisión de texto dentro de una citación, no como una forma de pausa.</p> <p>Cuando los puntos suspensivos se usan dentro de una frase, utilice tres.</p> <p>Cuando los puntos suspensivos se usan entre oraciones, utilice cuatro, que sirve como el período de la primera frase, dependiendo del lugar donde se produce el texto omitido.</p> <p>Siempre utilice espacios entre y alrededor de los puntos</p>	Suele señalar una interrupción del pensamiento, lo que indica que el sentido de la oración es incompleta, o una introducción de suspense, tras lo cual el sentido será completado.

		suspensivos.	
Signo de Interrogación	?	Se coloca al final de una frase que es una pregunta directa. Puede ser insertado en el medio de algo, dentro de paréntesis, para mostrar que algo es incierto.	Se coloca al final de una frase que es una pregunta directa. Puede ser insertado en el medio de algo, dentro de paréntesis, para mostrar que algo es incierto.
Paréntesis		Señala una observación o información incidental intercalada en el texto.	Señala una observación o información incidental intercalada en el texto.
Corchetes		Cuando hay paréntesis dentro de paréntesis, En interpolaciones editoriales; Sustituir palabras en las citas; Transcribir la forma fonéticas de palabras;	Más común el uso en los textos científicos.
Guión	-	Una regla general es que los guiones separan ítems; Utilice el guion doble para indicar una interrupción en el pensamiento y se puede utilizar dentro de una frase para insertar una frase entre paréntesis; Ni un doble guión ni un guión simple deben tener espacios en ambos lados.	Señala el principio y el final de un diálogo, para distinguir cada uno de los interlocutores; las oraciones intercaladas, los resúmenes al final de un texto. También puede sustituir los paréntesis. Separa los extremos de los intervalos.
Comillas	“””, «», ‘ ’	Los elementos que deben estar entre comillas: citas directas, títulos de canciones, poemas cortos, ensayos, programas de televisión y radio, títulos de cuentos, títulos de artículos, partes de libros (capítulos o secciones).	Se utilizan comillas para delimitar citas, para referir los títulos de las obras, para destacar una palabra o frase.

Anexo 4

Penn Treebank Tagset

POS Tag	Description	Símbolo
Fat	Signo de exclamación	!
Fc	Coma	,
Fd	Dos puntos	:
Fg	Guión	-
Fit	Signo de Interrogación	?
Fp	Punto final	.
Fs	Puntos suspensivos	...
Fx	Punto y coma	;

Anexo 5

Algoritmo para la corrección de los textos:

A. Correcciones generales:

- 1) Retirar comillas “ ”, ‘ ’, «» (no substituir por nada);
- 2) Substituir paréntesis y corchetes, de forma que:
 - a. (e [> substituir por _- , por lo que _ es un espacio y – es un guión.
 - b.) e] > substituir, considerando el signo de puntuación que viene después:
 - i. Si es una coma (,) se retira el), o] y la coma y se pone un guión (-).
 - ii. Si es un punto (.), punto y coma (;), dos puntos (:), guión (-), exclamación (!) o interrogación (?), se retira el paréntesis o el corchete y se mantiene el signo existente.

B. Abreviaciones y Acrónimos:

- 1) Las abreviaciones más comunes y que pueden provocar errores (al ser consideradas partes finales de una oración) son las siguientes:
 - a. Portugués:
 - i. Dr. -> Doutor;
 - ii. Sr. -> Senhor;
 - iii. Sra. -> Senhora;
 - iv. V. Exca. -> Vossa Excelência;
 - v. V. Exc.a. -> Vossa Excelência;
 - vi. V. Exc^a. -> Vossa Excelência;
 - vii. S. Exca. -> Sua Excelência;
 - viii. S. Exc.a. -> Sua Excelência;
 - ix. S. Exc^a. -> Sua Excelência;
 - x. V. Sa. -> Vossa Senhoria;
 - xi. V. S.-> Vossa Senhoria;
 - xii. V. S^a. -> Vossa Senhoria;

- xiii. V. Rv.ma -> Vossa Reverendíssima;
- xiv. S. -> São (nem sempre...);
- xv. D. -> Dom ou Dona ou Doutor (problema..);
- xvi. n. -> número;
- xvii. n.o -> número;
- xviii. n.º. -> número;

b. Inglés:

- i. Mr. -> Mister;
- ii. Mrs. -> Mistress;
- iii. Ms. -> Miss;
- iv. Dr. -> Doctor;
- v. St. -> Saint;

2) Los Acrónimos deben tratarse después de las abreviaciones, pues de esta forma ya se habrán eliminado la mayoría. El resto debe ser analizado una a una, sustituyéndolas de la forma siguiente:

- i. X. -> X;
- ii. X.X. -> XX;
- iii. X. X. -> XX;

Si el acrónimo tiene más que dos letras deberá ser tratado del mismo modo, uniendo las letras, eliminando los puntos y los espacios entre ellas. Es necesario observar que si el último punto es un punto final de frase deberá mantenerse.

3) Corrección de los párrafos:

a. Inicio:

- i. Verificar si hay guiones largos al principio de cada párrafo y eliminarlos.
- ii. Verificar si hay signos de puntuación al principio de cada párrafo y eliminarlos. Es necesario tener en cuenta que lo más común es la

presencia de puntos suspensivos (...), pero, en ocasiones, aparecen otros sin propósito.

iii. Retirar los espacios en blanco antes de cada párrafo.

b. Final:

i. Retirar los espacios en blanco después de los signos de puntuación: punto (.), exclamación (!) e interrogación (?) de cada párrafo.

ii. Verificar si existen estos signos de puntuación al final del párrafo: dos puntos (:), punto y coma (;) o coma (,). En estos casos, significa que el párrafo fue separado equivocadamente, de forma que hay que proceder a la unión de las dos partes, sin olvidar de poner un espacio entre el signo y la palabra siguiente.

iii. La tercera situación es encontrar párrafos que no posean ningún signo de puntuación al final. Es el caso, por ejemplo, de los títulos y subtítulos, que normalmente no deben llevar un signo de puntuación al final, pero que en el caso de nuestro análisis es importante delimitar. En el caso de no encontrar signos de puntuación al final del párrafo, se pueden hacer dos cosas:

1. Poner el punto (por ejemplo, en los finales de títulos y subtítulos);

2. Verificar si el texto siguiente forma parte del párrafo, separado equivocadamente. En este caso, unir las partes sin olvidar de insertar un espacio entre las palabras.

4) Corregir los puntos suspensivos. Este signo de puntuación a veces aparece equivocado y debe ser tratado, pues si no aparece de la forma correcta no se considera como un signo de puntuación, provocando graves errores en la estadística al ser un signo muy frecuente, principalmente en portugués:

a. Substituir . _ . (punto espacio punto) -> .. (punto punto).

b. Substituir ..(punto punto) -> ... (puntos suspensivos).

c. Substituir (cuatro puntos) -> ... (puntos suspensivos).

5) Eliminar las repeticiones de signos. La repetición de signos es el error más serio para nuestra estadística de oraciones y pausas lingüísticas, pues, por ejemplo, el ingenio retórico muy común de poner los signos de interrogación y exclamación

(?!) para denotar sorpresa, se contarán como dos signos de puntuación y, consecuentemente, dos pausas lingüísticas:

- a. La primera etapa es eliminar todos los espacios en blanco antes de los signos de puntuación (! ? ; : , .) y en el caso del guión, apenas si aparece si aparece entre dos guines (guión espacio guión).
- b. Luego, eliminar los guiones que aparecen después de los signos de puntuación (! ? ; : , .) separados con un espacio: por ejemplo, !_ – (exclamación espacio guión).
- c. Por último, eliminar las repeticiones de la forma siguiente:

1	2	3	4	5	6	7	8
!!	?!	;!	:!	,!	.!	–!	‘!
!	??	;?	:?	,?	.?	–?	‘?
!;	?;	::	::	,;	.;	–;	‘;
!:	?:	::	::	,:	.:	–:	‘:
!.	?.	;	::	,.	..	–,	‘,
!,	?,	;	::	,.	..	–.	‘.
!–	?–	;–	:–	,–	.–	––	‘–

En las situaciones descritas en las columnas 1, 2, 3, 4 y 5, mantener el primer signo de puntuación. En las situaciones siguientes descritas em las columnas 6, 7 ey8, mantener el último signo de puntuación.

- 6) Al final, es necesario substituir todas las ocurrencias de dos espacios entre cada palabra por un único espacio.

Anexo 6

Modelo de archivo de configuración del Freeling 2.0 para el idioma portugués.

```
##
#### default configuration file for Português analyzer
##

#### General options
Lang=pt
Locale=default

#### Trace options. Only effective if we have compiled with -
DVERBOSE
#
## Possible values for TraceModule (may be OR'ed)
#define SPLIT_TRACE      0x00000001
#define TOKEN_TRACE     0x00000002
#define MACO_TRACE      0x00000004
#define OPTIONS_TRACE   0x00000008
#define NUMBERS_TRACE   0x00000010
#define DATES_TRACE     0x00000020
#define PUNCT_TRACE    0x00000040
#define DICT_TRACE      0x00000080
#define SUFF_TRACE      0x00000100
#define LOCUT_TRACE     0x00000200
#define NP_TRACE        0x00000400
#define PROB_TRACE      0x00000800
#define QUANT_TRACE     0x00001000
#define NEC_TRACE       0x00002000
#define AUTOMAT_TRACE   0x00004000
#define TAGGER_TRACE    0x00008000
#define HMM_TRACE       0x00010000
#define RELAX_TRACE     0x00020000
#define RELAX_TAGGER_TRACE 0x00040000
#define CONST_GRAMMAR_TRACE 0x00080000
#define SENSES_TRACE    0x00100000
#define CHART_TRACE     0x00200000
#define GRAMMAR_TRACE   0x00400000
#define DEP_TRACE       0x00800000
#define UTIL_TRACE      0x01000000

TraceLevel=3
TraceModule=0x0000

## Options to control the applied modules. The input may be
partially
## processed, or not a full analysis may me wanted. The specific
## formats are a choice of the main program using the library, as
well
```

• • •

```

## as the responsibility of calling only the required modules.
## Valid input/output formats are: plain, token, splitted, morfo,
tagged, parsed
InputFormat=plain
OutputFormat=tagged

# consider each newline as a sentence end
AlwaysFlush=no

#### Tokenizer options
TokenizerFile=$FREELINGSHARE/pt/tokenizer.dat

#### Splitter options
SplitterFile=$FREELINGSHARE/pt/splitter.dat

#### Morfo options
AffixAnalysis=yes
MultiwordsDetection=no
NumbersDetection=yes
PunctuationDetection=yes
DatesDetection=yes
QuantitiesDetection=yes
DictionarySearch=yes
ProbabilityAssignment=yes
OrthographicCorrection=no
DecimalPoint=,
ThousandPoint=.
LocutionsFile=$FREELINGSHARE/pt/locucions.dat
QuantitiesFile=$FREELINGSHARE/pt/quantities.dat
AffixFile=$FREELINGSHARE/pt/afixos.dat
ProbabilityFile=$FREELINGSHARE/pt/probabilitats.dat
DictionaryFile=$FREELINGSHARE/pt/dicc.src
PunctuationFile=$FREELINGSHARE/common/punct.dat
ProbabilityThreshold=0.001
#NER options
NERecognition=basic
NPDataFile=$FREELINGSHARE/pt/np.dat
## --- comment lines above and uncomment those below, if you want
## --- a better NE recognizer (higer accuracy, lower speed)
#NERecognition=bio
#NPDataFile=$FREELINGSHARE/pt/ner/ner.dat

#Spelling Corrector config file
CorrectorFile=$FREELINGSHARE/pt/corrector/corrector.dat

## NEC options
NEClassification=no
NECFilePrefix=$FREELINGSHARE/pt/nec/nec

## Sense annotation options (none,all,mfs)
SenseAnnotation=none
SenseFile=$FREELINGSHARE/pt/senses30.src
DuplicateAnalysis=false
UKBConfigFile=$FREELINGSHARE/pt/wsd.dat

#### Tagger options
Tagger=hmm
TaggerHMMFile=$FREELINGSHARE/pt/tagger.dat

```

```
TaggerRelaxFile=$FREELINGSHARE/pt/constr_gram.dat
TaggerRelaxMaxIter=500
TaggerRelaxScaleFactor=670.0
TaggerRelaxEpsilon=0.001
TaggerRetokenize=yes
TaggerForceSelect=tagger

#### Parser options
GrammarFile=$FREELINGSHARE/pt/grammar-dep.dat

#### Dependence Parser options
DepTxalaFile=$FREELINGSHARE/pt/dep/dependences.dat

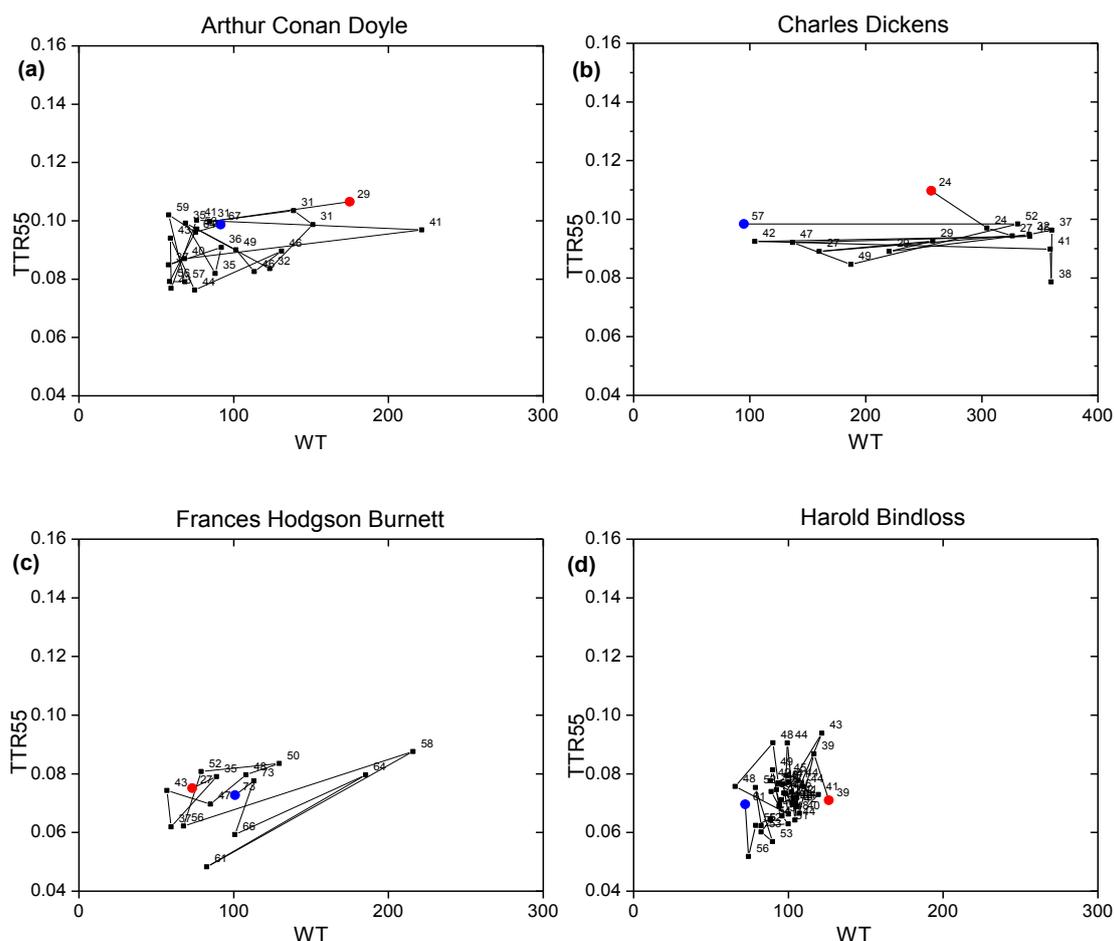
#### Coreference Solver options
CoreferenceResolution=no
CorefFile=$FREELINGSHARE/pt/coref/coref.dat
```



Escritores de Lengua Inglesa

A) TTR55

Para el restante de autores en lengua inglesa, Arthur Conan Doyle, Charles Dickens, Frances Hodgson Burnett, Harold Bindloss, Harry Collingwood, James Fenimore Cooper y Jacob Abbott no se ha encontrado ningun registro del comportamiento de la métrica TTR55 que se ha relacionado con la edad, como demuestra los gráficos de la Figura A7.1.



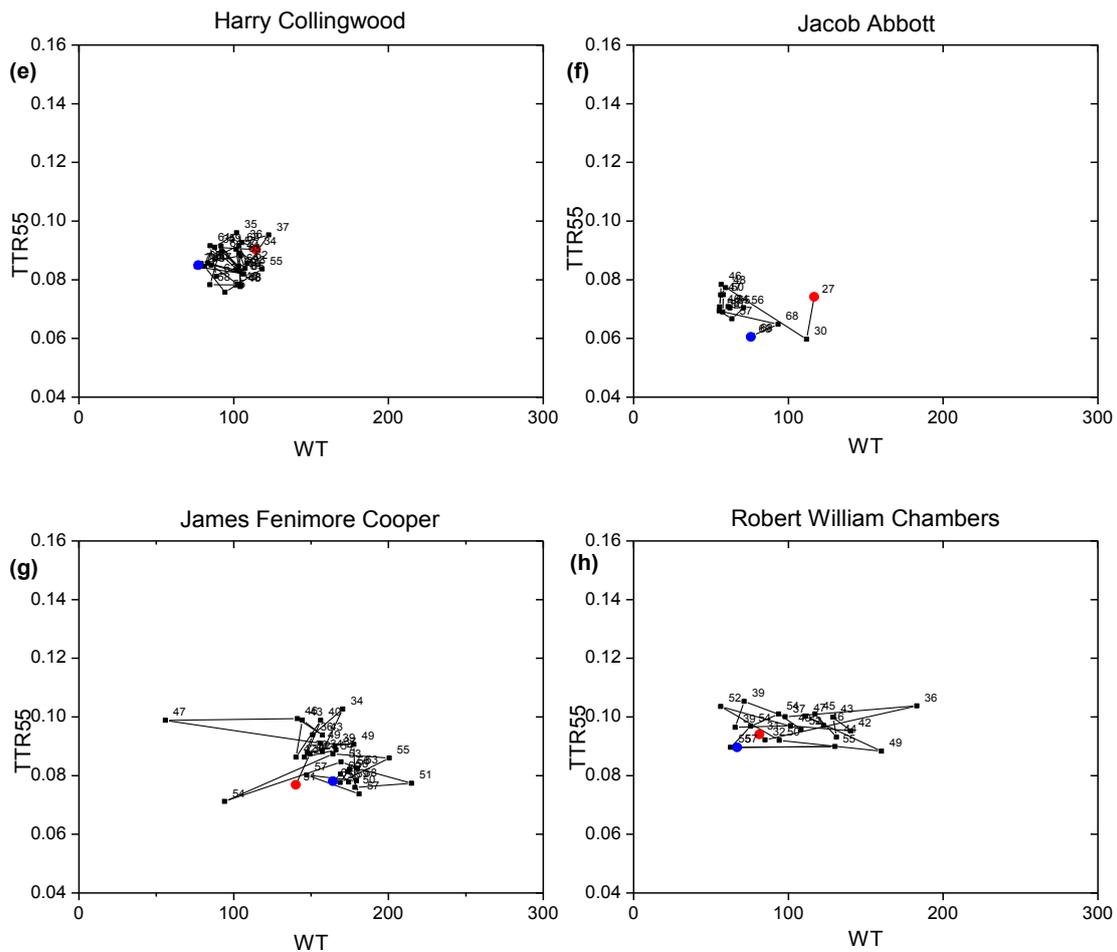


Figura A7.1 SPT en TTR55 y WT para ocho autores en lengua inglesa.

B) WTIR

Para la métrica WTIR, otros autores también tenían libros con diferentes tasas de incorporación de nuevas palabras, similares al libro *Jackson's Dilemma*, pero no se escribieron al final de sus carreras literarias, como el *Northwest* de Harold Bindloss, escrito a la edad de 56 años (Figura A7.2a) y el libro *Across the Spanish Main* de Harry Collingwood, escrito a la edad de 55 años (Figura A7.2b).

El gráfico de los libros de James Fenimore (Figura A7.2c) no mostró ninguna alteración en el trazado que sugiriera una disfunción en el comportamiento en la escritura: solo la concentración de los últimos libros escritos por el autor en la parte inferior del gráfico sugiere que al final de la vida el autor presentó una menor riqueza de vocabulario en los libros que escribió.

...

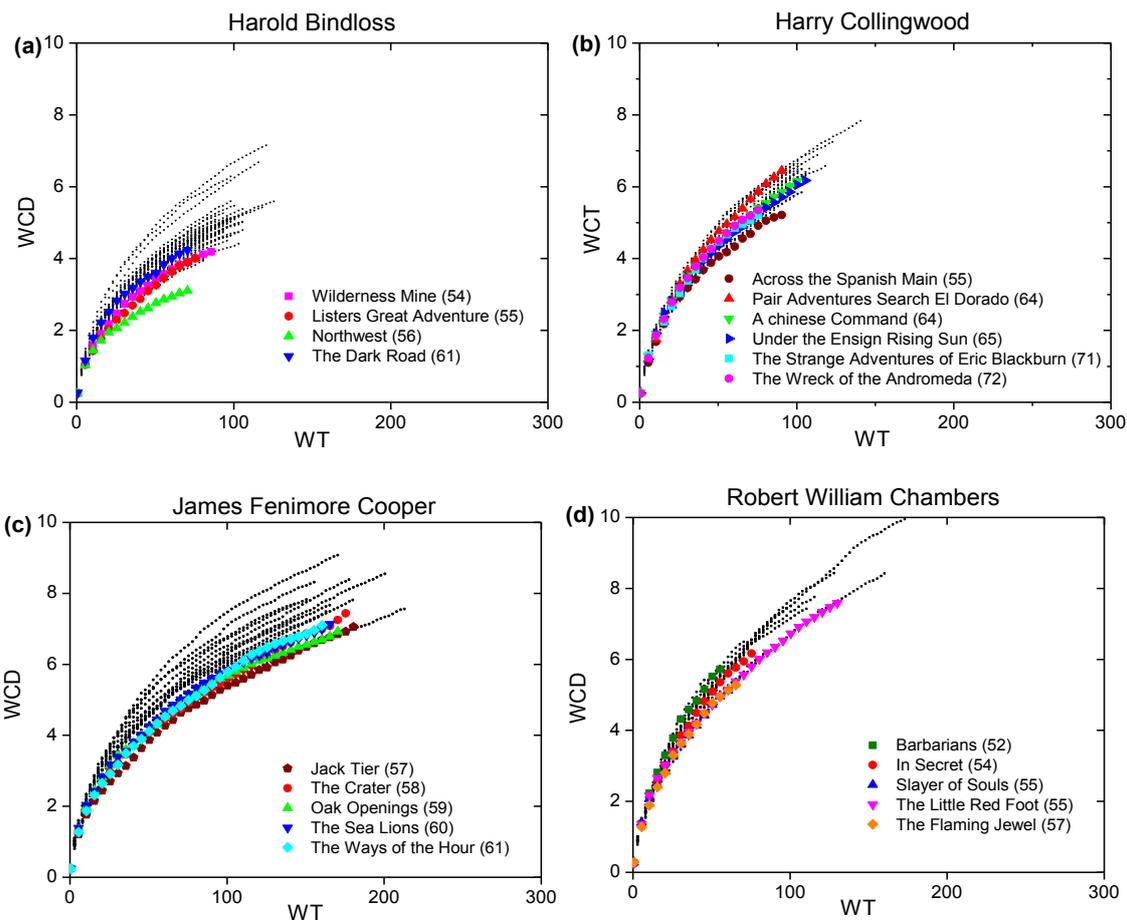


Figura A7. 2 WTIR para cuatro autores en lengua inglesa.

El gráfico de Robert William Chambers (Figura A7.2d) mostró un comportamiento normal, pero los libros analizados aquí no representan la totalidad del período productivo de este autor. El último libro examinado lo escribió a los 57 años, pero produjo, por lo menos, seis novelas más hasta su muerte.

C) WPS

En la Figura A7.3 se presentan los gráficos de SPT para los diez autores de lengua inglesa, Jacob Abbott, Harold Bindloss y Wilkie Collins. Como Iris Murdoch, los escritores Arthur Conan Doyle (Figura A7.3a), Charles Dickens (Figura A7.3b), Frances Hodgson Burnett (Figura A7.3c), Harry Collingwood (Figura A7.3d), James Fenimore Cooper (Figura A7.3e) y Robert William Chambers (Figura A7.3f) no presentaron ningún cambio en su trayectoria que pueda sugerir una tendencia de aumento o disminución de la longitud promedio de las oraciones durante toda su vida.

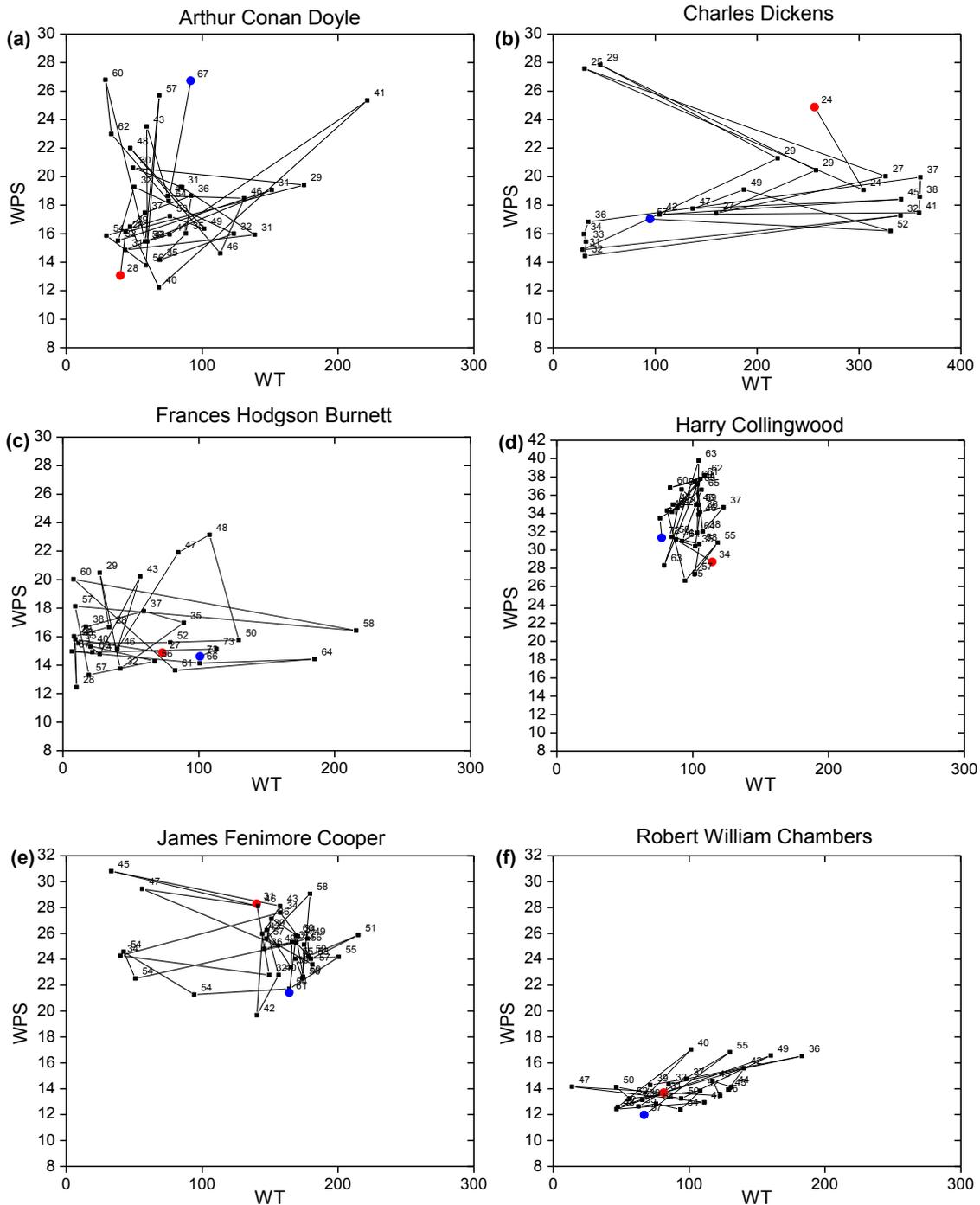


Figura A7.3 SPT del WPS para seis escritores en lengua inglesa.

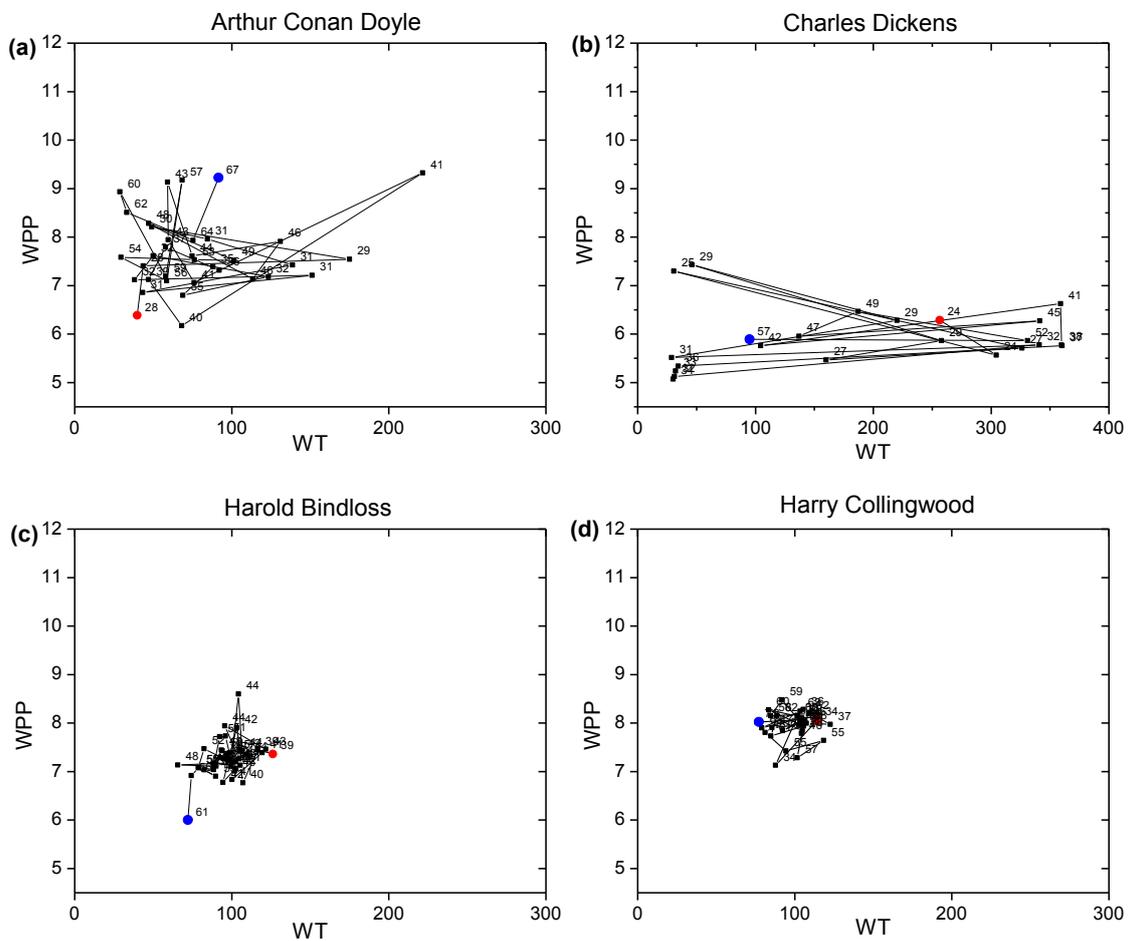
D) WPP

Para la trayectoria del WPP, Harold Bindloss (Figura A7.4c) fue el único escritor en el que se ha detectado el mismo comportamiento de iris Murdoch, o sea, la reducción del valor de WPP con la edad; sin embargo, este resultado debe ser analizado

• • •

con cautela, porque no ha sido posible realizar el estudio con todos los libros del autor porque no eran accesibles en formato electrónico cuando da realización de esta investigación. Tras el último texto incluido en este estudio, el autor escribió cerca de 15 obras más en nueve años de producción literaria, hasta su muerte.

Los escritores Arthur Conan Doyle (Figura A7.4a), Charles Dickens (Figura A7.4b), Harry Collingwood (Figura A7.4d), James Fenimore Cooper (Figura A7.4e), Robert William Chambers (Figura A7.4f) y Wilkie Collins (Figura A7.4g) no muestran ninguna tendencia relacionada con el cambio de estado de la longitud promedio de las oraciones con el envejecimiento.



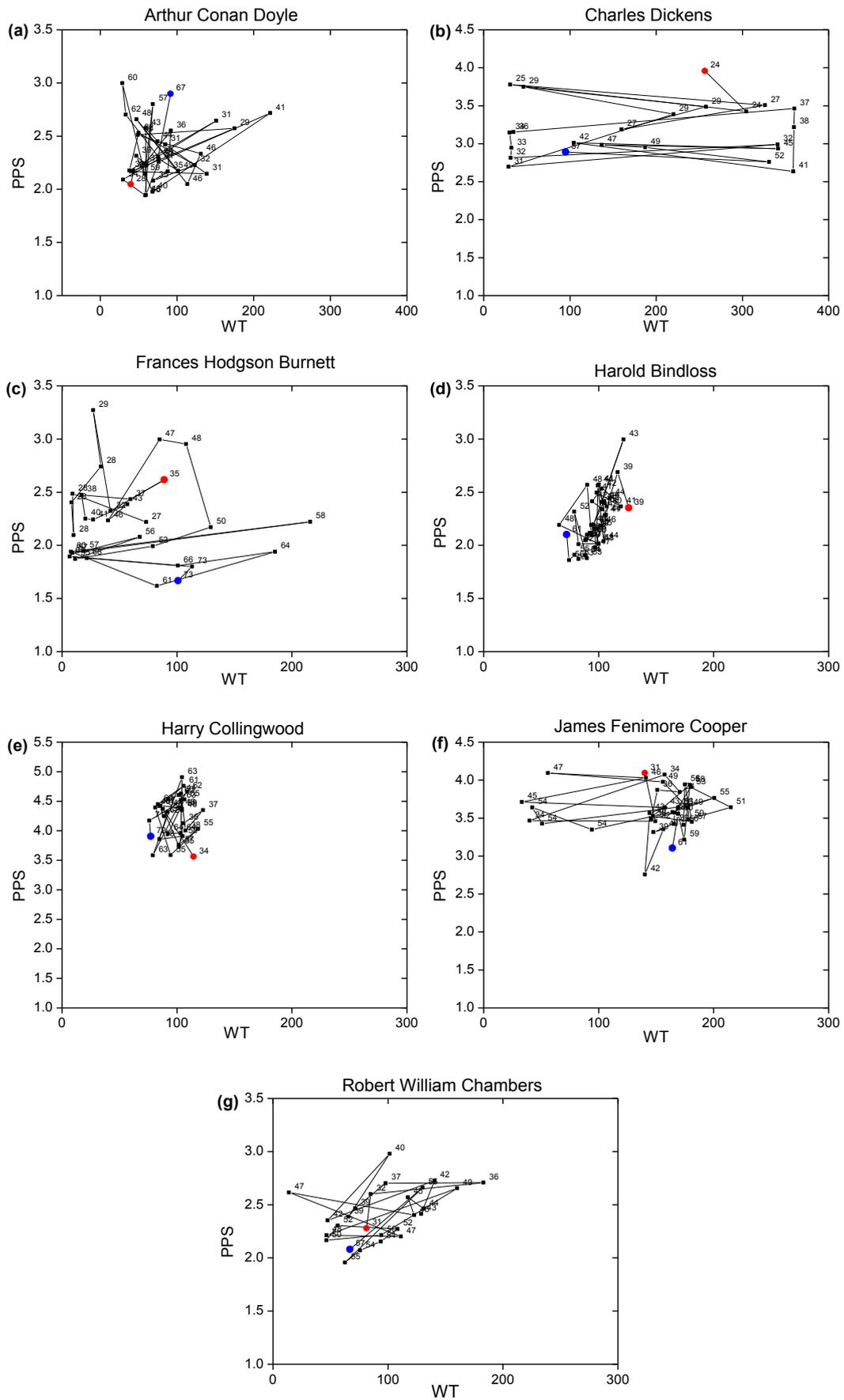


Figura A7.5 SPT del PPS para los escritores en lengua inglesa.

F) Comparación del WPS, WPP y PPS

ARTHUR CONAN DOYLE

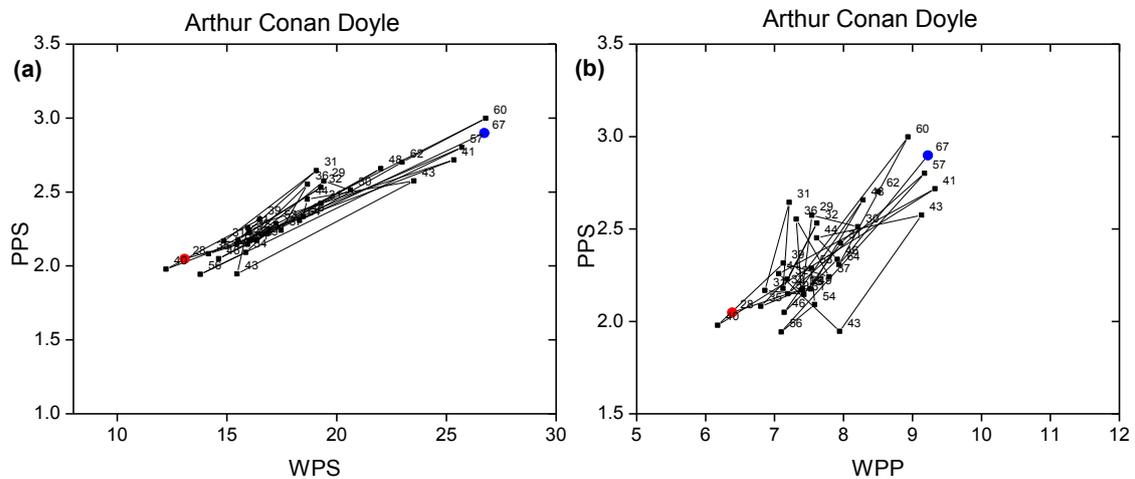


Figura A7.6 Comparación de WPP y WPS (a) y WPP y PPS (b) de Arthur Conan Doyle.

Para el escritor Arthur Conan Doyle, el aumento de la longitud de las oraciones significó un aumento en la cantidad de pausas lingüísticas, pero no se detecta ningún comportamiento de alteración relacionado con el envejecimiento (Figura A7.6a). Esto indicaría que cuando escribe textos con un mayor número de oraciones, incrementa el número de pausas dentro de estas y, además, muestra también pausas más grandes (Figura A7.6b).

CHARLES DICKENS

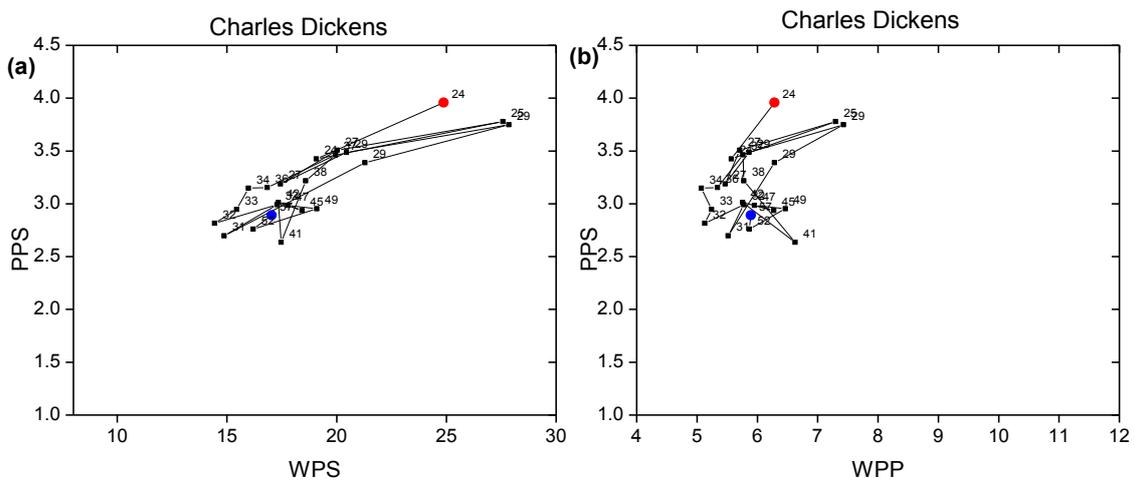


Figura A7.7 Comparación de WPP y WPS (a) y WPP y PPS (b) de Charles Dickens.

Charles Dickens tampoco presenta ningún cambio en su comportamiento respecto a la cantidad de pausas lingüísticas y a la longitud de las oraciones con el envejecimiento (Figura A7.7a). El aumento en la longitud de las oraciones está provocado por el aumento en el número de pausas lingüísticas que las componen, como se puede notar en la tendencia al crecimiento asociado de estos dos parámetros, en la misma figura; sin embargo, las pausas no aumentan su longitud, pues van desde 5 a 7 palabras, como promedio, durante toda su vida (Figura A7.7b). El cambio notado en el WPS y el WPP de la primera figura (tendencia al crecimiento) y la invariabilidad del PPS no tiene ninguna relación con el paso del tiempo y la secuencia con que los libros fueron escritos, lo que se refleja en una ausencia de linealidad en las edades registradas junto a los puntos en el grafico.

HAROLD BINDLOSS

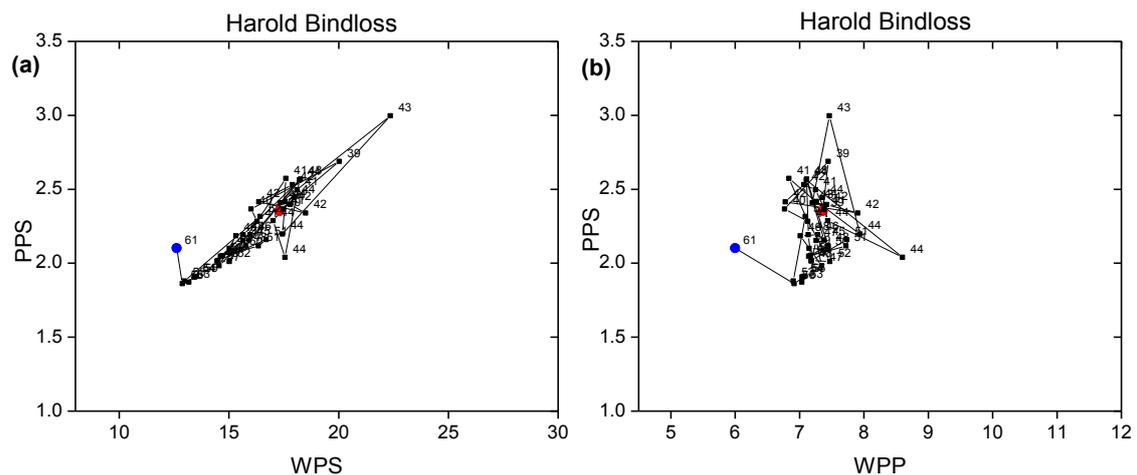


Figura 14.1 Comparación de WPS y WPP (a) y WPP y PPS (b) para Harold Bindloss.

El último libro de Harold Bindloss analizado, escrito a los 61 años, presenta un comportamiento diferente en relación con la fragmentación, porque es el texto con pausas de menor longitud (Figura 14.1a). Pero este no fue el último libro escrito por el autor y, por ello, no es posible extraer conclusiones sobre el comportamiento de su escritura con el envejecimiento.

HARRY COLLINGWOOD

No existe ningún indicio de fragmentación en los textos del escritor Harry Collingwood. El aumento de la longitud de las oraciones se relaciona con el aumento en

la cantidad de pausas (Figura A7.8a), pero estas no cambian en su longitud, quedando entre 7 y 8,5 para todos los textos (Figura A7.8b).

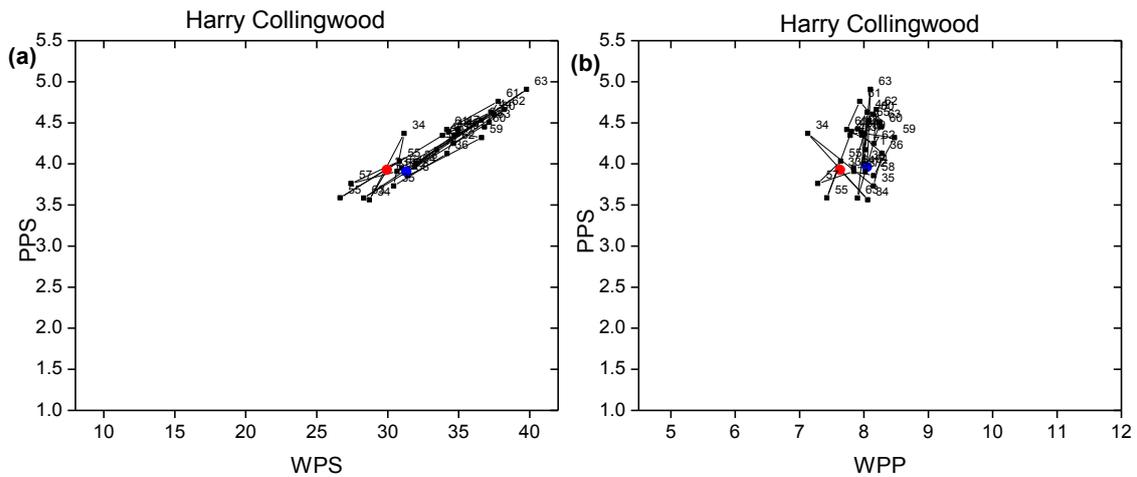


Figura A7.8 Comparación de WPS y WPP (a) y WPP y PPS (b) para Harry Collingwood.

JACOB ABBOTT

Los libros del escritor Jacob Abbott mantienen una relación de crecimiento entre la longitud de las oraciones y la cantidad de pausas que las componen, incluyendo incluso un comportamiento diferente y característico en los últimos escritos (Figura A7.9a). Al final de su vida el escritor publicó libros con oraciones mayores, compuestas por más pausas y con un aumento de longitud (Figura A7.9b).

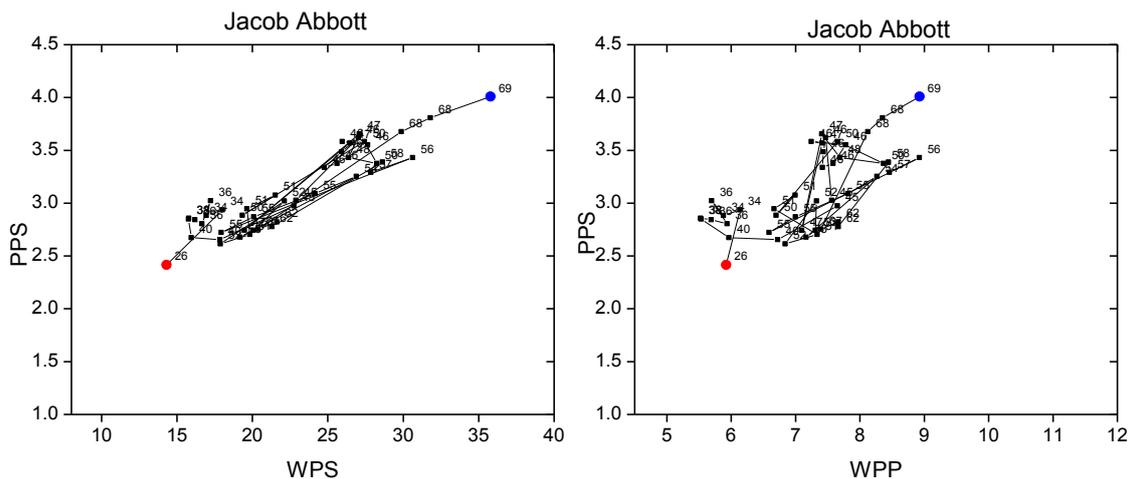


Figura A7.9 Comparación de WPS y WPP (a) y WPP y PPS (b) para Jacob Abbott.

Este comportamiento es claramente contrario a la fragmentación observada en los libros de Iris Murdoch. Pero es necesario advertir que es muy complicado sacar

cualquier conclusión sobre el envejecimiento del autor, ya que no existe una garantía completa de que estos libros sean los escritos originales (es decir, que podrían haber pasado por revisiones que hayan modificado sus características).

JAMES FENIMORE COOPER

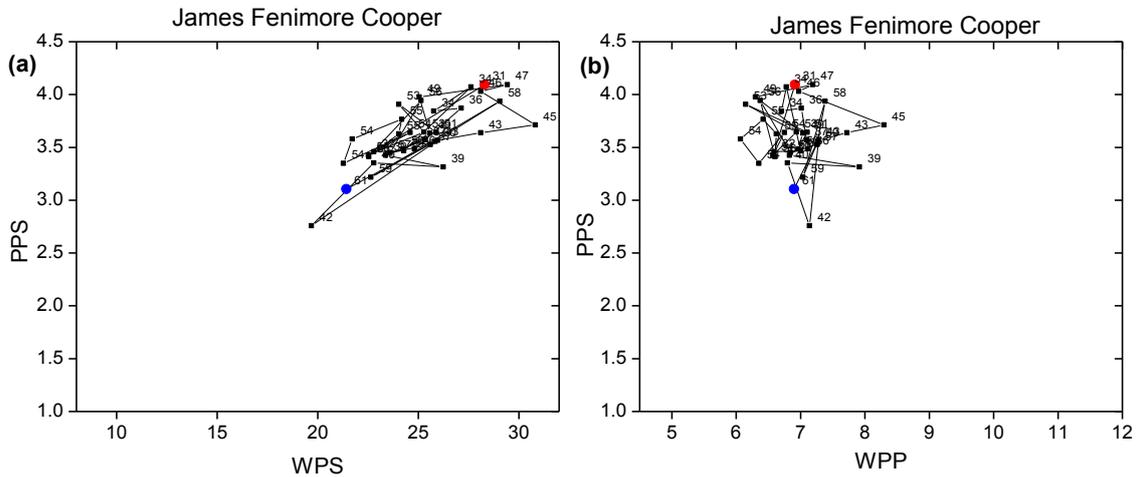


Figura A7.10 Comparación de WPS y WPP (a) y WPP y PPS (b) para James Fenimore Cooper.

El crecimiento de las oraciones se relaciona con un mayor número de pausas lingüísticas añadidas (Figura A7.10a), pero estas no presentan cambios en la longitud medio a lo largo de la carrera del escritor James Fenimore Cooper (Figura A7.10b).

ROBERT WILLIAM CHAMBERS

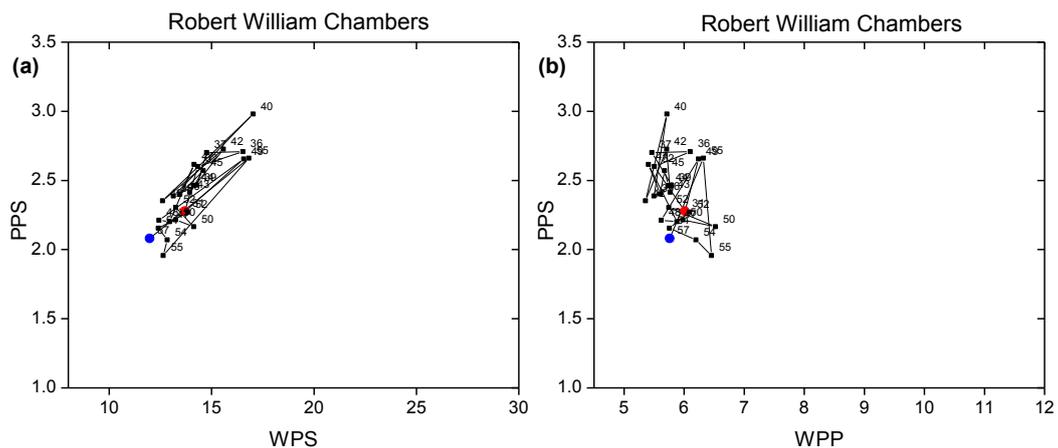


Figura A7.11 Comparación de WPS y WPP (a) y WPP y PPS (b) para Robert William Chambers.

No existen indicios de cambios con respecto al uso de la puntuación, ni tampoco con la longitud de las oraciones, relacionados con la edad, en los libros del autor Robert

William Chambers (Figura A7.11). Hay un crecimiento de la longitud de las oraciones asociado a un aumento en la cantidad de las pausas lingüísticas pero no hay asociación de estos cambios con la edad.

WILKIE COLLINS

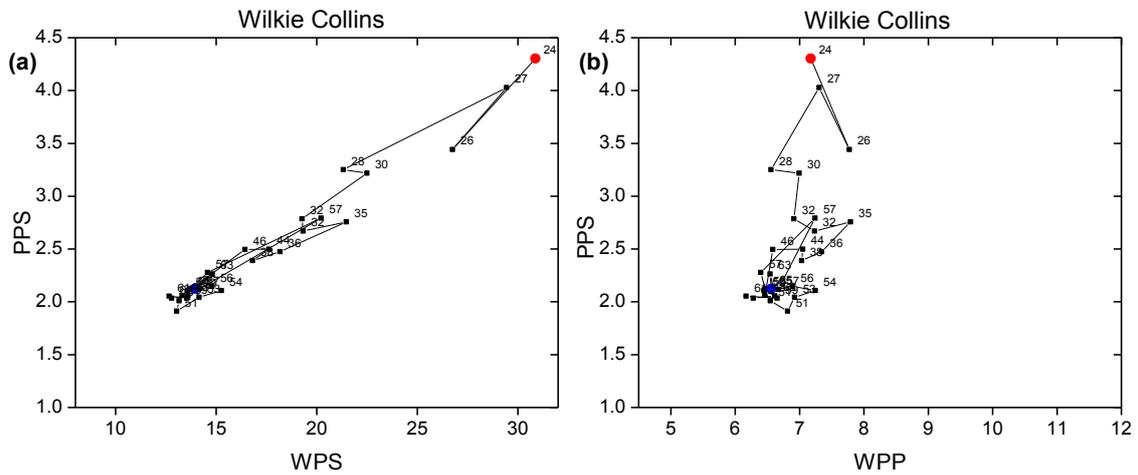


Figura A7.12 Comparación de WPS e WPP (a) e WPP e PPS (b) para Wilkie Collins.

El autor Wilkie Collins escribe textos con oraciones cada vez más pequeñas a lo largo de su vida, por lo que utiliza una menor cantidad de pausas lingüísticas (Figura A7.12a), sin cambiar también su longitud (Figura A7.12b).

la edad. El primero tuvo una vida literaria muy corta, lo que impide sacar conclusiones sobre su envejecimiento. No obstante, José Saramago y Machado de Assis escribieron hasta una edad avanzada (y todos sus libros escritos cuando eran mayores se han incluido en el conjunto de datos analizados).

Para el escritor Camilo Castelo Branco (Figura A7.13b) no se puede extraer conclusiones sobre el comportamiento del TTR55 en relación con el envejecimiento, ya que precipitó su muerte con un suicidio a los 65 años; sin embargo, hacemos hincapié en que su último libro sufrió un cambio considerable, con el valor más alto encontrado para el autor.

El gráfico del escritor Eça de Queirós (Figura A7.13c) sugiere que el autor pasó por dos fases de TTR: primero, una fase con valores cercanos a 0,10 y, segundo, una nueva fase con valores de TTR55 cercanos a 0,12; sin embargo, la cantidad de libros de este autor es pequeña para extraer una conclusión decisiva respecto a este comportamiento. Así, no es posible sacar conclusiones sobre el comportamiento de este parámetro en relación con el envejecimiento debido a que falleció muy joven, a los 55 años de edad.

Tampoco es posible extraer conclusiones para el escritor Érico Veríssimo (Figura A7.13d) en relación con el envejecimiento natural y con el deterioro cognitivo, pues murió repentinamente a los 70 años de un ataque al corazón cuando su producción literaria era activa. No obstante, sí se observa un aumento en el valor de la TTR55 en su último libro.

Los escritores Camilo Castelo Branco, Eça de Queirós y Érico Veríssimo, a su vez, presentan valores de TTR55 mayores con en el tiempo.

B) WTIR

Los autores, Camilo Castelo Branco (Figura A7.14b), Eça de Queirós (Figura A7.14c), Erico Verissimo (Figura A7.14d) y Jorge Amado (Figura A7.14e) demuestran un comportamiento diferente, ya que para todos ellos los libros que se encuentran en la parte inferior de la gráfica fueron escritos al inicio de sus carreras, mientras que los libros situados en la parte más alta son los del final de sus carreras. Esto indica un aumento en el vocabulario utilizado en la escritura y se explica, en parte, porque Camilo Castelo se suicidó a los 65 años, Eça de Queirós murió joven (55 años) y la vida de Érico Veríssimo se vio interrumpida por un ataque cardíaco a los 70 años. Así, todos

ellos vieron interrumpida su vida de forma temprana, sin experimentar ningún síntoma de deterioro cognitivo. La única duda la presenta Jorge Amado, pues no ha sido posible analizar los últimos libros de su carrera literaria.

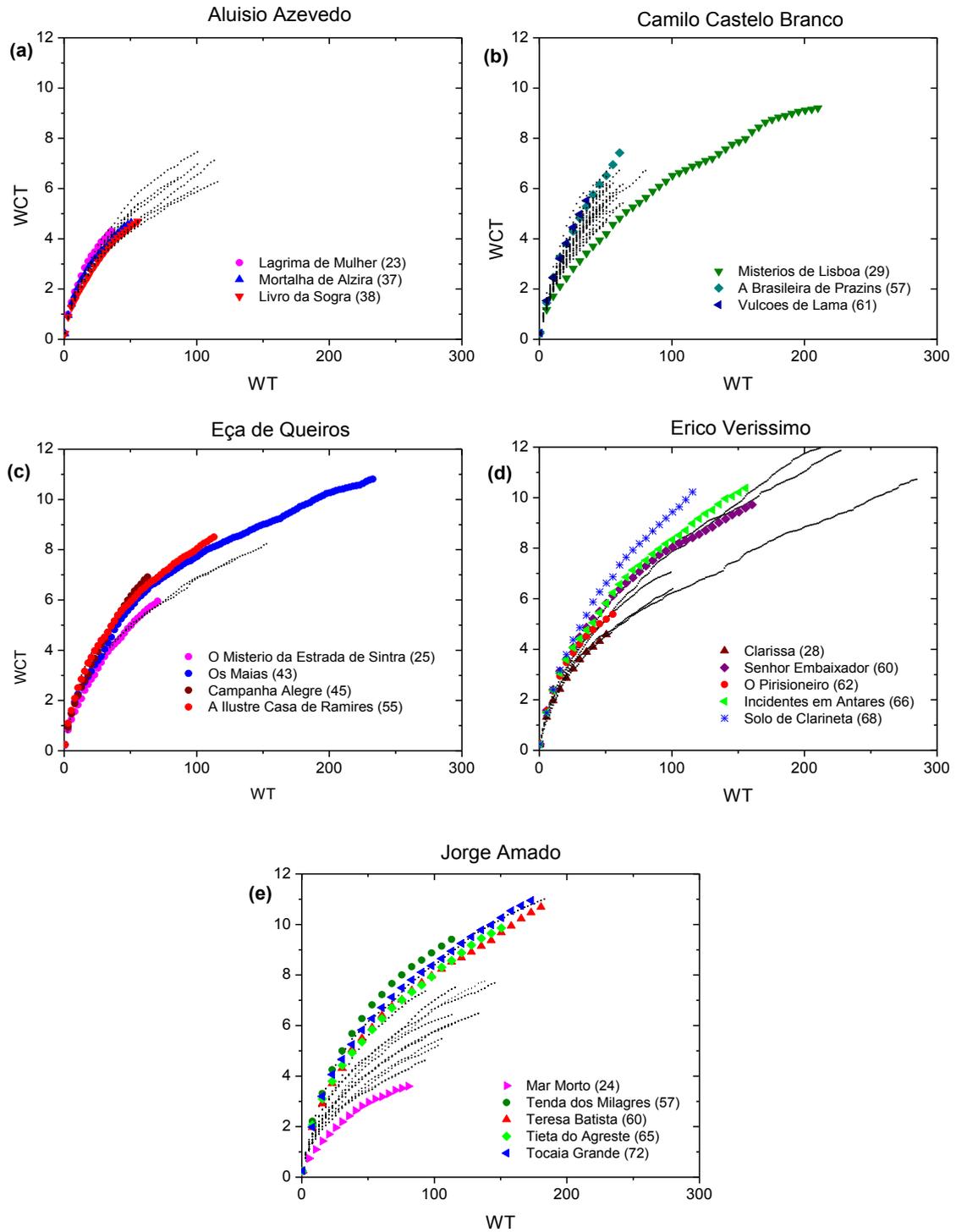


Figura A7.14 WTIR para cinco autores en lengua portuguesa.

C) WPS

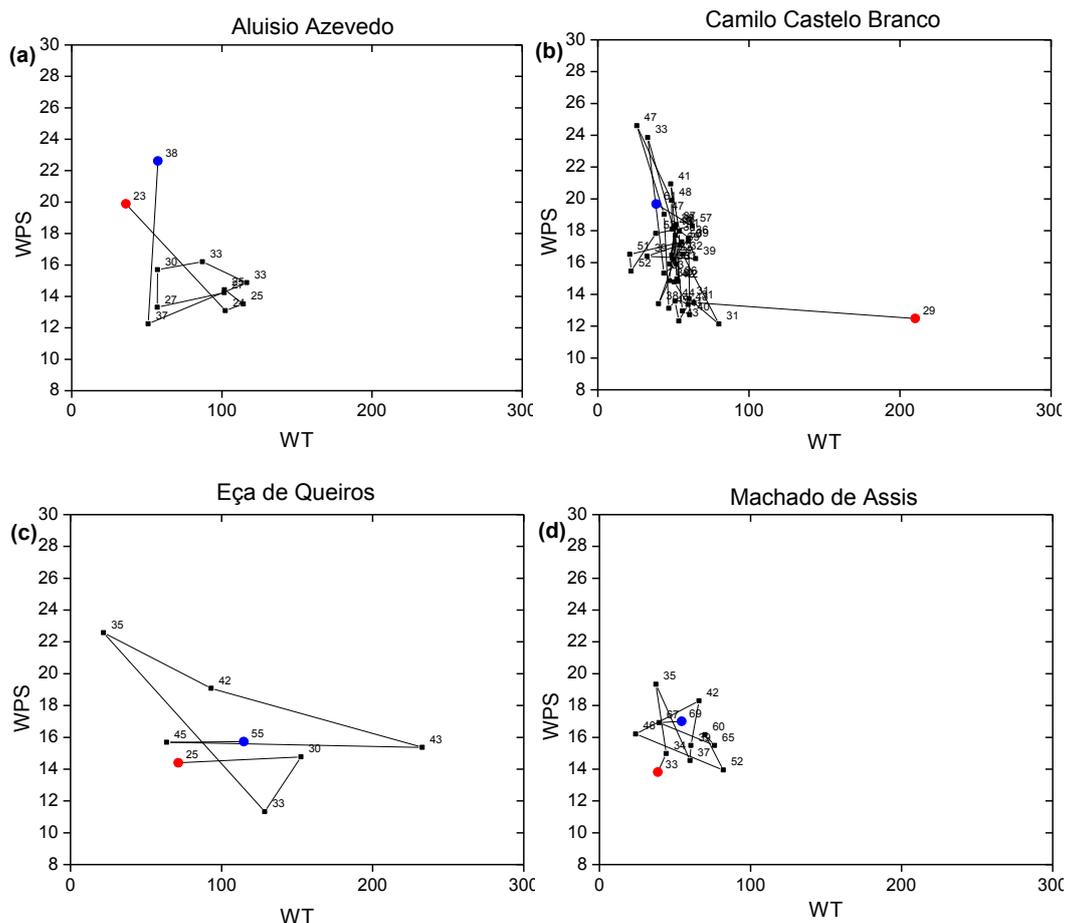


Figura A7.15 SPT de WPS para los escritores en lengua portuguesa.

Aluísio Azevedo (Figura A7.15a) presentó para su primer y su último libro valores superiores a la media de los otros libros del intermedio de su carrera. Esto puede sugerir una manera diferente de construir esos libros en relación con los demás. El comportamiento de los demás autores no sugiere nada especial.

D) WPP

La métrica WPP medida para los autores Aluisio Azevedo, Eça de Queirós, Érico Veríssimo, Jorge Amado y Machado de Assis (Figura A7.16), cuando analizada por medio de la trayectoria no presentó ningún comportamiento que sugiera deterioro cognitivo con la edad.

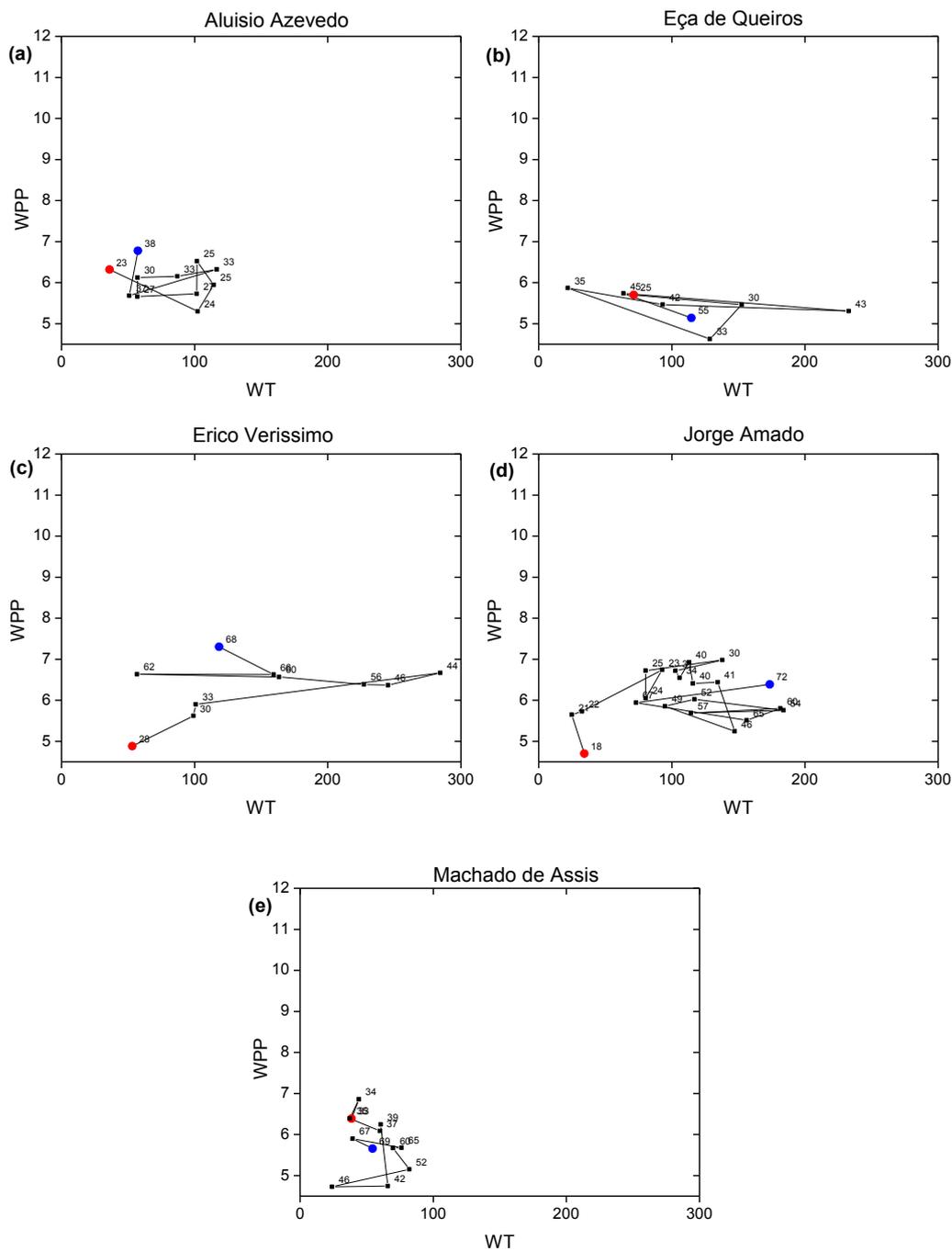


Figura A7.16 SPT del WPP para escritores en lengua portuguesa.

E) PPS

El primer y último libro de Aluísio Azevedo (Figura A7.17a) posee un valor más alto de pausas lingüísticas por oraciones que los demás. Pero como también existe un aumento en la longitud de las oraciones no es posible caracterizar este parámetro como una fragmentación de su escritura.

F) Comparación entre WPS, WPP y PPS

En el primer y último libros, el autor Aluisio Azevedo produjo oraciones más grandes, añadiendo un número mayor de pausas (Figura A7.18a), mientras que se mantiene la longitud de estas pausas, como en los otros libros (Figura A7.18b). Lo mismo ocurrió con las novelas de Eça de Queirós (Figura A.18 c y d).

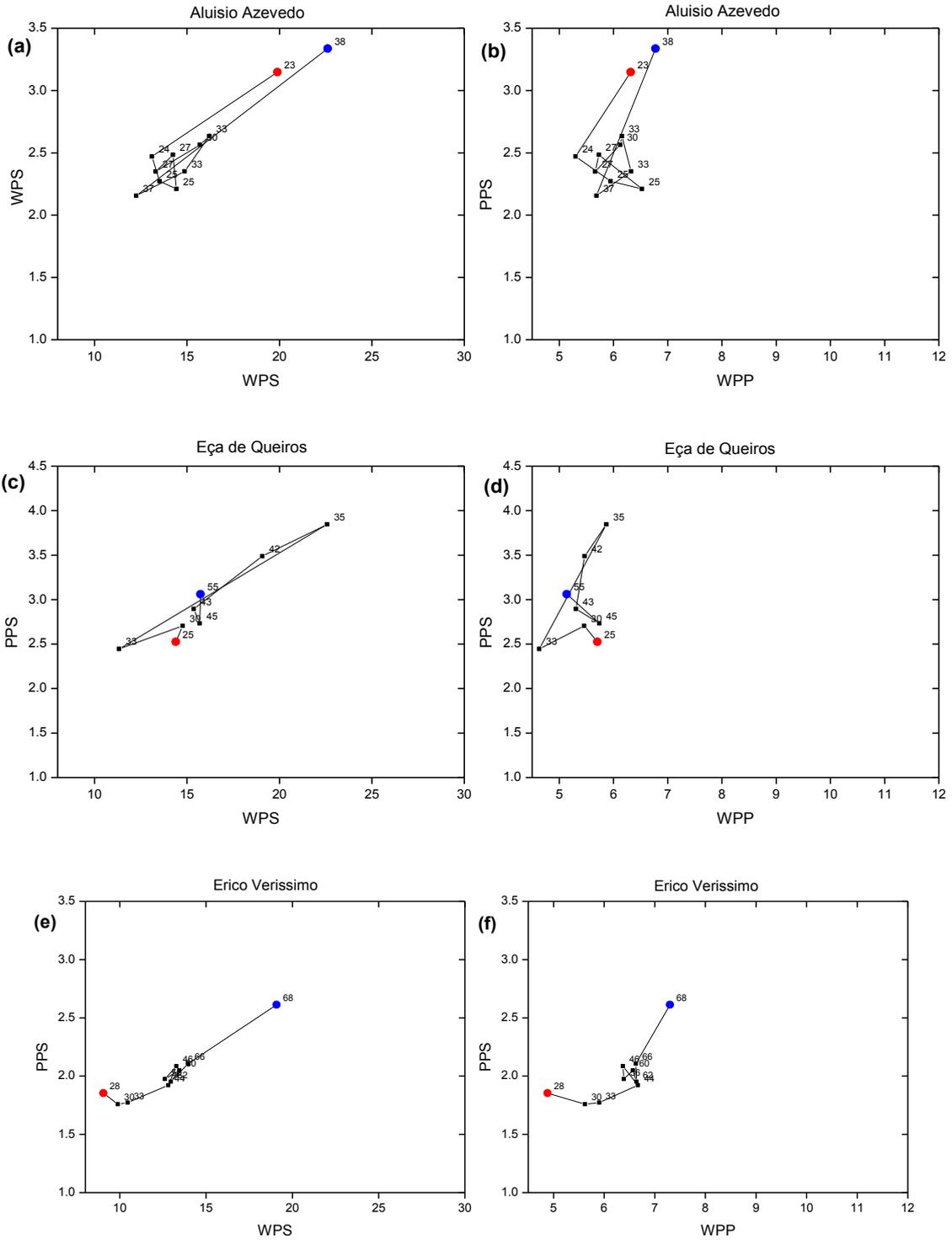


Figura A7.18 Comparación de WPS y WPP y WPP y PPS para libros en portugués.

Para Érico Veríssimo, hubo una tendencia a construir textos con una cantidad media de oraciones cada vez mayor con la edad y las oraciones mayores se acompañaron de un mayor número de pausas, y estas también crecieron con la edad, principalmente para su último libro (Figura A7.18e y f).



