

---

# KEYLOGGING PARA EL ESTUDIO DE LOS PROCESOS COGNITIVOS DEL TRADUCTOR

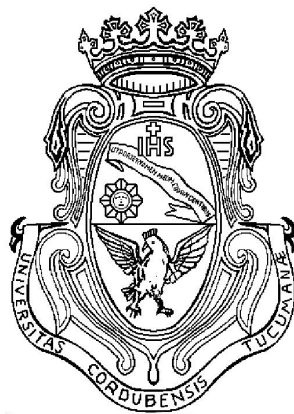
---

Trabajo especial presentado para acceder al título de  
Licenciada en Ciencias de la Computación

Roxana Anabel Lafuente  
[roxana.lafuente@gmail.com](mailto:roxana.lafuente@gmail.com)

Paula Estrella  
[pestrella@famaf.unc.edu.ar](mailto:pestrella@famaf.unc.edu.ar)

Adolfo Martín García  
[adolfomartingarcia@gmail.com](mailto:adolfomartingarcia@gmail.com)



30 de Julio de 2015

Facultad de Matemática, Astronomía y Física  
Universidad Nacional de Córdoba



Keylogging para el estudio de los procesos cognitivos del traductor por Roxana Anabel Lafuente se distribuye bajo una  
Licencia Creative Commons Atribución-NoComercial-SinDerivadas 2.5 Argentina.

# Índice general

|   |            |
|---|------------|
| <b>Índice General</b>                               | <b>I</b>   |
| <b>Agradecimientos</b>                              | <b>III</b> |
| <b>Resumen</b>                                      | <b>IV</b>  |
| <b>Índice de figuras</b>                            | <b>V</b>   |
| <b>1. Introducción</b>                              | <b>6</b>   |
| 1.1. Motivación . . . . .                           | 7          |
| 1.2. Objetivo . . . . .                             | 8          |
| 1.3. Hipótesis . . . . .                            | 8          |
| 1.4. Estructura de la tesis . . . . .               | 9          |
| <b>2. Áreas involucradas</b>                        | <b>10</b>  |
| 2.1. Traducción . . . . .                           | 10         |
| 2.2. Traductología . . . . .                        | 11         |
| 2.3. Cognición . . . . .                            | 12         |
| 2.4. Ciencias de la computación . . . . .           | 12         |
| <b>3. Marco teórico</b>                             | <b>13</b>  |
| 3.1. La traducción como proceso cognitivo . . . . . | 13         |
| 3.2. Técnicas existentes . . . . .                  | 15         |
| 3.3. Factores que afectan la traducción . . . . .   | 17         |
| <b>4. Metodología</b>                               | <b>22</b>  |
| 4.1. Herramientas desarrolladas . . . . .           | 22         |
| 4.1.1. ResearchLogger . . . . .                     | 22         |
| 4.1.2. ResearchAnalyzer . . . . .                   | 25         |
| 4.1.3. Comparación con otras herramientas . . . . . | 29         |
| 4.2. Sesiones de traducción . . . . .               | 31         |
| 4.2.1. Participantes . . . . .                      | 31         |
| 4.2.2. Condiciones . . . . .                        | 33         |

|  |              |
|--|--------------|
| 4.2.3. Textos usados . . . . .                     | 34           |
| 4.2.4. Protocolo . . . . .                         | 35           |
| <b>5. Resultados</b>                               | <b>36</b>    |
| 5.1. Subgrupos . . . . .                           | 36           |
| 5.2. Tipo de análisis . . . . .                    | 36           |
| 5.3. Direccionalidad . . . . .                     | 37           |
| 5.3.1. Tiempo . . . . .                            | 38           |
| 5.3.2. Pausas . . . . .                            | 40           |
| 5.4. Influencia de memoria de traducción . . . . . | 41           |
| 5.4.1. Tiempo . . . . .                            | 41           |
| 5.4.2. Pausas . . . . .                            | 42           |
| 5.5. Presión de tiempo . . . . .                   | 43           |
| 5.5.1. Tiempo . . . . .                            | 43           |
| 5.5.2. Pausas . . . . .                            | 45           |
| <b>6. Discusión General</b>                        | <b>47</b>    |
| 6.1. Fortalezas y limitaciones . . . . .           | 47           |
| 6.2. Trabajo futuro . . . . .                      | 48           |
| <b>7. Conclusiones</b>                             | <b>50</b>    |
| <b>Acrónimos</b>                                   | <b>LII</b>   |
| <b>Bibliografía</b>                                | <b>LV</b>    |
| <b>Anexos</b>                                      | <b>LVI</b>   |
| <b>Anexo A. Textos usados</b>                      | <b>LVII</b>  |
| <b>Anexo B. Cuestionarios usados</b>               | <b>LXII</b>  |
| <b>Anexo C. Detalles técnicos</b>                  | <b>LXIII</b> |
| C.1. ResearchLogger . . . . .                      | LXIII        |
| C.2. ResearchAnalyzer . . . . .                    | LXIII        |
| <b>Anexo D. Tablas de resultados</b>               | <b>LXV</b>   |

# Agradecimientos

*Quisiera agradecer especialmente a mi familia: mis padres Hugo y Elida, mi hermano Gabriel, mis tías Norma y Teresa y mi nona Juana. Gracias por su afecto y apoyo incondicional, y más que nada, gracias por acompañarme a seguir mis sueños y por alentarme a volar cada vez más alto. También gracias a las amigas de la familia: Teresa, Verónica y Alba.*

*Agradezco también a la comunidad de FaMAF, tanto a mis compañeros como a los profesores que nos acompañaron en estos años de cursado, de quienes me llevo muchos gratos recuerdos e importantes enseñanzas. A los traductores que voluntariamente nos ayudaron a recolectar los datos necesarios para poder realizar este estudio.*

*Quisiera también expresar mi gratitud hacia la Dra. Maria Inés Arrizabalaga y Laura Bruno quienes se mostraron interesadas en esta investigación y nos ayudaron a contactar alumnos de la Facultad de Lenguas. Especialmente gracias a la valiosa colaboración del Dr. Agustín Ibañez en la interpretación y análisis de los resultados. Finalmente, gracias a mis directores, Paula y Adolfo, quienes me encaminaron y aconsejaron para que esta investigación pueda llevarse a cabo. ¡Gracias guiarme en mis primeros pasos como investigadora!*

Roxana Anabel Lafuente.

# Resumen

En este trabajo se extiende una herramienta para adaptarla a la investigación de los procesos cognitivos involucrados en la traducción humana mediante el uso de técnicas de *key-logging* y de análisis estadístico de datos.

La técnica de *key-logging* genera un log de alta precisión temporal que contiene el registro de las acciones de *mouse* y teclado que el traductor realiza. Al analizar el log cualitativamente y cuantitativamente, podemos inferir algunos aspectos de los procesos cognitivos comprendidos en las fases generales del acto traductor. Esto permite que se pueda analizar el proceso de producción, permitiendo la identificación de las fases de traducción. Además, la técnica de *key-logging* permite que el análisis se haga no sólo en el texto fuente y en el texto meta (productos estáticos) sino también en el producto incremental.

Trabajos anteriores han recolectado datos con otras herramientas gratuitas pero para un entorno controlado en la plataforma Windows y han analizado las fases en la traducción usando mayormente estadística descriptiva. En este trabajo, realizamos la recolección de los datos utilizando una herramienta de código abierto y profundizamos el análisis de los datos con estadística inferencial, por ejemplo, usando *Mixed Repeated Measures Analysis of variance* (MRM-ANOVA). Estas son las principales características que destacan la originalidad del presente trabajo, además de ser uno de los pocos estudios en el área en el país: es el primero en el ámbito de la UNC y, según nuestro conocimiento, hay un sólo grupo trabajando en estudios similares en la Universidad del Aconcagua.

Los aportes de esta tesis son prácticos (las herramientas de recolección y análisis quedarán libremente disponibles) y teóricos (se estudia una serie de hipótesis sobre el proceso traductor y se define una metodología y protocolos de trabajo para sesiones experimentales de traducción).

**Palabras claves.** *key-logging*, cognición, traducción humana, fases en la traducción, python, R

# Índice de figuras

|   |    |
|---|----|
| 1.1. El proceso de traducción. . . . .  | 6  |
| 3.1. Fases en la traducción. . . . .  | 14 |
| 4.1. Ejemplo de log original de PyKeylogger. . . . .                              | 23 |
| 4.2. Ejemplo de log <i>keystrokes</i> de ResearchLogger. . . . .                  | 24 |
| 4.3. Ejemplo de log <i>clicks</i> de ResearchLogger. . . . .                      | 25 |
| 4.4. Captura de pantalla de la ventana de traducción . . . . .                    | 25 |
| 4.5. Diagrama de clases de ResearchAnalyzer . . . . .                             | 28 |
| 5.1. Tiempo de sesión total (en minutos) - Directa vs. Inversa. . . . .           | 38 |
| 5.2. Tiempo de fases (en minutos) - Directa vs. Inversa . . . . .                 | 39 |
| 5.3. Pausas en la sesión - Direccionalidad . . . . .                              | 40 |
| 5.4. Tiempo de sesión total (en minutos) - Sin memoria vs. Con memoria . . . . .  | 41 |
| 5.5. Tiempo de fases (en minutos) - Sin Memoria vs. Con memoria . . . . .         | 42 |
| 5.6. Pausas en la sesión - Uso de memoria de traducción . . . . .                 | 43 |
| 5.7. Tiempo de sesión total (en minutos) - Mucho tiempo vs. poco tiempo . . . . . | 44 |
| 5.8. Tiempo de fases (en minutos) - Mucho tiempo vs. poco tiempo . . . . .        | 45 |
| 5.9. Pausas en la sesión - Presión de tiempo . . . . .                            | 46 |

# Capítulo 1

## Introducción

Como explica Albir (2001), para definir la traducción es necesario diferenciar tres tipos de traducción: intersemiótica, intralingüística e interlingüística. La traducción intersemiótica o transmutación es una interpretación de los signos verbales mediante los signos de un sistema no verbal. La traducción intralingüística o reformulación es una interpretación de los signos verbales mediante otros signos de la misma lengua. Por último, la traducción interlingüística o traducción propiamente dicha es una interpretación de los signos verbales mediante cualquier otra lengua.

En este trabajo se estudiará la traducción interlingüística donde, en resumen, el trabajo del traductor consistirá en comprender el significado de un texto en un lenguaje dado, llamado texto origen o texto fuente, con el fin de producir un texto de significado equivalente, en otro idioma, llamado texto traducido o «texto meta».

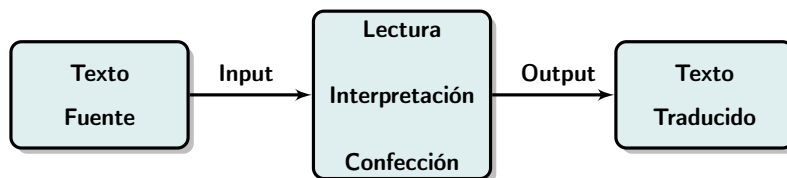


Figura 1.1: El proceso de traducción.

Los procesos cognitivos son aquellos procesos que involucran el uso de la memoria, la toma de decisiones, la inferencia, el razonamiento y el aprendizaje (Hutchins, 2000). Cada ser humano es único gracias a su sistema cognitivo, sin embargo, existen muchos atributos cognitivos que son compartidos por todos los seres humanos; esto también se aplica a la traducción.

El proceso de toma de decisiones es uno de los mecanismos más complejos del pensamiento humano dado que varios factores y cursos de acción intervienen en el. Más aun, el proceso entero se ve afectado por variables personales y ambientales. De hecho, los individuos pueden tomar diferentes decisiones dependiendo de si se sienten observados por sus jefes, en la cantidad de información que poseen para decidir algo o si ciertas motivaciones juegan un papel relevante en sus vidas, (María L. Sanz de Acedo Lizárraga, 2007). Para el caso de la traducción humana, algunos factores en el entorno que afectan los procesos cognitivos en la traducción son el uso de memoria

de traducción, la presión de tiempo o la presión por estar rindiendo un examen, entre otros.

El presente trabajo consiste en desarrollar las herramientas necesarias que permitan analizar los procesos cognitivos involucrados en la traducción humana para identificar patrones comunes en el comportamiento de los traductores. Como aplicación de los desarrollos realizados, se pretende estudiar cómo los procesos mentales de los traductores varían en diferentes situaciones dependientes del entorno por medio de sesiones de traducción que permitan recolectar datos de usuario.

Se conoce una gran variedad de técnicas y herramientas que pueden emplearse para abordar este tipo de investigaciones (Christensen, 2011a), en este caso, se eligió la técnica de *key-logging* por (a) proveer un acercamiento directo a la toma de decisiones progresiva sobre los aspectos cognitivos lingüísticos de la traducción (b) permitir la observación de la construcción del texto meta, a diferencia de otros métodos que solo se enfocan en el producto final (c) bajo costo económico (d) poseer alta validez ecológica, es decir, el entorno de prueba es muy parecido a la vida real de modo que no influye en el comportamiento normal del sujeto. Un *key-logger* es un programa que tiene acceso a la actividad de teclado y *mouse* lo que permite obtener un registro de la actividad del traductor mientras realiza su tarea. Estos datos se conocen como *User Activity Data (UAD)* (Michael Carl, 2008) que luego son sometidos a análisis estadísticos para encontrar patrones de actividad comunes en el comportamiento de los sujetos.

Para analizar cómo los distintos factores afectan los procesos cognitivos durante la traducción, se organizaron sesiones de traducción con diferentes condiciones, que se detallarán más adelante, siguiendo un protocolo metodológico establecido para tal fin. En cada una de estas sesiones, se usó un software de *key-logging* para guardar un registro de cada evento de *mouse* y de teclado que efectuaban los traductores mientras traducían textos controlados (de complejidad similar). Luego, dichos datos fueron analizados con una serie de *scripts* en Python y R desarrollados con el fin de facilitar la tarea de análisis a potenciales usuarios del sistema de *keylogging*.

## 1.1. Motivación

Se han publicado numerosos estudios sobre los procesos cognitivos en la traducción usando técnicas diferentes a la de *key-logging*. En contraste, el número de trabajos que usan esta técnica es mucho menor y por lo general hacen uso de las herramientas (Carl, 2012) o InputLog(M. Leijten, 2005). Lo novedoso de este trabajo es que se usa una versión modificada de un *key-logger open-source* existente<sup>1</sup>. La herramienta que se presenta en este trabajo tiene diferencias significativas con las anteriores; por ejemplo, Translog sólo loguea los eventos que ocurren en su ventana de traducción mientras que nuestro software loguea todos los eventos que ocurren en todas las ventanas abiertas.

ResearchLogger, la herramienta de toma de datos presentada en este trabajo, busca proveer a los investigadores con una herramienta de *key-logging* confiable y no se limita a la traductología sino que puede ser útil para cualquier investigación sobre el comportamiento de usuarios y/o procesos de escritura. Adicionalmente, se proveerá un módulo de análisis de la información recolectada llamado

---

<sup>1</sup>Disponible en <http://pykeylogger.sourceforge.net/>



ResearchAnalyzer.

Se elige la traducción como objeto de estudio ya que es un proceso cognitivo complejo que implica constantes revisiones y toma de decisiones. Además, existe un consenso respecto de una estructura canónica en fases para la producción del texto meta lo cual no se da en otros procesos de escritura tales como la escritura de *tweets*. Finalmente, la traducción es una actividad clave en el contexto local ya que es una de las carreras más populosas en la Universidad Nacional de Córdoba. Esta investigación propicia el acercamiento interdisciplinar entre la traductología, la cognición y las ciencias de la computación; así como también, interinstitucional entre la Facultad de Matemática, Astronomía y Física y (FaMAF) la Facultad de Lenguas (FL).

Para las sesiones experimentales de traducción se convocó a alumnos de la Facultad de Lenguas (UNC) y del Instituto Nueva Formación que se encuentran cursando entre 3er y 5to año de la carrera de Traductorado en Inglés. Esto nos permitió recolectar información sobre cómo reaccionan ante la direccionalidad en la traducción, el uso de herramientas de asistencia a la traducción (como lo son diccionarios, memorias de traducción o traductores automáticos), su contribución al producto final y la presión de tiempo.

## 1.2. Objetivo

Se desea extender una herramienta de *key-logging* (*open-source*) sensible a los procesos *online* en la traducción, (Christensen, 2011a). Esto permitirá inferir algunos aspectos de los procesos cognitivos subyacentes en la traducción.

Para confirmar la validez del sistema se estudiará (a) la modulación de variables tales como direccionalidad, uso de recursos y restricciones temporales (b) detección sistemática de la modulación de variables para medir la sensibilidad del instrumento propuesto usando los resultados obtenidos en estudios anteriores.

## 1.3. Hipótesis

### Direccionalidad

Los resultados de Augusto Buchweitz (2006) muestran que la traducción inversa requiere de mayor tiempo hasta llegar a la obtención del texto traducido a comparación de la traducción directa. Esto promovió la formulación de la siguiente hipótesis:

**H1:** La traducción directa se realiza con mayor rapidez que la traducción inversa en textos de complejidad comparable.

### Uso de memoria de traducción

Pym (2011) sugiere que la tecnología no necesariamente agiliza el proceso de traducción y ejemplifica como el uso de una memoria de traducción añade complejidad en el proceso de toma de decisiones. Los mencionados puntos inspiraron la formulación de la siguientes hipótesis:

**H2:** La traducción usando memoria de traducción es más lenta que cuando la misma no se usa.

**H3:** El uso de memoria de traducción complica la toma de decisiones lo que se ve reflejado en un aumento en las pausas para la condición con memoria.

## Presión de tiempo

Mientras que el efecto de presión de tiempo en interpretes ha sido ampliamente estudiado, poco se ha investigado acerca de los efectos de la presión de tiempo en traducciones escritas. Sharmin et al. (2008) menciona estudios anteriores donde se ve que la presión de tiempo causa que los traductores dediquen menos tiempo a la fase de orientación y omiten la revisión final en vez de cambiar el ritmo de traducción.

**H4:** Bajo presión de tiempo, los participantes reducen el tiempo dedicado a las fases de orientación y revisión.

## 1.4. Estructura de la tesis

La tesis está estructurada en capítulos de manera que los capítulos 2 y 3 presentan las áreas involucradas y marco teórico necesario para este proyecto interdisciplinario; en el capítulo 4 se detalla la metodología empleada mientras que el capítulo 5 detalla los resultados obtenidos. El capítulo 6 provee un informe sobre futuros hilos de investigación además de las limitaciones y fortalezas del trabajo. El trabajo concluye con el capítulo 7, en el cual se presentan las conclusiones obtenidas. El detalle de cada capítulo es el siguiente:

**Capítulo 2** Se introducen las áreas ajenas a la computación que son pertinentes en este proyecto: traducción, traductología y cognición.

**Capítulo 3** Se definen conceptos que serán necesarios para la interpretación de los resultados, los resultados obtenidos en estudios anteriores y los distintos tipos de análisis posibles para este tipo de experimentos.

**Capítulo 4** Se detalla la metodología empleada en este trabajo, las condiciones de experimentos que se desean estudiar y los aspectos técnicos del desarrollo de las herramientas.

**Capítulo 5** Se presentan los resultados obtenidos para la condición de traducción periodística.

**Capítulo 6** Discusión general sobre limitaciones de la investigación y futuros hilos de investigación a seguir.

**Capítulo 7** Se llega a una conclusión del presente trabajo y se provee de un resumen de los resultados obtenidos.

## Capítulo 2

# Áreas involucradas

Dado el carácter interdisciplinario del presente trabajo, dedicamos este capítulo a describir brevemente las áreas involucradas que no pertenecen a las ciencias de la computación.

### 2.1. Traducción

Se puede entender la traducción como la actividad que consiste en comprender el significado de un texto en un idioma, llamado *texto origen* o texto de partida, para producir un texto con significado equivalente, en otro idioma, llamado texto traducido o *texto meta*. El resultado de esta actividad, el texto traducido, también se denomina traducción y representa la obra de un traductor. Cuando la traducción se realiza de manera oral se conoce como interpretación.

Dado que hay dos idiomas involucrados en la actividad de traducir, ésta tiene una dirección o sentido, por ejemplo *del Inglés al Español* o *Inglés  $\rightarrow$  Español*, donde denominamos  $L1$  a la lengua materna y  $L2$  a la segunda lengua aprendida. Frecuentemente la dirección es  $L2 \rightarrow L1$ , ya que es la única forma en que el traductor podrá traducir naturalmente, con exactitud y máxima eficiencia (Newmark and Hall, 1988). Sin embargo, la dirección opuesta también se practica dado que el traductor para ejercer debe tener un avanzado nivel en  $L2$ . Según la dirección empleada se dice que la traducción es *directa* ( $L2 \rightarrow L1$ ) o *inversa* ( $L1 \rightarrow L2$ ).

El texto a traducir tiene un rol fundamental y da origen al tipo de traducción que se debe realizar. Aunque no hay una única clasificación de los tipos de traducción existentes, (Roberts, 1988) presenta un análisis de distintas propuestas que apuntan a crear tal clasificación, por ejemplo de acuerdo a la función del texto, a su grado de especialización, al propósito general de la traducción, etc. Para el presente trabajo nos limitaremos a nombrar sólo algunos tipos de traducción que consideramos más conocidos y que podrían ser de utilidad en capítulos posteriores:

- Traducción divulgativa: tienen por principal característica el uso del idioma con pocos tecnicismos puesto que van dirigidas al público en general, que no tiene necesariamente conocimientos previos sobre los temas tratados, por ejemplo en revistas, reportajes o libros, y tratan de temas igual de variados (cocina, las ciencias, la tecnología, etc.)

- Traducción literaria: consiste en la traducción de obras literarias (novelas, cuentos cortos, obras de teatro, poemas, etc.) y suele considerarse una actividad literaria en sí misma.
- Traducción científico-técnica: se encarga de la traducción de textos pertenecientes a ámbitos de los conocimientos científico-técnicos como pueden ser las ciencias de la salud (medicina, farmacia, etc.), la informática (software, *hardware*, etc.), las telecomunicaciones (telefonía, Internet, etc.), la biotecnología (manipulación genética de alimentos, nuevas variedades vegetales y animales, etc.), automoción y mecánica, etc.
- Traducción audiovisual: a diferencia de los otros tipos de traducción, la traducción audiovisual cambia el material de soporte permitiendo la traducción de diálogos, escenas, etc a través de la subtitulación, el doblaje y el *voice-over* (superposición de las voces traducidas en la banda sonora original).

## 2.2. Traductología

Según (Albir, 2001), "La traductología (también conocida como estudios sobre la traducción) es la disciplina que estudia sistemáticamente la teoría, la descripción y la aplicación de la traducción y la interpretación, así como las mismas en conjunto". Dada su juventud como área de investigación, co-existen al presente varias formas de nombrarla, a saber *Traductología*, *Estudios de la traducción*, *Translémica*, *Translatología*; en este trabajo usaremos *Traductología*, ya que es el más utilizado en la literatura. Esta nueva área de estudio e investigación se ha consolidado recientemente como una disciplina específica que estudia la traducción en sus diversas manifestaciones y su desarrollo ha sido enorme en sus tres direcciones: estudios teóricos, descriptivos y aplicados.

Por su carácter multidisciplinar, la traductología permite la convergencia de muchos conceptos, métodos y metodologías de otros campos de estudio, por ejemplo de la literatura comparada, la informática, la historia, la lingüística, la filosofía, la psicología o la terminología y, en su gran mayoría, los estudios en este área son empíricos.

Los estudios empíricos pueden subdividirse, según sea el objeto de estudio, en: *basados en el producto*, es decir, en el texto traducido o en *corpora*, y *basados en el proceso*, los cuales requieren de sesiones de traducción donde se observa a los sujetos realizar la tarea (Göpferich, 2008) (citado en (Christensen, 2011b)). Los estudios empíricos requieren tanto la descripción del fenómeno estudiado como una explicación del mismo. Utilizando los resultados de los estudios basados en el producto (o *corpora*), los investigadores pueden formular sus hipótesis y eventualmente sucede que se atribuyen algunos eventos o resultados al proceso de traducción, es decir, aquellos fenómenos que no pueden explicarse directamente desde los *corpora* se supone que surgen del proceso traductor, el cual se describe como una *caja negra* (Toury, 1982). Recientemente, ha surgido una nueva forma de investigar esta *caja negra* mediante experimentos de traducción, donde se monitorea a los traductores y se recolecta información de su actividad a través del teclado, el *mouse* y su actividad ocular; este tipo de estudios se le denomina en Inglés estudios de *keylogging* y *eye-tracking*.

### 2.3. Cognición

La cognición se define como la facultad de un ser vivo para procesar información a partir de la percepción, el conocimiento adquirido (experiencia) y características subjetivas que permiten valorar la información. Como se explica en (?), la cognición consiste en procesos tales como el aprendizaje, razonamiento, atención, memoria, resolución de problemas, toma de decisiones y procesamiento del lenguaje.

Las teorías cognitivas abordan principalmente los procesos internos que ocurren durante las acciones; en el área de la traducción se le denomina *proceso traductor* al conjunto de procesos cognitivos disparados durante la actividad de generar un texto meta. La investigación del proceso traductor generalmente se centra en el flujo de trabajo del traductor (*workflow*) o en los procesos mentales del traductor, haciendo referencia a procesos externos e internos, respectivamente (Schubert, 2009). Los procesos internos se definen como la actividad mental que da origen a las traducciones y que no se puede observar directamente en el sujeto; por otro lado, los procesos externos son todos aquellos que pueden ser observados a simple vista por quien conduce el experimento de traducción.

Ambos tipos de procesos están relacionados, ya que los procesos internos se desarrollan dentro (y por lo tanto son influenciados por) el contexto del flujo de trabajo que es externo. En otras palabras los procesos externos guían e incluso podrían restringir y determinar el proceso interno. Otra forma en que están relacionados es en que se supone que a cada proceso externo o acción observable tiene su contraparte como proceso interno en el proceso traductor.

En vista de lo anterior, la tendencia en la investigación en ciencias cognitivas es ver la actividad mental como la interacción entre la mente humana, el cuerpo y el entorno, estudiando tanto lo que sucede dentro de la mente como lo que sucede en el resto del cuerpo y del entorno de trabajo.

### 2.4. Ciencias de la computación

Las ciencias de la computación pueden aportar enormemente a las áreas mencionadas. Por ejemplo, las herramientas informáticas permiten estudiar los pasos intermedios de la traducción a través de los datos que se guardan, particularmente de las teclas presionadas, de los *clicks* de *mouse*, algunos también guardan datos de la actividad ocular de fijación y dilatación de pupila. Estos datos permiten inferir aspectos sobre los procesos cognitivos involucrados en la traducción y se utilizan para analizar el aspecto conductual y las traducciones intermedias antes de obtener el producto final.

Algunos de los estudios realizados utilizan estos datos para describir potenciales problemas de traducción y las estrategias utilizadas para resolverlos, y para estudiar las distintas estrategias de traducción empleadas, entre otras tareas. Como se describirá en detalle, en este trabajo se realizó un estudio de este tipo utilizando la adaptación de un *keylogger* de código abierto.

# Capítulo 3

## Marco teórico

En este capítulo repasaremos las bases teóricas tanto de la traducción como del área de cognición en general y las técnicas que se han usado en investigaciones anteriores.

### 3.1. La traducción como proceso cognitivo

Desde un punto de vista cognitivo psicológico, el proceso de traducción consiste de dos tipos de procesamientos: el procesamiento del texto fuente o comprensión del texto fuente y el procesamiento del texto traducido o producción del texto traducido (Hvelplund, 2011).

#### Procesamiento del texto fuente

En el procesamiento del texto fuente se diferencian dos procesos cognitivos: la lectura del texto fuente y la comprensión del texto fuente (Hvelplund, 2011).

Este tipo de procesamiento se estudia con el uso de las técnicas de *eye-tracking* que permiten obtener un log de coordenadas donde el sujeto fijó la mirada mientras leía el texto fuente.

#### Procesamiento del texto meta

Las técnicas de *key-logging* nos permiten inspeccionar como se crea el texto meta, de manera que se focalizará nuestro análisis en el procesamiento del texto meta.

Se ha determinado que la producción del texto meta se subdivide en tres sub-etapas o fases, (Hvelplund, 2011):

- **Orientación:** Inicia en el instante en el cual el traductor empieza a ver el texto fuente y termina cuando presiona la primer tecla para empezar a formar el texto traducido. El objetivo de esta fase es familiarizarse con el texto fuente.
- **Elaboración de un borrador:** Comienza inmediatamente luego de la etapa anterior y finaliza cuando el traductor presiona la tecla correspondiente de la última letra de la primera

versión completa del texto traducido; que en general, estará marcada por el punto final. La meta de esta fase es crear una traducción equivalente a la del texto fuente.

- **Revisión:** Inicia inmediatamente luego de la etapa anterior y finaliza cuando el traductor determina que la traducción es satisfactoria. Tiene como objetivo determinar si la traducción tiene la calidad deseada por el traductor y realizar los cambios correspondientes hasta que la traducción pase los criterios de calidad del traductor.

Pensando en el ciclo de vida del software e imaginando que la traducción es el producto final en vez del software, las fases se corresponderían con el análisis y diseño, implementación y testeo respectivamente.

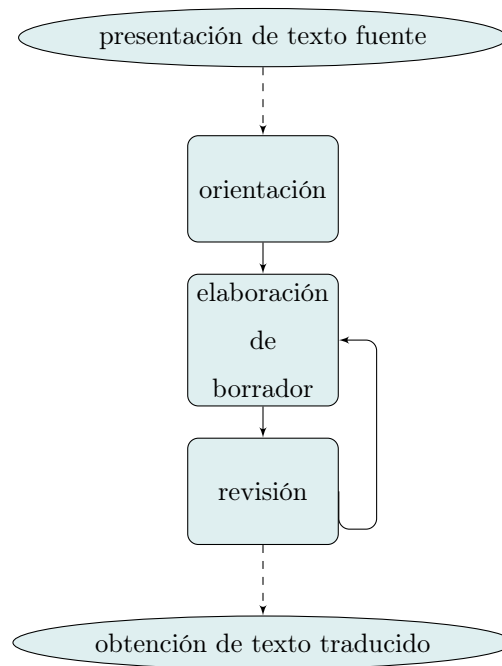


Figura 3.1: Fases en la traducción.

Koehn (2010) detalla la clasificación de pausas por su tiempo de duración y describe el comportamiento asociado a cada tipo de pausa.

1. Pausa inicial: La pausa al inicio de la traducción, si la hubiere. Por definición, esta primera pausa es la fase de orientación.
2. Pausa final: La pausa al final de la traducción, si la hubiere.
3. Pausa corta: Duración entre 2 y 6 segundos (exclusive). Indican un poco de indecisión.
4. Pausa mediana: Duración entre 6 y 60 segundos (exclusive). Indican que el traductor está pensando y planificando sus próximas acciones.
5. Pausa grande: Duración de 60 segundos o más. Indican que el traductor está trabado y trata de resolver un gran problema.

Notemos que no se clasifican las pausas menores a 2 segundos; esto se debe a que las pausas de menos de 2 segundos se consideran parte de una secuencia de acciones, por ejemplo, tiempos entre presiones de teclas al escribir una palabra. Esta clasificación está basada en la intuición del autor y su amplio conocimiento sobre el proceso de traducción. No es una definición estricta sino que intenta determinar los tipos de problemas con los que se encuentran los traductores.

### 3.2. Técnicas existentes

Existen varias técnicas que pueden ser usadas para estudiar los procesos cognitivos durante la traducción; a grandes rasgos, los métodos pueden ser intrusivos o no intrusivos. Brevemente se repasarán las técnicas no intrusivas, sus ventajas y desventajas. Los métodos invasivos suelen usarse solo para evaluaciones pre-quirúrgicas en pacientes que requieren extirpación de tejido cerebral, en su gran mayoría debido a tumores o epilepsia. Por lo tanto, los métodos intrusivos raramente utilizados para explorar funciones del lenguaje.

#### *Talk Aloud, Think Aloud y Verbalizaciones Retrospectivas*

De acuerdo a Nielsen (2004), las técnicas de *Talk Aloud* y *Think Aloud* se basan en la hipótesis de que todo lo que sabemos se almacena en algún momento en la memoria de corto plazo. Debido a esto, es posible verbalizar lo que se está pensando por un breve período de tiempo ya que la información aún se encuentra en la memoria de corto plazo. Por otro lado, cuando ocurre un lapso de tiempo considerable entre la verbalización y el pensamiento, la información almacenada en la memoria a corto plazo se pierde y en vez de reportar directamente los pensamientos, comenzamos a dar explicaciones y descripciones de los mismos, dando lugar a la técnica de Verbalizaciones retrospectivas.

En resumen, *Talk Aloud* consiste en que el sujeto exprese oralmente pensamientos que ya existen en su cabeza de manera verbal mientras que *Think Aloud* consiste en que el sujeto exprese oralmente la serie de pensamientos que van surgiendo a lo largo de la traducción, incluso si esos pensamientos no son verbales tales como pensamientos visuales, auditivos, etc. Finalmente, la técnica de verbalizaciones retrospectivas consiste en que el sujeto de explicaciones y descripciones de su comportamiento después de haber realizado una traducción.

Se ha demostrado que las verbalizaciones en las técnicas de *Talk Aloud* y *Think Aloud* añaden una carga cognitiva importante de manera que los procesos cognitivos que quieren estudiarse tendrían la carga extra del uso de estos protocolos. La técnica de verbalizaciones retrospectivas tiene la desventaja de que debido al lapso de tiempo que pasa hasta que se verbaliza el pensamiento, la verbalización es más propensa a errores además de que hay que confiar en lo que el sujeto recuerda haber pensado.



### Técnicas de *functional neuroimaging*

Las técnicas de *functional neuroimaging* tienen excelente resolución espacial pero carecen de buena resolución temporal ya que no revelan precisamente cuando ocurre el proceso; estas técnicas ofrecen información precisa sobre que parte del cerebro está implicada en un proceso dado.

*Positron emission tomography* (PET) detecta rayos gamas usando un *scanner* cerebral para medir el flujo de sangre, el metabolismo y los neurotransmisores de unión o absorción. PET tiene una buena resolución espacial y provee la oportunidad de medir aspectos de función cerebral que no pueden observarse con otros métodos no invasivos (por ejemplo, metabolismo de la serotonina o dopamina). Pero más allá de esta ventaja, requiere de grandes presupuestos, equipamiento y personal especializado y potencial peligro para los técnicos y sujetos expuestos a la radiación de iones.

*Functional magnetic resonance imaging* (fMRI) ofrece alta resolución espacial de manera no intrusiva aunque está caracterizada por su baja resolución temporal, altos presupuestos y requerimientos estrictos en cuanto a infraestructura. Más importante aún, la interpretación de los resultados de fMRI no es trivial, (A. M. García).

### Técnicas de *electroencephalography*

Los métodos de *electroencephalography* (EEG) poseen excelente resolución temporal pero resolución espacial pobre. EEG mide las oscilaciones de los potenciales eléctricos del cerebro en el cuero cabelludo que son producidas por la acción cooperativa entre neuronas. Un solo sensor (es decir, un electrodo) provee la estimación de entre 100 millones y 1 billón de neuronas promediadas. Las señales de EEG pueden ser analizadas a través de inspección visual, descomposición de señales en frecuencias, estimando las relaciones entre los sensores y frecuentemente promediando las respuestas evocadas, (A. M. García).

### Técnicas de *Eye-tracking* y *Key-logging*

Esta técnica consiste en guardar las coordenadas en la pantalla del movimiento de los ojos usando una máquina especial, llamada *eye-tracker*, así como también cada pulsación de *mouse* y teclado que realice el traductor durante el experimento con ayuda de herramientas de *key-logging*. Estos datos se conocen como *User Activity Data* (UAD) (Michael Carl, 2008).

La técnica de *Eye-Tracking* se basa en la hipótesis de que el sujeto está pensando en como traducir una sentencia que está observando. Un *eye-tracker* guarda información sobre el foco de atención visual del sujeto mediante coordenadas cartesianas (x, y) e información sobre las pupilas. Esto resuelve el problema de la carga añadida por los métodos anteriores ya que solo registra la interacción visual (Nielsen, 2004). Sin embargo, el *eye-tracker* tiene un gran costo económico por lo que no es accesible a todos los investigadores. Además, tiene la desventaja de que cuando el sujeto mueve la cabeza se ocasiona ruido en el aparato lo que dificulta el posterior análisis. En cuanto a desventajas de la técnica de *key-logging*, la principal se refiere al volumen de datos generados, el

cual no puede ser fácilmente

A pesar de estas desventajas, esta combinación de técnicas es la más utilizada hasta el momento por ser la más completa, no intrusiva y que no altera los procesos cognitivos que quieren estudiarse.

El programa más popular de registro de UAD en el ámbito de la traducción es Translog II, este permite configurar experimentos de traducción así como también *eye-trackers* con los cuales tomar los datos (Translog).

Para resumir, la información recolectada con *eye-tracking* permite evidenciar el procesamiento del texto fuente y el texto meta. Por otro lado, la información recolectada con *keylogging* pone en evidencia lo que ocurre al momento de producir el texto meta.

### 3.3. Factores que afectan la traducción

#### Direccionalidad

La direccionalidad en la traducción se refiere al hecho de que la traducción o interpretación se haga desde la lengua madre o hacia la lengua madre. También son posibles otras combinaciones de idiomas en el caso de las personas multilingües (Pavlovic, 2007).

Cuando la lengua madre y el idioma de origen coinciden, hablamos de la que dirección es inversa. En caso contrario, se dice que la dirección de la traducción es directa. Muchos traductores y estudiantes consideran que la traducción inversa es más dificultosa o desafiante que la traducción directa; sin embargo, hasta el momento no se ha encontrado evidencia significativa de que la carga cognitiva en la traducción inversa sea mayor cuando se traduce en dirección directa.

Nataša Pavlovic (2009) usa *eye-tracking* para estudiar a 8 estudiantes y 8 traductores profesionales al traducir 2 textos (L1 Danés y L2 Inglés). El estudio se focaliza en estudiar el esfuerzo cognitivo mediante el tiempo de fijación, la duración de fijación promedio, la dilatación de la pupila y pupilometría. Se confirmó parcialmente la hipótesis de que las traducciones L2 requieren mayor esfuerzo cognitivo que las traducciones L1; dos de los cuatro indicadores de esfuerzo cognitivo (duración total de la traducción y dilatación de la pupila). Para ambos estudiantes y profesionales, las traducciones inversas requirieron de períodos de tiempo mayores para su finalización que las directas y valores de dilatación de pupila mayores. Se comprobó que los estudiantes requieren de mayor esfuerzo cognitivo en ambas direcciones a comparación de los profesionales (3/4 indicadores); la justificación racional es que los estudiantes todavía no han desarrollado las estrategias necesarias que los ayuden a reducir los tiempos de producción.

Trece estudiantes del tercer año del “Departamento de Idioma Inglés y Literatura” de la “Facultad de Filosofía” en Tuzla, Bosnia participan del experimento de Pavlović (2013) que consiste en usar *key-logging*, *eye-tracking*, cuestionarios y protocolos verbales en sesiones de traducción que involucran textos fuentes no técnicos en L1 Inglés y L2 Bosnio / Croata / Serbio. La longitud de los textos es de alrededor de 140 palabras y fueron extraídos de guías de viaje. Sin usar presión de tiempo, permitiendo que los participantes usen sus propios diccionarios y definiendo un problema como cualquier palabra que haga dudar al sujeto sobre cual es la traducción correcta de

dicha palabra, los cuatro grupos encontraron dificultades en la categoría léxica / semántica (entre 50 % y 66 % de todos los problemas) y problemas sintácticos (20 % y 30 %). Para ciertos tipos de problemas, se encontró un alto grado de coincidencia entre las traducciones en ambas direcciones. La calidad de las traducciones L1 demostraron ser de mayor calidad que las traducciones L2. Sin embargo, el número de problemas total que debieron resolver los traductores fue el mismo en ambas direcciones. El autor infiere que, si el traductor está correctamente entrenado, puede producir una traducción L2 que es de calidad equivalente a una traducción L1. Aunque esto podría llevar más tiempo y podría requerir de mayor esfuerzo.

Pavlovic (2007) usa cuestionarios, historiales de internet, grabaciones de audio y vídeo y se concentran solo en el texto traducido. Doce sujetos (divididos en 4 grupos de 3) participan en las sesiones de traducción colaborativas y 54 sujetos (30 individualmente y 24 en grupos de 8) para los experimentos de control. Deben traducir dos textos, de Inglés a Croata y viceversa. El análisis se hace usando *choice network analysis* (CNA), análisis cuantitativo (mediciones, porcentajes, conteo, cocientes, valores medios, T-tests), análisis cualitativo, mapas mentales y diagramas. En el proceso de traducción que involucra textos de dificultad comparable de dominios no específicos no se encontró una diferencia de esfuerzo cognitivo sustancial en la dificultad con respecto a la direccionalidad en el texto no técnico y el par (Inglés, Croata). Se descubrió que los traductores novatos utilizan mayor cantidad de recursos internos en la traducción directa además de poseer mayor fluidez y obtener traducciones de mayor calidad.

## Memorias de traducción

Numerosas herramientas informáticas, conocidas como programas de traducción asistida, se han desarrollado con el fin de facilitar la tarea de los traductores. Estas herramientas ofrecen sugerencias de auto completado de palabras, opciones de traducción de sentencias y palabras, post-edición de texto, entre otros.

Dentro de este grupo de herramientas informáticas, existe un subconjunto denominado herramientas de memoria de traducción que son en general las más usadas. Una memoria de traducción es básicamente una Base de datos (BD) que almacena *translation units*. Las *translation units* o unidades de traducción, son pares de segmentos de textos fuente y su correspondiente traducción a otra lengua. Los segmentos o unidades de traducción pueden ser sentencias, párrafos o incluso unidades tales como títulos, encabezados o enumeraciones de elementos pero no palabras individuales ya que ese tipo de información se encuentra en las bases de terminología.

El software que usa las memorias de traducción descompone el texto fuente en segmentos. A medida que el traductor comienza a traducir un segmento, el programa busca coincidencias en la BD. Cuando hay una coincidencia, se presenta dicha potencial traducción al usuario. El traductor puede aceptar la sugerencia, modificarla o reemplazarla con una nueva traducción. En los dos últimos casos, la nueva posibilidad segmento se almacena en la BD para futuros usos. Cuando no hay una coincidencia, el traductor manualmente genera un nuevo segmento que es almacenado en la Memoria de traducción (MT).

En cuanto a los algoritmos de búsqueda, existen dos variantes principales. Algunas memorias de traducción buscan coincidencias exactas mientras que otros buscan coincidencias parciales. En el último caso, se emplean *fuzzy matching algorithms* para conseguir los segmentos similares.

En este estudio, se optó por usar la herramienta de memoria de traducción OmegaT ya que es una herramienta *open-source* escrita en Java y que, además, es usada por los estudiantes de “Traductorado de Inglés” de la “Facultad de Lenguas” de la “Universidad Nacional de Córdoba”, quienes conforman la mayoría de la muestra analizada.

Las investigaciones realizadas sobre el efecto real que este tipo de herramientas tienen sobre el proceso de traducción han descubierto que la traducción asistida por computadora acorta el tiempo necesario para llegar al producto final, además de mejorar la calidad de las traducciones finales.

Koehn (2010) usa Caitra, una web interactiva que registra *keystrokes* y provee asistencia en la traducción. El experimento involucra 10 traductores no profesionales, 5 franceses y 5 ingleses. Estos deben traducir textos de 192 sentencias en Francés al Inglés. Se analizan las pausas para ver como distribuyen el tiempo los traductores. Se observa que los textos traducidos son de mayor calidad y se obtienen un 27% más rápido cuando tienen habilitadas las opciones de predicción y un 39% más rápido cuando post-editan. La asistencia proporcionada elimina en gran medida la cantidad de pausas largas (>60 segundos).

(Mellinger, 2014) adapta el software web M. Denkowski (2012) para usarlo en un estudio en el cual participan 9 traductores profesionales. Ellos deben traducir textos periodísticos en español de aproximadamente 400 palabras. El análisis se hace con estadística inferencial y descriptiva, usando ANOVAs y *average pause ratio* para determinar el esfuerzo cognitivo. Todos los participantes trabajaron de manera secuencial; muchos participantes tradujeron de principio a fin sin volver a segmentos anteriores aunque la mayoría volvían a los segmentos anteriores cercanos para realizar correcciones antes de continuar con el resto de los segmentos. Cinco de los nueve participantes mostraron una correlación entre el número de *keystrokes* usados por segmento y la calificación de dificultad asociada a ese segmento. De manera que mientras más difícil era considerada la traducción, mayor era el número de *keystrokes* usados. Por otro lado, tres de los nueve participantes mostraron una correspondencia entre el número de *clicks* usados por segmento y la dificultad asociada al segmento. Tres de los nueve participantes mostraron una correlación entre el tiempo que les llevó traducir un segmento y la dificultad asociada al segmento. Al comparar el esfuerzo ejercido (medido con el *average pause ratio*) con el esfuerzo percibido (medido con las puntuaciones), se notó que existía una correlación entre ambos ya que cuando el *average pause ratio* era elevado también lo era el esfuerzo percibido.

Pym (2011) afirma que la tecnología no necesariamente hace que el proceso de traducción sea más eficiente. Para justificar esto, propone reducir el problema de traducción a tres simples partes (1) un problema se reconoce, por ejemplo: como digo “malestar” en Inglés? (2) se generan alternativas (*discomfort, unease, malaise, etc*) (3) se selecciona una solución. Se puede observar que la memorias de traducción ayudan en en la rápida producción de alternativas en la parte (2).

Bajo algunas circunstancias, el uso de memorias de traducción complican la toma de decisiones y luego podrían convertirse en un impedimento en el proceso de selección (paso 3).

### Presión de tiempo

El factor tiempo es de vital importancia para los intérpretes quienes deben convertir, de manera oral e instantánea, frases del lenguaje fuente al lenguaje meta. Sin embargo, el tiempo también afecta a aquellos traductores que trabajan en agencias de traducción. Por este motivo, resulta interesante analizar cuales son los efectos de la presión de tiempo en el proceso de traducción escrito.

Sharmin et al. (2008) usa el seguimiento ocular de 18 estudiantes del segundo año traductorado en Inglés de la Universidad de Tampere al traducir tres textos cortos, de alrededor de 70 palabras, en dirección Inglés → Finlandés bajo distintas condiciones. Las variables independientes a estudiar son la complejidad del texto y la presión de tiempo durante la tarea; el objetivo es analizar cómo co-varían ambas a través de la duración de la fijación y la cantidad de fijaciones. La complejidad de los tres textos se midió de acuerdo a los elementos léxicos de baja frecuencia y a la complejidad estructural dada por expresiones anafóricas y aposición. La condición de tiempo se estipuló en dar 6, 5 y 4 minutos para cada texto, donde la secuencia de tiempos es constante pero se varían los textos para cada sujeto. Los *tests* estadísticos aplicados no encontraron resultados significativos respecto a fijación en las distintas condiciones de tiempo, tomando en cuenta la pantalla completa. Analizando las pantallas del texto origen y del texto meta por separado, se encuentra que la fijación en la pantalla del texto origen decrece con la disminución de tiempo; sin embargo, en la pantalla meta hay diferencias insignificantes, que se supone han influenciado el resultado de la pantalla completa. En general, la duración de las fijaciones es más larga en la pantalla del texto meta, lo cual tiene mucho sentido al tratarse del objeto de atención de los sujetos.

En (Liparini Campos, 2005) se investiga el efecto de la presión de tiempo en 5 traductores novatos de la Facultad de Letras de la Universidad Federal de Minas Gerais que utilizan el par Alemán - Portugués, encontrando evidencia de que el tiempo influye tanto en el producto como en el proceso. Respecto al proceso se ve que la cantidad de teclas que indican revisión en tiempo real (*guardadas* con Translog) disminuye cuando hay menos tiempo para realizar la tarea. Esto muestra influencia en el producto también, donde se ven frases más segmentadas y con problemas de cohesión. Esta investigación se continúa en (Alves and Campos, 2009), donde se estudia el efecto de la presión de tiempo en relación a los apoyos (internos o externos) utilizados por traductores profesionales. Nuevamente se utiliza el par Alemán - Portugués, esta vez para traducir textos de alrededor de 550 palabras extraído de un manual de instrucciones de un cardiófrecuencímetro y poniendo límites de tiempo individuales, es decir, el límite es constante pero es un porcentaje del tiempo que les lleva copiar un texto a cada sujeto (70 % del tiempo). En este caso se puede ver que se reduce el tiempo destinado a la fase de orientación, siendo entre 0 % y 8,5 % del tiempo total y la mayoría de los sujetos le asignó sólo alrededor del 2 % del tiempo total. Mirando la media total se ve que el uso de mejorías de traducción o el cambio en el tiempo asignado no hace variar el tiempo

asignado a la fase de orientación. Esto se corresponde con la forma de trabajo de los profesionales, que tienden a orientarse gradualmente en la fase de producción, digamos a demanda. Esto además se confirma con algunos relatos de la entrevista retrospectiva realizada por los investigadores.

En el trabajo de doctorado de Hvelplund (2011) se realiza un estudio empírico de la asignación de recursos cognitivos durante la traducción, recolectando datos con un dispositivo de seguimiento ocular y de grabado de actividad de teclado, puntualmente Tobii Eyetracker y Translog. Se consideran cuatro factores principales: el tipo de texto a procesar (origen o meta), experiencia en traducción del sujeto, dificultad del texto origen y presión de tiempo. En las sesiones de traducción participaron 12 profesionales y 12 estudiantes, traduciendo tres textos de alrededor de 850 caracteres tomados de distintos periódicos Británicos en dirección Inglés → Danés. En la condición de nuestro interés (presión de tiempo) se tradujeron dos textos: uno sin límite de tiempo y otro con la restricción, donde los límites de tiempo se fijaron de manera individual; la condición sin restricción fija la cantidad de tiempo igual al utilizado en una tarea de copia (o *baseline*) y la restricción consiste en permitir un 85 % del tiempo empleado en la tarea de copia. Entre las hipótesis planteadas están la cantidad de *unidades de atención* - *AU* (cambios en movimientos oculares y de tipeo que indican un cambio de tarea cognitiva) y carga cognitiva en relación al tiempo asignado para traducir. La primera predecía que bajo presión de tiempo las AU son de menor duración, lo cual pudo ser confirmado por los datos recolectados, incluso hay evidencias de que las AU más afectadas son las que suceden cuando se procesa el texto origen, mientras que en el procesamiento del texto meta no hay cambios en las AU. La segunda hipótesis mencionada predice que la carga cognitiva es mayor cuando hay un tiempo limitado para terminar la tarea y también pudo ser confirmada con los datos de la dilatación de la pupila, la cual es mayor si hay restricciones de tiempo. Sin embargo, algunas otras hipótesis no se pudieron confirmar o refutar con los datos coleccionados, resaltando la gran dificultad que hay en este tipo de estudios, incluso utilizando tecnología de punta y una variada muestra de sujetos.

# Capítulo 4

## Metodología

### 4.1. Herramientas desarrolladas

Existen múltiples herramientas de *key-logging* que pueden usarse para el estudio de los procesos de escritura y comportamiento de usuario. En el ámbito de la traducción las más conocidas son Translog e InputLog. La primera permite la toma de datos tanto de *keylogging* como de *eye-tracking* y con ella se han desarrollado numerosas investigaciones sobre traductología. La segunda es una herramienta que guarda en un *log* la actividad de: teclado, *mouse* y sonido; principalmente es usada para describir procesos de escritura o traducción en detalle.

En este trabajo final, se optó por extender una herramienta de código abierto de manera que sea una alternativa libre y que al mismo tiempo provea las mismas características ofrecidas por sus competidores. Por este motivo, se desarrolló ResearchLogger, que está basada en un keylogger de código abierto escrito en el lenguaje de programación *Python* llamado PyKeylogger versión 1.2.1. ResearchLogger se usó para la toma de datos de esta investigación.

Los editores de texto sofisticados ofrecen una gran variedad de opciones de formato y atajos de teclado que crean distracciones innecesarias para lo que se desea estudiar. Por este motivo, se implementó desde cero una interfaz de traducción de texto plano que ofrezca un entorno controlado. Esta interfaz se usó en todas las condiciones con excepción de la condición con memoria, en la cual se utilizó OmegaT. Algunos usuarios pudieron usar una base terminológica o página de búsqueda de términos, que en nuestro caso fue Linguee.

Además, a lo largo de este trabajo se emplean diferentes *scripts* escritos en R y Python para analizar los datos de manera eficiente y buscar patrones en ellos. Ya que estos scripts podrían facilitar la tarea de otros investigadores, se decidió reunir los scripts generados en una *Application Programming Interface* (API). Esto dio lugar a la herramienta ResearchAnalyzer.

#### 4.1.1. ResearchLogger

La técnica de *key-logging* que usa ResearchLogger permite observar cómo un traductor produce una traducción más allá de la entrada (texto fuente) y la salida (texto traducido) del proceso, es

decir, permite la observación del proceso *online*. Por lo tanto, el log detallado nos permite saber como el traductor distribuyó el tiempo para cada tarea, cuando hubo pausa, etc.

## Base: PyKeylogger

Se mantuvo el diseño original de PyKeylogger añadiendo algunas modificaciones que consistieron en: (a) arreglar bugs (b) cambiar la estructura de los *logs*. Aunque, en teoría, PyKeylogger versión 1.2.1 guardaba información sobre los *clicks* y tomaba fotos del área alrededor de donde ocurre un *click*, estas funcionalidades no estaban presentes al usar el programa en Ubuntu o en Debian. Un análisis en profundidad reveló que el código estaba presente pero no funcionaba adecuadamente en el sistema operativo mencionado. Como la toma de datos se hizo bajo sistema operativo Ubuntu, la primera mejora al código consistió en arreglar varios *bugs*. En cuanto a la estructura de los *logs* de PyKeylogger versión 1.2.1, inicialmente solo guardaban información sobre las teclas presionadas (no sobre la liberación de las mismas).

Inicialmente, la información que PyKeylogger guardaba sobre los *keystrokes* era la siguiente:

- Fecha y hora.
- Programa de foco y título de la ventana donde se origina el evento.
- Nombre de usuario.
- Nombre univoco de tecla (solo presiones de tecla).

Es decir que un log original tenía el siguiente formato:

```
20150606|1653|terminal|Window|user|PCName:/home/user/Tesis/Writing|hola
20150606|1654|terminal|Window|user|PCName:/home/user/Tesis/Writing|chau
[KeyName:Control_L][KeyName:Control_R][KeyName:F12]
20150606|1655|message|Window|user|Enter Password|[KeyName:Return]
```

Figura 4.1: Ejemplo de log original de PyKeylogger.

## Agregados y modificaciones a PyKeylogger

Las modificaciones en el código de PyKeylogger se realizaron en los siguientes archivos: *baseeventclasses.py*, *controlpanel.py*, *myutils.py* y principalmente en *onclickimagecapture.py*, *timed-screenshot.py*, *pyxhook.py* y *keylogger.pyw* que fue renombrado a *keylogger.py*. Además, se añadió un módulo de constantes *constants.py* y otro para pruebas automatizadas *tests.py*.

El primer paso en el desarrollo de ResearchLogger consistió en modificar la estructura del log de *keystrokes* que proveía PyKeylogger originalmente para guardar mayor información y de manera que sea más fácil de parsear posteriormente. Como resultado, se agrega mayor información sobre los *keystrokes* para contar con los siguientes datos:



- Fecha y hora.
- Programa de foco y título de esa ventana.
- Nombre de usuario.
- Nombre univoco de tecla.
- Milisegundo en el que se presionó y liberó una determinada tecla.
- Tipo de evento  $\in \{key\_down, key\_up\}$
- Coordenadas donde estaba el mouse cuando se apretó dicha tecla.

Un log de *keystrokes* obtenido al usar ResearchLogger tiene el siguiente formato:

```
20141031|1835|tk|Window|user|Sesión de traducción|120485670,l,key_down,331,508
120485757,l,key_up,331,508 120485911,e,key_down,331,508 120485979,e,key_up,331,508
120488004,backspace,key_down,331,508 120488063,backspace,key_up,331,508
20141031|1836|tk|Window|user|Sesión de traducción|120524795,shift,key_down,225,495
120524870,e,key_down,225,495 120524898,shift,key_up,225,495 120524954,e,key_up,225,495
```

Figura 4.2: Ejemplo de log *keystrokes* de ResearchLogger.

Se consiguió que el programa original capture información sobre los clicks en un log especializado y que se tomen fotos del área alrededor de las coordenadas donde se origina un determinado *click*. La información que se almacena sobre los *clicks* es la siguiente:

- Fecha y hora.
- Programa de foco y título de esa ventana.
- Resolución actual.
- Coordenadas donde se efectuó el *click*.
- Milisegundo en el cual se efectuó el *click*.
- Tipo de evento  $\in \{left\_down, left\_up, middle\_down, middle\_up, right\_down, right\_up\}$

Como resultado, un log de *clicks* tiene el siguiente formato:

En resumen, ResearchLogger guarda información detallada sobre la actividad de teclado y *mouse* de todas las ventanas activas. Esto tiene la ventaja de que puede saberse cuanto tiempo se dedica en las diferentes ventanas, por ejemplo, navegador web, sesión de traducción, memoria de traducción, etc.

```

20141031|1834|tk|Window|user|Abrir|1280x1024|659,72,120419861,left_down,imagen1.png
659,72,120419999,left_up,imagen2.png
20141031|1835|tk|Window|user|Abrir|1280x1024|383,268,120479762,left_down,imagen3.png
383,268,120479892,left_up,imagen4.png

```

Figura 4.3: Ejemplo de log *clicks* de ResearchLogger.

## Ventana de traducción

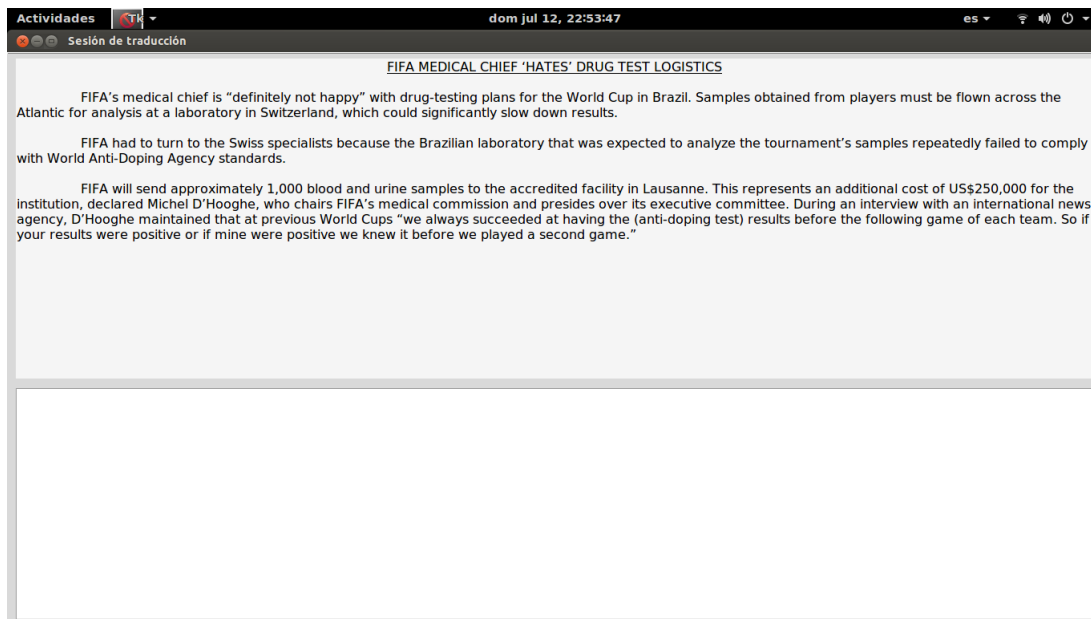


Figura 4.4: Captura de pantalla de la ventana de traducción

La ventana de traducción se ofrece junto al código de ResearchLogger; su uso es opcional e independiente de la herramienta de *key-logging*. Está escrita en Python, usando las librerías gráficas *Tkinter* y *tkSimpleDialog*.

Una de las características útiles que posee es que al cerrar la ventana de traducción, se guarda el texto traducido automáticamente. Esto evita la pérdida de información en caso de que el sujeto accidentalmente cierre el programa sin haber guardado el documento final.

### 4.1.2. ResearchAnalyzer

ResearchAnalyzer es una colección de *scripts* escritos en *Python* que permite que el usuario pueda analizar fácilmente los *logs* generados con ResearchLogger. En su primer versión se ofrece análisis de fases en la traducción, análisis de pausas, un *parser* e información sobre los *clicks* y *keystrokes* usados en la sesión.

## Descripción de la API

Las clases ClickParser, KeyParser y Parser son las encargadas del *parsing* los *logs*. ClickParser permite analizar los *clicks* independientemente del resto del log. Provee los siguientes métodos:

- **set\_clicks(title):** Método privado que carga en memoria el registro de todos los *clicks* con el título de ventana *title*. Si el parámetro es el *string* vacío, entonces nos interesan todos los títulos.
- **pretty\_printer():** Imprime en pantalla el log en crudo

KeyParser permite analizar los *keystrokes* independientemente del resto del log.

- **set\_keys(title):** Método oculto que carga los *keystrokes* en memoria. El *string* argumento indica que títulos de ventanas nos interesan, *string* vacío si nos interesan todas.
- **pretty\_printer():** Método que imprime en pantalla el log crudo de manera legible.

Parser analiza conjuntamente los *keystrokes* y *clicks*. De momento las capturas de pantallas y el área al rededor de los *clicks* son a modo de ayuda para el investigador y no son analizadas por ResearchAnalyzer.

- **merge\_clicks\_and\_keys\_logs():** Método oculto que trae a memoria la combinación de *clicks* y *keystrokes* guardados.
- **pretty\_printer():** Imprime en pantalla una versión legible del log de *keystrokes* y de *clicks*.

LogInfo brinda resúmenes de información sobre los logs en cuanto a: *clicks*, *keystrokes*, pausas y fases de la traducción.

- **get\_unique\_pressed\_clicks():** Devuelve el conjunto de los tipos de *clicks* que se usaron en la sesión. Los tipos de *clicks* pertenecen al conjunto {left, middle, right}.
- **get\_all\_pressed\_clicks():** Devuelve la lista de tuplas de todos los eventos de *click* detectados. Cada tupla contiene información sobre: (coordenada x, coordenada y, resolución de pantalla, tipo de presión, nombre del programa, título de ventana).
- **get\_click\_info():** Retorna una lista de tuplas con un resumen de la información de los *clicks*. Cada tupla es de la forma: (tipo de click, tiempo de presión, tiempo de liberación, tiempo total de presión, coordenada x al presionar, coordenada y al presionar, coordenada x al soltar, coordenada y al soltar).
- **print\_click\_summary():** Imprime en pantalla un breve resumen de información sobre los *clicks*: cantidad de *clicks*, tipos de *clicks* y ventanas en las cuales se hizo *clicks*, tiempo promedio de presión de *clicks*, varianza y desviación estándar de los tiempos de presión de *clicks*.
- **get\_unique\_pressed\_keys():** Devuelve el conjunto con todas las teclas (únicas) que se usaron en la sesión.

- **get\_all\_pressed\_keys():** Devuelve una lista de tuplas que representa todas las teclas que se presionaron en la sesión, en el orden en el que ocurrieron. Cada tupla tiene información sobre (id de tecla, tipo de evento, programa, título de la ventana)
- **get\_letter\_info():** Dada una estructura que contiene el log parseado, crea una lista con tuplas que resume la siguiente información: (id de tecla, tiempo de presión, tiempo de liberación, tiempo total de presión, coordenada x al presionar, coordenada y al presionar, coordenada x al soltar, coordenada y al soltar)
- **print\_key\_summary():** Imprime en pantalla un breve resumen de información sobre las teclas usadas en la sesión: cantidad de teclas presionadas, teclas únicas presionadas, ventanas donde se escribió, presión de tecla promedio, varianza y desviación estándar de presión de tiempo de tecla, teclas de función usadas, teclas de borrado usadas, teclas de movimiento usadas y combos de teclado usados.
- **get\_orientation\_info():** Devuelve un par (tiempo de inicio, tiempo de finalización) correspondiente a la fase de orientación.
- **get\_drafting\_info():** Devuelve un par (tiempo de inicio, tiempo de finalización) correspondiente a la fase de elaboración de un borrador.
- **get\_revision\_info():** Devuelve un par (tiempo de inicio, tiempo de finalización) correspondiente a la fase de revisión.
- **get\_total\_session\_time():** Devuelve el tiempo total de la sesión.
- **print\_phases\_summary():** Imprime en pantalla un resumen sobre las fases de la traducción: (a) orientación (b) elaboración de un borrador (c) revisión como se ve en "Procesamiento del texto meta". Para cada una de las fases, se calcula la duración de la fase en milisegundos y el porcentaje de tiempo de la sesión total. Requiere que manualmente se marque con el símbolo especial # el final de la fase de elaboración de un borrador, o lo que es equivalente, el inicio de la fase de revisión.
- **print\_pauses():** Muestra en pantalla el tiempo de duración en milisegundos de las pausas que caen en alguna de las siguientes categorías (a) pausas cortas: entre 2 y 6 segundos, se muestran en color verde (b) pausas medianas: entre 6 y 60 segundos, se muestran en amarillo (c) pausas largas: mayores a 60 segundos, se imprimen en rojo. Los tipos de pausas se determinan de acuerdo a Koehn (2010).
- **get\_pauses():** Devuelve una lista de tiempos que representan todas las pausas que hubo en la sesión en el orden en que ocurrieron.
- **print\_pause\_summary():** Imprime en pantalla un resumen sobre las pausas con la siguiente información: Cantidad de pausas total, Cantidad de pausas cortas, medianas y largas,

Cantidad de pausas no significativas (menores a dos segundos), Tiempo promedio de pausa, Varianza de las pausas.

Los métodos de las clases PauseInfo y PhaseInfo son usados por LogInfo para modularizar el código pero pueden usarse independientemente, según sea conveniente.

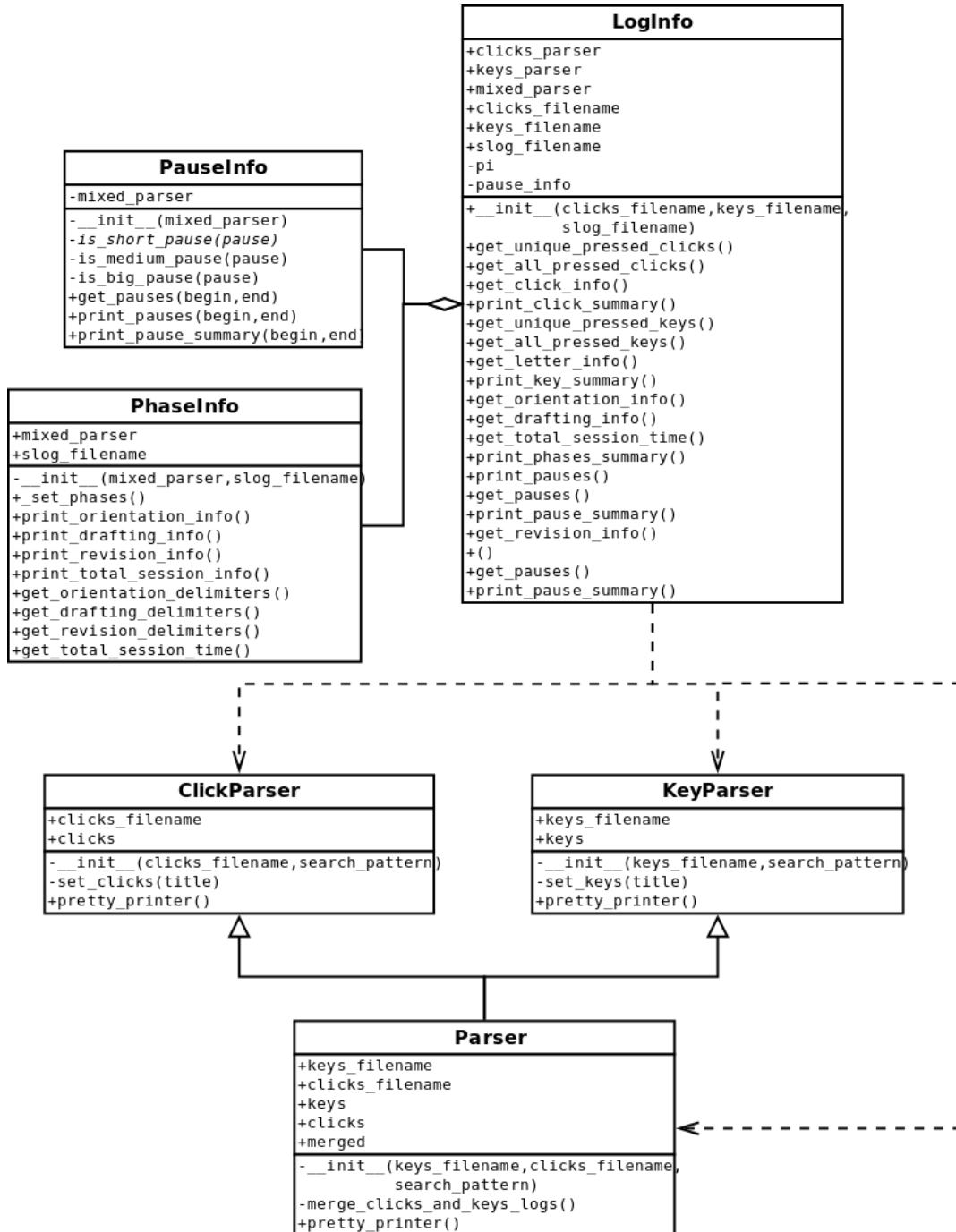


Figura 4.5: Diagrama de clases de ResearchAnalyzer

## Detalles de la implementación

Para automatizar la detección de las fases, el problema principal reside en que la computadora aprenda a discernir cual es el punto final que determina la finalización de la etapa de elaboración de borrador. Como intentar resolver este problema, excede este trabajo final, supusimos ciertas precondiciones para realizar el proceso automáticamente:

- El log está estructurado con el formato generado por ResearchLogger.
- No existe un evento up sin al menos un evento down relacionado.
- El punto de finalización de la fase de elaboración de borrador está señalizada con un símbolo especial: #, siempre respetando el mismo formato que el *log*.

Bajo estas suposiciones fue fácil automatizar el análisis de las fases. La detección de las fases a partir de los logs generados por ResearchLogger se determinó como sigue:

- Orientación: Desde el key\_up del click en el botón de abrir hasta el key\_down de la primer tecla que se aprieta.
- Elaboración de borrador: Desde el key\_down de la primer tecla que se aprieta hasta el key\_down del ultimo punto (marcado con el símbolo especial #).
- Revisión: Desde el key\_down del ultimo punto (marcado con el símbolo especial #) hasta que se cierra el programa.

### 4.1.3. Comparación con otras herramientas

Compararemos ResearchLogger con dos productos: Translog e InputLog. El primero es desarrollado en la *Copenhagen Business School* para el estudio de los procesos en la traducción (Carl, 2012). El segundo es desarrollado en la *University of Antwerp* y sus principales desarrolladores son Luuk Van Waes y Mariëlle Leijten. Inputlog ejecuta análisis específicos para los textos que se escriban usando Microsoft Word (M. Leijten, 2005).

De acuerdo al análisis realizado por Gabriela Saldanha (2013), Translog tiene fortalezas y limitaciones. En primer lugar, requiere que el traductor trabaje dentro del entorno del programa. Es decir, que no guarda la actividad que ocurre fuera de Translog. Esto limita la posibilidad de analizar la interacción con otras tecnologías de traducción tales como las memorias de traducción. Además, el restringir a los participantes a un ambiente que no es familiar para ellos acota sus rutinas de traducción usuales; esto obviamente tiene un impacto en la validez de la investigación. En contraste, ResearchLogger, al igual que InputLog, guarda en un *log* todo lo que ocurre en todas las ventanas mientras el programa está activo, esto da más flexibilidad que Translog.

En segundo lugar, como Translog no posee correctores de ortografía ni opciones de formato, estos aspectos de la producción de la traducción no pueden ser estudiados. En esta investigación usamos un editor de texto plano ya que no nos interesan las ediciones de formato; no obstante,

esto puede estudiarse con ResearchLogger al usar algún editor de textos como Libre Office Writer o MS Word. Lo mismo ocurre con InputLog.

Gabriela Saldanha (2013) también menciona que otras herramientas de *key-logging* como BB-FlashBack o InputLog funcionan independientemente del entorno en el que trabaje el traductor. Estas herramientas también permiten la grabación de audio, lo cual habilita la captura simultánea de reportes verbales. En este sentido, la versión actual de ResearchLogger no captura audio como su competidores pero podría añadirse esta característica a futuras versiones de la herramienta. Un punto a favor de Translog es que puede usarse en conjunto con *eye-trackers*. Esto, de momento, está fuera del alcance de ResearchLogger y no será considerado en el futuro cercano.

Finalmente, Gabriela Saldanha (2013) menciona que Translog es gratuito para la comunidad científica. Y si bien la herramienta es generalmente estable, como su desarrollo no es la principal preocupación de sus creadores, las mejoras y actualizaciones son lentas. De modo que algunas codificaciones de caracteres y sistemas operativos son lentas de ser introducidas y por lo tanto no son soportadas; los bugs también pueden ser lentos de corregir. Esto es un claro punto a favor de ResearchLogger ya que es de código abierto y gratuito. Esto posibilita que cualquier programador pueda adaptar el software de acuerdo a sus necesidades sin tener que esperar a que salga una nueva versión.

Tanto InputLog como ResearchLogger usan estadística descriptiva para mostrar un resumen de las pausas. En el caso de ResearchLogger, bajo ciertas suposiciones mencionadas anteriormente, es posible obtener un resumen sobre pausas; tal y como se describe en la API. Para concluir, combinando información de Writing PRO: Logging programs y con la experiencia propia de haber probado los programas manualmente, se puede resumir la información detallada en la siguiente tabla.

| <b>Programa</b><br><b>Característica</b>                  | <b>ResearchLogger</b> | <b>Translog 2</b> | <b>InputLog</b> |
|---|-----------------------|-------------------|-----------------|
| ¿Guarda <i>keystrokes</i> y movimientos de <i>mouse</i> ? | Sí                    | Sí                | Sí              |
| ¿Es combinable con <i>eye-tracking</i> ?                  | No                    | Sí                | Sí              |
| ¿Identifica aplicaciones?                                 | Sí                    | No                | Sí              |
| Sistemas operativos                                       | Windows, Linux        | Windows           | Windows         |
| Entorno de <i>key-logging</i>                             | Toda ventana          | Solo Translog     | Toda ventana    |
| Resumen de los datos                                      | Sí                    | No                | Sí              |
| Análisis de pausas  | Sí                    | No                | Sí              |
| Análisis de fases   | Sí                    | No                | No              |
| Replay de sesión  | No                    | Sí                | Sí              |
| ¿Código abierto?  | Sí                    | No                | No              |

Cuadro 4.1: Comparación general. ResearchLogger y otras herramientas

Adicionalmente, se puede profundizar la comparación entre Translog y ResearchLogger de acuerdo al tipo de *key-logging* realizado. En primer lugar, Translog no guarda información sobre las teclas de función que se presionan, por ejemplo, BLOQ MAYUS, SHIFT, CTRL. Esto impide que pueda determinarse cómo se capitalizó una letra (usando SHIFT o BLOQ MAYUS), pero más importante aún, usando este software no puede saberse que combos de teclas que se usaron en una sesión. Al usar un carácter con tilde tal como “á”, Translog guarda en el *Log* un único evento: “á”, la aparición de la letra con tilde. En cambio, ResearchLogger discrimina la presión de las dos teclas por separado. A nuestro entender, Translog estaría guardando información sobre la aparición de caracteres en pantalla y no el momento de presión de cada tecla por separado.

En segundo lugar, Translog no diferencia cuando una tecla se presiona reiteradamente o se aprieta una sola vez pero se mantiene presionada durante varios segundos. Esto afectaría el cálculo del tiempo de presión de una tecla. ResearchLogger discrimina entre presión y liberación de teclas, dando la mayor cantidad de información posible sobre una sesión y permitiendo la obtención de inferencias más robustas.

| <b>Programa</b><br><b>Característica</b>                | <b>ResearchLogger</b>          | <b>Translog 2</b>                              |
|---|--------------------------------|--|
| Tipo de evento que guarda                               | Presión y liberación de teclas | Momento de aparición de caracteres en pantalla |
| Guarda información sobre las combinaciones de teclas    | Sí                             | No   |
| Diferenciación entre pulsaciones continuas o reiteradas | Sí                             | No   |
| Distingue teclas de función                             | Sí                             | No   |

Cuadro 4.2: Comparación keylogging. ResearchLogger y otras herramientas

## 4.2. Sesiones de traducción

### 4.2.1. Participantes

Al finalizar cada sesión, se solicitaba que cada uno de los catorce participantes llenara tres cuestionarios: uno sobre conocimientos informáticos, otro de historial de idiomas y el último sobre bilingüismo y traducción. Todos los cuestionarios fueron auto-administrados y están disponibles en el anexo “Cuestionarios usados”.

### Conocimientos informáticos

El cuestionario de conocimientos informáticos incluye preguntas de auto-calificación sobre las principales variables de interés: conocimiento sobre sistemas operativos, programas de memoria de traducción y procesadores de texto además de hábitos al mecanografiar.



Todos los participantes tenían conocimientos suficientes para instalar y des-instalar programas, navegar por el sistema de archivos con el fin de crear carpetas o realizar operaciones sobre archivos bajo el sistema operativo Windows. Solo 9 de los 15 participantes, un 60 % del total, sabía realizar todas estas operaciones en Ubuntu. Aunque ninguno de los participantes había tomado cursos de mecanografía, todos consideraban tener una habilidad buena o muy buena para mecanografiar.

Finalmente, un 80 % de los participantes consideraba saber usar la herramienta OmegaT. Solo un 33 % de los participantes había usado SDL Trados y un 20 % había estado en contacto con Multiterm.

## Historial de idiomas

El *Language History Questionnaire* (LHQ) es una herramienta que permite la evaluación de los antecedentes lingüísticos en bilingües tales como hábitos de uso del lenguaje y competencia en múltiples lenguajes, (Li, 2013). El cuestionario usado en esta investigación es una adaptación de la herramienta mencionada ya que, al estar estandarizado, facilita la comparación de los resultados en diferentes experimentos.

Un 27 % de los participantes tuvo un contacto de 4 meses o más en un país extranjero de habla inglesa. Algunas particularidades a notar son las siguientes:

- El sujeto 5 es trilingüe, habiendo residido por 11 años en el exterior y hablando diariamente Inglés, Español y Francés.
- Los sujetos 7, 8 y 12 hablan Francés como lengua alternativa.
- El sujeto 14 tomó breves cursos de Alemán, Portugués e Italiano.

## Bilingüismo y Traducción / Interpretación

El nivel de competencia traductora depende críticamente de la cantidad de experiencia en la traducción además de habilidades lingüísticas en L1 y L2, (Adolfo M. García, 2014). Por eso, usamos el cuestionario sugerido en Adolfo M. García (2014) para evaluar las competencias traductorales de los sujetos. El cuestionario incluye preguntas demográficas e items de auto-evaluación con el fin de sus capacidades. Las variables de principal interés son: edad, método de aprendizaje de L2, cantidad de años de estudio y exposición en L2.

La mayoría de los participantes iniciaron su formación en el idioma inglés a una temprana edad. En general, aquellos con mayor cantidad de años dedicados al estudio del inglés (más de 10 años) cuentan con certificaciones en el idioma tales como el *Preliminary English Test* (PET) o el *First Certificate in English* (FCE) que son otorgados por Cambridge English School Of Languages (ESOL). En particular, el sujeto 12 aprobó el *Test of English as a Foreign Language* (TOEFL) con un puntaje de 103/120. Algo a notar en este aspecto es que sólo los sujetos 8 y el 12 están ejerciendo su profesión aún siendo estudiantes.

### 4.2.2. Condiciones

Se realizaron tres experimentos: direccionalidad, uso de recurso y presión de tiempo. Para esto, se organizaron cinco condiciones: *baseline*, traducción directa sin memoria y sin presión de tiempo, traducción inversa sin memoria y sin presión de tiempo, traducción directa con memoria y sin presión de tiempo y finalmente traducción directa sin memoria y con presión de tiempo. Los experimentos dejan que cada participante decida como actuar en caso de desconocer una palabra, si incluir el título del texto, entre otras decisiones. A continuación, se detallan las características de cada una de las condiciones:

#### Baseline

Los participantes debían transcribir un texto en idioma Inglés a su velocidad de escritura normal, es decir, no tenían límite de tiempo para completar la tarea. El objetivo de esta condición es saber a que velocidad de escritura trabajaba cada sujeto. El texto puede encontrarse en el anexo "Venezuela takes first steps toward dialogue".

#### SMD-20: Traducción directa sin memoria y sin presión de tiempo

Cada sujeto debía traducir el texto "FIFA medical chief 'hates' drug test logistics" de Inglés a Español en aproximadamente 20 minutos como máximo y sin el uso de una memoria de traducción. Este tiempo fue elegido de manera que no supusiera una presión de tiempo en los participantes.

#### SMI-20: Traducción inversa sin memoria y sin presión de tiempo

Con un límite de tiempo de 20 minutos, al igual que en las otras condiciones sin presión de tiempo, los participantes debían traducir el texto "La FIFA inicia procedimiento disciplinario contra México por racismo de aficionados" de Español a Inglés. Para esta condición tampoco debían usar memoria de traducción.

#### SMD-10: Traducción directa sin memoria y con presión de tiempo

Tal y como sugiere Hansen (2006), los límites de tiempo que presionan a cada participante dependen de su experiencia como traductores. La dificultad en este tipo de estudio es que la gente reacciona diferente a la presión de tiempo, por lo tanto, lo que no representa presión de tiempo para algunos podría representarlo para otros. En nuestro caso, se propuso un límite de 10 minutos para traducir el texto "IOC tightens control over Rio games". Se consideró este tiempo como adecuado dada la longitud del texto y la competencia traductora de los participantes.

#### CMD-20: Traducción directa con memoria y sin presión de tiempo

Se decidió utilizar OmegaT ya que (a) es la herramienta de memoria usada en la Facultad de Lenguas de la Universidad Nacional de Córdoba, por lo que los sujetos están familiarizados

con la misma (b) es una herramienta de código abierto y gratuita (c) funciona bajo Windows y sistemas operativos Linux. En este caso, los sujetos debían traducir “García Márquez left behind unpublished manuscript” en aproximadamente 20 minutos.

### Balanceo de condiciones

Para evitar obtener resultados obvios por el cansancio de los participantes en la condición que se lleve a cabo al finalizar las sesiones, los 15 sujetos realizaron las mismas tareas pero en diferente orden. Dicho orden puede observarse a continuación:

|    | Condición 1 | Condición 2 | Condición 3 | Condición 4 | Condición 5 |
|----|-------------|-------------|-------------|-------------|-------------|
| 1  | SMD-20      | CMD-20      | SMI-20      | SMD-10      | Baseline    |
| 2  | CMD-20      | SMD-20      | SMI-20      | SMD-10      | Baseline    |
| 3  | SMD-20      | CMD-20      | SMI-20      | SMD-10      | Baseline    |
| 4  | Baseline    | CMD-20      | SMD-10      | SMI-20      | SMD-20      |
| 5  | Baseline    | SMD-10      | SMI-20      | CMD-20      | SMD-20      |
| 6  | Baseline    | SMI-20      | CMD-20      | SMD-10      | SMD-20      |
| 7  | Baseline    | SMI-20      | SMD-20      | SMD-10      | CMD-20      |
| 8  | Baseline    | SMD-10      | SMD-20      | CMD-20      | SMI-20      |
| 9  | Baseline    | SMD-20      | SMI-20      | CMD-20      | SMD-10      |
| 10 | Baseline    | SMD-10      | CMD-20      | SMD-20      | SMI-20      |
| 11 | Baseline    | CMD-20      | SMI-20      | SMD-10      | SMD-20      |
| 12 | Baseline    | SMD-20      | SMD-10      | CMD-20      | SMI-20      |
| 13 | CMD-20      | SMD-20      | SMI-20      | SMD-10      | Baseline    |
| 14 | Baseline    | SMD-10      | SMI-20      | SMD-20      | CMD-20      |

Cuadro 4.3: Orden de condiciones para los participantes.

#### 4.2.3. Textos usados

Cinco textos semi-auténticos originalmente publicados en la versión *online* del Buenos Aires Herald entre Mayo y Junio de 2014 fueron usados para los experimentos. Los artículos tratan sobre temas no técnicos y están destinados al público en general. Cada texto fue ligeramente alterado para incrementar el nivel de similitud entre cada uno de ellos.

Todos los textos fueron cuidadosamente igualados para múltiples variables como puede verse en el apéndice “Selección de los textos”. Esto incluye siete métricas de legibilidad <sup>1</sup> (como propone

<sup>1</sup>Dos de los índices (*Flesch Reading Ease* y *LIX*) proveen estimaciones sobre el nivel de dificultad de los textos. Los otros cinco (*Automated Readability Index*, *Flesch-Kincaid*, *Coleman-Liau*, *Gunning Fog* y *SMOG*) evalúan el número de años (en términos del sistema de calificaciones de Estados Unidos) necesarios para leer los textos sin dificultad.

Jensen (2009)), cantidad de oraciones y tipos de oraciones; cantidad de palabras, longitud de palabras, frecuencia promedio de palabra, cantidad de caracteres y no literalidad (como sugiere Jensen (2009)). Además, siguiendo el procedimiento propuesto por Hvelplund (2011), los textos fueron evaluados por un panel de lectores expertos en términos de comprensibilidad, coherencia y correctitud gramatical.

Los resultados demuestran que los textos son similares en todas las variables. Los textos fueron evaluados como difíciles y adecuados para alumnos de secundario (con edades en el rango de 18 a 22) del sistema educativo de Estados Unidos. Todos los textos tienen 165 palabras distribuidas en ocho oraciones. Las similitudes entre los textos abarcan cantidad total de caracteres y caracteres por palabra, longitud de palabra promedio y sílabas por palabra. MRM-ANOVA reveló que la frecuencia por palabra promedio no era significativamente diferente entre los textos. Adicionalmente, se compararon los textos en términos de no linealidad (número de *idioms*, metáforas o expresiones metonímicas). Finalmente, siguiendo el procedimiento descrito por Hvelplund (2011), un panel de tres revisores expertos con L1 Español L2 Inglés se encargó de leer los textos. Una vez leídos los textos, los expertos calificaron cada texto del 1 (más fácil) al 5 (más difícil) de acuerdo al nivel de comprensibilidad, coherencia y correctitud gramatical. De acuerdo a los datos de los cuestionarios, los cinco textos obtuvieron una calificación de 4 en promedio.

#### 4.2.4. Protocolo

Las sesiones de traducción duraron entre 2 y 3 horas cada una, dependiendo del participante. Las mismas se llevaron a cabo en el laboratorio 30 de la Facultad de Matemática, Astronomía y Física de la Universidad Nacional de Córdoba. En cada sesión se siguió el siguiente protocolo:

1. Recepción de participantes: breve descripción del proyecto y de las tareas a realizar.
2. Entrega de formulario de consentimiento informado para que firmen (una copia para ellos y otra para nosotros).
3. Comienzo de la sesión de acuerdo al orden detallado en la subsección "Balanceo de condiciones".
4. Despedida de los participantes; guardado de los datos con SCP y back-up en pendrive.

## Capítulo 5

# Resultados

En este capítulo se detallan los resultados obtenidos al analizar los datos de *key-logging* de 14 estudiantes de traductorado de Inglés. Cada uno de ellos participó en una sesión de traducción dividida en tres experimentos independientes entre sí tal y como se detalló en el capítulo anterior.

### 5.1. Subgrupos

La muestra es un grupo heterogéneo de estudiantes de diferentes niveles de la Facultad de Lenguas de la Universidad Nacional de Córdoba; todos con L1 Español y L2 Inglés. Doce de los catorce participantes son de sexo femenino y todos entre 20 y 30 años de edad. Información más detallada sobre las competencias traductoradas de los participantes pueden encontrarse en la sección: “Participantes”. Se clasificó a los participantes en tres sub-grupos:

- Nivel 3: Estudiantes de 3er año. Sujetos 1, 2, 3, 4, 5, y 6. (6 sujetos)
- Nivel 4: Estudiantes de 4to año. Sujetos 12, 13 y 14. (3 sujetos)
- Nivel 5: Estudiantes de 5to año. Sujetos: 7, 8, 9, 10 y 11. (5 sujetos)

Como el tamaño de la muestra es pequeño (14 sujetos), en general, al hacer las comparaciones asumimos que el grupo es homogéneo (todos estudiantes). Sin embargo, cuando se crea conveniente, se harán comentarios sobre las comparaciones entre los sub-grupos mencionados.

### 5.2. Tipo de análisis

Los resultados presentados en este trabajo deben considerarse solo exploratorios y debido al tamaño pequeño de la muestra.

## Gráficos

Se generaron gráficos de puntos y líneas usando el lenguaje R para comparar los tiempos de sesión y de fases por participante y *bar plots* para analizar los distintos tipos de pausas en la sesión

y las fases para cada participante.

### *Mixed Repeated Measures Analysis of Variance (MRM-ANOVA)*

ANOVA es un método de estadística inferencial que prueba la igualdad de tres o más medias poblacionales por medio del análisis de varianzas muestrales. Para comparar varianzas se utiliza la distribución F como estadístico de prueba. Esta distribución decide si dos varianzas son iguales.

MRM-ANOVA combina el control de la varianza intra-sujeto de las medidas repetidas con un factor inter-sujeto (grupo). Dado que este método es sensible al número de participantes, hemos reportado el *partial eta* para estimar el tamaño del efecto de las comparaciones estadísticas. El *effect size* es un valor que permite determinar en que medida la Variable Independiente (VI) afecta a la Variable Dependiente (VD). Se usó *partial eta squared* ( $partial \eta^2$ ) como *effect size* para ANOVA. *Partial  $\eta^2$*  es una estimación de la proporción de variación de la variable dependiente que está explicada por cada efecto. Cuando no existieron variables inter-sujetos o de medidas repetidas, se usó ANOVA.

### Post hoc Tukey

MRM-ANOVA detecta si hay diferencias significativas entre grupos o condiciones pero no dice cuales son los grupos o condiciones que difieren; para determinar esto último, cuando hay más de dos comparaciones posibles en un análisis dado, se usan los *post hoc tests*. Estos *tests* solo deben usarse cuando se detecta una diferencia significativa entre las medias de los grupos.

Hay una gran cantidad que *post hoc tests* que pueden usarse: Fisher's LSD (*Least Significant Different*), Tukey's HSD (*Honestly Significant Difference*), Bonferroni, Scheffe, entre otros. Usualmente *Tukey's honestly significant difference* (HSD) es recomendado por los estadísticos pues representa un punto medio entre los métodos más conservadores (Bonferroni) y los más laxos (Fisher). Esto se hace usando el siguiente algoritmo:

1. Calcular el análisis de varianza, por ejemplo, usando MRM-ANOVA.
2. Seleccionar dos medias.
3. Calcular el HSD para cada comparación de medias.
4. Verificar si los resultados de HSD son significativos usando la tabla de probabilidad/valores críticos de Tukey.

### 5.3. Direccionalidad

Veremos como la direccionalidad afecta al tiempo y a la cantidad de pausas en el proceso de traducción.

### 5.3.1. Tiempo

#### Tiempo Total

Se desea averiguar que relación existe entre la VI direccionalidad y la VD tiempo total de sesión.

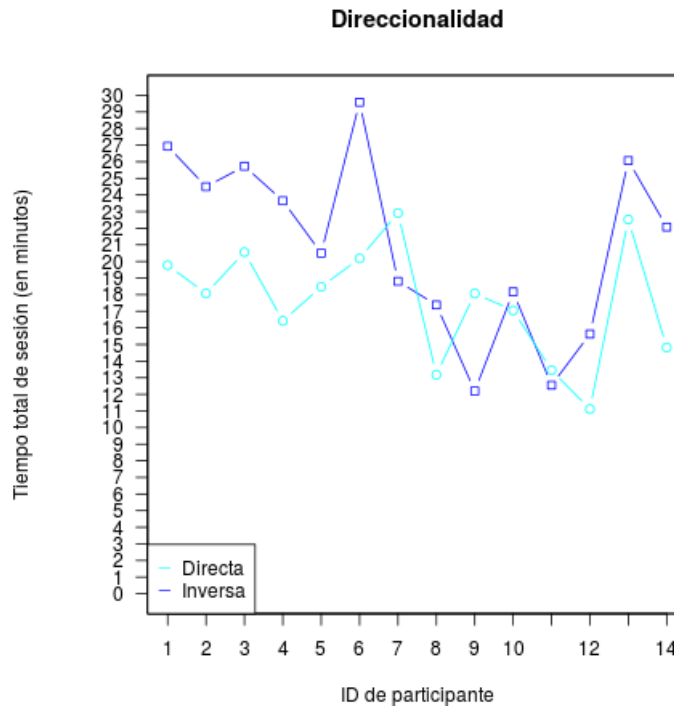


Figura 5.1: Tiempo de sesión total (en minutos) - Directa vs. Inversa.

Considerando a los participantes como un grupo homogéneo, el 79% de ellos tarda menos tiempo en traducir en dirección directa. Los sujetos 7, 9 y 11, todos estudiantes de 5° año, son los únicos que traducen más rápido en dirección inversa. El sujeto 7 tradujo en dirección inversa como segunda condición inmediatamente después de la condición de *Baseline*. Los tres realizan la condición de traducción inversa como segunda o tercera condición en la sesión. Sólo 7 y 11 traducen primero inversa y luego directa. Para el sujeto 10 la diferencia en el tiempo de traducción en ambas direcciones es irrelevante. Podría imaginarse que mientras más avanzado sea el traductor, la dirección en la traducción deja de ser un factor que añade dificultad o carga cognitiva extra.

Usando MRM-ANOVA y suponiendo que la muestra es homogénea se comprobó que hay un efecto significativo de la VI direccionalidad en la VD tiempo total de sesión al nivel  $p < 0.05$  [ $F(1, 11) = 15.925, p = 0.00212$ ].

Al considerar la muestra como heterogénea y usando nuevamente *repeated measures* MRM-ANOVA, se obtienen resultados significativos con [ $F(2, 11) = 8.4792, p = 0.00591$ ]. Sin embargo, HSD solo distingue que para el sub-grupo nivel 3 el tiempo de la traducción directa es menor que para la traducción inversa [ $MSE = 4682E7, df = 15.61, p = 0.004$ ]. Además distingue que en dirección inversa, el sub-grupo nivel 5 tarda menos tiempo que el nivel 3.

### Tiempo Por Fase

Se desea averiguar que relación existe entre la VI direccionalidad y las VD tiempo de fases de la traducción: orientación, elaboración de bosquejo y revisión.

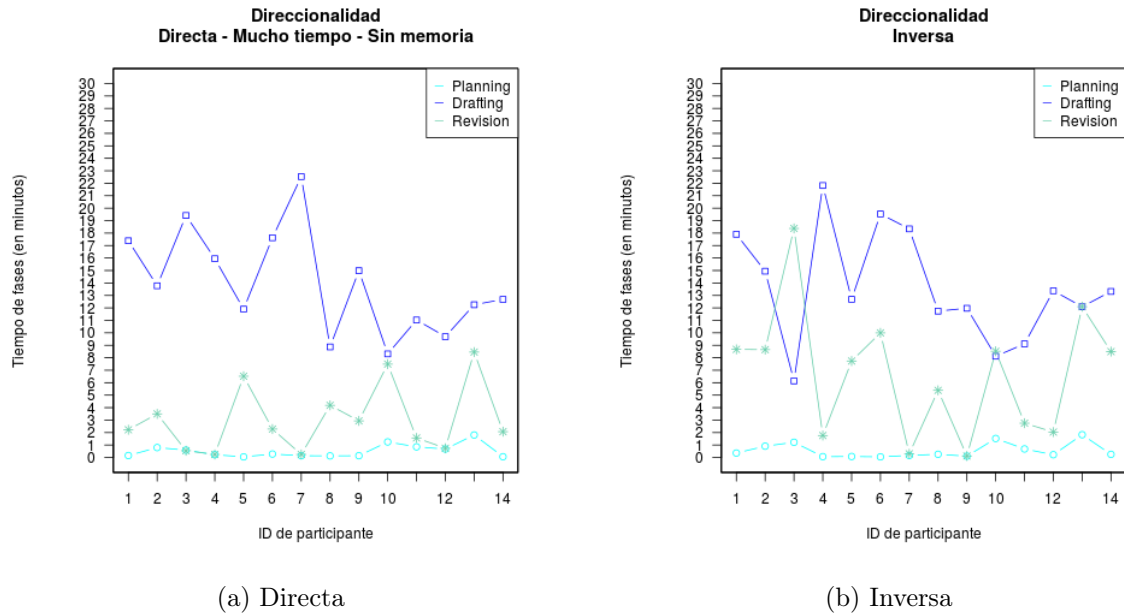


Figura 5.2: Tiempo de fases (en minutos) - Directa vs. Inversa

Al traducir en dirección directa, la mayor parte del tiempo de la sesión se dedica a la fase de elaboración de borrador. En general lo mismo ocurre para la dirección inversa, sin embargo, la excepción es el sujeto 3 quien dedica más tiempo a la fase de revisión. En el caso particular del sujeto 10, dedica casi el mismo tiempo para las fases de elaboración de borrador y revisión en ambas direcciones.

| DI         | Media     | Error estándar | N  |
|------------|-----------|----------------|----|
| D Revisión | 191833,99 | 48125,79       | 14 |
| I Revisión | 403157,08 | 79082,99       | 14 |

Cuadro 5.1: Revisión directa e inversa [F(1, 11) = 8.3871, p=0.01455]

Usando *repeated measures* MRM-ANOVA, no se detecta ningún resultado significativo para las fases de orientación y elaboración de un borrador. Por otro lado, para la fase de revisión en la muestra homogénea, se obtiene que el tiempo de revisión en la traducción directa es menor que en la inversa [F(1, 11) = 8.3871, p=0.01455]. Esto podría indicar que los participantes se sienten un poco más inseguros en la traducción inversa.

Analizando la figura observamos que en ambas direcciones se dedica entre 0 y 2 minutos a la fase de orientación. También verificamos el resultado obtenido con ANOVA con respecto al aumento de tiempo dedicado a la revisión para la dirección inversa.



### 5.3.2. Pausas

#### Pausas Total

Se desea ver si la direccionalidad afecta la cantidad de pausas durante la sesión según la clasificación usada por Koehn (2010).

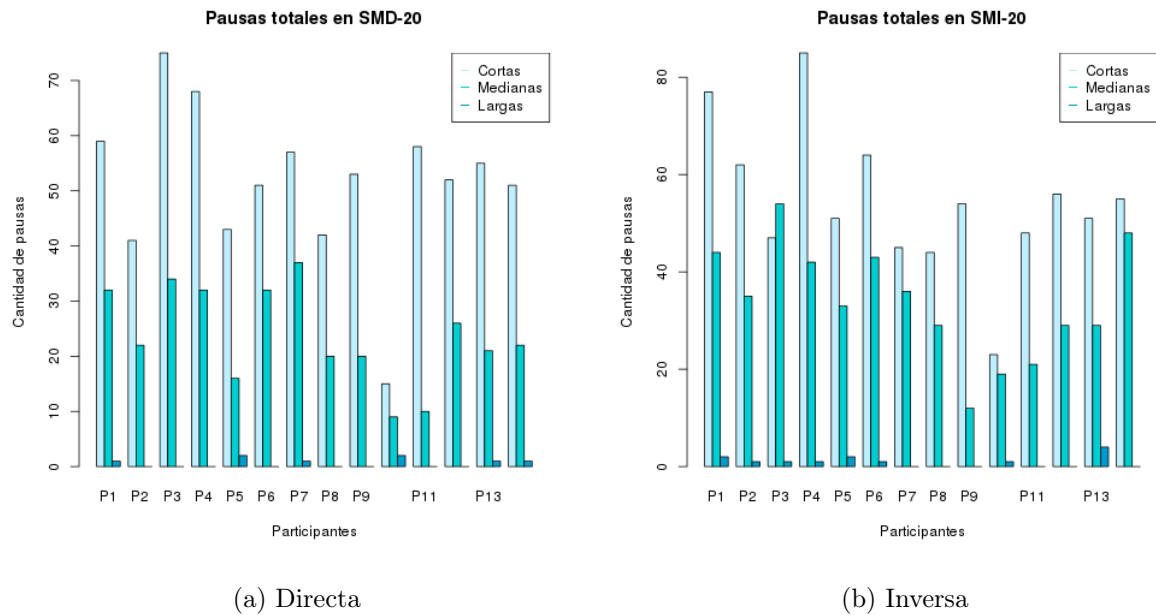


Figura 5.3: Pausas en la sesión - Direccionalidad

A lo largo de la sesión, considerando una muestra homogénea, para un 71% de la muestra la cantidad de pausas cortas es menor en la traducción directa. No obstante, no se encontraron resultados significativos con ANOVA.

Las pausas medianas son menores en la traducción directa en un 93% de los sujetos; ANOVA detecta resultados significativos con  $[F(1,13) = 21.043, p = 0.00051]$  y HSD solo distingue que en los alumnos de tercer año se observan más pausas que en los de quinto año al traducir en dirección inversa  $[MSE = 120.24, df = 11, p = 0.047]$ . El hecho de que haya más pausas medianas en la traducción inversa indica que, en este tipo de traducción, los participantes tienen que planear más cuales serán sus próximas acciones.

#### Pausas Por Fase

En la fase de elaboración de borrador, en dirección inversa la cantidad de pausas largas es menor que en la dirección directa. En un 79% de los participantes, la cantidad de pausas medianas es menor en la fase de elaboración de borrador para la dirección directa.

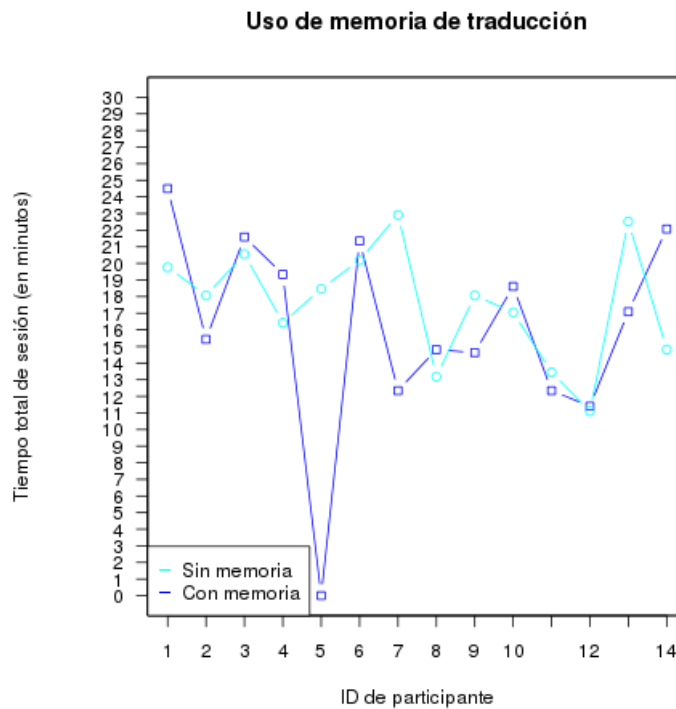
En la fase de revisión, en un 86% de los participantes la cantidad de pausas cortas, medianas y largas es menor en la dirección directa. Indicando que efectivamente en cada una de las fases dudan o se encuentran con más problemas en dirección inversa.

## 5.4. Influencia de memoria de traducción

### 5.4.1. Tiempo

#### Tiempo Total

Se desea ver que relación existe entre la VI uso de memoria y la VD tiempo total de sesión.



**Nota:** No hay información disponible sobre el participante 5 para la condición con memoria.

Figura 5.4: Tiempo de sesión total (en minutos) - Sin memoria vs. Con memoria

Los resultados del análisis con MRM-ANOVA no resultan significativos. Más aún, la mitad de los participantes traducen más rápido con memoria y la otra mitad más rápido sin memoria. Esto podría deberse a que el efecto que el uso de memoria tiene en la traducción es muy dependiente de como la persona aproveche las herramientas facilitadas.

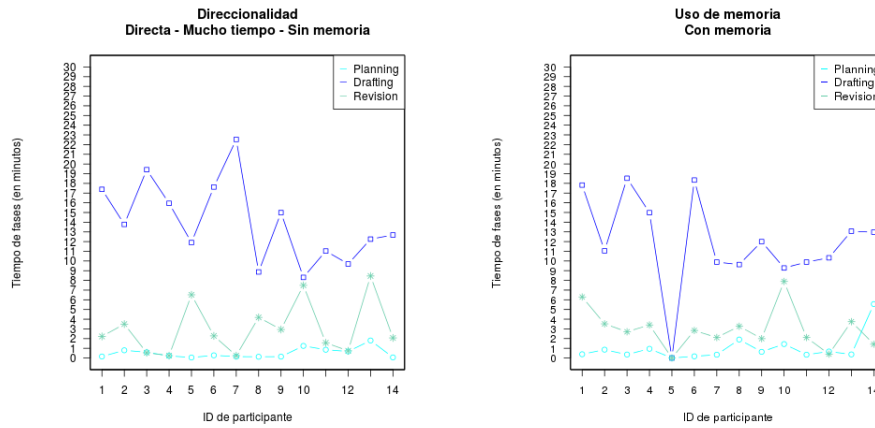
#### Tiempo Por Fase

| Level | Media     | Error estándar | N |
|-------|-----------|----------------|---|
| 3     | 989250,1  | 71918918,88    | 5 |
| 4     | 710016,33 | 92846,88       | 3 |
| 5     | 698488,1  | 71918,88       | 5 |

Cuadro 5.2: Elaboración de borrador y uso de memoria [F(2, 10) = 4.8857, p = 0.03310]

Para la fase de elaboración de borrador, MRM-ANOVA indica que existen resultados significativos  $[F(2, 10) = 4.8857, p = 0.03310]$  y  $\text{partial } \eta^2=0,49$ . El posterior análisis con HSD indica que el sub-grupo nivel 3 es más lento que el 5  $[MSE = 5172E7, df = 10,000, p=0,041]$ .

Notemos que, en general, el tiempo dedicado a la fase de orientación no se ve afectado por el uso o no de memoria de traducción.



(a) Sin memoria

(b) Con memoria

**Nota:** No hay información disponible sobre el participante 5 para la condición con memoria.

Figura 5.5: Tiempo de fases (en minutos) - Sin Memoria vs. Con memoria

### 5.4.2. Pausas

#### Pausas Total

En la condición con memoria de traducción se presentó la mayor cantidad de pausas largas, incluso más que en la condición de traducción inversa. El 64 % de los participantes tuvo al menos una pausa larga al usar memoria mientras que el 43 % de los participantes tuvo al menos una pausa larga al no usar memoria. En efecto, ANOVA detecta que hay resultados significativos para la muestra homogénea en cuanto a la mayor cantidad de pausas largas al usar memoria de traducción  $[F(1,11) = 12.571, p = 0.00459]$

Esto podría corresponderse con el trabajo de Pym (2011) quien sugiere que al tener mayor opciones, el tiempo de elección de la opción adecuada aumenta y por lo tanto la cantidad de pausas. Sería interesante determinar los tiempos en los que ocurren las pausas largas para determinar que ocurría en el registro de comportamiento de usuario aunque solo un *eye-tracker* podría determinar con exactitud cual era el punto de atención durante las pausas.

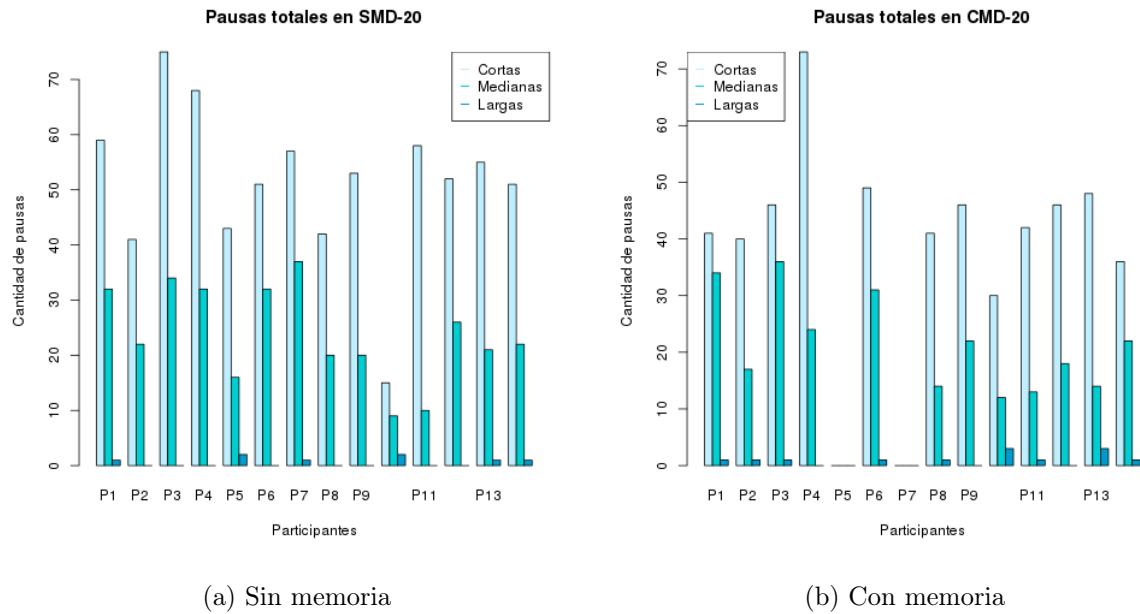


Figura 5.6: Pausas en la sesión - Uso de memoria de traducción

### Pausas Por Fase

Al traducir con memoria, en un 57% de los participantes se notó el uso de una estrategia de postergación de problemas. Eso se intuye ya que en esta condición prácticamente no hay pausas largas durante la fase de elaboración de borrador de modo que las pausas largas aparecen recién en la fase de revisión.

## 5.5. Presión de tiempo

Con este análisis se pretende determinar si bajo presión de tiempo los participantes son capaces de reducir el tiempo dedicado para cada una de las fases. Esto, por ejemplo, no ocurre en el caso en que un participante no tenga tiempo suficiente para terminar de traducir. En un caso como el mencionado, podría obviarse totalmente la fase de revisión y queremos determinar si esto ocurre con los participantes.

Además, si hiciéramos un análisis sobre la proporción de tiempo que se dedica a cada fase, podríamos determinar si aún con presión de tiempo se dedica el mismo porcentaje a cada fase en ambas condiciones.

### 5.5.1. Tiempo

#### Tiempo Total

Como era de esperarse, todos los participantes (salvo uno) tradujeron la condición de presión de tiempo más rápido que al no tener presión. Si bien este resultado parece trivial, es importante notarlo pues estos experimentos sirven para probar la herramienta por lo que es necesario que

esta intuición se cumpla al analizar los datos. El análisis con MRM-ANOVA confirma lo anterior [ $F(1, 11)=37.53, p=0.00007$ ] y  $partial \eta^2 = 0,77$ . Para el caso del sujeto 12, donde esto no ocurre, se intuye que el límite de tiempo propuesto como presión de tiempo no representa un desafío en cuanto a su velocidad de traducción normal.

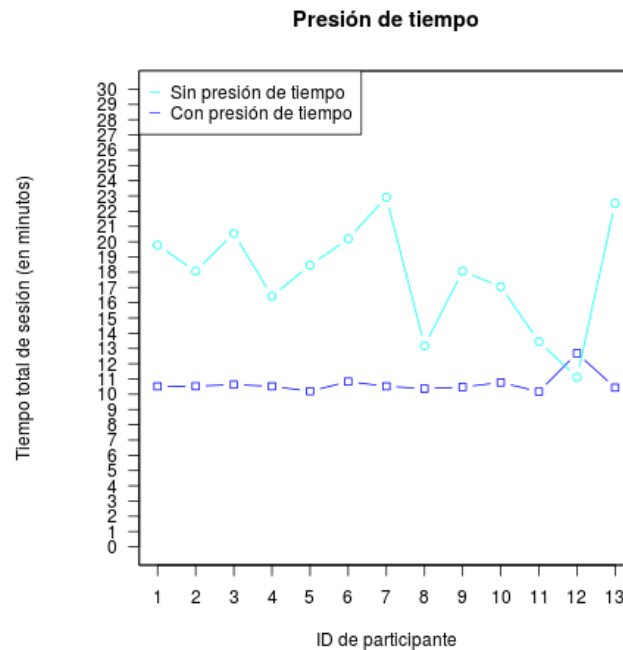


Figura 5.7: Tiempo de sesión total (en minutos) - Mucho tiempo vs. poco tiempo

## Tiempo Por Fase

Resulta interesante ver que, bajo presión de tiempo, el sub-grupo nivel 3 distribuye el tiempo de todas las fases de manera muy similar. Esto no se observa para los otros sub-grupos. Otra peculiaridad del sub-grupo nivel 3 es que dedican casi el mismo tiempo para las fases de orientación y revisión bajo presión de tiempo.

El análisis con MRM-ANOVA para cada una de las fases resulta significativo. Para la fase de orientación: [ $F(1, 11)=5.09, p=0.05$ ] y  $partial \eta^2 = 0,32$ . Para la fase de elaboración de borrador: [ $F(1, 11)=20.184, p=0.00091$ ] y  $partial \eta^2 = 0,65$ . Finalmente, para la fase de revisión: [ $F(1, 11)=6.70, p=0.03$ ] y  $partial \eta^2 = 0,38$ .

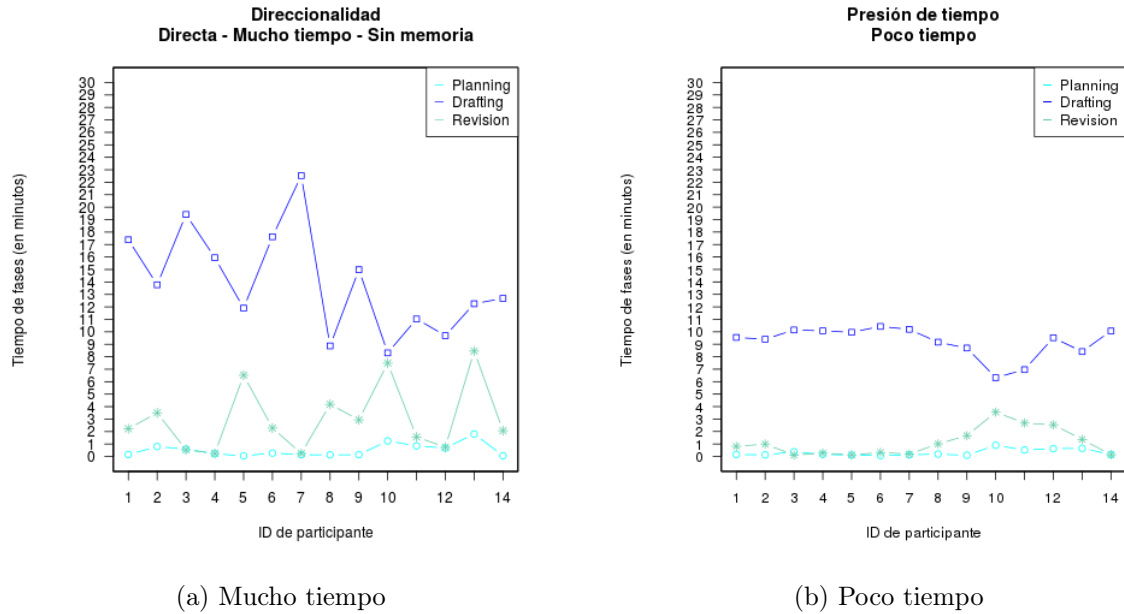


Figura 5.8: Tiempo de fases (en minutos) - Mucho tiempo vs. poco tiempo

### 5.5.2. Pausas

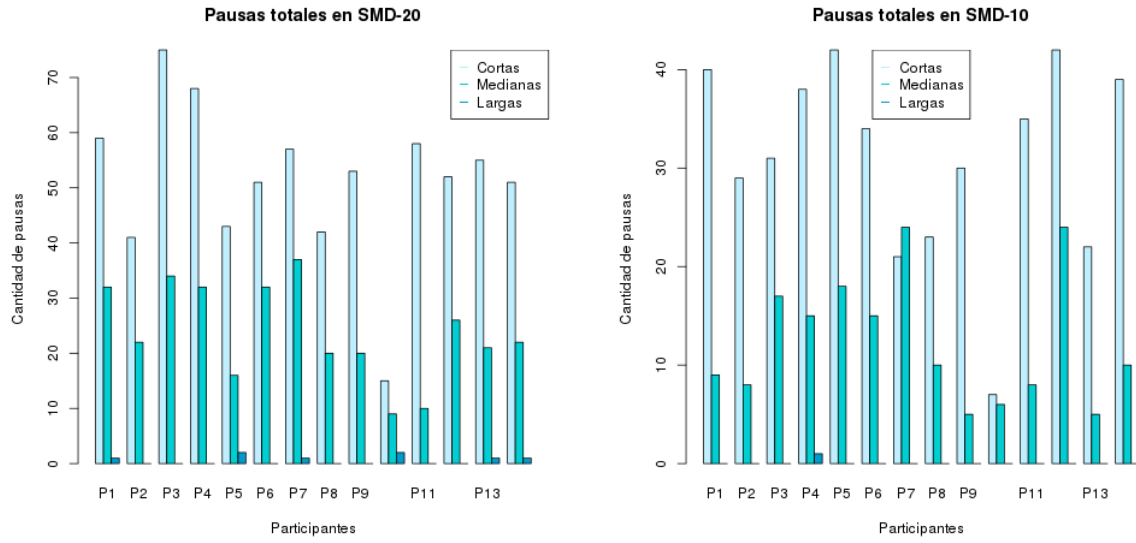
#### Pausas Total

Observemos que para el caso de traducción directa, el número de pausas largas en la sesión total fue superior a 0 en 6 de los participantes. En cambio, bajo presión de tiempo uno solo de los participantes tuvo una pausa larga. Esto podría indicar, que bajo presión de tiempo los problemas que pudieran llegar a surgir son postergados y hasta ignorados.

El análisis con ANOVA indica que hay significativamente menos pausas cortas en la traducción bajo presión de tiempo; el análisis con HSD indica que hay menos pausas cortas bajo presión de tiempo en los estudiantes de tercer año [MSE = 138.47, df = 19.005, p = 0.233] y en los estudiantes de quinto año [MSE = 138.47, df = 19.005, p = 0.28] pero no en los de cuarto.

En cuanto a pausas medianas, ANOVA también detecta que hay menos pausas de este tipo bajo presión de tiempo [F(1, 13) = 30.959, p = 0.00009] para la muestra homogénea. Para la muestra heterogénea, ANOVA detecta que hay diferencias significativas [F(2,11) = 0.44794, p = 0.6501]; HSD distingue a los estudiantes de tercero con los de quinto año [MSE = 59.129, df = 17.313, p = 0.032].

ANOVA detecta resultados significativos para las pausas largas en la muestra homogénea [F(1, 13) = 4.7895, p = 0.04749] pero no para la muestra heterogénea.



(a) Mucho tiempo

(b) Poco tiempo

Figura 5.9: Pausas en la sesión - Presión de tiempo

### Pausas Por Fase

La cantidad de pausas cortas, medianas y largas en la fase de revisión bajo presión de tiempo son mucho menores en comparación de las obtenidas para la condición sin presión de tiempo. En un 50% de los participantes la cantidad de pausas es nula en la fase de revisión. En cuanto a pausas largas, un 93% de los participantes no tiene pausas de este tipo ni durante la elaboración de borrador ni al revisar.

## Capítulo 6

# Discusión General

### 6.1. Fortalezas y limitaciones

El uso de ResearchLogger para la toma de datos se considera como una de las fortalezas más grandes de la investigación ya que presenta grandes ventajas a comparación de Translog en cuanto a la calidad de los datos recolectados. Un análisis detallado de la comparación entre sus competidores puede hallarse en: “Comparación con otras herramientas”

La combinación de datos de *eyetracking* y *keylogging* provee una visión más completa de los aspectos subyacentes a los procesos cognitivos en la traducción. No obstante, la Universidad Nacional de Córdoba todavía no posee un *eyetracker* propio y resulta imposible obtener uno debido al presupuesto limitado. Por este motivo, solo se trabaja con la técnica de *keylogging* de manera que se restringe el estudio solo al procesamiento del texto meta y resulta imposible saber cual es el foco de atención del traductor durante las pausas.

La cantidad de sujetos que participaron en las sesiones de traducción no fue lo suficientemente grande como para obtener inferencias estadísticas lo suficientemente robustas; por lo tanto, se sugiere tomar los datos como exploratorios debido al bajo poder estadístico.

El riguroso criterio para la selección de los textos también se considera una gran fortaleza de esta investigación ya que son muy pocos los estudios que han elegido los textos tan selectivamente. Más aún, el uso de los cuestionarios auto-administrados nos da información detallada sobre los perfiles de los participantes lo que representa una gran ventaja a la hora de elegir los sub-grupos o tratar de explicar las conductas más inusuales.

Finalmente, tal y como propone Hvelplund (2011), las condiciones se balancearon aleatoriamente de manera que los hallazgos no sean afectados negativamente por una presentación uniforme de las condiciones. El orden de las condiciones se detalla en la sub-sección “Balanceo de condiciones”.



## 6.2. Trabajo futuro

Para este trabajo final, se desarrollaron dos herramientas: una de toma de datos y otra para el análisis de los datos. En esta sección se detallará el trabajo futuro a desarrollarse para estas dos herramientas y para este proyecto en general.

### Mejoras a ResearchLogger (herramienta de toma de datos)

ResearchLogger posee una *Text User Interface* (TUI) que no resulta intuitiva para todo tipo de usuario. Con el fin de beneficiar a la mayor cantidad de investigadores, en futuro cercano, se pretende añadir una *Graphical User Interface* (GUI) de manera que el producto sea atractivo para todo tipo de investigadores.

Adicionalmente, se desea hacer pruebas exigentes sobre la precisión en milisegundos de la herramienta y sus competidores. Se ha comprobado que ResearchLogger no pierde información, sin embargo, todavía se desconoce la diferencia de precisión de milisegundos entre las herramientas. Futuras pruebas responderán este interrogante.

### Mejoras a ResearchAnalyzer (herramienta de análisis)

ResearchAnalyzer es actualmente una API que surgió de combinar todos los *scripts* de análisis generados para esta investigación. Al igual que en el caso de ResearchLogger, se pretende añadir una GUI que provea de manera intuitiva las mismas opciones proveídas por la API actual.

Además, se desea extender las opciones de análisis de datos de manera que pueda realizarse lo siguiente:

- *Replay* de sesión: Una repetición de la sesión de traducción permitiría que los investigadores puedan ver gráficamente como se construyó el texto traducido. Esta característica del programa todavía está en desarrollo.
- Generación automática de gráficos: Los gráficos usados a lo largo de este trabajo se hicieron usando *scripts* en R. Sería deseable que la herramienta de análisis sea capaz de generar determinados gráficos automáticamente, por ejemplo, un gráfico de puntos y líneas que muestre la relación entre *keystrokes* y pausas para un determinado participante.

### Otros hilos de investigación

En primer lugar, es de gran interés repetir la presente investigación para un mayor número de sujetos con el fin de obtener inferencias estadísticas más robustas. Otro próximo objetivo consiste en estudiar otros tipos de procesos de escritura. Por ejemplo, una de las principales áreas de interés para futuras investigaciones está relacionado con las enfermedades neurodegenerativas tales como el Parkinson, mediante el uso de las herramientas ResearchLogger y ResearchAnalyzer. El objetivo sería encontrar diferencias en el comportamiento de escritura entre pacientes sanos y aquellos que padecen la enfermedad.

Otro hilo de interés tiene que ver con estudiar el impacto que tiene la post-edición en el producto final, que estrategias de traducción se emplean, la predisposición o aceptación que tienen los traductores en formación en cuanto al rol de post-editores, entre otras.

Finalmente, con los principales resultados de esta investigación y potenciales investigaciones a futuro, se pueden mejorar los sistemas de traducción asistida. Esto es posible, ya que al investigar cuales son aquellos aspectos que más ralentizan o complejizan el proceso de traducción, es posible decidir que opciones ofrecerían mayor ayuda a los traductores.

## Capítulo 7

# Conclusiones

El objetivo general de este estudio es investigar los aspectos cognitivos en los estudiantes de traductorado de la Universidad Nacional de Córdoba. Con este fin se extendió una herramienta de código abierto llamada PyKeylogger dio lugar a la herramienta que nombramos ResearchLogger, además, se desarrolló desde cero un analizador automático de ciertas características de los registros generados por ResearchLogger.

A lo largo del desarrollo de la herramienta de toma de datos, se detectó una serie de importantes desventajas en el software que es usado actualmente por la comunidad científica. Las características ofrecidas por ResearchLogger hacen que sea una potencial gran herramienta para la toma de datos en investigaciones relacionadas con los procesos de escritura. Especialmente, la portabilidad del sistema y el hecho de que sea de código abierto hacen que esta herramienta una poderosa alternativa a sus competidores.

La meta de esta primer versión de ambos sistemas era verificar que los hallazgos obtenidos usando las herramientas mencionadas son acordes a los obtenidos en la literatura. Así surgió el diseño de un estudio interdisciplinario en el cual se juntan las áreas de traductología, traducción, cognición y ciencias de la computación. Esto es una prueba fehaciente de que las ciencias de la computación colaboran enormemente al desarrollo de otras áreas.

A continuación, se resumirá brevemente los resultados hallados:

**H1:** La traducción directa se realiza con mayor rapidez que la traducción inversa en textos de complejidad comparable. En la muestra observada, esta hipótesis se acepta para el sub-grupo de nivel 3. Para el resto de los participantes, no se obtuvo resultados estadísticamente significativos.

**H2:** La traducción usando memoria de traducción es más lenta que cuando la misma no se usa. La mitad de los participantes traduce más rápido al usar asistencia en la traducción. Por otro lado, ANOVA no detecta diferencias significativas de manera que no pudimos comprar esta hipótesis para la muestra dada.

**H3:** El uso de memoria de traducción complica la toma de decisiones lo que se ve reflejado en un aumento en las pausas para la condición con memoria. Un análisis estadístico descriptivo indica que la cantidad de pausas en la condición con memoria es mayor que en la traducción directa. Se

intuye que la hipótesis puede aceptarse para la muestra, sin embargo, para sostener tal afirmación es necesario realizar un análisis estadístico inferencial.

**H4:** Bajo presión de tiempo, los participantes reducen el tiempo dedicado a las fases de orientación y revisión. El análisis con ANOVA confirma que el tiempo dedicado a cada fase es menor al traducir con presión de tiempo. Si bien esto parece trivial, si un traductor dedicara la misma cantidad de tiempo a la fase de orientación independientemente de la presión de tiempo. Y quisiera repetir esto para la fase de elaboración de borrador, entonces, se quedaría sin tiempo para la fase de revisión. Pero esto demuestra que ante presión de tiempo los participantes reducen el tiempo dedicado a cada fase. A partir de estos resultados, sería conveniente determinar si la proporción de tiempo dedicada a cada fase es la misma en ambas condiciones.



# Bibliografía

- A. M. García, E. Mikulan, A. I. *A neuroscientific toolkit for cognitive translology. En R. Muñoz (ed.), Reembedding Translation Process Research. Ámsterdam: John Benjamins. (Aceptado).*
- Adolfo M. García, Agustín Ibáñez, D. H. A. L. H. M. M. C. G. L. S. C. A. R.-R. (2014). *Word reading and translation in bilinguals: the impact of formal and informal translation expertise.*
- Albir, A. H. (2001). *Traducción y traductología.* Cátedra.
- Alves, F. and Campos, T. L. (2009). Translation technology in time: investigating the impact of translation memory systems and time pressure on types of internal and external support. *Copenhagen Studies in Language*, pages 191–218.
- Augusto Buchweitz, F. A. (2006). Cognitive adaptation in translation: an interface between language direction, time, and recursiveness in target text production.
- Buenos Aires Herald. <http://www.buenosairesherald.com/>. Último acceso: 30-07-2015.
- Caitra. <http://tool.statmt.org/>. Último acceso: 30-05-2015.
- Carl, M. (2012). *Translog-II: a Program for Recording User Activity Data for Empirical Reading and Writing Research.*
- Christensen, T. P. (2011a). *Studies on the Mental Processes in Translation Memory-assisted Translation – the State of the Art.*
- Christensen, T. P. (2011b). Studies on the mental processes in translation memory-assisted translation—the state of the art. *trans-kom. Zeitschrift für Translationswissenschaft und Fachkommunikation*, 4(2):137–160.
- Gabriela Saldanha, S. O. (2013). *Research Methodologies in Translation Studies.*
- Göpferich, S. (2008). *Translationsprozessforschung: Stand - Methoden - Perspektiven.* Translations-Wissenschaft. Narr.
- Hansen, G. (2006). *Time pressure in translation teaching and translation studies.*
- Hutchins, E. (2000). Distributed cognition. *International Encyclopedia of the Social and Behavioral Sciences. Elsevier Science.*

- Hvelplund, K. (2011). *Allocation of cognitive resources in translation an eye-tracking and key-logging study*. PhD thesis - Department of International Language Studies and Computational Linguistics Copenhagen Business School.
- InputLog. <http://www.inputlog.net/>. Último acceso: 31-05-2015.
- Jensen, K. (2009). Indicators of text complexity. *Copenhagen Studies in Language*, (37):61–80.
- Koehn, P. (2010). *A process study of computer-aided translation*.
- Li, P., Z. F. T. E. P. B. (2013). *Language history questionnaire (LHQ 2.0): A new dynamic web-based research tool*. *Bilingualism: Language and Cognition*, 17(3), 673-680. DOI: 10.1017/S1366728913000606.
- Linguee. <http://www.linguee.es/>.
- Liparini Campos, T. (2005). O efeito da pressão de tempo na realização de tarefas de tradução: uma análise processual sobre o desempenho de tradutores em formação. *Unpublished MA Thesis*. Federal University of Minas Gerais, Brazil.
- M. Denkowski, A. L. (2012). *TransCenter: Web-Based Translation Research Suite*. AMTA 2012 Workshop on Post-Editing Technology and Practice Demo Session.
- M. Leijten, L. V. W. (2005). *Inputlog: A logging tool for the research of writing processes*.
- María L. Sanz de Acedo Lizárraga, María T. Sanz de Acedo Baquedano, M. C.-E. (2007). Factors that affect decision making: gender and age differences. *International Journal of Psychology and Psychological Therapy*.
- Mellinger, C. D. (2014). *Average Pause Ratio as an Indicator of Cognitive Effort in Post-Editing: A Case Study*.
- Michael Carl, Arnt Lykke Jakobsen, K. T. H. J. (2008). *Studying Human Translation Behavior with User-Activity Data*.
- Nataša Pavlovic, K. T. H. J. (2009). *Eye tracking translation directionality*.
- Newmark, P. and Hall, P. (1988). *A textbook of translation*, volume 1. Prentice Hall New York.
- Nielsen, J. (2004). *Capturing Thoughts, Capturing Minds? - from Think Aloud to Participatory Analysis*.
- OmegaT. <http://www.omegat.org/>. Último acceso: 30-05-2015.
- Pavlovic, N. (2007). *Directionality in collaborative translation processes*.
- Pavlović, T. (2013). *Exploring directionality in translation studies*.
- PyKeylogger. <http://pykeylogger.sourceforge.net/>. Accessed: 10-05-2015.

- Pym, A. (2011). What technology does to translating. *Translation & Interpreting Vol 3, No 1*.
- ResearchAnalyzer. <https://github.com/roxana-lafuente/ResearchAnalyzer>. Último acceso: 13-07-2015.
- ResearchLogger. <https://github.com/roxana-lafuente/ResearchLogger>. Último acceso: 13-07-2015.
- Roberts, R. P. (1988). Towards a typology of translations. *studies*, 31:35.
- Schubert, K. (2009). Positioning translation in technical communication studies. *Journal of Specialised Translation*, 11:17–30.
- SDL Trados. <http://www.sdl.com/>. Último acceso: 04-06-2015.
- Sharmin, S., Špakov, O., Rähä, K.-J., and Jakobsen, A. L. (2008). Effects of time pressure and text complexity on translators' fixations. In *Proceedings of the 2008 symposium on Eye tracking research & applications*, pages 123–126. ACM.
- Toury, G. (1982). A rationale for descriptive translation studies. *Dispositio*, pages 23–39.
- Translog. <http://bridge.cbs.dk/platform/?q=Translog-II>. Último acceso: 01-02-2015.
- Writing PRO: Logging programs. [http://www.writingpro.eu/logging\\_programs.php](http://www.writingpro.eu/logging_programs.php). Último acceso: 12-07-2015.



# Anexos

## Anexo A

# Textos usados

### **Venezuela takes first steps toward dialogue**

Leaders of Venezuela’s opposition agreed to meet with President Nicolás Maduro’s government to discuss ways to defuse the country’s ongoing political crisis. Opposition spokesmen have openly agreed to a preliminary meeting. A mission of Unasur diplomats assured them that the Socialist government is open to discussing a four-point agenda it has put forth as a starting point for talks. Their proposals include an amnesty for imprisoned government opponents, the creation of an independent commission to determine who is to blame for violence, and the disarmament of groups loyal to the President. It is unclear whether the meeting— which was proposed to take place later yesterday —will usher in formal talks aimed at reconciliation. Still, just the opposition’s willingness to resume dialogue with the government represents a potential breakthrough. This gesture could help ease tensions on the streets, especially after two months of deadly protests that have rocked Venezuela but so far pose little threat to Maduro’s mandate.

### **FIFA medical chief ‘hates’ drug test logistics**

FIFA’s medical chief is “definitely not happy” with drug-testing plans for the World Cup in Brazil. Samples obtained from players must be flown across the Atlantic for analysis at a laboratory in Switzerland, which could significantly slow down results. FIFA had to turn to the Swiss specialists because the Brazilian laboratory that was expected to analyze the tournament’s samples repeatedly failed to comply with World Anti-Doping Agency standards. FIFA will send approximately 1,000 blood and urine samples to the accredited facility in Lausanne. This represents an additional cost of US\$250,000 for the institution, declared Michel D’Hooghe, who chairs FIFA’s medical commission and presides over its executive committee. During an interview with an international news agency, D’Hooghe maintained that at previous World Cups “we always succeeded at having the (anti-doping test) results before the following game of each team. So if your results were positive or if mine were positive we knew it before we played a second game.”

## **García Márquez left behind unpublished manuscript**

Nobel-prize laureate novelist Gabriel García Márquez left behind an unpublished manuscript which he decided not to print while alive, an editor told reporters yesterday as the writer's admirers offered a musical tribute to him in Colombia.

Cristóbal Pera, editorial director of Penguin Random House Mexico, said that García Márquez's relatives have not yet decided whether to allow the volume to be released posthumously.

Several publishing companies have expressed their interest in obtaining the publication rights.

García Márquez died at his Mexico City home on April 17.

The manuscript's working title is We'll See Each Other in August (En agosto nos vemos).

An excerpt of the manuscript published in Spain's La Vanguardia newspaper contains what appears to be a first chapter, describing a trip taken by a 50-ish married woman who visits her mother's grave on a tropical island every year.

In this introductory chapter, she has a love affair with an elderly gentleman at the hotel where she stays.

## **La FIFA inicia procedimiento disciplinario contra México por racismo de aficionados**

El miércoles pasado la FIFA inició un procedimiento disciplinario contra México. Las acciones comenzaron luego de registrarse presuntas expresiones de racismo de sus aficionados durante el partido frente a Camerún en la ciudad brasileña de Natal. México y Camerún se enfrentaron el 13 de junio en las eliminatorias del Mundial de Fútbol. En un contundente comunicado, la FIFA explicó que la investigación se inició después de que se produjeran lo que calificó de "conductas impropias" por parte de los aficionados mexicanos. Además, fuentes de la FIFA informaron que están investigando informes de cánticos homofóbicos de parte de simpatizantes del anfitrión, Brasil. También ha habido reclamos porque los seguidores de Croacia y Rusia exhibieron pancartas con contenidos antisemitas y racistas. Desde Río de Janeiro, el corresponsal deportivo de la BBC, Richard Conway, informó que los problemas de racismo y discriminación en el fútbol continúan, incluso en Brasil, en el transcurso del emocionante Mundial de Fútbol.

## **IOC tightens control over Rio games**

The International Olympic Committee announced a series of measures including a stronger presence in Rio de Janeiro to monitor progress and speed up lagging preparations for the 2016 Olympics, president Thomas Bach said yesterday. Preparations for the 2016 Games, the first on the South American continent, have been plagued by delays and rising costs. Communication was ineffective between different levels of the Brazilian government and organizers. Throughout this week, international federations also criticized the Brazilian government for what they said were constant delays and problems in communicating efficiently with the Games organizing committee. “We had the meeting (yesterday) in a very constructive atmosphere with our partners from Rio,” Bach told a news conference in Turkey at the end of a two-day executive board meeting. “We then took some decisions about how we can accelerate one way or another the works in Rio. We now need to work even closer with the organizing committee and the different levels of government.”

## Selección de los textos

Tres traductores profesionales con (L1: Español; L2: Inglés) revisaron los textos seleccionados y asignaron a cada uno de ellos un puntaje según el siguiente criterio: (1) Muy pobre (2) Pobre (3) Aceptable (4) Bueno (5) Muy bueno.

| <b>Medida / Textos</b> | <b>VENEZUELA</b> | <b>FIFA</b> | <b>GARCÍA<br/>MARQUÉZ</b> | <b>RACISMO</b> | <b>IOC</b> |
|------------------------|------------------|-------------|---------------------------|----------------|------------|
| Comprensibilidad       | 3.66             | 4           | 4                         | 4.66           | 4.33       |
| Coherencia             | 3.66             | 4.33        | 3.66                      | 4.33           | 4.33       |
| Gramaticalidad         | 3.33             | 4           | 3.66                      | 4              | 4          |

Cuadro A.1: Puntajes de comprensibilidad, coherencia y gramaticalidad de los textos

| Parámetro  | VENEZUELA | FIFA  | GARCÍA MARQUÉZ | RACISMO | IOC   |
|--|-----------|-------|----------------|---------|-------|
| LIX Fórmula  | 53        | 53    | 53             | -       | 54    |
| <i>Flesch Reading Ease score / Szigriszt-Pazos</i> | 43,4      | 43,4  | 43,4           | 43,8    | 43,4  |
| <i>Automated Readability Index</i>                 | 14,3      | 14,2  | 13,3           | -       | 14,3  |
| <i>Flesch-Kincaid Index</i>                        | 12,3      | 12,3  | 12,3           | -       | 12,3  |
| <i>Coleman-Liau Index</i>                          | 14,6      | 14,4  | 13,3           | -       | 14,5  |
| <i>Gunning fog Index</i>                           | 17,9      | 16,5  | 16,3           | -       | 17,7  |
| <i>SMOG Index</i>                                  | 15,2      | 14,3  | 14,1           |         | 15,1  |
| Oraciones  | 8         | 8     | 8              | 8       | 8     |
| Oraciones simples                                  | 2         | 3     | 3              | 3       | 3     |
| Oraciones compuestas                               | 0         | 0     | 0              | 0       | 0     |
| Oraciones complejas / complejas-compuestas         | 6         | 5     | 5              | 5       | 5     |
| Palabras   | 165       | 165   | 165            | 165     | 165   |
| Palabras complejas                                 | 40        | 34    | 33             | -       | 39    |
| Palabras por oración                               | 20,63     | 20,63 | 20,63          | 20,63   | 20,63 |
| Silabas por palabra                                | 1,68      | 1,68  | 1,68           | 2,29    | 1,68  |
| Caracteres   | 1077      | 1081  | 1053           | 1099    | 1074  |
| Caracteres por palabra                             | 5,41      | 5,37  | 5,19           | 6,66    | 5,39  |
| <i>Non-literality</i>                              | 10        | 7     | 6              | 4       | 10    |
| Comprensibilidad                                   | 3.66      | 4     | 4              | 4.66    | 4.33  |
| Coherencia   | 3.66      | 4.33  | 3.66           | 4.33    | 4.33  |
| Correctitud gramatical                             | 3.33      | 4     | 3.66           | 4       | 4     |

Cuadro A.2: Resumen de las variables controladas en los textos

## Anexo B

# Cuestionarios usados

Los cuestionarios usados en estos experimentos fueron auto-administrados y se completaron a través de Internet.

### Conocimientos Informáticos

El cuestionario de conocimientos informáticos consta de quince preguntas donde los participantes deben auto-calificar sus conocimientos en el uso de memorias de traducción, Internet y habilidades mecanográficas.

[https://docs.google.com/forms/d/1fUXXLXDzyx9IRvWVU-5t3m8ZoQ1uwFpuLVdLvt\\_Q1Rk/viewform?c=0&w=1](https://docs.google.com/forms/d/1fUXXLXDzyx9IRvWVU-5t3m8ZoQ1uwFpuLVdLvt_Q1Rk/viewform?c=0&w=1)

### Historial de idiomas

El cuestionario de historial de idiomas consta de 24 preguntas sobre la segunda lengua de los participantes: edad a la que comenzaron a hablarla, si han vivido en el exterior y por cuanto tiempo, auto-evaluación sobre el nivel de la segunda lengua, entre otros.

[https://docs.google.com/forms/d/1b84ghF8RMR0oJ\\_F8r6sHPyGFjB58F9KP27tp\\_xkX20E/viewform?c=0&w=1](https://docs.google.com/forms/d/1b84ghF8RMR0oJ_F8r6sHPyGFjB58F9KP27tp_xkX20E/viewform?c=0&w=1)

### Bilingüismo y traducción

El cuestionario de bilingüismo y traducción está formado por 27 preguntas sobre los datos demográficos de los participantes, certificaciones en la segunda lengua y estimaciones sobre el tiempo dedicado al uso de la segunda lengua.

<https://docs.google.com/forms/d/1FdoS2Qm4wIy15UxaZS4WMLX2Xj4gdhj8uQTA35ZrAGU/viewform?c=0&w=1>

# Anexo C

## Detalles técnicos

### C.1. ResearchLogger

#### Librerías requeridas

PyKeylogger requiere la instalación de ciertas librerías para su uso. Entre ellas se encuentran: python-xlib, ConfigObj, Python Imaging Library (PIL), pyHook, py2exe, PyWin32, pygtk.

#### Modo de uso

El directorio de ResearchLogger contiene un archivo *ResearchLogger.py* que al ser ejecutado inicia el proceso de *keylogging*. Desde el momento en el que se ejecuta, toda acción de *mouse* y teclado será guardada en un directorio especial. Por defecto, el directorio se crea en el mismo directorio donde se encuentra el código de ResearchLogger.

Para terminar la sesión de *keylogging*, es necesario presionar el combo de letras CTRL IZQUIERDO + CTRL DERECHO + F12, esto abrirá una ventana de dialogo en la que se preguntará por una contraseña. Por defecto, la contraseña es el *string* vacío. Esto abrirá una nueva ventana con opciones de configuración de la herramienta. Para cerrar, ir a *Accions > Quit*. La ventana de traducción simple puede accederse desde *ResearchLoggerInterface.py*, su uso es recomendado pero totalmente opcional.

### C.2. ResearchAnalyzer

#### Librerías requeridas

El único requerimiento que tiene ResearchAnalyzer es la librería termcolor que es usada para mostrar resultados con diferentes colores de manera que sean más legibles.



## Modo de uso

Para esta primera versión, será necesario crear un *script* en Python que use los métodos provistos por la API, sin embargo, se pretende desarrollar una interfaz gráfica para que la herramienta también sea útil para usuarios que no sepan programar.

# Anexo D

## Tablas de resultados

### Tiempos

#### Baseline

| <b>BASELINE</b>       | <b>Nivel</b> | <b>Tiempo total</b> | <b>Orientación</b> | <b>Elab. de borrador</b> | <b>Revisión</b> |
|-----------------------|--------------|---------------------|--------------------|--------------------------|-----------------|
| Sujeto 1              | 3            | 462000              | 4448               | 451833                   | 5719            |
| Sujeto 2              | 3            | 338000              | 2265               | 240903                   | 94832           |
| Sujeto 3              | 3            | 474000              | 3958               | 418692                   | 51350           |
| Sujeto 4              | 3            | 572000              | 20018              | 529131                   | 22851           |
| Sujeto 5              | 3            | 386968              | 17461              | 298419                   | 71088           |
| Sujeto 6              | 3            | 642000              | 2730               | 448219                   | 191051          |
| Sujeto 7              | 5            | 500000              | 11806              | 459313                   | 28881           |
| Sujeto 8              | 5            | 659000              | 7563               | 297658                   | 353779          |
| Sujeto 9 <sup>1</sup> | 5            | -                   | -                  | -                        | -               |
| Sujeto 10             | 5            | 323096              | 19575              | 261504                   | 42017           |
| Sujeto 11             | 5            | 404172              | 3716               | 304368                   | 96088           |
| Sujeto 12             | 4            | 352000              | 13594              | 334326                   | 4080            |
| Sujeto 13             | 4            | 404000              | 4552               | 312809                   | 86639           |
| Sujeto 14             | 4            | 732000              | 3554               | 404164                   | 324282          |

Cuadro D.1: Tiempo en milisegundos para las fases en el Baseline.

| <b>BASELINE</b>          | <b>Tiempo total</b> | <b>Orientación</b> | <b>Elab. de borrador</b> | <b>Revisión</b> |
|--------------------------|---------------------|--------------------|--------------------------|-----------------|
| Promedio                 | 480710,46           | 8864,62            | 366256,85                | 105589          |
| Desviación estándar (DE) | 127808,19           | 6467,67            | 86431,79                 | 110410,12       |

Cuadro D.2: Resumen estadístico de los datos recolectados en el Baseline.

**SMD-20**

| <b>SMD-20</b> | <b>Nivel</b> | <b>Tiempo total</b> | <b>Orientación</b> | <b>Elab. de borrador</b> | <b>Revisión</b> |
|---------------|--------------|---------------------|--------------------|--------------------------|-----------------|
| Sujeto 1      | 3            | 1186000             | 9322               | 1043184                  | 133494          |
| Sujeto 2      | 3            | 1084000             | 47681              | 826470                   | 209849          |
| Sujeto 3      | 3            | 1233000             | 35542              | 1164965                  | 32493           |
| Sujeto 4      | 3            | 985000              | 13399              | 956878                   | 14723           |
| Sujeto 5      | 3            | 1108145             | 2940               | 714148                   | 391057          |
| Sujeto 6      | 3            | 1211000             | 16627              | 1057548                  | 136825          |
| Sujeto 7      | 5            | 1375000             | 8541               | 1351787                  | 14672           |
| Sujeto 8      | 5            | 791000              | 7646               | 532268                   | 251086          |
| Sujeto 9      | 5            | 1084000             | 8483               | 899266                   | 176251          |
| Sujeto 10     | 5            | 1023000             | 74919              | 498665                   | 449416          |
| Sujeto 11     | 5            | 807000              | 51219              | 661742                   | 94039           |
| Sujeto 12     | 4            | 666657              | 40581              | 581691                   | 44385           |
| Sujeto 13     | 4            | 1351000             | 108382             | 735306                   | 507312          |
| Sujeto 14     | 4            | 889000              | 3768               | 760922                   | 124310          |

Cuadro D.3: Tiempo en milisegundos para las fases de la condición SMD-20.

| <b>SMD-20</b> | <b>Tiempo total</b> | <b>Orientación</b> | <b>Elab. de borrador</b> | <b>Revisión</b> |
|---------------|---------------------|--------------------|--------------------------|-----------------|
| Promedio      | 1056700,14          | 30646,43           | 841774,29                | 184279,43       |
| DE            | 204092,63           | 30221,42           | 241455,69                | 155674,15       |

Cuadro D.4: Resumen estadístico de los datos recolectados en la condición SMD-20.

**SMI-20**

| <b>SMI-20</b> | <b>Nivel</b> | <b>Tiempo total</b> | <b>Orientación</b> | <b>Elab. de borrador</b> | <b>Revisión</b> |
|---------------|--------------|---------------------|--------------------|--------------------------|-----------------|
| Sujeto 1      | 3            | 1616000             | 21355              | 1074253                  | 520392          |
| Sujeto 2      | 3            | 1470000             | 54376              | 896567                   | 519057          |
| Sujeto 3      | 3            | 1544000             | 73186              | 368424                   | 1102390         |
| Sujeto 4      | 3            | 1419000             | 4055               | 1309242                  | 105703          |
| Sujeto 5      | 3            | 1230190             | 4737               | 761152                   | 464301          |
| Sujeto 6      | 3            | 1774820             | 3072               | 1171966                  | 599782          |
| Sujeto 7      | 5            | 1128000             | 9890               | 1100348                  | 17762           |
| Sujeto 8      | 5            | 1043000             | 15453              | 704437                   | 323110          |
| Sujeto 9      | 5            | 732000              | 6392               | 718309                   | 7299            |
| Sujeto 10     | 5            | 1090000             | 91407              | 487349                   | 511244          |
| Sujeto 11     | 5            | 753000              | 41047              | 546891                   | 165062          |
| Sujeto 12     | 4            | 937953              | 13247              | 802248                   | 122458          |
| Sujeto 13     | 4            | 1561000             | 109652             | 725592                   | 725756          |
| Sujeto 14     | 4            | 1324000             | 15105              | 799194                   | 509701          |

Cuadro D.5: Tiempo en milisegundos para las fases de la condición SMI-20.

| <b>SMI-20</b> | <b>Tiempo total</b> | <b>Orientación</b> | <b>Elab. de borrador</b> | <b>Revisión</b> |
|---------------|---------------------|--------------------|--------------------------|-----------------|
| Promedio      | 1258783,07          | 33069,57           | 818998                   | 406715,5        |
| DE            | 312859,83           | 34175,71           | 258886,60                | 295324,43       |

Cuadro D.6: Estadísticas del tiempo en milisegundos para las fases en la condición SMI-20.

**SMD-10**

| <b>SMD-10</b> | <b>Nivel</b> | <b>Tiempo total</b> | <b>Orientación</b> | <b>Elab. de borrador</b> | <b>Revisión</b> |
|---------------|--------------|---------------------|--------------------|--------------------------|-----------------|
| Sujeto 1      | 3            | 631000              | 9695               | 572777                   | 48528           |
| Sujeto 2      | 3            | 632000              | 7831               | 564462                   | 59707           |
| Sujeto 3      | 3            | 638000              | 22393              | 608756                   | 6851            |
| Sujeto 4      | 3            | 631000              | 10073              | 604947                   | 15980           |
| Sujeto 5      | 3            | 611906              | 6631               | 597672                   | 7603            |
| Sujeto 6      | 3            | 650000              | 4744               | 625796                   | 19460           |
| Sujeto 7      | 5            | 631000              | 7766               | 611716                   | 11518           |
| Sujeto 8      | 5            | 622000              | 11538              | 550011                   | 60451           |
| Sujeto 9      | 5            | 628000              | 5899               | 521832                   | 100269          |
| Sujeto 10     | 5            | 646249              | 53653              | 379435                   | 213161          |
| Sujeto 11     | 5            | 610000              | 31332              | 417714                   | 160954          |
| Sujeto 12     | 4            | 761332              | 37733              | 571292                   | 152307          |
| Sujeto 13     | 4            | 626000              | 39262              | 504733                   | 82005           |
| Sujeto 14     | 4            | 622000              | 9679               | 603986                   | 8335            |

Cuadro D.7: Tiempo en milisegundos para las fases de la condición SMD-10.

| <b>SMD-10</b> | <b>Tiempo total</b> | <b>Orientación</b> | <b>Elab. de borrador</b> | <b>Revisión</b> |
|---------------|---------------------|--------------------|--------------------------|-----------------|
| Promedio      | 638606,21           | 18444,93           | 552509,21                | 67652,07        |
| DE            | 35678,71            | 15140,37           | 71584,18                 | 64231,63        |

Cuadro D.8: Estadísticas del tiempo en milisegundos para las fases en la condición SMD-10.

**CMD-20**

| <b>CMD-20</b> | <b>Nivel</b> | <b>Tiempo total</b> | <b>Orientación</b> | <b>Elab. de borrador</b> | <b>Revisión</b> |
|---------------|--------------|---------------------|--------------------|--------------------------|-----------------|
| Sujeto 1      | 3            | 1470120             | 23224              | 1069422                  | 377474          |
| Sujeto 2      | 3            | 925605              | 51623              | 662849                   | 211133          |
| Sujeto 3      | 3            | 1295249             | 21442              | 1111631                  | 162176          |
| Sujeto 4      | 3            | 1160100             | 56944              | 898955                   | 204201          |
| Sujeto 5      | 3            | -                   | -                  | -                        | -               |
| Sujeto 6      | 3            | 1281145             | 10482              | 1100599                  | 170064          |
| Sujeto 7      | 5            | 740316              | 20379              | 593289                   | 126648          |
| Sujeto 8      | 5            | 888366              | 114271             | 577929                   | 196166          |
| Sujeto 9      | 5            | 877502              | 38090              | 719804                   | 119608          |
| Sujeto 10     | 5            | 1116495             | 86190              | 556842                   | 473463          |
| Sujeto 11     | 5            | 740176              | 20379              | 593289                   | 126508          |
| Sujeto 12     | 4            | 683816              | 40320              | 619588                   | 23908           |
| Sujeto 13     | 4            | 1025483             | 22323              | 778443                   | 224717          |
| Sujeto 14     | 4            | 1202959             | 333670             | 784148                   | 85141           |

Cuadro D.9: Tiempo en milisegundos para las fases de la condición CMD-20.

| <b>CMD-20</b> | <b>Tiempo total</b> | <b>Orientación</b> | <b>Elab. de borrador</b> | <b>Revisión</b> |
|---------------|---------------------|--------------------|--------------------------|-----------------|
| Promedio      | 1031333,23          | 64564,38           | 774368,31                | 192400,54       |
| DE            | 235898,97           | 82766,11           | 198788,57                | 114377,16       |

Cuadro D.10: Estadísticas del tiempo en milisegundos para las fases en la condición CMD-20.

## Pausas

### Baseline

| Sujeto    | Pausas                  |         |       |          |         |       |                |         |       |
|-----------|-------------------------|---------|-------|----------|---------|-------|----------------|---------|-------|
|           | Elaboración de borrador |         |       | Revisión |         |       | Sesión (total) |         |       |
|           | Corta                   | Mediana | Larga | Corta    | Mediana | Larga | Corta          | Mediana | Larga |
| Sujeto 1  | 22                      | 0       | 0     | 0        | 0       | 0     | 22             | 0       | 0     |
| Sujeto 2  | 3                       | 0       | 0     | 0        | 0       | 1     | 3              | 0       | 1     |
| Sujeto 3  | 24                      | 2       | 0     | 0        | 0       | 0     | 24             | 2       | 0     |
| Sujeto 4  | 31                      | 2       | 0     | 0        | 1       | 0     | 31             | 3       | 0     |
| Sujeto 5  | 6                       | 2       | 1     | 1        | 3       | 0     | 7              | 5       | 1     |
| Sujeto 6  | 26                      | 4       | 0     | 1        | 8       | 0     | 27             | 12      | 0     |
| Sujeto 7  | 23                      | 4       | 0     | 0        | 0       | 0     | 23             | 4       | 0     |
| Sujeto 8  | 4                       | 0       | 0     | 0        | 3       | 0     | 4              | 3       | 0     |
| Sujeto 9  | 33                      | 9       | 0     | 0        | 0       | 0     | 33             | 9       | 0     |
| Sujeto 10 | 0                       | 0       | 0     | 0        | 0       | 0     | 0              | 0       | 0     |
| Sujeto 11 | 4                       | 0       | 0     | 0        | 6       | 0     | 4              | 6       | 0     |
| Sujeto 12 | 25                      | 1       | 0     | 0        | 0       | 0     | 25             | 1       | 0     |
| Sujeto 13 | 2                       | 0       | 0     | 0        | 0       | 0     | 2              | 0       | 0     |
| Sujeto 14 | 34                      | 2       | 1     | 5        | 12      | 0     | 39             | 14      | 1     |

Cuadro D.11: Pausas en las distintas fases para Baseline.

## SMD-20

| Sujeto    | Pausas                  |         |       |          |         |       |                |         |       |
|-----------|-------------------------|---------|-------|----------|---------|-------|----------------|---------|-------|
|           | Elaboración de borrador |         |       | Revisión |         |       | Sesión (total) |         |       |
|           | Corta                   | Mediana | Larga | Corta    | Mediana | Larga | Corta          | Mediana | Larga |
| Sujeto 1  | 58                      | 30      | 1     | 1        | 2       | 0     | 59             | 32      | 1     |
| Sujeto 2  | 39                      | 17      | 0     | 2        | 5       | 0     | 41             | 22      | 0     |
| Sujeto 3  | 75                      | 34      | 0     | 0        | 0       | 0     | 75             | 34      | 0     |
| Sujeto 4  | 68                      | 23      | 0     | 0        | 0       | 0     | 68             | 23      | 0     |
| Sujeto 5  | 41                      | 9       | 1     | 2        | 7       | 1     | 43             | 16      | 2     |
| Sujeto 6  | 51                      | 30      | 0     | 0        | 2       | 0     | 51             | 32      | 0     |
| Sujeto 7  | 57                      | 37      | 1     | 0        | 0       | 0     | 57             | 37      | 1     |
| Sujeto 8  | 29                      | 12      | 0     | 13       | 8       | 0     | 42             | 20      | 0     |
| Sujeto 9  | 49                      | 16      | 0     | 4        | 4       | 0     | 53             | 20      | 0     |
| Sujeto 10 | 15                      | 2       | 0     | 0        | 7       | 2     | 15             | 9       | 2     |
| Sujeto 11 | 57                      | 5       | 0     | 1        | 5       | 0     | 58             | 10      | 0     |
| Sujeto 12 | 48                      | 13      | 0     | 4        | 13      | 0     | 52             | 26      | 0     |
| Sujeto 13 | 52                      | 9       | 0     | 3        | 12      | 1     | 55             | 21      | 1     |
| Sujeto 14 | 50                      | 19      | 0     | 1        | 3       | 1     | 51             | 22      | 1     |

Cuadro D.12: Pausas en las distintas fases para SMD-20.



## SMI-20

| Sujeto    | Pausas                  |         |       |          |         |       |                |         |       |
|-----------|-------------------------|---------|-------|----------|---------|-------|----------------|---------|-------|
|           | Elaboración de borrador |         |       | Revisión |         |       | Sesión (total) |         |       |
|           | Corta                   | Mediana | Larga | Corta    | Mediana | Larga | Corta          | Mediana | Larga |
| Sujeto 1  | 75                      | 34      | 0     | 2        | 10      | 2     | 77             | 44      | 2     |
| Sujeto 2  | 54                      | 23      | 0     | 8        | 12      | 1     | 62             | 35      | 1     |
| Sujeto 3  | 17                      | 12      | 0     | 30       | 42      | 1     | 47             | 54      | 1     |
| Sujeto 4  | 82                      | 41      | 0     | 3        | 1       | 1     | 85             | 42      | 1     |
| Sujeto 5  | 40                      | 22      | 0     | 11       | 11      | 2     | 51             | 33      | 2     |
| Sujeto 6  | 61                      | 34      | 0     | 3        | 9       | 1     | 64             | 43      | 1     |
| Sujeto 7  | 45                      | 36      | 0     | 0        | 0       | 0     | 45             | 36      | 0     |
| Sujeto 8  | 40                      | 19      | 0     | 4        | 10      | 0     | 44             | 29      | 0     |
| Sujeto 9  | 54                      | 12      | 0     | 0        | 0       | 0     | 54             | 12      | 0     |
| Sujeto 10 | 21                      | 5       | 0     | 2        | 14      | 1     | 23             | 19      | 1     |
| Sujeto 11 | 43                      | 11      | 0     | 5        | 10      | 0     | 48             | 21      | 0     |
| Sujeto 12 | 52                      | 23      | 0     | 4        | 6       | 0     | 56             | 29      | 0     |
| Sujeto 13 | 46                      | 16      | 0     | 5        | 13      | 4     | 51             | 29      | 4     |
| Sujeto 14 | 44                      | 26      | 0     | 11       | 22      | 0     | 55             | 48      | 0     |

Cuadro D.13: Pausas en las distintas fases para SMI-20.

## SMD-10

| Sujeto    | Pausas                  |         |       |          |         |       |                |         |       |
|-----------|-------------------------|---------|-------|----------|---------|-------|----------------|---------|-------|
|           | Elaboración de borrador |         |       | Revisión |         |       | Sesión (total) |         |       |
|           | Corta                   | Mediana | Larga | Corta    | Mediana | Larga | Corta          | Mediana | Larga |
| Sujeto 1  | 40                      | 9       | 0     | 0        | 0       | 0     | 40             | 9       | 0     |
| Sujeto 2  | 28                      | 5       | 0     | 1        | 3       | 0     | 29             | 8       | 0     |
| Sujeto 3  | 31                      | 17      | 0     | 0        | 0       | 0     | 31             | 17      | 0     |
| Sujeto 4  | 38                      | 15      | 1     | 0        | 0       | 0     | 38             | 15      | 1     |
| Sujeto 5  | 42                      | 18      | 0     | 0        | 0       | 0     | 42             | 18      | 0     |
| Sujeto 6  | 34                      | 15      | 0     | 0        | 0       | 0     | 34             | 15      | 0     |
| Sujeto 7  | 21                      | 24      | 0     | 0        | 0       | 0     | 21             | 24      | 0     |
| Sujeto 8  | 20                      | 10      | 0     | 3        | 0       | 0     | 23             | 10      | 0     |
| Sujeto 9  | 27                      | 1       | 0     | 3        | 4       | 0     | 30             | 5       | 0     |
| Sujeto 10 | 4                       | 1       | 0     | 3        | 5       | 0     | 7              | 6       | 0     |
| Sujeto 11 | 26                      | 0       | 0     | 9        | 8       | 0     | 35             | 8       | 0     |
| Sujeto 12 | 39                      | 15      | 0     | 3        | 9       | 0     | 42             | 24      | 0     |
| Sujeto 13 | 21                      | 2       | 0     | 1        | 3       | 0     | 22             | 5       | 0     |
| Sujeto 14 | 39                      | 10      | 0     | 0        | 0       | 0     | 39             | 10      | 0     |

Cuadro D.14: Pausas en las distintas fases para SMD-10.

**CMD-20**

| Sujeto    | Pausas                  |         |       |          |         |       |                |         |       |
|-----------|-------------------------|---------|-------|----------|---------|-------|----------------|---------|-------|
|           | Elaboración de borrador |         |       | Revisión |         |       | Sesión (total) |         |       |
|           | Corta                   | Mediana | Larga | Corta    | Mediana | Larga | Corta          | Mediana | Larga |
| Sujeto 1  | 41                      | 34      | 1     | 0        | 0       | 0     | 41             | 34      | 1     |
| Sujeto 2  | 38                      | 11      | 0     | 2        | 6       | 1     | 40             | 17      | 1     |
| Sujeto 3  | 44                      | 36      | 0     | 2        | 0       | 1     | 46             | 36      | 1     |
| Sujeto 4  | 68                      | 18      | 0     | 5        | 6       | 0     | 73             | 24      | 0     |
| Sujeto 5  | -                       | -       | -     | -        | -       | -     | -              | -       | -     |
| Sujeto 6  | 45                      | 28      | 0     | 4        | 3       | 1     | 49             | 31      | 1     |
| Sujeto 7  | -                       | -       | -     | -        | -       | -     | -              | -       | -     |
| Sujeto 8  | 34                      | 7       | 0     | 7        | 7       | 1     | 41             | 14      | 1     |
| Sujeto 9  | 43                      | 16      | 0     | 3        | 6       | 0     | 46             | 22      | 0     |
| Sujeto 10 | 28                      | 6       | 0     | 2        | 6       | 3     | 30             | 12      | 3     |
| Sujeto 11 | 40                      | 9       | 0     | 2        | 4       | 1     | 42             | 13      | 1     |
| Sujeto 12 | 46                      | 17      | 0     | 0        | 1       | 0     | 46             | 18      | 0     |
| Sujeto 13 | 45                      | 10      | 1     | 3        | 4       | 2     | 48             | 14      | 3     |
| Sujeto 14 | 36                      | 22      | 0     | 0        | 0       | 1     | 36             | 22      | 1     |

Cuadro D.15: Pausas en las distintas fases para CMD-20.