

Otro hecho que corrobora la elección de los tres elementos como indicadores de la competencia traductora es que todos constan, de manera explícita o implícita, en la mayoría de modelos de funcionamiento la competencia traductora que hemos expuesto (cf. supra 2.2. *Modelos de funcionamiento de la competencia traductora*), lo cual indica su validez como indicadores. En cualquier caso, uno de los objetivos de esta tesis es realizar pruebas de validez de los instrumentos creados, y una de las pruebas diseñadas se refiere específicamente a la validez de contenido, (cf. infra. 5.3. *Diseño para el estudio de la validez*), de manera que los datos aclararán hasta qué punto funciona la operacionalización propuesta.

Una vez elegidos los tres elementos indicadores de la competencia traductora, el siguiente paso es definir cómo observarlos, es decir, qué tipo de instrumentos hay que diseñar para valorar cada indicador, y dada la complejidad de los tres elementos, hemos preferido tratarlos por separado.

3. La concepción general de la traducción

La concepción general de la traducción es la idea abstracta que se tiene de lo que es y de lo que implica la traducción. Creemos que existen dos grandes bloques conceptuales dentro de la noción general de traducción, uno es la traducción y el otro es la competencia traductora.

Así pues, el instrumento creado para medir la concepción general de la traducción es un cuestionario multidimensional, que considera los diferentes factores que, a nuestro entender, son más relevantes dentro de la concepción abstracta de la traducción:

- (a) definición de la traducción,
- (b) elementos que intervienen en la traducción,
- (c) concepción de los problemas de traducción,
- (d) concepción de la unidad de traducción,
- (e) concepción de la equivalencia de traducción,
- (f) concepción de la función de la traducción,
- (g) concepción de la competencia traductora,
- (h) concepción de la subcompetencia instrumental,
- (i) concepción de la subcompetencia lingüística,
- (j) concepción de la subcompetencia extralingüística,
- (k) concepción de la subcompetencia estratégica,

De estos doce elementos, los seis primeros responden a la concepción de la traducción y los otros seis a la concepción de la competencia traductora, de manera que las dos secciones conceptuales están equilibradas en cuanto a número de preguntas, dado que cada uno de estos elementos está representado mediante un ítem en el cuestionario elaborado (cf. páginas 39-41 del volumen II).

Por otra parte, en este instrumento añadimos dos preguntas para indagar sobre las estrategias utilizadas para la resolución de problemas de comprensión en la lengua extranjera y de reexpresión en la lengua materna durante el acto de traducción. Sin embargo, las respuestas a estas preguntas no afectan a la puntuación de este instrumento sino que sirven a otro propósito, el de contrastar una de las hipótesis complementarias respecto al uso de obras de consulta (cf. infra 5.1.3. *Hipótesis complementarias*).

Por último, nos gustaría señalar que los factores elegidos para observar la concepción general de la traducción responden a nuestra visión de la traducción y de la

competencia traductora, ya que el instrumento debe intentar medir el constructo y en nuestro caso partimos de una definición y de una descripción del funcionamiento de la competencia traductora. Así pues, las pruebas de validez y generalizabilidad deben mostrar si los resultados obtenidos mediante este instrumento valoran la concepción general de la traducción de los sujetos tal como la entendemos nosotros. De esta manera, para valorar la concepción general de la traducción tal como la entienden otros autores habría que elaborar otro cuestionario.

4. Los problemas de traducción

4.1. Teoría de resolución de problemas

Existen multitud de teorías acerca del modo en que funciona la resolución de problemas, especialmente en Psicología, disciplina que se ha ocupado con gran énfasis del tema. Sin pretender ser exhaustivos, podemos decir que encontramos desde autores que definen una habilidad general y un método de resolución distinto para cada tipo de problema, hasta autores que opinan que no existe una habilidad de resolución de problemas general y unitaria, sino que ésta se desarrolla de manera diferente cada vez que a una persona se le presenta un problema, dado que existe una variedad infinita de problemas:

“In dealing with problem solving, psychologists have generally fallen victim of the (...) fallacy of thinking that one term means one thing. Factor analysis has never demonstrated a unitary problem-solving ability. The reason is that there is almost an infinite variety of problems. If we define a problem as a situation to which the individual has no ready, adequate response, we see that the variety of problem situations is enormously large and that problem solving is essentially as broad as behavior itself.” (Guilford, 1968:63)

Respecto a los autores que describen las fases del proceso de resolución de problemas, éstos opinan que se trata de una tarea que requiere procesos de razonamiento complejos, y suelen proponer diferentes etapas. Sternberg (1996:346-350), por ejemplo, distingue siete etapas: (a) identificación del problema, (b) definición del problema, (c) formulación de una estrategia, (d) organización de la información, (e) distribución de recursos, (f) supervisión del proceso y (g) evaluación.¹⁶

Nosotros adoptamos esta última tendencia y aceptamos la idea de que existen fases definidas dentro del proceso de resolución de problemas, de manera que nuestro propósito es aplicar la distinción de fases a la resolución de problemas de traducción.

4.2. La definición de problema de traducción

A la hora de abordar los problemas de traducción existen dos obstáculos básicos; el primero es que no contamos con una definición de problema de traducción específica que sea aceptada mayoritariamente en nuestra disciplina y el segundo obstáculo es que tampoco disponemos de una clasificación de problemas de traducción

¹⁶ En este sentido, cf. también Kluwe, 1987; Mayer, 1986:50-54, 74-81, 405-439; Pérez Echiverría y Pozo, 1994; Pozo, 1994; Puente Ferreras, 1998:408-417; Sternberg, 1996:345-374.

que haya sido probada empíricamente o que sea aceptada en general por la disciplina. Como señala Wilss, la Traductología tiene serias dificultades con los problemas:

“But the understanding of problems that we find in such publications is not uniform, and the investigation of the differences between translation problems and translation difficulties as initiated by Nord (1987) has not been continued. This statement is not a negative judgement; it is intended merely to express the view that Translation Studies apparently has had, and still does have, great trouble in defining a suitable and reliable conceptual framework for problem-solving.” (1996:47)

En efecto, existen multitud de posturas respecto a la definición y a la clasificación de problemas de traducción¹⁷. Sin embargo, dado que no es nuestro propósito exponer todas las posibilidades sino aclarar la definición y la clasificación en las que nos basamos para crear el instrumento de medida de la actuación frente a los problemas de traducción, nos limitamos a presentar la definición de problema de traducción que hemos adoptado, incidiendo en la clasificación de problemas que aplicamos.

Entendemos por problema de traducción, de forma general, “an objective problem which every translator (irrespective of his level of competence and of the technical conditions of his work) has to solve during a particular translation task” (Nord, 1991:151). Sin embargo, Nord va más allá de la definición de problema de traducción y efectúa una distinción entre problemas y dificultades de traducción, especificando que éstas no se pueden prever, dado que son específicas de cada individuo:

“Translation difficulties, on the other hand, are subjective and have to do with the translator himself and his specific working conditions. A particular translation problem which seems very difficult to the beginner will remain a translation problem, even when the student has learned to cope with it. It can turn into a difficulty again, though, if the translator has to solve it without the necessary technical resources.” (Nord, 1991:151)

En este sentido, Nord distingue entre cuatro tipos de dificultades de traducción: las específicas del texto, las que dependen del traductor, las pragmáticas y las técnicas, es decir, las relacionadas con la especificidad del tema que trata el texto. Para nuestra investigación necesitamos clasificar y prever algunos de los problemas que presenta un texto en concreto, con el fin de poder observar la actuación de los sujetos frente a dichos

¹⁷ En este sentido, cf. Martínez Melis y Hurtado, 2000; Nord, 1991, 1992 y 1996; Presas, 1996; Waddington, 1999; Wilss, 1996.

problemas y analizar de este modo su aplicación de estrategias para resolverlos, pero no pretendemos en absoluto crear un instrumento para medir las dificultades individuales de cada estudiante, ya que nuestro interés es la competencia traductora de los estudiantes de traducción en general, y analizar las dificultades individuales sería otra investigación diferente.

Por otra parte, a pesar de que la distinción entre varios tipos de dificultad escapa a los propósitos de nuestra investigación, nos parece que éste es un tema que convendría estudiar con mayor detalle y, quizás, hacerlo objeto de una investigación dedicada a recoger datos empíricos que muestren las diferencias entre los problemas y las dificultades de traducción.

En cuanto a la clasificación de los problemas, Nord define cuatro tipos de problemas:

There are four categories of translation problems: (a) problems arising from the particular features of the source text (text-specific translation problems, e.g. a play on words), (b) problems arising from the nature of the translation task (pragmatic translation problems, e.g. the recipient orientation of a text), (c) problems arising from the differences in norms and conventions between the source and the target culture (cultural translation problems, e.g. text-type conventions), (d) problems arising from the structural differences between source and target language (linguistic translation problems, e.g. the translation of the English gerund into German)." (Nord, 1991:151)

Al determinar los cuatro tipos de problemas que contienen los textos que utilizamos en nuestro instrumento de medida nos basamos, en parte, en esta clasificación de los problemas de traducción en textuales, pragmáticos, culturales y lingüísticos. Sin embargo, existe una diferencia entre la clasificación de Nord y la elección de problemas que hemos realizado para nuestro instrumento, ya que, en primer lugar, incluimos los problemas textuales en los problemas lingüísticos y, en segundo lugar, observamos la actuación de los sujetos frente a un problema de transferencia.

Preferimos no utilizar la categoría de problema textual tal como la describe Nord por dos razones: en primer lugar, el problema textual se solapa fácilmente con el problema lingüístico, si entendemos problema lingüístico como problema que afecta a la competencia lingüística, ya que a nuestro entender la competencia textual se incluye en la competencia lingüística. Los ejemplos que propone la propia autora como problemas textuales -juegos de palabras, metáforas, palabras creadas por el autor del TO- recaen claramente, según nuestra definición, en la categoría lingüística (Nord, 1992:47). En segundo lugar, si, tal como dice Nord (1992:47) cualquier tipo de

problema que no sea lingüístico, cultural o pragmático es textual, resulta que la categoría de problema textual es muy difusa, porque incluiría, por ejemplo, desde los problemas de creatividad hasta los problemas de búsqueda de información. Así pues, prescindimos de esta categoría y proponemos otro tipo de problema que no queda cubierto por los otros tres (lingüístico, cultural y pragmático) y que es muy frecuente en cualquier tipo de traducción, ya que obedece a la subcompetencia de transferencia, tal como la define el modelo de PACTE (cf. supra 2.3. *El modelo de PACTE*). Nos referimos al problema de transferencia, que viene causado por alguna unidad o segmento del TO que requiere una equivalencia funcional y dinámica, ya que no existe una equivalencia preestablecida o, en caso de existir, no se puede utilizar en esa situación en concreto; el problema de transferencia puede derivarse de las interferencias entre las dos lenguas de trabajo, y en cualquier caso no está marcado por las dificultades lingüísticas ni extralingüísticas.

Por último, otra diferencia entre la clasificación de Nord y nuestra elección de problemas es que ampliamos el concepto de problema cultural a problema extralingüístico, ya que consideramos que éstos incluyen a aquellos. De este modo, los problemas extralingüísticos pueden ser culturales, temáticos o enciclopédicos, dado que todos ellos exigen las mismas habilidades por parte del traductor: razonamiento lógico, creatividad (en ocasiones) y conocimiento del mundo, de las dos culturas y del tema tratado por el texto.

Nos gustaría aclarar que nuestra propuesta de elegir cuatro tipos de problemas de traducción es simplemente una solución operativa para crear un instrumento de medida, y no pretende en absoluto ser una clasificación general de problemas de traducción; es decir que nuestra elección de problemas solamente es aplicable al instrumento de medida de la actuación frente a los problemas de traducción tal como lo concebimos en esta investigación.

Así pues, nuestra elección incluye cuatro tipos de problemas:

- (a) lingüístico; aquel constituido por un elemento léxico, sintáctico, semántico o textual que exige para su resolución los conocimientos incluidos en nuestro modelo de competencia comunicativa en las dos lenguas (cf. supra 2.3. *El modelo de PACTE*), es decir, los sistemas subyacentes de conocimientos y habilidades necesarios para la comunicación lingüística, como la competencia de comprensión en la LO y la de producción en la LM;

- (b) pragmático, aquel que se deriva del encargo, de la función de la traducción o de las expectativas del lector final de la traducción;
- (c) de transferencia, aquel constituido por un elemento léxico, sintáctico, semántico o textual que requiere una equivalencia dinámica;
- (d) extralingüístico, aquel constituido por un elemento que pone de relieve un conocimiento enciclopédico y un modelo concreto del universo, una diferencia cultural entre las dos culturas que intervienen en la traducción o la especificidad del campo temático que trata el TO y que para su resolución requiere una movilización de los conocimientos enciclopédicos, culturales y/o temáticos del traductor.

Con el fin de aclarar cada categoría, a continuación citamos un ejemplo de cada tipo de problema, basándonos en el texto "Would you like to be a writer?", que incluimos a continuación.

Encargo de traducción:

Una compañía española ha decidido abrir una empresa similar a "The Writers Bureau", con sede en Madrid, para enseñar a escribir en español, y te pide que traduzcas el siguiente anuncio (ver página 4 de este cuadernillo), extraído de la primera página de *The Guardian*, para publicarlo en el periódico *El País*. Tu texto será publicado tal como lo entregues en este periódico español.

Advertisement
Would you like to be a writer?

by NICK DAWS

Freelance writing can be creative,
5 fulfilling and a lot of fun, with excellent
 money to be made as well. What's more,
 anyone can become a writer. No special
 qualifications or experience are required.
 The market for writers is huge. In Britain
10 alone there are around 1,000 daily,
 Sunday and weekly papers, and

more than 8,000 magazines.
 The Writers Bureau runs a
 comprehensive correspondance course
15 covering every aspect of fiction and non-
 fiction writing.
 The 140,000 word course is written by
 professional writers and has been
 acclaimed by experts

20 Why not be a writer?

First-class home study course gets you a flying start. Earn while you learn . Expert tutors,
 personal guidance, help to sell your writing, and much more! It's ideal for beginners.
 Details free. No cost. No obligation. Send the coupon.

Name

25 *Address*

Telephone *Postcode*

- (a) Problema lingüístico (*flying*; línea 21). Presenta un problema lingüístico de comprensión léxica, porque el término no está utilizado según su acepción habitual, y también plantea un problema lingüístico de reexpresión, ya que hay que buscar un equivalente adecuado, que no se encuentra en los diccionarios bilingües, como “prometedor” o “brillante”, etc.;
- (b) Problema extralingüístico (*Sunday and weekly papers*; línea 11). Presenta un problema cultural porque en España no existen este tipo de periódicos (tenemos revistas dominicales, pero no son exactamente lo mismo) de modo que el traductor, de acuerdo con el encargo, debe intentar adaptar culturalmente estos elementos u optar por otra solución, como por ejemplo generalizar y resumir todos los tipos de publicaciones mencionados en el texto escribiendo “revistas y periódicos”, “publicaciones diarias o semanales”, etc.;
- (c) Problema de transferencia (*It's ideal for beginners. Details free. No cost. No obligation. Send the coupon*; líneas 22-23). Presenta un problema de transferencia porque la cohesión de estas frases en inglés no es la misma que la que presenta un texto de este tipo en castellano, lo habitual en español sería una sola frase, o a lo sumo dos, que presenten las ideas unidas entre sí, utilizando los conectores y referentes pertinentes. Una posible solución a este problema sería “Este curso de primera clase, ideal para principiantes, le asegura un comienzo brillante y le permite ganar dinero mientras aprende. Le ofrecemos (...) Conozca más detalles de manera gratuita y sin ningún compromiso enviando el siguiente cupón.”;
- (d) Problema pragmático, presentado por el encargo de traducción. (*Britain*; línea 9) este segmento resulta problemático porque el encargo de traducción especifica que el texto debe servir como publicidad de una empresa con sede en Madrid que va a ofrecer un curso de escritura en español, por lo tanto no tiene sentido dar información sobre lo que ocurre en Gran Bretaña al lector, y en cambio es necesario adaptar la información para que sea válida en el nuevo contexto y sustituir “Britain” por “España” (o Latinoamérica, etc.).

Somos conscientes de que esta elección de problemas no es exhaustiva ni definitiva, ya que se podría incluir, por ejemplo, algún problema relacionado con la creatividad, pero nos parece que, dado que los sujetos a quienes va dirigido este instrumento son estudiantes de primer curso, estos cuatro tipos de problemas son

suficientes para ofrecernos información respecto a su actuación frente a los problemas de traducción.

4.3. Cómo medir la actuación respecto a los problemas de traducción

El primer obstáculo que encontramos al intentar diseñar un instrumento de medida de la actuación frente a los problemas de traducción es que no contamos con una descripción del método de resolución de problemas para esta situación en concreto, que sin duda tiene sus características propias, como señala Wilss (1996:47).

Al hablar de problemas de traducción, la mayoría de autores suelen distinguir entre problemas resueltos y no resueltos, dado que es relativamente sencillo observar si un problema concreto ha sido resuelto o no por una persona que ha traducido un texto. Sin embargo, nosotros nos planteamos desde un principio la necesidad de poder observar y constatar todo el proceso de resolución de un problema, que a nuestro entender consta esencialmente de tres pasos: (a) detectar el problema, (b) aplicar las estrategias necesarias para resolverlo, (c) resolverlo.

Esta idea no es nueva en nuestra disciplina; Hansen, por ejemplo, incluye en su modelo de funcionamiento de la competencia traductora una subcompetencia de transferencia en la que habla de “*strategies for recognising and solving translation problems*” (Hansen, 1997:206), refiriéndose a la detección del problema como un paso necesario que el traductor debe ser capaz de dar. Dancette, por su parte, también está de acuerdo con las tres etapas implicadas en la resolución de problemas de traducción:

“For example, most writers identify major steps in problem solving such as problem identification (reception of initial data, first interpretation, awareness of goals and constraints), data processing (comparisons, classifications, inferences, deductions, construction of propositions and images, transformations, and so on), and response (monitoring, revision, and decision making).” (Dancette, 1997:86).

Por consiguiente, para elaborar el instrumento de medida de la actuación frente a los problemas de traducción hemos creado una escala ordinal. Ésta permite valorar los tres pasos descritos, si bien el segundo paso solamente se mide de forma indirecta, a través de la no resolución del problema. Los puntos de la escala se otorgan del siguiente modo:

0 puntos: el sujeto no ha detectado el problema.

1 punto: el sujeto ha detectado el problema pero no ha sabido resolverlo.

2 puntos: el sujeto ha resuelto bien el problema.

Con el fin de saber si un sujeto ha detectado o no un problema y si lo ha resuelto, es decir, para poder puntuar su actuación frente a un problema de traducción, en el instrumento creado se encarga a cada sujeto que traduzca un texto, de acuerdo con un encargo de traducción, y que rellene un cuestionario acerca del texto traducido. No nos extendemos en este punto, ya que más adelante dedicamos un apartado a explicar el método de puntuación del instrumento, así como el desarrollo del instrumento, desde la idea inicial hasta su formato actual (cf. *infra 4.3. Desarrollo del instrumento de medida de la actuación frente a los problemas de traducción, IPT*).

5. Los errores de traducción

Al igual que ocurre con los problemas de traducción, existen infinidad de puntos de vista respecto a la definición y a la clasificación de los errores de traducción. Sin embargo, en este caso las perspectivas son, si cabe, más numerosas, dado que la mayoría de docentes de la traducción corrigen las traducciones de sus estudiantes marcando los errores, y por consiguiente cuentan con una idea propia al respecto, que va sufriendo modificaciones de acuerdo con los cambios en la práctica docente. Una clara muestra de este hecho se puede ver en la tesis doctoral de Waddington (1999), dedicada a investigar diferentes métodos de evaluación en traducción escrita, donde el autor diferencia entre dos tipos de evaluación: la holística y la analítica; los resultados de una encuesta realizada por el autor, a la que respondieron 52 docentes de 20 centros españoles, ponen de manifiesto que la mitad de estos profesores se basaban enteramente en los errores de traducción de los estudiantes en una traducción para evaluar a éstos, si bien cada docente defendía un concepto y una clasificación de errores diferentes (cf. Waddington, 1999: 274-279).

Esta gran variedad de conceptos y clasificaciones nos lleva, nuevamente, a limitarnos a dejar clara nuestra postura respecto a los errores de traducción, sin entrar en disquisiciones respecto a las ventajas o desventajas de los modelos existentes. Para conocer el estado de la cuestión, nos remitimos a Martínez-Melis y Hurtado, 2000; respecto a la definición y clasificación del error en traducción, cf., p.e., Dancette, 1989; Delisle, 1993; Dollerup, 1982; Gile, 1992; Gouadec, 1981 y 1989; House, 1981; Hurtado 1995a y 1999; Kupsch-Loreseit, 1985; Kussmaul, 1995; Martínez Melis, 1997; Nord, 1996; Pym, 1993; Séguinot, 1989b, 1990 y 1991.

5.1. Definición y clasificación del error de traducción

El error de traducción es, pues, un concepto complejo en nuestra disciplina, dado que no existe un consenso sobre la denominación, clasificación y delimitación de cada caso concreto de posible error de traducción. Por consiguiente, consideramos que se debe analizar el error en una situación dada, es decir, dentro del marco de factores que influyen en una traducción (el encargo de traducción, las expectativas del lector final, la intencionalidad del autor original, etc.). Nord define error del siguiente modo: “Si, en el marco del concepto funcional, la finalidad del proceso traslativo se define mediante el encargo de traducción, un no-cumplimiento de tal encargo, con respecto a determinados

aspectos funcionales, debe considerarse como error o falta.” (1996:95). De esta definición se deduce que un error de traducción es un no cumplimiento de alguna de las características que debe reunir el texto meta, y este es el concepto de error que adoptamos en lo sucesivo.

Dejando de lado la definición del error de traducción, lo que sí parece claro es que el origen de un error suele ser un problema de traducción (o una dificultad de traducción) no resuelto o mal resuelto. Nord (1996:96-100) distingue entre (a) errores pragmáticos, que serían el resultado de no resolver bien un problema pragmático o de no respetar las indicaciones pragmáticas, (b) errores culturales, que no respetan las convenciones específicas de la cultura meta, y que serían el resultado de no resolver bien un problema o dificultad cultural y (c) errores lingüísticos, que Nord limita al contexto didáctico, ya que opina que en el contexto profesional no se dan. Estos últimos errores son, según la autora, el resultado de los problemas lingüísticos mal resueltos o de las dificultades particulares de un traductor en formación (el desconocimiento o la no observación de alguna regla gramatical, léxica, de ortografía o puntuación, etc.)

Como se desprende de esta clasificación, los errores de traducción están estrechamente ligados a los problemas de traducción, punto que apoya nuestra propuesta de operacionalización de la competencia traductora. Sin embargo, la clasificación de errores de Nord se nos antoja poco práctica a la hora de definir las categorías de errores que queremos observar en nuestro instrumento, ya que no se contemplan varios tipos de errores que son muy comunes en los estudiantes de primer curso como, por ejemplo, los falsos sentidos, es decir, aquellos segmentos del TM que dicen algo diferente de lo que dice el TO, aunque el TM tenga sentido y por consiguiente este error no constituya un incumplimiento del encargo o de la función de traducción.

De este modo, preferimos aplicar el baremo que propone Hurtado para corregir traducciones de los estudiantes (1995a y 1999). Este baremo se divide en cuatro apartados:

(a) Inadecuaciones que afectan la comprensión del texto original; este apartado incluye las siguientes categorías de error: contrasentido, falso sentido, sinsentido, adición innecesaria de información, supresión innecesaria de información, alusiones extralingüísticas no solucionadas, no mismo sentido y, por último, inadecuación de variación lingüística;

- (b) Inadecuaciones que afectan la expresión en la lengua de llegada: ortografía y puntuación, gramática, léxico, inadecuación textual e inadecuación estilística;
- (c) Inadecuaciones funcionales: inadecuación a la función textual prioritaria del original e inadecuación a la función de la traducción;
- (d) Aciertos, que incluye las categorías de buena equivalencia y muy buena equivalencia.

Este baremo es completo y permite marcar claramente cada tipo de error, lo cual facilita tanto la anotación del evaluador como la posterior comprensión del estudiante respecto al tipo de error en que ha incurrido. Por otra parte, esta propuesta cuenta con un apartado de aciertos, lo cual nos parece especialmente interesante tanto desde el punto de vista didáctico como para nuestro instrumento, ya que de esta manera se pueden utilizar puntos positivos (derivados de los aciertos) y puntos negativos (derivados de los errores) para obtener una puntuación total.

Sin embargo, la gran cantidad de categorías de error que incluye esta propuesta constituye un problema para nuestro instrumento, dado que éste debe contar con un método de corrección de las traducciones que requiera el mínimo esfuerzo por parte del evaluador. Así pues, a efectos del instrumento de la actuación respecto a los errores de traducción, solamente observamos tres categorías de las mencionadas:

- (a) aciertos: segmentos del texto en que la traducción es especialmente buena, ya sea porque propone una solución original, creativa –y adecuada al encargo, por supuesto- a alguno de los problemas que plantea el texto o porque la traducción suena como un texto original en castellano, gracias a la cohesión y coherencia, por ejemplo, etc.;
- (b) falsos sentidos: segmentos del texto en que la traducción no tiene el mismo sentido que el texto original;
- (c) sinsentidos: segmentos del texto traducido que no tienen sentido, es decir que no serían comprendidos por el lector final.

La elección de estos tres elementos no es aleatoria, sino que responde a dos razones. En primer lugar, se trata de categorías fácilmente observables por los evaluadores, es decir que la diferencia de criterios de los evaluadores puede afectar de modo reducido a la corrección de traducciones. Este punto es importante porque queremos crear instrumentos de medida cuyos resultados sean generalizables y, por consiguiente, el grado de variabilidad que depende del evaluador debe ser reducido,

como veremos más adelante (cf. infra 5.2. *Diseño para el estudio de la generalizabilidad*).

En segundo lugar, los sinsentidos y los falsos sentidos son errores específicos de la traducción, mientras que muchas otras categorías de errores se dan también en la redacción en lengua materna, sin que se trate necesariamente de una traducción; por otra parte, opinamos que tanto los falsos sentidos como los sinsentidos son errores especialmente graves para el traductor.

En tercer lugar, dado que nos interesa construir un instrumento de medida que sea sensible al cambio, es decir, que permita observar los cambios que tienen lugar en las traducciones de los estudiantes de primer curso, que constituyen la muestra de nuestra investigación, es conveniente elegir categorías de errores en las que normalmente se observan cambios a lo largo del primer curso. Si bien no existen estudios empíricos al respecto, nuestra experiencia como docentes nos lleva a pensar que los estudiantes de primer curso suelen cometer más falsos sentidos y sinsentidos a principio de curso que a final, mientras que siguen cometiendo una cantidad parecida de inadecuaciones de otros tipos, especialmente las que están relacionadas con la función del TM y con la lengua meta. Por otra parte los aciertos muestran claramente la evolución de los estudiantes, ya que a final del primer curso advierten problemas que al principio no detectaban, y por consiguiente solucionan más problemas y esas soluciones suelen constituir aciertos.

6. *Instrumentos de medida válidos y fiables*

6.1. Necesidad de crear instrumentos de medida válidos y fiables

Para llevar a cabo una investigación experimental, los instrumentos de medida son no sólo indispensables sino también determinantes, ya que si no reúnen las garantías necesarias de validez y fiabilidad podrían dar al traste con la investigación: podrían no medir lo que el investigador quiere medir o estarlo midiendo pero obtener datos desviados a causa de la gran variabilidad de alguno de los factores que intervienen en el proceso de medición. Por todo ello se hace indispensable crear instrumentos de medida válidos y fiables y realizar las pruebas necesarias para asegurarse de que los datos obtenidos mediante estos instrumentos serán útiles y servirán a su propósito.

Estas consideraciones, unidas a la novedad que suponen los instrumentos de medida en el campo de la traducción, nos hicieron advertir desde un principio la necesidad de tomar prestados conceptos y técnicas de la psicometría, la ciencia cuyo objeto de estudio es precisamente la medida, como se desprende de la siguiente definición:

“La meta de la Psicología, como de las otras ciencias empíricas, es la descripción, predicción y explicación de los fenómenos empíricos de su área de interés (...) Para lograr este objetivo, como señala Torgerson (1958), las ciencias deben recoger y comparar datos para establecer correlaciones, ecuaciones, etc. que permitan fundamentar las teorías. La medida es un componente fundamental de este proceso y es la que permite que los instrumentos de las matemáticas puedan aplicarse a la ciencia (Torgerson 1958:1) Esta afirmación acerca del papel fundamental de la medida es aplicable tanto a la Psicología como a otras ciencias empíricas y no es extraño pues que una materia importante en esta ciencia sea la **Psicometría**. Esta denominación aglutina todo el conjunto de modelos formales que posibilitan la medición de variables psicológicas, centrándose en las condiciones que permiten llevar a cabo todo proceso de medición en Psicología y en establecer las bases para que estos procesos se realicen de forma adecuada.” (Martínez 1995:21, nuestro subrayado)

Esta definición, especialmente la parte subrayada, es perfectamente aplicable a la Traductología, si la entendemos como ciencia empírica, y es especialmente útil en nuestro caso, ya que tratamos de describir fenómenos empíricos (la adquisición de la competencia traductora) y medir variables psicológicas.

6.2. Características de los instrumentos de medida

Hemos mencionado en varias ocasiones el término “instrumento de medida”, pero la psicometría suele hablar con más frecuencia de “tests”; de hecho, la disciplina especializada que estudia los problemas que surgen cuando se crean instrumentos de medida y uno de los métodos para su solución se denomina Teoría de los Tests, y su cometido es esencialmente el de estimar el grado en que los aspectos marginales (en nuestro caso, la combinación lingüística, el centro donde se imparte la enseñanza de traducción, el evaluador, etc.) influyen en un instrumento de medida en una situación determinada y diseñar métodos para minimizar dichos aspectos (Martínez 1995:30). La palabra inglesa *test*, cuyo significado como sustantivo equivale al término español prueba, procede del latín *testis*, y ha sido adoptada internacionalmente para designar un cierto tipo de examen de uso extendido en Psicología y Educación. Se podrían dar muchas definiciones de test, he aquí algunas: Cronbach, en 1971, lo definió como “técnica sistemática para comparar la conducta de dos o más personas”; Anstey, en 1976, lo describió del siguiente modo: “instrumento de evaluación cuantitativa de los atributos psicológicos de un sujeto”; Anastasi, en 1984, afirmó que “un test es una medida objetiva y estandarizada de una muestra de conducta”; Graham y Lilly, en 1984 ofrecieron la siguiente definición, “un test es una muestra estandarizada de conductas de la que pueden inferirse o predecirse otras conductas importantes”. Sin embargo, las dos definiciones más completas las ofrecen Martínez y Yela:

“En general, podemos considerar el test como un reactivo que, aplicado a un sujeto, revela y da testimonio del tipo o grado de su aptitud, de su forma de ser o del grado de instrucción que posee. Estos reactivos o tests constan de preguntas, tareas, estímulos, situaciones, etc. Que intentan poner de relieve una muestra de las conductas del sujeto, representativa de la característica que se quiere apreciar o medir.” (Martínez 1995:31)

“Un test es una situación problemática, previamente dispuesta y estudiada a la que el sujeto ha de responder siguiendo ciertas instrucciones y de cuyas respuestas se estima, por comparación con las respuestas de un grupo normativo (o un criterio u objetivo), la calidad, índole o grado de algún aspecto de su personalidad. (Yela, 1980:)

En el momento en que se plantea la necesidad de crear un instrumento de medida, hay que hacer frente a toda una serie de decisiones que irán dando forma a las características finales de los tests. A continuación describimos ese proceso, especificando las razones que nos han llevado a crear los instrumentos de medida del modo en que lo hemos hecho. En primer lugar, especificamos la clasificación de los

instrumentos de acuerdo con una serie de criterios, aclarando las categorías en que se integran nuestros instrumentos de medida.

En cuanto al método, y citando la terminología de Martínez (1995), nuestro instrumento es psicométrico, ya que para su evaluación seguiremos normas o criterios estandarizados. Respecto a la finalidad, se trata de un test de investigación, por oposición a un test de diagnóstico, ya que lo que pretendemos es obtener medidas, analizar diferencias individuales (de la segunda medida respecto a la primera) y calcular índices estadísticos grupales, además de contrastar hipótesis (cf. infra capítulos 5 y 6).

Por lo que se refiere al planteamiento del problema, las respuestas que obtendremos serán de ejecución máxima, es decir que la situación planteada al sujeto requiere que ponga en funcionamiento su capacidad en alto grado para poder resolverla, y respecto al área del comportamiento acotada se trata de un test de rendimiento, ya que está destinado a medir adquisición de un aprendizaje, en este caso adquisición de la competencia traductora.

La modalidad de aplicación del instrumento es de “papel y lápiz”, ya que todo el sistema de administración, instrucciones y realización se especifica en protocolos, cuadernillos y hojas de respuesta de forma escrita. Además, el instrumento de medida es de aplicación colectiva y por las demandas exigidas al sujeto se trata de un test de potencia o dificultad, ya que el tiempo, aunque se controla, no es determinante en el resultado, y en cambio es la calidad de ejecución lo que crea las diferencias individuales.

Según el modelo conceptual, el test que nos ocupa es un test referido al criterio, dado que el acento está en el nivel o el grado de ejecución sobre un dominio previamente definido: la competencia traductora. Respecto al modelo estadístico, la psicometría nos advierte claramente de un importante problema que afecta al diseño de instrumentos para la medición de constructos psicológicos:

“La medida obtenida siempre tiene error. Puesto que se basan en muestras limitadas en contenidos y recogidas en un momento en el tiempo, están afectadas por todos los errores de muestreo. Si se examina dos veces al sujeto en las mismas condiciones, es poco probable que las puntuaciones sean idénticas, por múltiples razones no predecibles. Estas inconsistencias ligadas a un muestreo de tareas, ocasiones o situaciones, pueden considerarse errores de medida. Uno de los problemas de la medida en Psicología es estimar el grado de error presente en un conjunto de observaciones” (Martínez 1995:29)

Estas consideraciones nos llevaron a tener en cuenta como factor importante la variabilidad de los datos obtenidos, y nos hemos esforzado en conocer el grado de variabilidad que es atribuible a las fuentes de variancia potenciales, a saber, los evaluadores, los centros universitarios o los diferentes planes de estudios y las combinaciones lingüísticas (cf. infra 5.2. *Diseño para el estudio de la generalizabilidad*), ya que sólo de esta manera podremos examinar los datos obtenidos mediante los instrumentos de medida de forma objetiva y por lo tanto llegar a conclusiones reales, y no desviadas por no tener en cuenta el margen de error habitual.

Como se desprende de todo lo dicho, hemos seguido el modelo estadístico de la Teoría de la Generalizabilidad, basada en el Modelo Lineal, ya que lo que mide el test son conceptos que varían a lo largo del tiempo y se trata de tres instrumentos que deben ser sensibles al cambio para poder observar los progresos efectuados por los estudiantes.

La siguiente figura resume las características psicométricas de los instrumentos de medida que hemos creado. La columna izquierda apunta el criterio según el cual el instrumento es del tipo que indica la columna derecha:

Método	Test psicométrico
Finalidad	Test de investigación
Planteamiento del problema	Test de ejecución máxima
Área de comportamiento acotada	Test de rendimiento
Modalidad de aplicación	Test de papel y lápiz
Tipo de demanda realizada al sujeto	Test de potencia o dificultad
Modelo conceptual	Test referido al criterio
Tipo de aplicación	Test de aplicación colectiva

Figura 8. *Características psicométricas de los instrumentos de medida creados.*

CAPÍTULO 4- CREACIÓN Y DESARROLLO DE LOS INSTRUMENTOS DE MEDIDA

1. La primera versión de los instrumentos

La idea de crear tres instrumentos de medida para medir la adquisición de la competencia traductora surgió en junio de 1996, cuando empezó a tomar forma el trabajo de investigación de la autora en el doctorado “Teoría de la Traducción”, (cf. Orozco, 1997). En aquel trabajo de investigación propusimos una metodología concreta para estudiar el proceso de adquisición de la competencia traductora basándonos en un experimento que combinaba varias pruebas, y realizamos una propuesta de operacionalización de la competencia traductora, eligiendo los tres elementos que hemos tenido en cuenta (concepción de la traducción, errores y problemas de traducción). Sin embargo, en las conclusiones de dicho trabajo de investigación quedó claro que el estudio propuesto no se podía llevar a cabo todavía porque no contábamos con instrumentos de medida de la adquisición de la competencia traductora, de manera que esa fue la tarea que nos propusimos para la presente tesis doctoral.

Por consiguiente, creamos tres instrumentos de medida que hemos ido perfeccionando a raíz de las diferentes pruebas piloto que se han realizado durante los últimos tres años. El proceso ha culminado con las pruebas de validez y fiabilidad de los instrumentos, que expondremos más adelante (cf. *infra 5.2 Diseño para el estudio de la generalizabilidad y 5.3. Diseño para el estudio de la validez*). Cada uno de los instrumentos ha sufrido su propia evolución, y en ocasiones se han realizado pruebas piloto solamente de uno o dos instrumentos, según la idiosincrasia de cada instrumento.

La figura 10 presenta las diferentes pruebas, ordenadas cronológicamente. En la primera columna se especifica qué instrumentos fueron los que se probaron en esa ocasión (resaltados en negrita) y en qué página del Volumen II se pueden consultar (entre paréntesis), además del número de versión (v.) y el idioma, ya que a cada lengua le corresponde un texto diferente, con los cambios pertinentes en el instrumento de medida de la actuación frente a los problemas de traducción (IPT); en lo sucesivo nos referiremos a cada instrumento por sus siglas: instrumento de medida de los errores de traducción (IET) e instrumento de medida de la concepción general de la traducción (ICGT). Las versiones definitivas se señalan con la abreviatura ‘v. def.’; en la segunda

columna consta la fecha en que tuvieron lugar y se asigna un número a cada prueba piloto (resaltado en negrita), para facilitar la referencia posterior, ya que al hablar de la evolución de cada instrumento en los apartados siguientes sólo nos referiremos a las pruebas en las que han intervenido; la tercera columna informa sobre el centro donde se realizaron las pruebas, Universitat Autònoma de Barcelona (UAB), Universitat de Vic (UV) o Universitat Jaume I (UJI), así como la cantidad de estudiantes que participaron y el idioma del que tradujeron, inglés, francés o alemán; por último, la cuarta columna especifica los cambios que tuvieron lugar en cada instrumento.

Instrumento/s probado/s, versión, idioma y p. del Vol. II donde se pueden consultar	Fecha / Número asignado	Centro y cantidad de estudiantes	Cambios efectuados en los instrumentos
IPT 1 ^a v. inglés (Vol. II, p.55) IET 1 ^a v. (Vol. II, p.119) ICGT 1 ^a v. inglés (Vol. II, p.31)	Octubre 1996 /1	UAB , 25 inglés	Creación de los tres instrumentos Texto : A.1. (Vol. II, p.9)
IPT 2 ^a v. inglés (Vol. II, p.56) ICGT 2 ^a v. inglés (Vol. II, p.35)	Octubre 1997 /2	UAB , 27 inglés	IPT : Eliminación de la entrevista, cambio del sistema de puntuación: se establecen cuatro problemas representativos. ICGT : redistribución de bloques conceptuales y adición de preguntas.
IPT , 3 ^a v., inglés (Vol. II, p.61)	Febrero 1998 /3	UAB , 33 inglés	Texto : A.3. (Vol. II, p.11) IPT : cambios de formato
IPT , 4 ^a v. inglés A (Vol. II, p.66 e inglés B (Vol. II, p.69); IPT , 4 ^a v. francés A (Vol. II p.72) y francés B (Vol. II, p.75); IPT , 4 ^a v. alemán A (Vol. II p.78) y alemán B (Vol. II, p.81).	Abril 1998 /4	UAB , 28 inglés, 14 francés 13 alemán	Textos utilizados : Inglés A: A.6. (Vol. II, p.14) Inglés B: A.4. (Vol. II, p.12) Francés A: B.2. (Vol. II, p.18) Francés B: B.4. (Vol. II, p.20) Alemán A: C.2. (Vol. II, p.24) Alemán B: C.4. (Vol. II, p.26) IPT : versión reducida, sólo preguntas con información cuantitativa.
IPT , 5 ^a v. inglés (Vol. II, p.84)	Mayo 1998 /5	UAB , 19 inglés	Texto : A.1. (Vol. II, p.9) IPT : cambio de formulación de algunas preguntas
Pruebas de validez y fiabilidad: IPT v. def. inglés (Vol. II, p.88) francés (Vol. II, p.92) alemán (Vol. II, p.96); IET v. def. (Vol. II, p.121); ICGT v. def. (Vol. II, p.42).	Octubre 1998 /6	UAB , 68 inglés 28 francés 21 alemán 34 ingl. (UV) 73 ingl. (UJI)	Textos : inglés (A.2., Vol. II, p.10) francés (B.4., Vol. II, p.20) alemán (C.2., Vol. II, p.24). IPT : eliminación de algunas preguntas, adición de otras y cambios de formulación. ICGT : Eliminación de preguntas. IET : reducción de las categorías, se tienen en cuenta sólo tres (FS, SS y aciertos).
Pruebas de validez y fiabilidad: IPT e ICGT , v. def. inglés (Vol. II, p. 125), francés (Vol. II, p.134) y alemán (Vol. II, p. 143); IET (Vol. II, p.121);	Mayo 1999 /7	UAB , 45 inglés 18 francés 8 alemán 27 ingl. (UV) 49 ingl. (UJI)	IPT : cambios en la formulación de algunas instrucciones y ligeros cambios de formato.

Figura 9. *Pruebas piloto realizadas con cada uno de los instrumentos de medida.*

Los alumnos a los que se hace referencia en la figura estaban cursando primero de la carrera de Traducción, en todos los casos, tanto en la UAB como en la UV y en la UJI. Las pruebas se realizaron siempre con la totalidad del grupo de clase, cuyo número varía porque, además de las diferencias de un año a otro, hay que tener en cuenta que solamente realizaron las pruebas piloto los alumnos que estaban ese día en concreto en clase, ya que en ningún caso se comunicó a los estudiantes que en una fecha concreta se iban a realizar las pruebas. También en todos los casos tanto los profesores como los estudiantes dieron su consentimiento para realizar las pruebas y para que los datos obtenidos fueran utilizados en un estudio, siempre y cuando sus datos fueran confidenciales, requisito que hemos respetado en todo momento.

1.1. Muestras de las pruebas piloto

En este apartado nos limitamos a hablar de las muestras, y al referirnos a los instrumentos y textos utilizados remitimos a las páginas correspondientes del volumen II donde se pueden encontrar; respecto a los cambios realizados en cada instrumento tras cada una de las pruebas piloto, se detallan más adelante (cf. *infra 4.3. Desarrollo del instrumento de medida de la actuación frente a los problemas de traducción 'IPT'*, *4.4. Desarrollo del instrumento de medida de errores de traducción 'IET'* y *4.5. Desarrollo del instrumento de medida de la concepción de la traducción 'ICGT'*).

La primera prueba piloto se realizó en octubre de 1996, con los instrumentos originales (cf. volumen II, páginas 31-34, 55 y 119-120, para ver el ICGT, el IPT y el IET, respectivamente). La muestra estaba constituida por 30 alumnos, que formaban uno de los dos grupos de traducción del inglés al español que hay cada año en el primer curso de la FTI de la UAB. El día que se llevaron a cabo las pruebas, se pidió la colaboración de los estudiantes, informándoles de que se iba a llevar a cabo una investigación con los datos de las pruebas que se iban a realizar en la clase ese día y se les informó asimismo de que no era obligatorio participar en la investigación. Dado que estas pruebas iban a ocupar dos clases enteras, más una entrevista de diez minutos que se debía concertar, se comunicaron los horarios a los estudiantes y se les recordó que, en caso de decidir colaborar, debían asistir a las dos sesiones y a la entrevista.

Una vez descartados los sujetos que tenían más de veinte años, los estudiantes de intercambio extranjeros, y los sujetos que no podían participar en el ensayo piloto por incompatibilidad de horarios, ya que el resto de alumnos decidieron tomar parte en las pruebas, quedaron 25 alumnos, de los cuales 21 acudieron a las tres sesiones.

La segunda prueba piloto se realizó justo un año después, en octubre de 1997, con algunas modificaciones en dos de los instrumentos (cf. volumen II, páginas 56-60 y 35-38 para ver el IPT y el ICGT, respectivamente) y la muestra fue elegida de la misma manera que en la ocasión anterior, entre los estudiantes de primer curso, con la salvedad de que esta vez se realizaron todas las pruebas en dos sesiones y participaron un total de 27 estudiantes, todos diferentes a los que habían tomado parte en las pruebas de octubre de 1996 (que para entonces ya estaban en segundo curso).

En febrero de 1998 se realizó la tercera prueba piloto, con el mismo grupo que había participado en octubre de 1997 y con pequeñas modificaciones en el IPT, ya que esta vez debían traducir otro texto (cf. volumen II, páginas 61 y 11 para ver el IPT y el texto, respectivamente). Participaron 33 estudiantes, de cuyos datos solamente fueron válidos 27, los mismos que habían realizado la prueba piloto número dos.

La prueba piloto número cuatro se realizó con una muestra más amplia, ya que se probaron por primera vez diferentes textos en los tres idiomas (alemán, francés e inglés). Esta prueba piloto corresponde al diseño creado para probar si la dificultad de los dos textos elegidos en cada lengua eran equivalentes respecto al grado de dificultad, que describimos con detalle más adelante (cf. infra 4.2. *La selección de textos para traducir*). La muestra en esta ocasión estuvo formada por la totalidad de estudiantes de alemán de primer curso de la FTI de la UAB, es decir, tanto los que traducen al castellano como los que traducen al catalán (13 en total, de cuyos datos solamente 11 fueron válidos), un grupo de estudiantes de traducción del francés al español (14 en total, de cuyos datos solamente 12 fueron válidos finalmente) y dos grupos de traducción del inglés, uno al español y otro al catalán (28 en total, de cuyos datos solamente 22 fueron válidos). El grupo de estudiantes de traducción del inglés al español que participó en esta prueba fue el grupo que no había participado en las pruebas piloto número dos y tres, puesto que el primer grupo ya conocía los textos que se debían traducir, y, como explicaremos más adelante, queríamos que fuera el primer contacto de los estudiantes con los dos textos, ya que se trataba justamente de probar si los dos textos elegidos tenían el mismo nivel de dificultad.

En mayo de 1998 (prueba piloto número cinco) se repitieron las pruebas, con algunas modificaciones en el IPT (cf. volumen II, página 84), y de los 27 estudiantes que habían realizado las pruebas en octubre de 1997 y en febrero de 1998 solamente repitieron 19, es decir, un setenta por ciento del total. Esta cantidad de “repeticiones” del setenta por ciento volverá a aparecer cuando analicemos los resultados de las

pruebas de validez y fiabilidad (pruebas 6 y 7), y no deja de ser un hecho curioso, ya que se trata de un dato frecuente en los procesos poco controlados, como la asistencia a clase o los experimentos de campo, en los que se pronostica un treinta por ciento de “pérdida” de sujetos en una prueba que se repite un tiempo después con la misma muestra, por razones diversas.

Por último, las muestras de las pruebas número seis y siete, que corresponden a las pruebas de validez y fiabilidad de los tres instrumentos de medida se analizarán con detalle más adelante (cf. *infra 5.5. Sujetos*). De momento baste decir que la muestra total de la primera prueba de fiabilidad y validez, llevada a cabo en octubre de 1998, fue de 224 estudiantes, procedentes de la UAB, la UV y la UJI, y la muestra de la segunda prueba de fiabilidad y validez fue de 147 sujetos, pertenecientes a las facultades de Traducción de las mismas universidades.

2. La selección de textos para traducir

Un elemento determinante tanto para el instrumento de medida de la actuación frente a los problemas de traducción (IPT) como para el instrumento de medida de los errores de traducción (IET) es el texto que deben traducir los estudiantes, ya que toda la evaluación posterior gira en torno al texto traducido. Por lo tanto, desde el primer momento se planteó la cuestión de la elección del texto o, mejor dicho, de los textos, ya que, por una parte, queríamos realizar pruebas en tres idiomas (inglés, francés y alemán) y por otra parte nos pareció que debíamos encontrar dos textos, uno para la primera medición y otro para la segunda medición, siempre y cuando estos dos textos tuvieran el mismo nivel de dificultad para los estudiantes.

Cuando se realiza un diseño experimental que consta de dos medidas, es un requisito psicométrico esencial que en las dos ocasiones se apliquen instrumentos equivalentes, de manera que se puedan comparar los resultados y atribuir los cambios en las dos medidas a factores como el paso del tiempo, la intervención del investigador, etc., pero en ningún caso a la inexactitud o al cambio propiciado por los instrumentos. Dado que nuestro proyecto experimental futuro (cf. infra 7.1. *Diseño de la investigación*) se basa en la repetición de las medidas, es necesario cumplir este requisito psicométrico, y por consiguiente el diseño de las pruebas de validez y fiabilidad responde asimismo a la repetición de medidas, antes y después de la intervención.

La selección de textos de dificultad equivalente es actualmente un problema en nuestra disciplina, porque no contamos con estudios normativos que permitan asegurar dicha equivalencia. De hecho, nos hallamos ante una de las dificultades más generalizadas en las investigaciones traductológicas, ya que la inmensa mayoría de estudios empíricos sobre la traducción se basan en la traducción de un texto. Hasta la fecha cada investigador ha resuelto este problema de diferente forma, aunque una parte importante recurre a su experiencia, como traductor o como docente, para elegir el texto o los textos en cuestión, como señala Neunzig, que se refiere también a las soluciones más frecuentes:

“La selección de textos para el uso experimental suele realizarse desde un punto de vista pragmático: se buscan textos partiendo de la propia experiencia y que evidencian el problema en cuestión [sobre el que se quiere investigar] (textos utilizados en clase en los que los alumnos muestran ciertas dificultades o que obligan a realizar ciertas operaciones, p.e. que obligan a ser ‘creativos’, como en el estudio descrito de Kussmaul) o se manipulan textos para que resulten

‘típicos’ para ciertos problemas de traducción, o bien se determinan *a posteriori* los puntos interesantes a analizar (donde los probandos cometan errores). Pero esto no resuelve una dificultad esencial de experimentos de seguimiento (los llamados *repeated measurement*) en Traductología: necesitamos textos ‘iguales’ para las sucesivas mediciones a lo largo del tiempo o para comparar diferentes idiomas. La solución de repetir siempre el mismo texto o de buscar a expertos que validen textos –que han sido los procedimientos más comunes- acarrea obviamente problemas: la correlación entre expertos ha resultado muy baja siempre que se ha medido y, en una repetición, puede influir la memoria (...)" (Neunzig, 1999b:30-31)

De hecho, en nuestro caso, la primera idea que tuvimos fue la de seleccionar los textos mediante el consenso entre expertos mencionado por Neunzig, de manera que elegimos diez textos en inglés, que habían sido utilizados en exámenes de primer curso en la Facultat de Traducció i d'Interpretació de la UAB en los dos años anteriores (entre primeras y segundas convocatorias de los dos semestres en que se divide la asignatura de traducción directa general B) y los enviamos a diez expertos, que eran profesores de traducción en la Universidad y traductores al mismo tiempo. Les pedimos que eligieran los dos textos que más se parecían en su opinión respecto al nivel de dificultad que presentaban para los alumnos de primer curso. El resultado de esta operación fue que ni siquiera dos de los diez expertos coincidieron en los textos elegidos, es decir que se cumplió claramente la generalización de Neunzig respecto al consenso de expertos.

Una de las conclusiones que sacamos de esta experiencia es que es muy difícil conseguir que haya acuerdo entre varios expertos si se realiza una pregunta abierta, ya que cada experto tiene sus propios criterios, y el abanico de posibles respuestas es muy amplio. Sin embargo, pensamos que si realizábamos una pregunta directa, indagando sobre un texto ya elegido, quizás era más fácil obtener el acuerdo, de manera que cambiamos de estrategia y esta vez elegimos nosotros mismos dos textos que *a priori* reunían las mismas dificultades en inglés, francés y alemán. La idea era, en primer lugar, conseguir el consenso de expertos para cada par de textos en cada una de las tres lenguas, y después realizar prueba para ver si efectivamente el nivel de dificultad era el mismo.

El procedimiento que seguimos para elegir los textos fue el siguiente: pedimos a tres profesores de traducción directa de primer curso de la FTI que eligieran dos textos que funcionaran en clase y que contuvieran problemas de traducción del nivel que se suele trabajar en primer curso de la carrera. Debemos mencionar que a nuestro parecer el nivel de dificultad de un texto se deriva de los problemas de traducción que presenta,

si bien dentro de esta afirmación general habría que realizar algunas puntualizaciones.

Una de las autoras que ha definido criterios para obtener el perfil de dificultad de un texto para compararlo con otro texto propone una serie de parámetros como el tema tratado, el registro, la dimensión tiempo-lugar, el receptor y el uso (Reiss, 1982). Neunzig y Solivellas (1984) añaden a estos valores la intención del autor, la informatividad, la densidad de exposición, el grado de elaboración lingüística, el léxico, las convenciones lingüísticas en la traducción entre el par de lenguas implicadas, el proceso de traducción global, las unidades de traducción y la presencia de interferencias referentes a cuestiones contrastivas entre las lenguas. Podríamos citar otros parámetros, ya que existen muchos estudios sobre análisis textual, pero se trata de estudios que no ofrecen una solución práctica a nuestro problema, ya que el nivel de dificultad de un mismo texto puede ser, y de hecho suele ser, diferente para dos traductores (debido, por ejemplo, a que uno de ellos está familiarizado con el tema que trata el texto, o conoce a fondo la tipología textual en cuestión y el otro traductor no).

El hecho de que la dificultad de un texto sea diferente para cada traductor, según sus características personales, hace que describir perfectamente las características de dos textos previamente no asegure su equivalencia respecto al grado de dificultad. Desde luego, se podría pensar en analizar también a los traductores, pasándoles cuestionarios relacionados con su experiencia anterior, su conocimiento del tema que trata el texto que van a traducir, etc., pero en nuestro caso, tratando con grupos enteros de estudiantes que no deben percibir en lo posible que son sujetos de un estudio, de manera que no modifiquen su conducta habitual, es impensable.

Por consiguiente, decidimos analizar los textos en cuestión para que todos tuvieran un nivel equivalente *a priori*, sin profundizar excesivamente en este análisis, y prestamos especial atención a los problemas que presentaba cada texto para encontrar niveles de dificultad equivalentes.

Cuando diseñamos la prueba en la queríamos ver si para los estudiantes los dos textos elegidos eran equivalentes entre sí respecto al nivel de dificultad, advertimos que debíamos “recortar” los textos elegidos, ya que los originales eran demasiado largos para que los estudiantes pudieran traducir dos textos en dos horas, que era el tiempo máximo que podían prolongarse las pruebas si pretendíamos reunir en una misma aula a estudiantes con tres combinaciones lingüísticas diferentes sin alterar su horario de clases. Por consiguiente, los textos originales debían tener un total de entre 100 y 125 palabras, que permitía que cada estudiante tradujera dos textos en dos horas.

Los textos originales elegidos fueron seis, dos para cada combinación lingüística, y se pueden consultar en su versión completa en el volumen II, en las siguientes páginas: textos en inglés, A.3. (p.11) y A.5. (p.13), textos en francés, B.1. (p.17) y B.3. (p.19) y textos en alemán, C.1. (p.23) y C.3. (p.25). Estos seis textos tienen varias características comunes: son de temática general, de registro medio, de tipologías textuales conocidas por los estudiantes (textos periodísticos y comerciales) con un estilo sencillo y directo. Por otra parte, todos estos textos obligan al estudiante a movilizar conocimientos enciclopédicos y culturales para traducirlos de manera aceptable y muestran un léxico adecuado al nivel de estudiantes de traducción de primer curso, es decir que si bien es probable que tengan que buscar algunas palabras o expresiones en un diccionario, no deberían tener grandes problemas para comprender el texto de manera global. Un último elemento común a los seis textos es que todos contienen problemas de varios tipos. Sin embargo, en la mayoría de casos no existía un problema pragmático proporcionado directamente por el texto, de manera que lo introdujimos mediante el encargo de traducción.

A pesar de la gran controversia que existe en nuestra disciplina respecto a la manipulación de textos con fines pedagógicos, nosotros consideramos que en este caso era absolutamente necesario manipular los textos, dado que teníamos unas restricciones de tiempo nada despreciables y que la traducción de estos textos no iba a repercutir en ninguna actividad pedagógica, solamente se trataba de recoger datos que nos interesaban para una investigación.

La técnica utilizada para “modificar” los textos originales inglés, francés y alemán y reducirlos fue bastante simple: en primer lugar, se identificaron los cuatro problemas en cada texto, uno de tipo lingüístico, otro extralingüístico, otro de transferencia y finalmente otro pragmático y a continuación se procedió a cortar todo aquello que no era necesario para conferir coherencia y cohesión al texto. Es decir que la condición para recortar los textos fue que la versión reducida continuara siendo un texto coherente y mostrara los elementos de cohesión necesarios. Sería tedioso explicar con detalle las partes de los textos que se recortaron y en qué consisten los problemas de cada uno de los seis textos, de manera que nos limitamos a incluir los seis textos, en su versión reducida, donde hemos subrayado los cuatro problemas identificados. Junto a cada segmento problemático hemos incluido, entre paréntesis, una letra que identifica el tipo de problema: (L) es el problema lingüístico, (E) el problema extralingüístico, (T) el problema de transferencia y (P) el problema pragmático. Para ver los cuestionarios

correspondientes a cada texto reducido, se puede consultar el volumen II (páginas 66 y 69 para los IPT de los textos en inglés, páginas 72 y 75 para los de francés y páginas 78 y 81 para los de alemán).

Tanto para elegir los textos como para “modificarlos” y para identificar los problemas que consideramos en cada texto contamos con la ayuda de varios expertos en traducción de cada lengua (inglés, francés y alemán). Consultamos al menos con dos expertos en cada lengua por separado, hasta asegurarnos de que sus opiniones coincidían tanto respecto al nivel de dificultad de los dos textos de su combinación lingüística como respecto a los problemas que identificamos en cada uno de los textos.

Texto inglés A:

Encargo de traducción: Una compañía española ha decidido abrir una empresa similar a “Safeway”, con sede en Madrid y te pide que traduzcas el siguiente anuncio, extraído de *The Economist*, de manera que tu traducción pueda ser publicada como un anuncio en el periódico *El País*.

CRÈCHE SPECIALISTS

Exciting opportunities to be part of a fast growing operation

Competitive salary + car + benefits

Safeway is not only one of Great Britain's (P) largest retailers (L), it is also one of the most innovative. We have helped pioneer the concept of in-store crèche facilities (T) and are now fully committed to expanding this popular service by introducing crèche facilities to an increasing number of our stores.

We are looking to recruit two Crèche Specialists to be responsible for a number of crèches in designated geographical areas. You will be expected to display a high level of creativity and will be actively encouraged to use your initiative.

You must have at least five years' management experience gained within a group care environment catering for young children. A child care or social work qualification is essential as you will liaise closely with the Department of Health and Social Security (E).

Texto inglés B:

Encargo de traducción: *La Vanguardia* te pide que traduzcas al español el siguiente artículo, procedente de la revista *Time*, del 1 de abril de 1996, con el fin de publicar la traducción en la revista *Magazine* que acompaña al periódico los domingos.

According to folklore, that merriest of monarchs, Charles II, was once so delighted by a succulent roast that he drew his sword and dubbed the joint “Sir Loin” (P). The story is a fable, of course, but it perfectly illustrates England's longstanding love affair with the cow. John Bull, the well-fed personification of Empire, looks as if he is just toddling off (L) from a pint at the pub to his midday Sunday meal of roast beef, Yorkshire pudding and plum pie.

Roast beef may now be killing. For the first time, the British government openly acknowledged last week a plausible link between “mad-cow disease” and a new, unexplained form of fatal brain degeneration in people. After a decade of denials, the admission was like a gunshot at a dinner party (E). It set off a consumer panic, crisis meetings in the European Union and overnight bans (T) on British beef around the world.

Texto francés A:

Encargo de traducción: El suplemento dominical de *La Vanguardia* te pide que traduzcas el siguiente artículo que apareció en *Le Monde*.

ELITISME SCIENTIFIQUE

Les causes de la stagnation, voire dans certains lycées (P), de la régression des classes scientifiques (T), qui est un désastre national, mériteraient un enquête plus sérieuse. J'en vois la raison principale dans le fait que, triés vers le haut par les exigences de l'enseignement supérieur et des classes préparatoires (E), les programmes scientifiques sont devenus trop denses, trop lourds à assimiler, écrasants pour la très grande majorité des élèves, même pour ceux qui arrivent en classe de deuxième avec un bagage satisfaisant. Il est douteux que des horaires plus corsés permettraient (L) de redresser véritablement la situation.

Texto francés B:

Encargo de traducción: El suplemento dominical de *El País* te pide que traduzcas el siguiente artículo que apareció en *Le Monde*.

SECOURS D'URGENCE

A partir du 16 mai, les animateurs (L) d'Antenne 2 seront sur le pied de guerre pour livrer une nouvelle bataille humanitaire, avec "Passeport pour la vie", une opération lancée par la deuxième chaîne (E) et l'Association des paralysés de France. Celle-ci a pour but de répandre l'usage d'une petite carte de santé d'urgence, format de poche, contenant des renseignements sur les traitements médicaux en cours, les vaccins, le groupe sanguin et l'adresse du médecin traitant du détenteur (T). Action de collecte de fonds certes, mais avant tout action de prévention, cette campagne d'intérêt général bénéficie du haut patronnage du président de la République, du soutien du cabinet du premier ministre, de six ministères, du Sénat et du Palais Bourbon (P). Fin de cette opération: le 17 juin.

Texto alemán A :

Encargo de traducción: traducir el texto como folleto informativo para turistas españoles

FERIENQUARTIER VON DER MITWOHNZENTRALE (E)

Hanne Kotmann gründete 1987 die erste "Mit-Wohn-Börse" in München. Inzwischen (L) gibt es 45 solcher Mitwohnzentralen in Deutschland. Sie vermitteln Unterkünfte an jeden, der vorübergehend ein Zimmer, eine Wohnung oder ein Haus sucht. Die Anbieter sind überwiegend Privatleute, die ihre Wohnung oder nur ein möbliertes Zimmer für eine begrenzte Zeit untervermieten wollen. Die Mietdauer kann von drei Tagen bis zu mehreren Monaten reichen. Den Preis für das Quartier setzt der Vermieter fest.

Zu den Kunden zählen längst nicht mehr nur Studenten, sondern ebenso Geschäftsleute. Auch für Urlauber (T) eignen sich die Mitwohnzentralen als Makler für eine günstige Unterkunft bei Stadtreisen. In München kostet ein Zimmer beispielsweise 50 Mark (P) pro Nacht. In Berlin werden für ein Doppelzimmer 70 bis 80 Mark Miete berechnet, in Hamburg beginnen die Preise bei 60 Mark.

Texto alemán B.

Encargo de traducción: traducir el texto para un folleto publicitario de una compañía ferroviaria española.

FAMILIEN REISEN

Wie wäre es, wenn Sie Adventszeit (E) zum Familienbesuch nutzen?

Mit den liebsten Menschen unterwegs (L) sein, Zeit füreinander haben, spielen, lachen, Eindrücke sammeln. Für Eltern und Kinder sind das kostbare Tage und Wochen. Reisen macht Freude und kann mit Kindern ganz einfach und entspannend sein. Wenn Sie mit der Bahn fahren (T).

Kinder brauchen Bewegung und Abwechslung, auch auf Reisen – mit dem Zug kein Problem. Kinder wollen die Aufmerksamkeit der Eltern – die können sich im Zug ganz auf den Nachwuchs konzentrieren. Familien wollen preiswert reisen – das können sie mit der Bahn.

Sie sehen, die Bahn ist das ideale Verkehrsmittel für Familien. Dafür, daß auch drum herum alles stimmt und Sie sich, von der Reiseplanung an über den Gepäcktransport bis hin zum günstigen Ticket in der Ferienregion auf uns verlassen können, haben wir gesorgt.

Gute Reise mit der Bundesbahn! (P)

2.1. La prueba de equivalencia de textos

Para probar el nivel de dificultad de los estudiantes con cada uno de los textos elegidos, dos en inglés, dos en francés y dos en alemán, decidimos emplear nuestro instrumento de medida de la actuación frente a los problemas de traducción, del cual hicimos una versión “reducida” para estas pruebas, ya que en esta ocasión no nos interesaban los datos cualitativos que recoge el IPT, sino los cuantitativos, es decir, la puntuación de 0, 1 o 2 según el estudiante no hubiera detectado el problema, lo hubiera detectado pero no resuelto o lo hubiera resuelto respectivamente. En realidad tuvimos que diseñar seis versiones del IPT, una para cada texto probado, ya que los problemas de cada texto eran diferentes; con el fin de poder referirnos con claridad a cada texto, llamaremos a los dos textos de cada lengua A y B, tal como consta en las dos páginas precedentes.

El diseño que ideamos para probar si la dificultad de los textos era la misma fue el siguiente: se dividieron las muestras de alumnos (había tres muestras, una por cada combinación lingüística, inglés-español, francés-español y alemán-español) en dos grupos al azar, de manera que existieran seis grupos, y se pidió a una mitad de cada combinación lingüística que tradujera en primer lugar el texto A (y contestara al cuestionario del IPT) y en segundo lugar el texto B (y contestara después el IPT correspondiente). La otra mitad de alumnos de cada combinación lingüística debía traducir primero el texto B (con su correspondiente cuestionario IPT) y después el texto A. De esta manera, si los resultados en el IPT de los dos grupos de alumnos de cada

combinación lingüística eran equivalentes para los dos textos, independientemente del orden en que los hubieran traducido, quedaría demostrado que el nivel de dificultad de los dos textos en cuestión era el mismo.

Las pruebas para los tres idiomas se realizaron en abril de 1998, al mismo tiempo y en una misma aula, en presencia de varios profesores para asegurar que los estudiantes no intercambiaran opiniones. Los estudiantes tuvieron acceso a diccionarios y obras de consulta (gramáticas, guías de estilo, etc.) que fueron provistos por los profesores, es decir que todos tuvieron acceso a las mismas obras, que debían consultar en la mesa del profesor. Se controló el tiempo, de forma que cada estudiante tuvo 60 minutos para traducir cada texto y llenar el correspondiente cuestionario del IPT, y en total tardaron dos horas en traducir los dos textos y llenar los dos cuestionarios. Del total de la muestra de 55 estudiantes, se pudieron utilizar los datos de 22 pruebas de inglés, 12 de francés y 11 de alemán (el resto fueron descartadas por quedar incompletas o porque los sujetos que las habían realizado tenían como lengua materna el inglés, francés o alemán).

La manera de corregir estas pruebas fue la siguiente: se pidió a un evaluador por cada combinación lingüística que llenara una tabla en la que debía valorar la actuación de cada estudiante en relación a cada uno de los cuatro problemas del texto, otorgando 0, 1 o 2 puntos en cada ocasión. Con este fin debía consultar, en primer lugar, la traducción del alumno, de manera que, si había resuelto el problema, obtenía directamente un “2”, mientras que si no lo había resuelto, entonces el evaluador miraba la respuesta de ese alumno a las preguntas 1, 2 y 4 del cuestionario (cf. volumen 2, p. 66-68) para decidir entre un “0”, en caso de que no hubiera detectado el problema, y un “1” si lo había detectado. Al comparar los puntos obtenidos por cada sujeto en cada texto se puede ver la diferencia en la dificultad entre los dos textos, ya que a más puntos, más fácil (porque el estudiante detectó y/o resolvió más problemas) y a menos puntos, más difícil (porque el estudiante detectó y/o resolvió menos problemas). Los resultados de estas pruebas mostraron que en ningún caso se dio una equivalencia total entre los dos textos. A continuación mostramos tres gráficos, uno por cada combinación lingüística, donde se pueden observar los resultados.

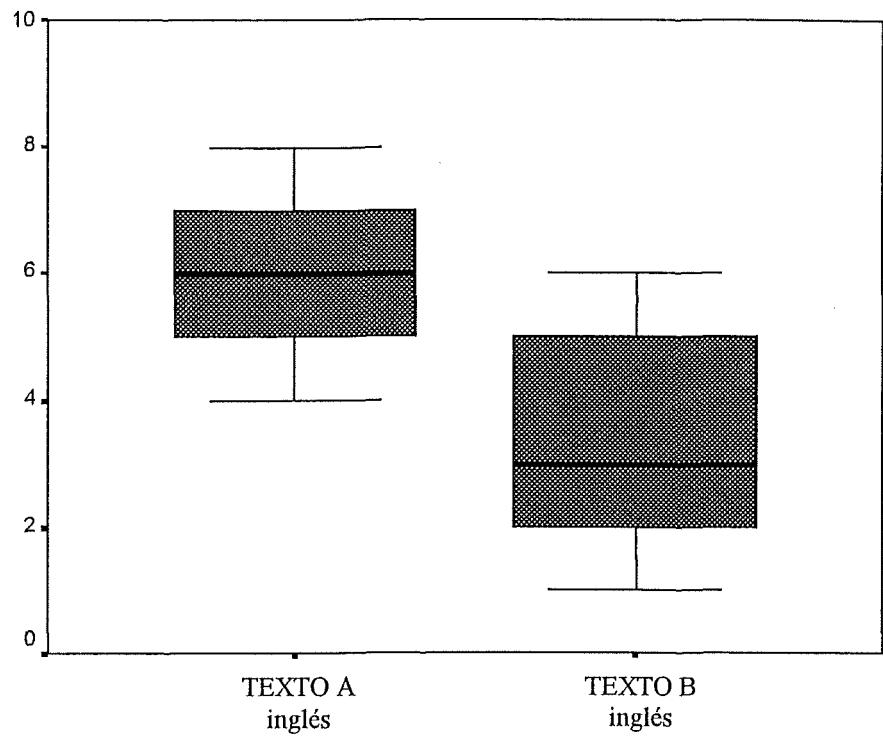


Gráfico 1. *Diagrama de caja de los resultados globales de las pruebas de equivalencia de textos en inglés.*

Para elaborar el gráfico 1 se han sumado los puntos obtenidos por cada estudiante en los cuatro problemas, es decir que, por ejemplo, el estudiante de inglés 1, que ha obtenido para el texto A un dos en el problema lingüístico, un dos en el problema de transferencia, un dos en el problema extralingüístico y un uno en el problema pragmático, tiene un total de siete puntos en el texto A.

De esta manera, el eje vertical muestra las cantidades totales de puntos obtenidas por todo el grupo, que están representadas por sendos **diagramas de caja**. Para obtener un diagrama de caja se divide el conjunto de datos, ordenados, en cien partes iguales. Cada uno de esos 100 valores se denominan percentiles. La **mediana** corresponde al valor central de los datos, que se corresponde con el percentil 50, y en el gráfico está representada por la línea gruesa horizontal que parte en dos el cuadrado o caja gris. Las líneas que marcan el límite superior e inferior de la caja, corresponden al percentil 25 y al percentil 75; la diferencia entre ambos es la **amplitud intercuartil**, correspondiente a los resultados obtenidos por la mitad central de la muestra. Por último, los bastones que sobresalen de la caja muestran las puntuaciones más alta y más baja obtenidas por individuos del grupo (percentiles 100 y 0 respectivamente). Por consiguiente, para el texto A en inglés, el estudiante que peor resultado ha obtenido ha sumado 4 puntos, y el

que mejor resultado, 8 puntos, mientras que en el texto B la peor puntuación obtenida ha sido 1 punto y la mejor 6 puntos. Los datos más útiles para comprobar la equivalencia de los textos son las medianas, que muestran el punto exacto que divide los resultados por la mitad, es decir que el 50% de la muestra está por encima y el 50% de la muestra está por debajo de esa puntuación, y las amplitudes intercuartiles. Comparando las medianas y las amplitudes intercuartiles, es decir, el espacio comprendido dentro de la caja de los textos A y B, se observa claramente que el texto A ha sido, en general, más fácil para los sujetos que el texto B.

Si analizamos los datos desde otro punto de vista, caso por caso, resulta que los dos textos en inglés solamente fueron exactamente equivalentes para un 14 % de los sujetos, mientras que la mayoría, el 72 %, han tenido más dificultades en el texto B que en el texto A. Entre este 72 %, un 9 % tiene un solo punto de diferencia entre los dos textos.

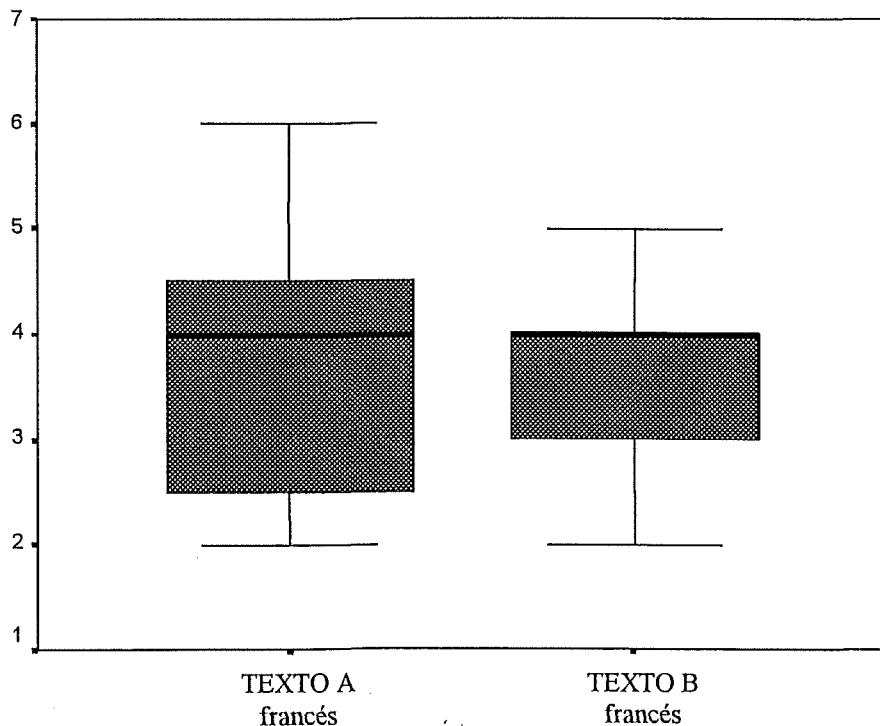


Gráfico 2. *Diagrama de caja de los resultados globales de las pruebas de equivalencia de textos en francés.*

Observando las medianas de los resultados obtenidos por los estudiantes en los dos textos en francés, se puede ver que su nivel de dificultad es mucho más parecido que los textos en inglés, pero aún así, las amplitudes intercuartiles nos muestran una

importante diferencia entre ellos. Observando los datos desde otra perspectiva, caso por caso, estos textos han resultado ser equivalentes respecto al nivel de dificultad para un 8 % de los sujetos, mientras que un 50 % de los sujetos para quienes el texto A es más difícil que el texto B, y un 42 % para quienes el texto B es más difícil que el A. Sin embargo, entre estos últimos (92 %), un 58 % solamente tiene un punto de diferencia entre los dos textos.

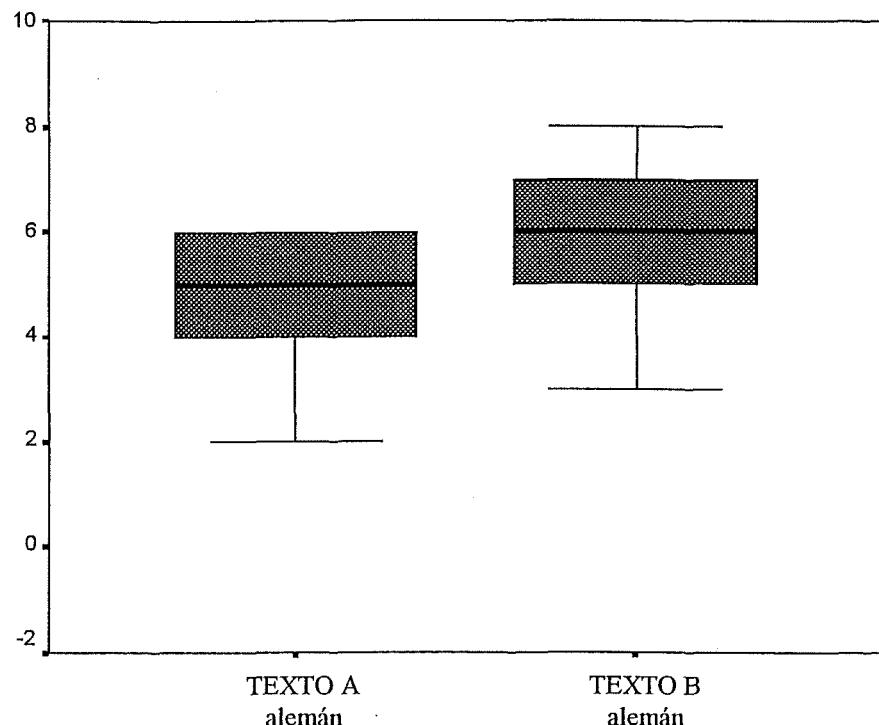


Gráfico 3. *Diagrama de caja de los resultados globales de las pruebas de equivalencia de textos en alemán.*

Los textos en alemán tampoco son equivalentes respecto al nivel de dificultad, como muestran las medianas y las amplitudes intercuartiles de los resultados obtenidos por los estudiantes en los problemas de estos dos textos. Los datos de análisis caso por caso nos indican que, en el caso del alemán, no hay ningún sujeto para quien los textos hayan tenido el mismo nivel de dificultad, un 18 % ha tenido más dificultades en el texto B que en el A y un 82 % ha encontrado más dificultades en el texto A que en el texto B. De este 82 %, un 27 % ha tenido un solo punto de diferencia entre los dos textos.

Dado que ha habido casos, en las tres combinaciones lingüísticas, en que el texto que globalmente ha resultado más difícil para el grupo en general (el texto B en inglés y los textos A en francés y en alemán) ha sido sin embargo más fácil para algunos sujetos,

tampoco se podría suponer que un texto es sistemáticamente más difícil que el otro, porque hay estudiantes para los que ocurre al contrario.

Estos resultados no dejan lugar a dudas sobre la complejidad de encontrar textos equivalentes respecto al nivel de dificultad, y a pesar de poder llegar a aproximarse, como en el caso de los textos en francés, sigue existiendo una diferencia, que para algunos sujetos es mayor que para otros. Por consiguiente, utilizar los dos textos elegidos para cada combinación lingüística hubiera significado arriesgarnos a no saber si los cambios producidos en la segunda medición se debían a la formación en traducción de los sujetos o a la diferencia marcada por un texto más sencillo o más difícil de traducir, con problemas más o menos fáciles de detectar y resolver.

La decisión que tomamos tras ver estos resultados fue la de descartar esta técnica de elección de textos y concluimos que hacía falta un estudio profundo sobre la nivelación de textos para traducir. Dado que este objetivo se apartaba bastante del que nos habíamos propuesto para la presente tesis doctoral, decidimos optar por otra solución que sí nos aseguraba que los dos textos tendrían el mismo nivel de dificultad para cualquier estudiante: utilizar el mismo texto para la primera y la segunda medición.

Somos conscientes de que esta decisión es polémica, ya que es evidente que un traductor que se enfrenta a un texto por segunda vez lo hace de manera diferente y utiliza menos tiempo para comprender el TO debido a los efectos de la memoria (cf. Neunzig, 1999b:31). Sin embargo, en este caso el efecto de la memoria queda paliado en gran parte porque, en primer lugar, no se trata de traductores profesionales sino de estudiantes, que no traducen sistemáticamente ni tienen un método de trabajo concreto que les haga recordar necesariamente una tarea realizada en el pasado; en segundo lugar, habrán transcurrido ocho meses entre la primera traducción y la segunda, sin que se haya vuelto a nombrar o leer el texto en cuestión; en tercer lugar, la primera medida se realiza el primer día de clase de los alumnos en la Universidad, un día ya de por sí bastante desconcertante; este último elemento ayuda a que los sujetos prácticamente olviden el texto traducido cuando, ocho meses después, se les pida que lo vuelvan a traducir. De hecho, la técnica de servirse de un mismo texto en dos momentos diferentes con la misma muestra fue utilizada por Séguinot (1991) con buenos resultados.

De cualquier forma, con el fin de asegurar que estos argumentos eran ciertos, decidimos realizar otra prueba piloto (número cinco) en mayo de 1998 y hacer traducir a la misma muestra que había realizado la prueba piloto número dos, el mismo texto que ya habían traducido en octubre de 1997, ocho meses antes, tras lo cual fuimos

preguntando a cada estudiante, en una breve entrevista, si pensaban que el hecho de haber traducido ese mismo texto a principio de curso les había influido en cualquier aspecto. La gran mayoría de estudiantes replicaron que no recordaban más que vagamente el texto, y aquellos que lo recordaban no pudieron decir si su manera de traducir el texto había cambiado en algún aspecto, ya que no recordaban cómo lo habían traducido.

En vista de estos resultados, decidimos utilizar este mismo texto para la versión definitiva del IPT en inglés (cf. volumen II, página 9), si bien finalmente recurrimos a la versión reducida del texto (cf. volumen II, página 10), para que la traducción se realizara en una hora como máximo y el total de las pruebas tuviera un límite de tiempo de dos horas.

Por obvio que parezca, no está de más observar que, para poder utilizar el mismo texto en las dos mediciones de manera eficaz, hay que tomar la precaución de asegurarse de que, tras la primera medición, los estudiantes no conserven el TO, ni su traducción ni ningún tipo de borrador que hayan podido hacer para realizar la traducción.

Finalmente, cabe destacar que los textos francés y alemán que se utilizaron en las pruebas 6 y 7, es decir, en las versiones definitivas del IPT, fueron las versiones que se habían utilizado en la prueba piloto cuatro, que por consiguiente también fue útil para probar los textos francés y alemán (cf. volumen II, páginas 20 y 24 respectivamente), y comprobar que el tiempo que necesitaban los estudiantes para traducir cada texto era de una hora como máximo.

3. Desarrollo del instrumento de medida de la actuación frente a los problemas de traducción (IPT)

En este apartado nos referimos al desarrollo del IPT, que cuenta con seis versiones diferentes. Para consultarlas, basta con buscar en el volumen II el número de página que se indica entre paréntesis bajo cada título de sección.

3.1. Primera versión del IPT

(cf. Vol. II, página 55)

La primera idea, al crear este instrumento, era la de poder observar cómo se enfrentaban los estudiantes a los problemas de traducción que encontraban al traducir un texto. Si bien en el presente trabajo defendemos nuestra postura respecto a la tipología de problemas de traducción elegida (cf. supra 3.4.2. *La definición de problema de traducción*), en la etapa inicial de este proyecto albergábamos muchas dudas respecto a qué enfoque adoptar, ya que teníamos un gran interés en reflejar la actuación de los estudiantes frente al mayor número de problemas posible, y asimismo pretendíamos abarcar el mayor abanico de tipologías de problemas posible.

Por lo tanto, para realizar el primer IPT buscamos un texto que nos diera el mayor juego posible en lo que a problemas se refiere, y elegimos el texto titulado “Would you like to be a writer?” (cf. volumen II, página 9), cuyas características se adaptaban a nuestros requisitos: tipología textual conocida por los estudiantes, temática general, estilo claro, conciso y directo y gran cantidad de posibles problemas de traducción de diferentes tipos. Para asegurarnos de que el texto funcionaba, en primer lugar realizamos una prueba exploratoria, para la cual pedimos a cinco personas recién licenciadas en Traducción e Interpretación que tradujeran el texto y nos comentaran los problemas que habían tenido al traducir, además de analizar la traducción que nos entregaron. Los sujetos encontraron una serie de problemas en el texto y tomamos nota de todos los segmentos o elementos del texto que habían resultado problemáticos.

Una vez valorados los problemas de los que nos habían informado los sujetos, decidimos añadir a esa lista todos los puntos que nosotros consideramos que podían ser problemáticos para los estudiantes de primer curso de la carrera, de manera que obtuvimos una extensa lista final de 70 problemas (cf. volumen II, páginas 111-113).

El siguiente paso fue clasificar los problemas, para ver si eran lo suficientemente variados. El sistema de clasificación de problemas que aplicamos fue el de Nord (1991),

es decir, que dividimos los problemas en tres grandes grupos: (a) pragmáticos, (b) culturales y (c) lingüísticos. A su vez, dividimos los problemas lingüísticos en seis apartados: (1) problemas de puntuación, (2) problemas léxicos, (3) problemas sintácticos y/o gramaticales, (4) problemas textuales, (5) problemas de registro y tono y (6) problemas de estilo.

El resultado de esta clasificación fue alentador, ya que en el texto había problemas de todas las categorías, si bien algunas incluían un mayor número de problemas que otras: 12 problemas pragmáticos, que en su mayoría tenían que ver con la tipología textual, 2 problemas culturales y 56 problemas lingüísticos. Para analizar detalladamente los problemas, se puede consultar el volumen II (páginas 111-113), ya que en la plantilla de corrección del IPT consta, en la primera columna, el fragmento “problemático” y, en la tercera columna, el código que identifica cada problema de acuerdo con la clasificación que hemos mencionado en el párrafo anterior, de manera que un “A1” es un problema pragmático, mientras que un “C3” es un problema lingüístico sintáctico y/o gramatical, y un “C6” es un problema lingüístico de estilo.

En la plantilla de corrección del IPT de esta primera versión se puede observar asimismo un espacio reservado para que el evaluador anote la traducción del sujeto y, en la última columna, el evaluador debe anotar la puntuación obtenida en ese problema en concreto. En caso de que la traducción del segmento en cuestión indicara que el estudiante había resuelto el problema, la puntuación debía ser “2”, pero en caso de que la traducción del segmento en cuestión no fuera aceptable, había que discernir entre un “0” (si el sujeto no había detectado el problema) o un “1” (en caso de que el sujeto hubiera detectado el problema), y para ello se diseñó una entrevista, que debía mantener el evaluador con cada estudiante por separado.

La entrevista con cada estudiante se introdujo de la siguiente manera: “a continuación te voy a hacer una serie de preguntas sobre cómo ha sido el proceso de traducción de algunos segmentos del texto que has traducido, y se trata de que tú me digas cómo ocurrieron las cosas, qué pensaste cuando leíste el segmento en el texto original y cómo llegaste hasta la traducción que escribiste finalmente”. Para que el estudiante pudiera contestar, se le entregaba el texto original en inglés y su traducción, además de los borradores que hubiera podido realizar, con el fin de ayudarle a recordar lo que había traducido y cómo lo había hecho.

Los segmentos sobre los que el evaluador llamaba la atención al sujeto eran únicamente los que había señalado previamente en la plantilla de corrección, es decir,

aquellos en que se debía discernir entre un “0” y un “1”. La pregunta que realizó el evaluador ante dichos segmentos fue la siguiente: “¿recuerdas cómo ha sido el proceso de traducción de este segmento en concreto?”

La traducción del texto y la entrevista posterior eran los elementos que proporcionaban la información cuantitativa del IPT, pero además existía un cuestionario que ofrecía información cualitativa sobre la actuación de los sujetos frente a los problemas de traducción, y en esta primera versión del IPT (cf. volumen II, página 55) el cuestionario constaba de diez preguntas, todas abiertas, que intentaban recabar información acerca del proceso de comprensión del TO y el análisis textual (ítems 1, 2 y 3), la adquisición de información por parte de los sujetos (ítems 4, 5 y 6), la percepción de los problemas y los puntos problemáticos del texto (ítems 7 y 9), el proceso de transferencia y, de forma indirecta, el uso de las estrategias empleadas para solucionar los problemas que habían encontrado (ítem 8). Por último, la pregunta 10 indagaba sobre la autopercepción de la competencia traductora de cada sujeto.

La prueba piloto de esta primera versión del IPT, que se realizó en octubre de 1996, mostró muchos puntos débiles en este instrumento. En primer lugar, no se trataba de un instrumento práctico, ni para los sujetos ni para los evaluadores, por las razones que exponemos a continuación:

- (a) Para aplicarlo hacían falta tres sesiones (una sesión previa para pasar el cuestionario que descartaba a los sujetos que no reunían las características de la muestra del estudio, que se puede ver en el volumen II, página 155, una segunda sesión de tres horas para traducir el texto y llenar el cuestionario del IPT y una tercera sesión de 10 minutos para la entrevista con el evaluador, cosa que automáticamente hacía que se perdieran sujetos) de hecho, de los 25 sujetos que participaron en las pruebas, solamente 20 asistieron a las tres sesiones, de manera que se perdió a un 20 % de los estudiantes por este motivo.
- (b) Los sujetos acababan muy cansados en la segunda sesión, ya que se trataba de tres horas seguidas de concentración, lo cual hacía que no prestaran tanta atención al cuestionario final como a la traducción, consecuencia nada deseable.
- (c) El evaluador debía llenar 70 casillas por cada sujeto y copiar la traducción de cada segmento, además de pensar si esa traducción era aceptable o no, lo cual llevaba mucho tiempo y, si tenemos en cuenta que para realizar la entrevista las plantillas de corrección ya debían estar llenadas, se trataba de un trabajo arduo e intensivo, ya que entre la segunda sesión (traducción del texto) y la entrevista debía pasar el

mínimo tiempo indispensable, de manera que los sujetos recordaran lo mejor posible el proceso de traducción.

En segundo lugar, los datos que ofreció la corrección del IPT no fueron los que esperábamos, ya que hubo muchos problemas que aparecían en la plantilla de corrección y que no habían sido problemáticos para la mayoría de estudiantes, e incluso hubo cuatro problemas que no fueron tales para ningún sujeto: el número 34 (“experts”, línea 29), el número 36 (“tutors”, línea 31), el número 42 (“from the start”, línea 33) y el número 69 (“beginners”, línea 62). Por otra parte, apareció un problema que no habíamos previsto: “First class” (línea 60).

En tercer lugar, los datos ofrecidos por el cuestionario posterior al texto eran pobres y muy difíciles de analizar, ya que las preguntas no estaban bien formuladas y, al ser abiertas, daban la opción al estudiante de contestar con un simple “sí”, “no sé” o “no entiendo”.

Tras analizar todos estos hechos llegamos a dos conclusiones: (a) teníamos que modificar el instrumento en dos sentidos: por un lado hacerlo más práctico, tanto para los estudiantes como para los evaluadores, y por otro lado cambiar el sistema de identificación de los problemas en el texto, ya que no podíamos pretender prever todos los problemas que iban a encontrar los estudiantes; y (b) debíamos aprender a elaborar cuestionarios, de manera que las preguntas estuvieran bien formuladas y nos dieran información útil. Con estas ideas en mente, realizamos la segunda versión del IPT, que describimos a continuación.

3.2. Segunda versión del IPT

(cf. Vol. II, páginas 56-60)

La prueba piloto número dos se realizó en octubre de 1997, con una muestra de 32 sujetos, y esta vez se utilizó una sola sesión para el IPT, si bien en realidad hubo dos sesiones, una primera sesión de una hora para el ICGT y otra sesión de dos horas para el IPT. Sin embargo, la reducción de una sesión no fue suficiente, ya que el tiempo requerido de los estudiantes seguía siendo demasiado, como demostró el hecho de que volvimos a perder sujetos que no asistieron a una de las dos sesiones: de los 32 sujetos que participaron en las pruebas, solamente 27 realizaron las dos sesiones, es decir que se perdió un total del 16 % de la muestra.

Los cambios realizados en esta segunda versión del IPT fueron bastante radicales, ya que decidimos cambiar tanto el sistema de detección de problemas en el

texto como la manera de saber si los sujetos habían detectado o no los problemas. En vez de intentar prever *a priori* todos los problemas del texto, método que había demostrado no dar buen resultado, optamos por buscar un solo problema de cada tipo en el texto, si bien nos aseguramos de que se tratara de un problema generalizado, es decir, que representara un problema para todos los sujetos, o al menos para la inmensa mayoría. A las tres categorías que habíamos descrito en el primer IPT (problemas lingüísticos, culturales y pragmáticos) añadimos un problema de transferencia, y decidimos que la categoría de problema cultural se podía ampliar a “problema extralingüístico”, de manera que también se pudieran incluir otros problemas enciclopédicos, no necesariamente culturales. Puesto que las definiciones de todos estos tipos de problemas han quedado expuestas con anterioridad (cf. supra 3.4.2. *La definición de problema de traducción*), no nos detendremos en este punto.

Para identificar los cuatro problemas en cuestión analizamos los cuestionarios que habían llenado los sujetos en la prueba piloto 1, en concreto las preguntas 7 y 9 (cf. Vol. II, página 55) donde especificaban los segmentos que habían sido más problemáticos, y de esta manera elegimos el problema de cada tipo que habían encontrado y reconocido como tal la mayor parte de los sujetos. La única excepción fue el problema pragmático, ya que la mayoría de sujetos no había detectado ningún problema de este tipo, de manera que decidimos modificar el encargo de traducción y especificar de forma aún más clara que la función de la traducción obligaba a realizar una adaptación al contexto español, ya que el texto debía servir como reclamo publicitario de una empresa española.

Los cuatro problemas elegidos para el texto “Would you like to be a writer?” (cf. Vol. II, página 9) fueron los siguientes: (a) problema lingüístico: *flying*, línea 60; (b) problema extralingüístico: *Sunday and weekly papers*, línea 11; (c) problema de transferencia: *It's ideal for beginners. Details free. No cost. No obligation. Send the coupon.*, líneas 61-62; (d) problema pragmático, presentado por el encargo de traducción: *Britain*, línea 9. Para ver una explicación detallada de estos problemas, cf. supra 3.4.2. *La definición de problema de traducción*.

Por otra parte, modificamos todo el sistema de recogida de datos para saber si los sujetos habían detectado o no los problemas. Suprimimos la entrevista, que implicaba una pérdida de sujetos y en su lugar decidimos elaborar un cuestionario para que los estudiantes reflejaran por escrito, justo al terminar de traducir, si habían

detectado o no cada uno de los cuatro problemas. El resultado de estos cambios fueron los ítems del 10 al 13 del cuestionario (cf. Vol. II, páginas 57-60).

El ítem 10 de este cuestionario daba opción a nombrar hasta cinco problemas encontrados por los estudiantes al traducir el texto, es decir que informaba de los problemas que había detectado cada sujeto. De esta forma, si un estudiante citaba alguno de los cuatro problemas mencionados, quedaba claro que lo había detectado. Además, esta pregunta también buscaba información cualitativa, ya que indagaba sobre la naturaleza del problema y sobre su resolución, para poder conocer mejor el punto de partida de cada sujeto (si reconocían o se planteaban a qué tipo de problema se enfrentaban) y qué tipo de estrategias utilizaban, o si eran conscientes de que, en ocasiones, no habían resuelto un problema a pesar de haberlo intentado.

En el caso de que el sujeto no citara alguno de los cuatro problemas en cuestión en el ítem 10, el evaluador debía consultar el ítem 11, en que se le preguntaba al estudiante directamente si encontraba “muy difícil”, “poco difícil” o “fácil” cada uno de los cuatro segmentos problemáticos. Si después de esta pregunta todavía se albergaran dudas respecto a si el sujeto había detectado el problema o no, se podía consultar el ítem 13, donde se preguntaba directamente por el proceso seguido para traducir cada uno de los cuatro segmentos.

Respecto a los datos cualitativos, tras consultar con varios psicómetras, dedicados a la elaboración de tests, mejoramos los ítems dejando una sola pregunta abierta (número 6), de manera que el resto de preguntas eran todas de elección múltiple; de este modo, lográbamos que los sujetos se sintieran obligados a rodear alguna de las respuestas posibles y después añadieran algo si era necesario, en vez de poder contestar con un simple “sí” o “no”. También añadimos varios ítems nuevos que recababan información acerca de los conocimientos de los sujetos sobre el análisis textual: la tipología textual (ítem 1), la función y el género textual (ítem 2), y varias preguntas para conocer mejor el proceso de comprensión (ítems 3, 4 y 5). Los ítems 6, 7, 8 y 9 iban dirigidos a indagar sobre la adquisición de información por parte de los sujetos, es decir, sus conocimientos sobre las posibles fuentes de documentación y su uso de ellas. El objetivo de los ítems 14 y 15 era conseguir datos sobre el proceso de reexpresión y finalmente la pregunta 16 iba destinada a obtener datos sobre la autopercepción de la competencia traductora que tenían los sujetos. La última pregunta tenía por objeto evaluar si habíamos mejorado el cuestionario y si los sujetos entendían bien todas las preguntas.

Por último, dado que los ítems 4, 5 y 6 del cuestionario del primer IPT (cf. Vol. II, página 55) no nos habían ofrecido suficientes datos sobre el uso de diccionarios, reformulamos las preguntas, que se convirtieron en los ítems 6, 7, 8 y 9 (cf. Vol. II, página 56) y añadimos una segunda fuente de información cualitativa: durante la segunda sesión, en la que los sujetos debían traducir el texto, se informó a los sujetos de que podían utilizar todas las obras de consulta que había sobre la mesa del profesor, sin especificar qué obras eran. La única condición para utilizarlas era que debían anotar en una hoja en blanco todas las obras de consulta que utilizaran. La finalidad de estas instrucciones era obtener datos que ilustraran cómo utilizan los estudiantes las obras de consulta (diccionarios monolingües de uso y normativos, diccionarios bilingües, de sinónimos, gramáticas, guías de estilo, etc.) y si este uso sufría cambios significativos a medida que pasa el tiempo, así como observar la influencia que ejercía el uso de unas u otras obras en la traducción resultante.

El análisis de los datos obtenidos en la segunda prueba piloto nos llevó a varias conclusiones. En primer lugar, como ya hemos comentado al comienzo de este mismo apartado, quedó muy claro que todos los instrumentos debían poder administrarse en una sola sesión de dos horas, con el fin de no perder sujetos entre una y otra sesión.

En segundo lugar, tras estudiar la distribución de las respuestas de todas las preguntas, descubrimos que había cuatro preguntas poco informativas, es decir que más de un 75 % de los sujetos, en la primera medida, contestaban bien, lo cual no es aceptable para un instrumento que pretende ser sensible al cambio, ya que si estas respuestas eran contestadas bien en la primera medida, y también eran contestadas bien en la segunda medición, ¿qué conclusiones podríamos sacar sobre el proceso de adquisición gracias a ellas? Nos referimos a las preguntas 1, 3, 14 y 15. De todas maneras, decidimos mantener estas preguntas en las siguientes pruebas piloto que se iban a realizar con el mismo grupo de sujetos (pruebas número 3 y número 5) para ver si, efectivamente, en posteriores mediciones también las contestaban bien una mayoría de estudiantes.

La distribución de las respuestas a las otras preguntas del cuestionario fue adecuada (un 70 % o más de los sujetos las contestaron mal o hubo muchas diferencias entre las respuestas de unos y otros) u ofrecieron otros datos interesantes, como aparentes contradicciones entre lo que los sujetos decían que habían hecho y lo que habían hecho de verdad. Por ejemplo, en el ítem 7, un 50 % contestó que sí, es decir, que un solo diccionario había resuelto todas su dudas, pero después en el ítem 8 hubo un

11 % de ellos que escribieron el nombre de más de un diccionario. Por lo tanto, decidimos conservar estos ítems.

La otra fuente de información cualitativa, la hoja donde los estudiantes debían especificar qué obras de consulta habían utilizado, tampoco dio buenos resultados, ya que la mayoría de los estudiantes anotaron cosas como “bilingüe”, sin decir de qué editorial, de manera que era muy laborioso descubrir de qué obra se trataba, y otros estudiantes, en vez de decir qué entrada habían consultado, pusieron la página consultada, lo cual hizo evidente que este sistema de “control” de obras de consulta no funcionaba. Sin embargo, sí pudimos observar que, en cuanto se hubieron repartido los textos a los sujetos, la mayoría se levantó para ir en busca de un diccionario; eligieron preferentemente los bilingües y, como segunda opción, los monolingües en inglés, pero ningún estudiante utilizó monolingües en español ni otro tipo de obras de consulta (gramáticas, manuales de estilo, diccionarios de dudas, de sinónimos, de redacción y estilo, etc.).

3.3. Tercera versión del IPT

(cf. Vol. II, páginas 61-65)

La tercera prueba piloto tuvo lugar en febrero de 1998, y en ella probamos un nuevo texto en inglés (cf. Vol. II, página 11), ya que en ese momento todavía albergábamos la idea de utilizar dos textos diferentes para las dos mediciones (cf. supra 4.2. *La selección de textos para traducir*). Al cambiar de texto, lógicamente cambió también el contenido de los ítems 11 y 12 del IPT, que hacen referencia directa a los segmentos problemáticos. Los problemas identificados, que aparecen en las preguntas 11 y 12 del IPT, se pueden consultar en el volumen II, página 64.

El método utilizado para elegir los problemas del texto fue el de la consulta a expertos y a estudiantes de segundo curso, que no iban a participar en las pruebas. Además de estos cambios, también realizamos algunas modificaciones en el formato del cuestionario. Así, por ejemplo, en el ítem 10, debajo de “problema I” se añadieron líneas de puntos para que los sujetos anotaran allí el segmento problemático, ya que en octubre de 1997 algunos estudiantes habían dicho que no entendían qué tenían que contestar en esa pregunta.

Esta segunda prueba con el grupo de estudiantes que había realizado la prueba piloto número dos en octubre de 1997 nos mostró que los cambios realizados eran positivos y que el nuevo texto funcionaba, ya que los segmentos elegidos resultaron ser

problemáticos para la mayor parte de los estudiantes. Por otra parte, también observamos nuevos fallos: quedó claro que el ítem 11 no ofrecía suficiente información como para discernir si el sujeto había detectado el problema o no, ya que en muchas ocasiones los estudiantes rodeaban la respuesta “b” (poco difícil), que no era clara. Por consiguiente, decidimos modificar este ítem, así como el sistema de anotación de las obras de consulta, que continuaba sin funcionar. Sin embargo, estas modificaciones no tuvieron lugar hasta más adelante, dado que en este punto de la investigación nos preocupaba especialmente cómo demostrar que los dos textos utilizados en las dos mediciones presentaban un grado de dificultad equivalente, y nos concentraríamos en diseñar las pruebas que tuvieron lugar en abril de 1998.

3.4. Cuarta versión del IPT

(cf. Vol. II, páginas 66-68)

El objetivo de estas pruebas era claramente diferente de las anteriores, ya que se trataba de probar si las parejas de textos elegidos para cada combinación lingüística presentaban un grado de dificultad equivalente; por lo tanto, se dio prioridad a que los estudiantes fueran capaces de realizar las pruebas en dos horas, ya que lo que interesaba era que pudieran traducir dos textos y llenar los cuestionarios en una sola sesión, que además coincidía con una clase de traducción para no trastocar su horario de clases. Este punto era importante porque queríamos reunir en una sola aula y en la misma sesión a estudiantes de cinco grupos de clase diferentes, es decir que el problema logístico cobraba una gran importancia.

Para realizar la traducción de dos textos en dos horas tuvimos que reducir los textos elegidos (cf. supra 4.2. *La selección de textos para traducir*), y también realizamos una “versión reducida” del IPT. La reducción del instrumento consistió básicamente en eliminar todos los ítems que ofrecían información cualitativa, que en esta ocasión no nos era de utilidad, y por la misma razón redujimos el ítem 10, que en esta versión es el ítem 1 (cf. Vol. II, página 66).

Dado que trabajamos con seis textos, dos por cada combinación lingüística, existen seis versiones de esta cuarta versión del IPT, que se pueden consultar en las páginas 66-68, 69-71, 72-74, 75-77, 78-80, 81-83 del volumen II y corresponden, respectivamente, a los textos de las páginas 14, 12, 18, 20, 24 y 26, que son los mismos textos que hemos reproducido en las páginas 152, 153 y 154 de este volumen. Los resultados de esta prueba, de los que ya hemos hablado (cf. supra 4.2.1. *La prueba de*

equivalencia de textos) evidenciaron que los textos no eran equivalentes en cuanto a nivel de dificultad y por lo tanto decidimos utilizar en lo sucesivo un mismo texto para realizar las dos mediciones.

Respecto al IPT, dado que esta versión reducida no era la que pensábamos utilizar en el futuro, la cuarta prueba no nos llevó a ninguna conclusión, salvo que los textos reducidos funcionaban como textos, es decir que mostraban los rasgos característicos textuales (tipología reconocible, cohesión y coherencia, etc.) y que los sujetos detectaban y resolvían de igual manera los cuatro problemas identificados.

3.5. Quinta versión del IPT

(cf. Vol. II, páginas 84-87)

Tras la decisión de utilizar un mismo texto en las dos mediciones, nos vimos en la obligación de probar de nuevo el texto inicial, que habíamos usado en la prueba piloto número dos, para asegurarnos de que efectivamente los sujetos mejoraban no solamente porque ya habían traducido ese texto sino por el efecto del paso de ocho meses en los que habían recibido formación sobre traducción, de manera que en mayo de 1998 procedimos a efectuar la prueba piloto número cinco.

En esta ocasión realizamos pocos cambios en el IPT, en concreto modificamos la formulación del ítem 11, de manera que la respuesta de los sujetos a este ítem fuera suficiente para discernir claramente si habían detectado o no el problema en cuestión. Las posibles respuestas quedaron reducidas a (a) “sí, me ha resultado problemático” y (b), “no, no me ha resultado problemático”.

La muestra volvió a ser la misma de las pruebas piloto número dos y tres, ya que este grupo de estudiantes no había realizado la prueba número cuatro en abril, que incluía los nuevos textos, y tras la sesión de mayo de 1998, se realizaron una serie de entrevistas con los sujetos donde se les preguntó si recordaban el texto que habían traducido en octubre de 1997 y si este hecho les había influido en la traducción actual. Las respuestas fueron alentadoras y de esta manera decidimos utilizar definitivamente el mismo texto para las pruebas de fiabilidad y validez de los instrumentos.

Por último, la modificación del ítem 11 resultó todo un éxito, ya que la nueva formulación permitía discernir claramente entre la detección o no del problema.

Tras un curso académico muy activo, en el que llevamos a cabo cuatro pruebas piloto, nos dedicamos a analizar datos y reflexionar, y finalmente decidimos realizar varias modificaciones que reflejamos en el siguiente apartado.

3.6. Versión definitiva del IPT

(cf. Vol. II, páginas 88-91)

Esta sexta versión del IPT se puede considerar la definitiva, ya que los cambios realizados entre la primera administración en octubre de 1998 y la segunda administración, en mayo de 1999, son completamente menores y no se modifica ningún ítem, es decir que los cambios que comentamos a continuación, que se realizaron en julio de 1998 para poder aplicar el instrumento “finalizado” en octubre de 1998, en las tres universidades participantes en las pruebas, fueron los últimos.

En primer lugar, en lo que respecta al texto, decidimos utilizar el primer texto en inglés, “Would you like to be a writer?”, ya que era el que habíamos probado en más ocasiones, si bien esta vez lo redujimos con la misma técnica utilizada con los textos en la prueba piloto número cuatro, dejando los cuatro problemas ya comentados, con el fin de que los sujetos pudieran traducirlo en una hora, en vez de en dos. El texto, ya reducido, puede consultarse en la página 10 del volumen II.

Realizamos cambios bastante significativos respecto a la quinta versión del IPT, especialmente en las preguntas que corresponden la parte cuantitativa del IPT:

- (a) decidimos eliminar el ítem 12, que no ofrecía ningún dato interesante porque, según comentaron en las entrevistas posteriores, la mayoría de estudiantes no sabían qué contestar y por lo tanto decían cualquier cosa, ya que no distinguían claramente los conceptos de tipología de problema;
- (b) eliminamos asimismo el ítem 13, que no aportaba una información determinante para discernir entre la detección o no de cada problema, dato que por otra parte ya quedaba claro con la nueva formulación del ítem 11, que en la versión definitiva es el ítem número 12 (cf. Vol. II, página 91);
- (c) añadimos un nuevo ítem, el número 11, en el que preguntamos si, para el sujeto, los cuatro segmentos problemáticos representan un problema de traducción en general, más allá de que lo sean para él en concreto. Este ítem responde a una posibilidad en la que no habíamos pensado antes y que merece la pena tener en cuenta: hemos observado que se ha dado el caso de que un estudiante resuelve bien un problema (y por lo tanto, de acuerdo con nuestra escala, recibe un “2”) y sin embargo contesta en el ítem 12 que ese segmento no ha sido problemático para él/ella, en este caso, ¿debemos pensar que el sujeto ha resuelto ese problema por casualidad, o se trata más bien de que el proceso de resolución de ese tipo de problemas está tan automatizado que conscientemente no lo considera problemático porque no

reconoce haber realizado ningún esfuerzo para resolverlo? Para salir de esta duda, decidimos incluir este ítem, que nos ayudará a discernir claramente entre las dos opciones.

Las preguntas del cuestionario que ofrecían datos únicamente cualitativos también sufrieron cambios: volvimos a analizar la distribución de las respuestas a los ítems 1, 3, 14 y 15, que en la prueba piloto número dos no había sido satisfactoria, y al comparar la distribución obtenida en octubre de 1997 con la obtenida en mayo de 1998 por la misma muestra, observamos que se repetía. Decidimos eliminar los ítems 1 y 14, ya que solamente aportaban información “binaria” (sí / no), y si la respuesta era “sí” en los dos casos, la información no era nada útil. Sin embargo, no quisimos eliminar los ítems 3 y 15 (que en esta versión corresponden a las preguntas 3 y 13, respectivamente), ya que no ofrecían solamente información binaria, sino también cualitativa: en la pregunta 3, los sujetos deben especificar cuáles son las ideas principales del texto, y de este modo se puede observar su capacidad de síntesis, y evidentemente, al comparar las respuestas de las dos mediciones, se observa la progresión, si es que la hay, de la capacidad de síntesis. En cuanto al ítem 13, ocurre lo mismo, aunque en esta ocasión lo que los sujetos especifican es qué cambios han realizado en su traducción antes de entregarla, con lo cual se puede observar si hay un cambio de hábitos en la segunda medida respecto a la primera.

El ítem 2 también sufrió un cambio, ya que la formulación, que preguntaba por la función y el género del texto era algo ambigua, y muchos estudiantes habían señalado que no sabían muy bien cómo contestar, de manera que decidimos desdoblarlo en dos ítems (cf. Vol. II, página 88, ítems 1 y 2), además de aclarar los conceptos por los que preguntábamos.

Por último, respecto a los problemas que habíamos observado con la segunda fuente de información cualitativa, es decir, la “lista” de obras de consulta utilizadas, diseñamos una nueva manera de preguntar por ellas, integrada dentro del cuestionario general y de manera que las columnas mostraran claramente que debían anotar la editorial y el tipo de obra consultada además de la entrada consultada (cf. Vol. II, página 128).

Otra novedad de la sexta prueba piloto es que por primera vez se unificaron el ICGT y el IPT en un solo cuadernillo de pruebas (cf. Vol. II, páginas 125-133), de manera que resulta mucho más sencillo entender el orden de los cuestionarios, ya que solamente hay que ir siguiendo las instrucciones escritas en el cuadernillo, y se evita

asimismo que los estudiantes lean el texto o se centren en él antes de terminar de responder al cuestionario del ICGT. Por otra parte, el hecho de que las instrucciones se den por escrito evita que diferentes investigadores puedan modificarlas ligeramente o que algunos sujetos no escuchen alguna parte.

En el cuadernillo consta el tiempo que se debe utilizar, como máximo, para cada parte de la prueba, y el total es de dos horas. De esta manera nos aseguramos de que no se perdieran sujetos entre varias sesiones y de repartir y recoger con más facilidad todas las hojas de preguntas y respuestas de los diferentes instrumentos. Estos cambios facilitaron en gran medida la administración de los instrumentos de medida a los 224 sujetos que participaron en las pruebas de octubre de 1999, repartidos en seis grupos de tres universidades diferentes.

Finalmente, nos gustaría destacar que en esta ocasión se elaboraron, por una parte, los cuestionarios completos para las tres combinaciones lingüísticas, inglés, francés y alemán (cf. Vol. II, páginas 125-133, 134-142 y 143-151, respectivamente), así como las instrucciones para los evaluadores acerca de la corrección de los instrumentos. Las instrucciones de corrección de la parte cualitativa del IPT, que es igual para las tres combinaciones lingüísticas, se puede consultar en la página 107 del volumen II, y las correcciones de la parte cuantitativa, que son diferentes para cada combinación, se encuentran en las páginas 103, 105 y 106 del volumen II para inglés, francés y alemán, respectivamente.

3.7. Características y puntuación del IPT

El instrumento definitivo consta de dos partes. La primera es una tarea: se presenta un texto con un encargo de traducción a los sujetos, que deben cumplir el encargo y traducir dicho texto. Dentro de ese texto se han identificado previamente cuatro problemas de traducción, uno pragmático, uno extralingüístico, uno de transferencia y otro lingüístico. Cuando los sujetos han traducido el texto, pasan a contestar un cuestionario que permite medir, en una escala ordinal, la actuación de los sujetos frente a los cuatro problemas de traducción especificados. Dicha escala de medición es la siguiente:

- 0 - el sujeto no ha detectado el problema.
- 1 - el sujeto ha detectado el problema pero no ha sabido resolverlo.
- 2 - el sujeto ha resuelto bien el problema.

Para saber si calificar a un sujeto con un “0” o no, existe un ítem (pregunta 10, página 131, volumen II) que combina una pregunta de elección múltiple con una pregunta abierta para saber qué problemas ha detectado el sujeto en el texto (hasta un máximo de cinco problemas). Otros dos ítems, de elección múltiple (preguntas 11 y 12, páginas 132 y 133, volumen II) completan la información del ítem 10 para saber si el sujeto ha detectado o no los cuatro problemas elegidos por el investigador.

Para diferenciar entre un 1 y un 2 se debe observar la traducción del texto que ha realizado el sujeto en cuestión, de manera que, si ha resuelto el problema de forma satisfactoria, obtendrá un 2, y en caso contrario, la puntuación para ese problema será un 1. Hasta aquí llega la parte cuantitativa del IPT, es decir que los resultados cuantitativos de cada sujeto en este instrumento estarán comprendidos entre los 0 puntos, en caso de que no detecten ni resuelvan ninguno de los cuatro problemas, y los 8 puntos, en el supuesto de que detecten y resuelvan los cuatro problemas. Las instrucciones de corrección referentes a las tres combinaciones lingüísticas se pueden consultar en las páginas 103, 105 y 106 del volumen II para inglés, francés y alemán, respectivamente.

La segunda parte del IPT se realiza una vez traducido el texto, y consiste en otros 11 ítems del cuestionario, que sirven para ofrecer información cualitativa acerca del proceso de traducción seguido por los sujetos. La utilidad de estas respuestas se basa en la comparación con las respuestas del mismo sujeto en la medición posterior (que se realiza con el mismo texto), que informa del progreso realizado por cada individuo. No existe una escala ordinal cuantitativa para estas 11 preguntas, dado que no hay respuestas correctas o incorrectas, sino tendencias que deseamos observar, más allá de las pruebas de la generalizabilidad y de la validez de los instrumentos. Las instrucciones de corrección de esta parte del IPT se pueden consultar en la página 107 del volumen II.

4. Desarrollo del instrumento de medida de los errores de traducción (IET)

La elaboración inicial de este instrumento puede parecer, a simple vista, más sencilla que la de los otros dos, puesto que los docentes de traducción estamos familiarizados con la evaluación sumativa en traducción, que en cierta manera es la medida de los errores de traducción de los estudiantes. Sin embargo, la evaluación en el campo de la traducción dista mucho de ser satisfactoria, tal como demuestra el hecho de que ésta sea el objeto de una tesis doctoral recién publicada (cf. Waddington, 1999) y otra en proceso¹⁸, así como la carencia de exámenes, cuestionarios y/o ejercicios de traducción normalizados o utilizados por más de una institución que se dedique a la docencia de la traducción.

Por consiguiente, este es el instrumento menos original de los tres, ya que partimos de un sistema de evaluación ya creado (cf. Hurtado, 1995 y 1999); también es el menos satisfactorio, dado que se le puede atribuir un cierto grado de subjetividad, según el criterio utilizado por el evaluador. Es evidente que esta subjetividad, que caracteriza a todos los sistemas de calificación de errores de traducción conocidos hasta la fecha (cf. Waddington, 1999) no viene marcada por el sistema de evaluación, sino por los criterios de los evaluadores, y por consiguiente es, hasta la fecha, inevitable, dado que cada evaluador cuanta con un bagaje profesional y docente diferente. La razón por la que incluimos este instrumento en nuestras pruebas a pesar de este hecho es que pensamos que los errores de traducción son un elemento indicador esencial del nivel de la competencia traductora de los estudiantes y no podemos prescindir de ellos. Por otra parte, las pruebas de fiabilidad y validez de este instrumento nos permitirán contar con datos reales sobre el grado de variabilidad que es atribuible al evaluador, utilizando un mismo sistema de calificación de errores.

4.1. Primera versión del IET

(cf. Vol. II, páginas 119 y 120)

Antes de decidirnos por la que finalmente fue la primera versión del IET, pensamos en crear un nuevo sistema de corrección de las traducciones, realizando una clasificación de los errores de traducción que dividía éstos según el problema que los había causado, ya que, siguiendo la definición de error de Nord (1996:96-100), un error

¹⁸ Nos referimos a la Tesis Doctoral de N. Martínez Melis, que se está realizando en el Departament de Traducció i d'Interpretació de la Universitat Autònoma de Barcelona en la actualidad, bajo la dirección de la Dra. Hurtado Albir.

es causado por un problema de traducción no resuelto. De este modo, nuestra clasificación incluía cinco grandes bloques:

- (a) Problemas no resueltos de comprensión o reexpresión; en esta categoría incluíamos los errores de los que no se podía saber a ciencia cierta si eran debidos a un problema de comprensión o de reexpresión, ya que estos dos problemas pueden estar imbricados. En este apartado estarían los errores debidos a diferencias culturales y a la traducción literal o incorrecta de unidades léxicas.
- (b) Problemas no resueltos de comprensión; en este apartado incluíamos, por ejemplo, los falsos sentidos, los sinsentidos, la traducción literal de falsos amigos potenciales, la mala traducción de tiempos verbales cuyo uso si se traducen literalmente no funciona igual en español, etc.
- (c) Problemas no resueltos de reexpresión; en esta categoría había cinco bloques: las faltas de ortografía y puntuación; las faltas gramaticales, sintácticas o léxicas; las faltas de cohesión; las faltas de coherencia y las faltas de estilo. Dentro del segundo bloque, las faltas sintácticas, gramaticales o léxicas, se incluían la no omisión de un elemento que no es natural en la lengua de llegada, la no adición de un elemento que es natural en la lengua de llegada, la utilización de una estructura que no es natural en la lengua meta, la mala colocación en una unidad que debe ocupar otro lugar para formar una proposición natural en la lengua meta y una elección léxica no apropiada.
- (d) Problemas no resueltos de tipología textual o funcionalidad; esta categoría incluía errores de registro y tono textual, además de la omisión de fórmulas habituales del tipo de texto en cuestión.
- (e) Problemas no resueltos de informatividad; en este apartado se incluía la omisión de información relevante del TO y la adición de información innecesaria en el texto de llegada, así como las redundancias innecesarias.

Con este sistema, analizamos todos los problemas que presentaba el texto que íbamos a utilizar en la primera prueba piloto, "Would you like to be a writer?", y, para cada uno de los setenta problemas descritos (cf. vol. II, páginas 111-113) clasificamos un posible error, de manera que detectamos 70 posibles errores, 19 de la primera categoría (problemas no resueltos de comprensión y/o reexpresión), 4 de la segunda (problemas no resueltos de comprensión), 30 de la tercera categoría (problemas no resueltos de reexpresión), 17 de la cuarta (problemas no resueltos de tipología textual y funcionalidad) y 5 de la quinta categoría (problemas de informatividad).

Una vez realizada la primera prueba piloto vimos que era absurdo aplicar este sistema de corrección porque, para empezar, muchos de los segmentos que habíamos previsto como problemáticos no lo habían sido, de manera que no había errores relacionados, y además había errores de otros tipos (por ejemplo, sinsentidos) que no aparecían en nuestra clasificación porque no estaban causados por ninguno de los 70 problemas descritos (cf. supra 4.3.1. *Primera versión del IPT*). Por lo tanto, decidimos que, ya que las pruebas se habían realizado y contábamos con 21 traducciones que podíamos volver a corregir, utilizaríamos otro método de corrección. El método elegido fue el sistema de puntuación desarrollado por Hurtado (1995), que modificamos levemente para nuestro propósito, tal como se puede observar en la página 120 del volumen II. Este es el sistema que utilizamos en la corrección habitual de las traducciones de clase, de manera que a primera vista no parecía nada problemático, y decidimos utilizarlo para corregir las 21 traducciones.

La aplicación de este instrumento consistió en corregir las traducciones realizadas por los estudiantes, partiendo de 20 puntos y restando o sumando tantos puntos como fuera necesario según los errores y los aciertos fueran más o menos importantes (cf. vol. II, página 120). Dado que en esta prueba piloto solamente había un evaluador, no hubo ningún problema para aplicar este instrumento, pero la corrección de cada texto llevó unos 10 minutos, lo cual, sumado a los otros dos instrumentos y multiplicado por 21 sujetos, supuso mucho tiempo, lo que nos llevó a pensar en otras posibles soluciones menos exigentes para el evaluador.

En julio de 1998, cuando estábamos planeando las pruebas de validez y fiabilidad, para las que queríamos contar con la ayuda de diez evaluadores, decidimos que el tiempo invertido por cada evaluador en corregir las traducciones debía ser menor, especialmente si queríamos que este instrumento fuera práctico, de manera que se nos ocurrió una nueva idea, que se materializó en el IET definitivo, que describimos a continuación.

4.2. Versión definitiva del IET

(cf. Vol. II, página 121)

La idea que concebimos para simplificar la corrección del IET, fue la de observar solamente algunos tipos de errores y los aciertos, es decir que en vez de sumar y restar puntos por todos los supuestos que teníamos en cuenta en la versión anterior del

IET, decidimos elegir tres categorías de las que allí aparecían. De esta manera, la corrección resulta mucho menos costosa, tanto en tiempo como en esfuerzo invertido.

Las tres categorías elegidas son las siguientes:

- (a) falsos sentidos: segmentos del texto en que la traducción no tiene el mismo sentido que el texto original;
- (b) sinsentidos: segmentos del texto traducido que no tienen sentido, es decir que no serían comprendidos por el lector final;
- (c) aciertos: segmentos del texto meta en que la traducción es especialmente buena, ya sea porque propone una solución original, creativa –y adecuada al encargo, por supuesto- a alguno de los problemas que presenta el texto, o porque la traducción suena como un texto original en LM, gracias a la cohesión y coherencia, etc.

En las instrucciones de corrección definitivas de este instrumento (cf. vol. II, página 121) se puede observar que la manera de puntuar respecto a las tres categorías es muy simple, se trata de corregir el texto buscando falsos sentidos, sinsentidos y aciertos y después anotar en la cuadrícula correspondiente la cantidad de cada una de las tres categorías encontradas.

Si bien las categorías de errores que nosotros hemos elegido son aciertos, falsos sentidos y sinsentidos, en realidad se podría utilizar otras categorías; sin embargo, como ya hemos dicho, la elección de estas tres opciones responde a nuestro criterio de que estas tres categorías son especialmente representativas del nivel de la competencia traductora de los estudiantes. De todos modos, si un docente quisiera modificar estas categorías y elegir otras, podría hacerlo perfectamente, siempre y cuando se asegurase de realizar nuevamente pruebas de fiabilidad, dado que los resultados que hemos obtenido en esta investigación se refieren al IET tal como lo hemos presentado. Evidentemente, cuantas más categorías se incluyan en la corrección, más completa será ésta, pero también requerirá más tiempo y dedicación por parte del evaluador o evaluadores, además de que quizás la fiabilidad del instrumento se vea reducida al aumentar las posibles categorías de error.

Por último, nos gustaría destacar que este instrumento no se ha modificado en lo más mínimo en las pruebas de octubre de 1998 y mayo de 1999, es decir que la segunda versión del IET es también la definitiva.

4.3. Características y puntuación del IET

Este instrumento consta de dos partes. La primera coincide con la primera parte del IPT, es decir, la traducción de un texto por parte del sujeto siguiendo un encargo de traducción (se trata del mismo texto, es decir que el sujeto realiza en total una sola traducción). La segunda parte consiste en la aplicación de una plantilla para corregir la traducción y saber cuántas veces se produce un tipo de error concreto. Los tipos de error que se han tenido en cuenta en las pruebas de fiabilidad y validez de los instrumentos son los falsos sentidos, los sinsentidos y los aciertos de cada traducción, cuyas definiciones están especificadas en las instrucciones de corrección (cf. vol. II, página 121).

La puntuación de este instrumento es la siguiente: se parte de 0 puntos, se resta un punto por cada error cometido (falso sentido o sinsentido) y se suma un punto por cada acierto. De este modo, los resultados obtenidos por cada sujeto en este instrumento no se mueven dentro de una escala limitada, ya que si un sujeto comete 20 errores y ningún acierto obtiene un -20, mientras que un estudiante que comete 2 errores y 2 aciertos obtiene un 0. Por consiguiente, los resultados pueden ser puntos negativos o positivos, y no existe un límite preciso en ninguno de los dos extremos.

A pesar de que nosotros medimos las tres categorías mencionadas (falsos sentidos, sinsentidos y aciertos), cada investigador podría elegir la cantidad de tipos de error que va a observar a lo largo de las mediciones en cada caso, según su interés. De este modo, podría decidir, por ejemplo, observar solamente la cantidad de falsos sentidos que comete el estudiante en las distintas mediciones, u observar además los contrasentidos, o los calcos léxicos y sintácticos, etc. Para dar una idea de la tipología de errores que se podrían observar, ofrecemos una plantilla extensiva en la página 119 del volumen II. Si el investigador prefiere observar todos los tipos de error mencionados en el plantilla también puede hacerlo, simplemente deberá dedicar más tiempo al proceso de puntuación de este instrumento y verificar de nuevo la fiabilidad de este instrumento.

5. Desarrollo del instrumento de medida de la concepción general de la traducción (ICGT)

Este instrumento fue concebido desde el primer momento como un cuestionario, cuyo objetivo es ofrecer datos sobre la concepción abstracta que tienen los estudiantes de la traducción y de toda una serie de nociones relacionadas.

También desde un primer momento pretendimos que se tratara de un instrumento sensible al cambio, es decir que los resultados debían permitirnos no tan solo saber qué piensan los estudiantes de la traducción y del resto de conceptos sino más bien conocer la evolución de estos conceptos, la mayoría abstractos, a medida que avanzan en su formación como traductores.

5.1. Primera versión del ICGT

(cf. Vol. II, páginas 31-34)

La elaboración de la primera versión de este instrumento fue muy compleja, ya que no teníamos ningún otro cuestionario en que basarnos, de manera que lo primero que nos planteamos fue qué queríamos averiguar mediante este test. La respuesta fue que deseábamos conocer de la manera más completa posible la concepción de la traducción que tenían los sujetos, y dentro de esa idea abstracta de traducción incluimos una serie de nociones básicas relacionadas con el concepto central, que pasamos a enumerar, junto con una breve explicación:

- (a) la noción de traducción en tanto que operación; queríamos saber si la entendían como una acción relacionada principalmente con el sentido o con la forma, si la situaban en el plano del habla o de la lengua, si la consideraban una acción entre lenguas, culturas, etc.;
- (b) la noción de un buen traductor, es decir, de la competencia traductora; queríamos saber si tenían la competencia traductora asimilada a la competencia lingüística bilingüe o si pensaban que había otras subcompetencias dentro de la competencia traductora, además de conocer su visión del mundo del traductor, su función social, etc.;
- (c) la unidad de traducción; deseábamos conocer su idea de lo que es y de si existe la unidad de traducción, si ésta es fija o variable y si la asimilaban con la palabra, con la frase, con el texto, etc.;

- (d) el proceso de traducción; queríamos saber si los estudiantes eran conscientes de que había un proceso y qué idea tenían de él;
- (e) la búsqueda de la equivalencia; queríamos conocer las técnicas de traducción que conocían, si tenían una noción de equivalencia dinámica o no, si conocían estrategias de reformulación, etc.;
- (f) el método de traducción; queríamos saber cuál era su método de trabajo global, qué criterios tenían, si respetaban la función de la traducción, etc.;
- (g) la documentación y las referencias extralingüísticas; deseábamos conocer hasta qué punto conocían las fuentes de documentación y la técnica que seguían cuando se enfrentaban a alguna referencia extralingüística.

Cuando elaboramos esta lista de nociones relacionadas cuya concepción por parte de los estudiantes queríamos conocer, también elaboramos una serie de hipótesis, que eran nuestras ideas de lo que pasaría entre una medición y la otra. Por ejemplo, pensábamos que en un primer momento los estudiantes tendrían una noción de traducción escolar o pedagógica, es decir, subordinada a la lengua, mientras que en una segunda medición mostrarían una concepción más cercana a la profesión del traductor, situada en el plano del habla; del mismo modo, creímos que en una primera medición tendrían el concepto de un buen traductor ligado al de un bilingüe, mientras que en una segunda medición ampliarían este concepto añadiendo otras habilidades y conocimientos a la idea de un buen traductor, que es en realidad su concepción de la competencia traductora.

Estas ideas nos llevaron a elaborar la primera versión del ICGT, donde constan los apartados ya mencionados. Evidentemente, cuando probamos esta versión con un grupo de estudiantes en octubre de 1996, eliminamos los títulos de cada bloque y cambiamos el orden de las preguntas, de manera que ni los títulos ni el orden pudieran influir de modo alguno en sus respuestas.

Una vez realizada la prueba piloto número uno, nos encontramos con el gran problema de cómo corregir este cuestionario, ya que, aunque habíamos previsto asignar un punto a cada pregunta si la respuesta era correcta y 0 puntos si no lo era (en total había 35 preguntas, por lo que el máximo de puntos posibles era de 35 y el mínimo era de 0) no era tan fácil discernir entre un “0” y un “1”, porque muchas de las preguntas daban opción a “0’25”, “0’5”, “0’75” puntos, etc. Si bien habíamos previsto las respuestas que considerábamos correctas y algunas posibilidades, no habíamos elaborado una “plantilla” de corrección donde se contemplaran todas las posibilidades,

cosa que por otra parte veíamos casi imposible, dado que había varias preguntas abiertas. Para poner un ejemplo, tomemos el ítem E.5. del cuestionario, “Cuando encuentras una palabra que comprendes pero que no sabes reexpresar en castellano, ¿qué haces?” (cf. Vol. II, página 33) nosotros habíamos previsto dar un punto entero si el sujeto mencionaba dos obras de consulta o más o alguna estrategia de reformulación, medio punto si mencionaba un diccionario bilingüe y 0 puntos si la respuesta era otra, pero obviamente hubo respuestas que no se adaptaron a ninguna de estas opciones (p.e., respuestas del tipo “preguntaría a alguien”, “depende del texto” o simplemente “depende”), por lo que no pudimos obtener una puntuación concreta para este instrumento y se hizo evidente que hacían falta cambios para que el ICGT fuera aplicable en la práctica.

Por otra parte, tras una reflexión acerca de los contenidos del ICGT, advertimos que necesitábamos más preguntas en el apartado de “técnicas”, ya que las que habíamos incluido no eran suficientes, que faltaban preguntas sobre el ámbito profesional del traductor y, finalmente, que había que reformular muchas de las preguntas, ya que el metalenguaje utilizado en ellas era desconocido para muchos de los estudiantes. En este sentido, hay que tener en cuenta que esta prueba piloto se realizó el primer día de clase, con lo cual era de esperar que los alumnos no estuvieran familiarizados con los términos propios de la Traductología.

5.2. Segunda versión del ICGT

(cf. Vol. II, páginas 35-38)

Para la segunda prueba del ICGT, que tuvo lugar en octubre de 1997, se realizaron bastantes cambios en el instrumento, empezando por una redistribución de los bloques conceptuales: creamos un nuevo bloque sobre técnicas, separándolas de este modo de la “equivalencia” y eliminamos el apartado de documentación y referencias extralingüísticas, ya que incluimos los ítems en otros bloques. Asimismo, cambiamos de apartado varias preguntas que respondían mejor a la nueva clasificación.

Aparte de redistribuir los bloques, añadimos nuevos ítems, especialmente en el apartado de “un buen traductor”, que en realidad hacía referencia a la competencia traductora. De esta manera, la segunda versión del ICGT constaba de 38 ítems, de los cuales solamente dos eran completamente abiertos (ítems 6 de los bloques A y D). Al igual que en la ocasión anterior, para poder administrar el cuestionario a los estudiantes,

creamos una versión sin títulos y aleatorizada, es decir, con las preguntas distribuidas al azar.

Tras la prueba piloto número dos realizamos un análisis de la distribución de las respuestas de los estudiantes, y resultó que, de las preguntas de elección múltiple y verdadero/falso había 25 ítems que presentaban una distribución poco satisfactoria, es decir que al menos el 75 % de los estudiantes había contestado bien a las preguntas, cosa que no es aceptable para la primera medición de un instrumento que debe ser sensible al cambio. Las preguntas en cuestión sumaban un 66 % del cuestionario.

Solamente nueve ítems de elección múltiple o verdadero/falso mostraron una distribución adecuada, es decir que o bien más de un 70 % de los sujetos las contestaron mal o hubo mucha diferencia entre las respuestas de los sujetos. Esto ocurrió con el 24 % del cuestionario. Las preguntas abiertas (A.6 y D.6, páginas 35 y 37 del volumen II) tuvieron una distribución muy variada, lo cual era positivo. Tras analizar estos resultados, decidimos eliminar todos los ítems con distribución poco satisfactoria, ya que no nos ofrecían información interesante. Sin embargo, hubo una pregunta en concreto, la D.5. (cf. página 37, vol. II), que decidimos conservar, ya que a pesar de la distribución de las respuestas, éstas se contradecían con la práctica en muchos casos.

Además de analizar la distribución de las respuestas, también analizamos los contenidos del cuestionario, y llegamos a la conclusión de que era mejor reestructurar completamente el ICGT, ya que al eliminar tantas preguntas había bloques para los que no teníamos ningún ítem, además de no estar plenamente satisfechos con los apartados existentes.

5.3. Versión definitiva del ICGT

(cf. Vol. II, páginas 39-41)

Esta tercera versión del ICGT, que se utilizó en las pruebas de octubre de 1998, se puede considerar la definitiva ya que los cambios que ha sufrido para la segunda administración de mayo de 1999 solamente afectan a las instrucciones iniciales, referidas a los dos instrumentos, el ICGT y el IPT.

Una novedad importante en esta versión definitiva es que por primera vez se realizó un cuestionario para cada combinación lingüística, aunque en realidad estas versiones solamente difieren en tres ítems, en que se hace referencia al idioma o la cultura de partida: las preguntas 5, 6 y 10. Estos ligeros cambios se pueden observar en las tres versiones definitivas del ICGT, que se pueden consultar en los cuadernillos de

pruebas definitivos (cf. páginas 125-127, 134-136, 143-145 para inglés, francés y alemán respectivamente).

En esta ocasión, dado el fracaso de la estructura anterior, decidimos restringir los bloques conceptuales a dos, que podían reunir el resto de conceptos: la traducción y la competencia traductora. Ambos bloques debían tener la misma importancia en la puntuación final, y para ello debían contener el mismo número de preguntas, de manera que el resultado en puntos fuera equilibrado. Prácticamente todas las preguntas que incluye el instrumento definitivo se basan en preguntas cuyas respuestas dieron una distribución adecuada en el análisis que hemos mencionado en el apartado anterior, si bien algunas han sido modificadas ligeramente para ser todavía más claras.

Dentro de los dos bloques incluimos dos nuevas preguntas abiertas sobre la concepción general de la traducción y de la competencia traductora, en las que se pide al sujeto que defina cada uno de estos conceptos de manera breve. Nos referimos a los ítems 1 de los bloques A y B (cf. Vol. II, páginas 39-41). Dado que las preguntas abiertas nos habían planteado una dificultad a la hora de puntuarlas en el pasado, esta vez decidimos especificar al máximo las respuestas que buscábamos. Las instrucciones de corrección de este instrumento, que se pueden ver en la página 47 del volumen II, aclaran las respuestas correctas para cada pregunta, así como la manera exacta de puntuar cada pregunta, de modo que nos limitaremos a explicar cuál es el objetivo de cada ítem.

En el bloque de traducción hay seis preguntas, y aparte de la primera, que indaga sobre la noción misma de traducción, las otras cinco tienen por objeto recabar información sobre los conceptos básicos relacionados con la traducción: los elementos que intervienen en una traducción (ítem2), los tipos de problemas que pueden surgir (ítem 3), la unidad de traducción (ítem 4), la equivalencia de traducción (ítem 5) y la función de la traducción (ítem 6). Respecto al bloque de la competencia traductora, intentamos incluir preguntas sobre los aspectos más importantes de las subcompetencias que forman la competencia traductora según el modelo en que nos basamos (cf. supra 2.3. *El modelo de PACTE*), a excepción de la subcompetencia psicofisiológica. Así, el ítem 2 indaga sobre la subcompetencia profesional, en concreto sobre la concepción del mercado laboral, mientras que el ítem 3 indaga sobre el aspecto de la búsqueda de información, que está incluido en la misma subcompetencia. El ítem 4 hace referencia a la subcompetencia lingüística, el ítem 5 a la subcompetencia extralingüística y,

finalmente, el ítem 6 se centra en un aspecto de la subcompetencia estratégica, las estrategias de comprensión.

Por último, existen dos ítems (el 7 y el 8, página 41, volumen II) que no corresponden a ninguno de los dos bloques y que indagan sobre las estrategias de comprensión y de reexpresión utilizadas por los sujetos. Estas dos preguntas pretenden recoger información cualitativa, y no puntúan en el ICGT, de manera que a efectos de puntuación, es como si hubiera solamente 12 ítems.

5.4. Características y puntuación del ICGT

Este instrumento está constituido por un test referido al criterio que incluye 14 ítems de tres clases: ítems de elección múltiple, ítems de verdadero/falso e ítems abiertos. Se trata de un cuestionario multidimensional, ya que dentro de la concepción “abstracta” de la traducción existen varios factores relevantes, y la mayoría de ellos están representados en el cuestionario: concepción de la traducción, elementos que intervienen en la traducción, concepción de los problemas de traducción, concepción de la unidad de traducción, concepción de la equivalencia de traducción, concepción de la función de la traducción, concepción de la competencia traductora, concepción de la subcompetencia instrumental, concepción de la subcompetencia lingüística, concepción de la subcompetencia extralingüística, concepción de la subcompetencia estratégica y, por último, las técnicas para la resolución de problemas de comprensión en la lengua extranjera y de reexpresión en la lengua materna durante el acto de traducción.

Dada la variedad de factores representados en el cuestionario, se han agrupado en torno a dos conceptos esenciales: la concepción de la traducción y la concepción de la competencia traductora, de manera que existen seis ítems dedicados a cada uno de estos dos bloques. Sin embargo, para que los sujetos no vean estos dos bloques conceptuales, ya que esto podría influir en sus respuestas, las preguntas se han aleatorizado, es decir que el orden que siguen los ítems en el cuestionario tal como se presenta a los estudiantes está escogido al azar. Esta versión aleatorizada se puede consultar en las páginas 42-44 del volumen II.

Los resultados se obtienen de manera diferente para cada uno de los tipos de preguntas mencionados:

- (a) En las preguntas abiertas, se otorga un punto para las respuestas correctas, 0 puntos para las respuestas totalmente erróneas y 0'25, 0'50 o 0'75 para las preguntas que no sean totalmente correctas ni erróneas pero contengan un 25%, un 50% o un 75%

de acierto respectivamente; las décimas otorgadas pueden variar según cada pregunta, es decir que para algunos ítems se otorgan 0 puntos, 0'1, 0'2, 0'3, 0'4, 0'5 0'6, 0'7, 0'8, 0'9 décimas o 1 punto. Estas preguntas son las siguientes: 1, 2, 3, 4, 8 (cf. páginas 42-44 del volumen II).

- (b) En la pregunta 7, de elección múltiple, se otorga un punto por respuesta correcta y se resta 0,3 puntos por cada respuesta incorrecta (hay cuatro respuestas posibles en este ítem, y de este modo se suprime el posible efecto de las respuestas dadas al azar).
- (c) Las preguntas cuya respuesta es “verdadero o falso” se puntúan asignando un punto positivo a las respuestas correctas y un punto negativo a las respuestas incorrectas. Estas preguntas son las siguientes: 9, 10, 11, 12, 13 y 14 (cf. páginas 42-44 del volumen II).
- (d) Las respuestas en blanco se cuentan como un 0 pero no restan puntos, ya que se advierte a los sujetos de que deben contestar sólo aquellas preguntas que entiendan y de las que crean saber la respuesta.

Este sistema de puntuación hace que los resultados de un sujeto en el ICGT oscilen entre los -6,3 puntos (suponiendo que un sujeto conteste mal a las seis preguntas de “verdadero o falso” y al ítem 7 y no conteste a los ítems 1, 2, 3, 4 y 8) y los 12 puntos, en el caso de que el estudiante conteste bien a todos los ítems mencionados. Las instrucciones de corrección del ICGT se pueden consultar en las páginas 47-49 del volumen II.

Al igual que ocurre con el IPT, el ICGT cuenta con una parte cualitativa, formada por los ítems 5 y 6 (cf. página 43, volumen II), que ofrecen información referente a las estrategias empleadas para resolver problemas de compresión y reexpresión, cuya puntuación no se tiene en cuenta en el cálculo de la puntuación total del instrumento, ya que la información que se obtiene es cualitativa.

6. Administración de los instrumentos de medida

Los tres instrumentos de medida están ideados para ser administrados simultáneamente, en el lapso de un máximo de dos horas. De esta forma, para cada medición se contará con tres resultados, los del IPT (puntuación total posible entre 0 y 8 puntos), el IET (un punto negativo por cada error y un punto positivo por cada acierto realizado) y el ICGT (puntuación total posible entre -6'3 y 12 puntos). La administración se debe realizar en el siguiente orden: ICGT, IPT e IET, ya que si los estudiantes realizan la traducción del texto antes de contestar al cuestionario sobre concepción general de la traducción, sus respuestas a dicho cuestionario pueden variar, según hemos comprobado en las pruebas piloto, ya que se produce un efecto de sensibilización que es preferible controlar.

La aplicación del ICGT no tiene una duración determinada, ya que depende de cada sujeto y si queremos obtener respuestas sinceras y meditadas no se debe presionar al sujeto con el tiempo. De todas maneras, existe un límite de tiempo de dos horas, que es lo suficientemente amplio para que los sujetos contesten a todas las preguntas y realicen la traducción, tal como ha quedado patente en las diversas pruebas piloto y en las pruebas de fiabilidad y validez de los instrumentos, ya que los sujetos han tardado entre 5 y 25 minutos en contestar al ICGT y entre 45 y 90 minutos en realizar la traducción y contestar al IPT (el IET no requiere dedicación adicional de tiempo por parte de los estudiantes, ya que la puntuación se otorga de acuerdo con la traducción que incluye el IPT). Por lo tanto, la dedicación total a los tres instrumentos oscila entre los 50 minutos y las 2 horas. Para traducir el texto se pueden ayudar de cualquier herramienta (obras lexicográficas, textos paralelos, etc.) que ellos hayan provisto (este es el método que se suele emplear en los exámenes de traducción, por lo que los sujetos están familiarizados con él); sin embargo, para la primera administración, el investigador debe proveer las obras de consulta que puedan utilizar los sujetos, ya que las pruebas se realizan el primer día de clase, y no se puede advertir a los sujetos de que traigan obras, menos aún cuando todavía no están familiarizados con las obras de consulta que utilizarán para traducir. La única condición que se pone a los estudiantes que realizan las pruebas es que deben anotar en un espacio del cuadernillo (cf. página 128, volumen II) todas las obras consultadas y la entrada o página que han mirado; de este modo también se obtiene información cualitativa sobre la utilización de diccionarios.

CAPÍTULO 5- VALIDACIÓN DE LOS INSTRUMENTOS DE MEDIDA

1. Hipótesis

Una vez creados los tres instrumentos de medida a que venimos refiriéndonos como ICGT, IPT e IET, era el momento de ver si eran válidos y fiables, es decir, si, por un lado, estos instrumentos son útiles para la finalidad con que se han diseñado (validez) y hasta qué punto se pueden generalizar los resultados que se obtienen con los instrumentos. En este apartado exponemos nuestras hipótesis empíricas acerca del funcionamiento de los tres instrumentos de medida, que son las que dieron lugar al diseño de las pruebas de fiabilidad y validez que se llevaron a cabo en octubre de 1998 y mayo de 1999.

Existen dos grupos de hipótesis, las principales y las complementarias. Las hipótesis principales se dividen a su vez en dos bloques; el primero, al que denominamos “A”, reúne hipótesis referidas a la sensibilidad al cambio de los tres instrumentos, mientras que las hipótesis “B” tratan sobre la generalizabilidad de los tres instrumentos. Las hipótesis “C”, o hipótesis complementarias, expresan nuestra opinión acerca de las tendencias de los estudiantes, que se analizan a partir de sus respuestas a las preguntas cualitativas del IPT y del ICGT.

1.1. Instrumentos sensibles al cambio

Desde el primer momento concebimos nuestros instrumentos de medida de la adquisición de la competencia traductora como instrumentos sensibles al cambio, ya que si creemos que existe un proceso de adquisición de una serie de competencias, debemos suponer que los sujetos variarán su conducta con el transcurso del tiempo, lo cual se puede formular en una hipótesis global a la que llamaremos “A”: si comparamos los resultados de cada sujeto en la primera y la segunda medición, habrá cambios en las tres medidas (una por instrumento).

Respecto al tipo de cambios que predecimos entre la primera y la segunda administración de los instrumentos, nuestras tres **hipótesis** son las siguientes:

A1- Los resultados de los estudiantes mejorarán en las tres medidas.

- A2-** Los resultados del IPT mostrarán una clara correlación con los resultados del IET desde el principio, y esta correlación se mantendrá en la segunda medición.
- A3-** Los resultados del IPT y del IET no mostrarán una correlación con los resultados del ICGT en la primera medición, y los del IET continuarán sin correlacionar con los del ICGT en la segunda medición, pero los resultados del IPT aumentarán su correlación con el ICGT en la segunda medición.

La explicación teórica de las hipótesis A y A1 es bastante sencilla y lógica, ya que responden a todas las teorías del aprendizaje, que defienden que un sujeto expuesto a una enseñanza concreta adquiere una serie de conocimientos, tanto operativos como declarativos. En este caso, los conocimientos operativos quedarán patentes en los resultados del IET y del IPT, mientras que los conocimientos declarativos se podrán ver en los resultados del ICGT, y si los instrumentos miden lo que deberían medir, se puede esperar un cambio en los resultados de los sujetos al comparar la segunda medición con la primera, teniendo en cuenta que la primera administración de los instrumentos se realiza previamente al contacto de los sujetos con la traducción (con la excepción de la traducción pedagógica en la enseñanza de lenguas que puedan haber realizado en la secundaria) y la segunda administración tiene lugar tras ocho meses de exposición a la enseñanza de la traducción en una facultad.

Las otras dos hipótesis son algo más complejas, y responden a observaciones realizadas por los profesores de traducción y a suposiciones generalizadas en nuestra disciplina que en esta ocasión podremos comprobar hasta qué punto son ciertas. La hipótesis A2 muestra la suposición general de los docentes de traducción de que la actuación de un estudiante frente a los problemas de traducción es la misma que frente a los errores de traducción, dado que si se resuelven satisfactoriamente los problemas no existen tantos errores en su traducción, mientras que si no se resuelven bien los problemas es probable que se cometan errores de traducción. Es evidente que esta suposición podría no ser cierta, ya que en teoría un sujeto puede resolver bien los problemas que plantea el texto y en cambio cometer muchos errores de redacción o de cohesión en la lengua de llegada que no están causados por un problema planteado por el texto o el encargo de traducción sino por su falta de competencia en la lengua de llegada, y también puede cometer errores en la traducción debido a su falta de competencia en la lengua extranjera, que no le ha permitido comprender bien el texto original. Por consiguiente, esta hipótesis dista de estar confirmada, y esperamos que con

los resultados obtenidos en la correlación entre los resultados del IPT y del IET en las dos mediciones se esclarezca la relación entre los problemas y los errores de traducción.

La hipótesis A3 se refiere al proceso de adquisición de conocimientos, y nuestra postura es que a medida que los estudiantes avanzan en su formación adquieren conocimientos declarativos sobre la traducción, y por lo tanto la concepción que tienen de la traducción empieza a hacerse consciente y deja de ser algo completamente abstracto, relacionándose así con la práctica. Pero esta relación entre los conocimientos declarativos y operativos no existe en un primer momento, de modo que opinamos que podremos observar cómo la correlación entre el IPT y el IET (que muestran los conocimientos operativos) y el ICGT (que muestra conocimientos declarativos) no existe en la primera medición. Respecto a la segunda medición, tras ocho meses de formación esperamos que los conocimientos operativos empiecen a correlacionar con los declarativos, si bien esperamos un comportamiento diferente en los datos del IPT y del IET. El IPT ofrece información acerca de la actuación de los sujetos respecto a los problemas de traducción, y valora positivamente el hecho de que los estudiantes detecten más problemas, aunque no los resuelvan correctamente; nuestra opinión es que la detección de problemas aumentará considerablemente a final de curso, y por consiguiente empezará a correlacionar con los resultados del ICGT, es decir, con los conocimientos declarativos. Sin embargo, la evolución de los estudiantes respecto a los errores de traducción es distinta; nuestra experiencia como docentes nos muestra que los estudiantes cometen menos errores de un tipo: cada individuo tiene un “punto débil” a principio de curso, ya sea la redacción en lengua materna, la comprensión de la lengua extranjera, etc., lo cual le lleva a cometer preferentemente un tipo concreto de error, que es el que primero suele remitir a lo largo del primer curso. Esta evolución en un tipo de errores en concreto hace que el número total de errores descienda con menos velocidad que aumenta la detección de problemas y la adquisición de conocimientos declarativos. Por consiguiente, no esperamos que el IET correlacione con el ICGT en la segunda medición.

Más allá de la contrastación de estas tres hipótesis, lo importante para nuestros instrumentos es que sean capaces de reflejar los cambios que se producen en los estudiantes entre principio y final de curso, es decir que aunque en un caso hipotético un estudiante empeorara, un instrumento sensible al cambio debería detectar ese empeoramiento.

1.2. Instrumentos generalizables

Los tres instrumentos pretenden medir la adquisición de la competencia traductora de los estudiantes, y es evidente que existen diferencias entre sus niveles de competencia traductora, porque cada uno cuenta con un bagaje distinto, tanto en lo que respecta a la educación como a las circunstancias personales, que hacen que cada cual tenga unos conocimientos y habilidades que difieren del resto de sujetos. Esta es una fuente de variabilidad sistemática, que no podemos controlar si queremos medir la realidad, como ocurre en nuestro caso. De hecho, esta es exactamente la variabilidad que pretendemos medir en los exámenes académicos, cuya pretensión no es otra que ordenar a los estudiantes en relación con el grado en que dominan los contenidos y habilidades que se supone que han adquirido mediante la enseñanza.

Decíamos que los distintos grados de competencia traductora de los estudiantes de traducción deben constituir la principal fuente de variabilidad de nuestros resultados de evaluación académica, pero existen otros tres factores, externos al sujeto, que también conocemos y que son difíciles de evitar: (a) la formación recibida mientras dura el experimento, que suele ser distinta según el profesor que imparte la asignatura de traducción, según el método de enseñanza utilizado y según el centro de formación, ya que los planes de estudios y por tanto las asignaturas que cursan los sujetos durante los meses que transcurren entre las dos mediciones suelen variar; (b) los evaluadores, que pueden valorar de maneras muy distintas una misma actuación según su criterio; (c) la combinación lingüística de la traducción, que puede implicar una dificultad añadida para el sujeto, según la distancia existente entre las lenguas y las culturas de origen y meta.

La diversidad de fuentes de variabilidad hace que sea muy difícil pensar en instrumentos estáticos, y por consiguiente consideramos más importante poder cuantificar el grado de variabilidad atribuible a estos tres elementos que intervienen en la traducción académica, que intentar controlarlos, algo muy alejado de la práctica docente real y que nos forzaría a realizar un experimento de laboratorio sin ninguna validez externa. Por consiguiente, hemos diseñado un estudio de la generalizabilidad (cf. *infra 5.2. Diseño para el estudio de la generalizabilidad*) mediante el cual contrastaremos una serie de hipótesis acerca de la variabilidad que confieren a los resultados de los instrumentos los tres elementos mencionados: evaluadores, centros de formación y combinaciones lingüísticas.

Para diferenciar las hipótesis referidas a la sensibilidad al cambio de los instrumentos, ya mencionadas, de las hipótesis relacionadas con la generalizabilidad, llamaremos a estas últimas hipótesis “B”. La hipótesis global B es la siguiente: la combinación lingüística, el evaluador y el centro donde se imparten los estudios de traducción causarán una variabilidad reducida (en comparación con la variabilidad atribuible a la competencia de los estudiantes) en los resultados obtenidos por los tres instrumentos. Esta hipótesis global se desglosa en las siguientes **hipótesis**:

- B1-** El IPT obtendrá resultados con muy poca variabilidad atribuible a la combinación lingüística, al evaluador y al centro donde se imparte la enseñanza.
- B2-** El IET obtendrá resultados con poca variabilidad atribuible a la combinación lingüística y al centro donde se imparte la enseñanza, mientras que la variabilidad atribuible al evaluador será mayor.
- B3-** El ICGT obtendrá resultados con muy poca variabilidad atribuible a la combinación lingüística y al centro donde se imparte la enseñanza.

Cuando hablamos de combinaciones lingüísticas, debemos advertir que nos referimos siempre a la traducción directa, como ya hemos aclarado en varias ocasiones, y solamente a las tres combinaciones lingüísticas que existen como primera opción en los planes de estudio de las Licenciaturas de Traducción en España, es decir, inglés español, francés español y alemán español. Hacemos la salvedad de otras lenguas maternas que sí se eligen como primera opción en algunos planes de estudio del estado español, como el catalán y el gallego, dado que por las dimensiones del estudio realizado no nos podíamos permitir incluir muestras lo suficientemente amplias de otras lenguas maternas.

En lo que concierne a los grados de variabilidad de los tres factores mencionados, evaluadores, centros y combinaciones lingüísticas, nuestra opinión es que el mayor grado de variabilidad se dará en los evaluadores, dado que por más que se intente unificar el criterio de corrección de una traducción cada uno tiene unas ideas ya formadas que son difíciles de cambiar. En este sentido, nos gustaría aclarar que opinamos que el IPT ofrecerá un grado de variabilidad mínimo, mientras que el IET mostrará una mayor variabilidad, debido precisamente a que los evaluadores están acostumbrados a corregir traducciones buscando errores y/o aciertos y por lo tanto será difícil que utilicen criterios compartidos (cambiando los suyos habituales) al utilizar el IET. En la hipótesis B3, referente al ICGT, no mencionamos al evaluador puesto que,

como se verá en el diseño de las pruebas de fiabilidad, solamente existirá un evaluador para este instrumento, cuya variabilidad en este sentido es nula.

Respecto al hecho de cursar estudios en un centro u otro, opinamos que en la primera medición la variabilidad será mínima en los tres instrumentos, ya que al llegar a la Universidad, sea cual sea ésta, los estudiantes difieren en sus conocimientos y habilidades, pero estas diferencias son normales e imprevisibles. Sin embargo, en la segunda medición influyen en los resultados otras variables que no podemos controlar, como la enseñanza recibida por los estudiantes en lengua y cultura extranjera, en lengua y cultura materna, en Lingüística, en Teoría de la Traducción, etc. a causa de las diferencias entre los distintos planes de estudio y la metodología empleada en la enseñanza de la traducción por cada profesor, y puesto que no controlamos estos factores, esperamos que exista cierta variabilidad entre los estudiantes de los diferentes centros en la segunda medición en los tres instrumentos.

No pensamos que la combinación lingüística elegida muestre una variabilidad significativa en ninguno de los tres instrumentos, pero quizás éste sea el factor menos conocido, por la inexistencia de estudios al respecto, de manera que solamente queremos adelantar que, *a priori*, parece que la traducción del alemán al español sea la más difícil para los estudiantes, por ser éstas las dos culturas y lenguas más alejadas; del mismo modo, podemos pensar que la traducción del inglés al español está más alejada y por lo tanto presenta una mayor dificultad para los estudiantes que la del francés. En cualquier caso, el análisis de los datos obtenidos indicará si existen o no estas diferencias en las muestras de las diferentes combinaciones lingüísticas.

1.3. Hipótesis complementarias

Las dos administraciones de los instrumentos de medida proporcionan también datos que informan sobre tendencias y acerca de los cuales se pueden formular hipótesis y contrastarlas.

En primer lugar, podemos conseguir información acerca del tiempo utilizado por los estudiantes para realizar las pruebas. Nuestra hipótesis en este sentido, a la que llamaremos **hipótesis C1**, es que en la primera medición, que se realiza el primer día de clase, los estudiantes saben que el evaluador no puede esperar de ellos que sepan más de lo que saben, de manera que invierten solamente el tiempo necesario para contestar a todas las preguntas y realizar la tarea que se les encomienda, pero no emplean más tiempo del estrictamente necesario. Sin embargo, en la segunda medición, que se realiza

a final de curso, esperamos que los estudiantes tarden más tiempo en completar las pruebas, por dos razones. La primera es que al tener más conocimientos, también dudan más sobre las posibles opciones de respuesta, cosa que no ocurre cuando solamente se piensa en una posibilidad; la segunda razón es que los estudiantes ya tienen unas expectativas que cumplir, quieren hacer lo que ellos creen que el evaluador espera que hagan.

Un segundo factor observable es el uso que hacen los estudiantes de las obras de consulta. Nuestra **hipótesis C2** es que, en la primera medición, cuando acaban de llegar a la Universidad, debido a la concepción de la traducción que tienen, que les hace pensar que se trata de una simple operación entre lenguas y que ya poseen los conocimientos necesarios de su lengua materna como para traducir bien, solamente utilizarán los diccionarios bilingües y, como mucho, los monolingües en la lengua extranjera, mientras que en la segunda medición, tras ocho meses de formación como traductores, esperamos que conozcan muchas más obras de consulta y que estén habituados a utilizarlas, por lo que suponemos que consultarán diccionarios de sinónimos, de uso o normativos en español, guías de estilo, etc. y quizás incluso echarán en falta algunas obras que no se encuentren en la mesa del profesor. Para contrastar esta hipótesis contamos con los siguientes datos:

- (a) la observación directa en las sesiones de pruebas, ya que al tener que acudir los sujetos a la mesa del profesor para coger cualquier obra de consulta que necesiten, se puede observar y anotar qué obras consultan y cuáles no;
- (b) la comparación de las respuestas de los ítems 6,7,8 y 9 del IPT en las dos mediciones: el análisis las respuestas a las preguntas 6,8 y 9 del IPT, que indagan sobre la obra de consulta más utilizada, sobre el resto de obras que han consultado y sobre otras “herramientas” u otro tipo de ayuda que hubiera resultado útil en caso de no estar en clase, corroborará si en la segunda medición existe más información que en la primera, mientras que la respuesta a la pregunta 7, que indaga acerca de si la obra de consulta más utilizada ha resuelto todas las dudas del sujeto, nos informará sobre si el estudiante piensa que es suficiente con consultar un diccionario o si cree conveniente consultar otras obras;
- (c) la comparación de las respuestas a los ítems 5 y 6 del ICGT en las dos mediciones, que piden al sujeto que ordene las acciones que lleva a cabo cuando encuentra un término o expresión que no comprende en el TO o una expresión que comprende pero no sabe cómo reexpresar en el TM, respectivamente. Las opciones que dan los

ítems 5 y 6 son (1) consultar el diccionario bilingüe, (2) consultar el monolingüe y (3) intentar extraer el sentido por el contexto (en caso del problema de comprensión) o (3) reformular el segmento hasta conseguir un equivalente (para el problema de reexpresión).

Un tercer factor cualitativo que se puede observar es el proceso de comprensión del TO, ya que comparando las respuestas a los ítems 1,2,3,4,5 y 10 del IPT (cf. Vol. II, páginas 88-91) en las dos mediciones, podemos observar los cambios de hábitos en dicho proceso. Comparando las respuestas en la primera y en la segunda administración de cada estudiante, estos ítems nos ofrecen información acerca de la evolución, a lo largo de ocho meses, de cuatro factores:

- (a) su capacidad de análisis del TO (se les pregunta cuál es la función del texto en el ítem 1 y cuál su género en el ítem 2);
- (b) su capacidad de síntesis (deben escribir en una frase la idea principal del texto, en el ítem 3 y las ideas secundarias en el ítem 4);
- (c) sus hábitos de lectura del TO (ítem 5);
- (d) su capacidad para detectar problemas de traducción (ítem 10).

Respecto al proceso de comprensión (entendiendo éste como la suma de los cuatro factores enumerados) nuestra **hipótesis C3** es que éste será más completo y profundo en la segunda administración, es decir que su capacidad de análisis, de síntesis y de detección de problemas aumentará, además de mejorar sus hábitos de lectura, leyendo, por ejemplo, el texto entero en vez de un párrafo antes de empezar a traducir o leyéndolo dos veces en vez de una.

El ítem 13 del IPT nos permite observar un cuarto factor, la verificación del TM, proceso en el que tiene una gran influencia la capacidad de autocritica y de distanciarse del texto escrito para poder releerlo como el lector final, prescindiendo del TO. La comparación de las respuestas de los sujetos a esta pregunta nos permitirá ver si existen cambios en sus hábitos tras ocho meses de formación como traductores. La **hipótesis C4** es que los sujetos tendrán más capacidad de autocritica y de verificación del TM en la segunda medición.

Por último, el ítem 14 del IPT indaga sobre la percepción de la dificultad de la traducción de los estudiantes. Se les pide que valoren, en una escala del 0 al 10, la dificultad del TO. La comparación de su valoración en las dos mediciones, teniendo en cuenta que se trata del mismo texto, nos informará acerca de la evolución de su percepción de la dificultad de la traducción. La **hipótesis C5** es que en la segunda

medición los sujetos considerarán el TO más difícil que en la primera, porque al tener más conocimientos sobre la traducción y más experiencia serán capaces de detectar problemas y dificultades que ocho meses antes no advertían, y el mismo texto ya traducido les parecerá más difícil.

1.4. Recapitulación

Dada la diversidad de las hipótesis planteadas, consideramos conveniente recapitular y exponer de forma clara las hipótesis que hemos formulado, así como aclarar el método que hemos utilizado para contrastar cada una, datos que muestran las figuras 10 y 11.

Hipótesis	Método de contrastación
A- Sensibilidad al cambio: si comparamos los resultados de cada sujeto en la primera y la segunda medición, habrá cambios en las tres medidas.	
A1 -Los resultados de los estudiantes mejorarán en todos los casos.	Comparación de las medias de las puntuaciones totales obtenidas en el ICGT, el IET y la parte cuantitativa del IPT de la muestra en la primera y la segunda administración.
A2 -Los resultados del IPT mostrarán una clara correlación con los resultados del IET desde el principio, y esta correlación se mantendrá en la segunda medición.	Correlación de Pearson entre los resultados obtenidos en el IET y en la parte cuantitativa del IPT por la muestra en cada una de las dos mediciones.
A3 - Los resultados del IPT y del IET no mostrarán una correlación con los resultados del ICGT en la primera medición, y los del IET continuarán sin correlacionar con los del ICGT en la segunda medición, pero los resultados del IPT aumentarán su correlación con el ICGT en la segunda medición.	Correlación de Pearson entre los resultados obtenidos en el ICGT y en la parte cuantitativa del IPT y el IET por la muestra en cada una de las dos mediciones.
B- Generalizabilidad: la combinación lingüística, el evaluador y el centro donde se imparten los estudios de traducción causarán una variabilidad reducida en los resultados obtenidos en los tres instrumentos.	
B1 - El IPT obtendrá resultados con muy poca variabilidad atribuible a la combinación lingüística, al evaluador y al centro donde se imparte la enseñanza.	Ánálisis de variancia, estimación de diversos componentes de la variancia y cálculo de correlaciones intraclasa a partir de los resultados obtenidos en cada medición del IPT (parte cuantitativa), analizando por separado la combinación lingüística, los evaluadores y la Universidad.
B2 - El IET obtendrá resultados con poca variabilidad atribuible a la combinación lingüística y al centro donde se imparte la enseñanza, mientras que la variabilidad atribuible al evaluador será mayor.	Ánálisis de variancia, estimación de diversos componentes de la variancia y cálculo de correlaciones intraclasa a partir de los resultados obtenidos en cada medición del IET, analizando por separado la combinación lingüística, los evaluadores y la Universidad.
B3 - El ICGT obtendrá resultados con muy poca variabilidad atribuible a la combinación lingüística y al centro donde se imparte la enseñanza.	Ánálisis de variancia, estimación de diversos componentes de la variancia y cálculo de correlaciones intraclasa a partir de los resultados obtenidos en cada medición del ICGT (parte cuantitativa), analizando por separado la combinación lingüística y la Universidad.

Figura 10. *Hipótesis principales acerca del funcionamiento de los tres instrumentos de medida y método de contrastación.*

La figura 10 se refiere a las seis hipótesis principales, y los métodos de contrastación propuestos son básicamente la comparación de las respuestas obtenidas por los estudiantes de modo global en la primera y en la segunda medición, realizando análisis de variancia, estimación de diversos componentes de la variancia y cálculo de correlaciones intraclase y de Pearson a partir de los resultados obtenidos en cada medición con los tres instrumentos. Todos los datos necesarios para estos análisis se han consignado en el programa de ordenador Excel, en su versión 97, mientras que los análisis estadísticos se han realizado mediante el paquete estadístico SPSS, versión 9.0.

La figura 11 recoge las hipótesis complementarias y los métodos de contrastación, que son la observación directa y la comparación de los porcentajes y las frecuencias de las respuestas a ciertos ítems de los instrumentos de medida en las dos mediciones, es decir que el resultado de estos análisis será la descripción de las tendencias de los sujetos respecto a varios factores, como el tiempo utilizado por los sujetos, su utilización de las obras de consulta o su proceso de comprensión del TO. Los análisis de estos datos, consignados en Excel 97, también se han realizado a través del paquete estadístico SPSS, que contempla tanto datos cualitativos (es decir, categorías) como cuantitativos.

Hipótesis	Método de contrastación
1-Tiempo utilizado. En la primera medición los estudiantes utilizarán menos tiempo para realizar las pruebas que en la segunda administración.	Observación directa (anotando el tiempo utilizado por cada estudiante cuando entrega el cuadernillo de pruebas).
2-Uso de obras de consulta. En la primera medición los estudiantes utilizarán especialmente los diccionarios bilingües y los monolingües en lengua extranjera, en detrimento del resto de obras de consulta, situación que cambiará en la segunda medición.	Observación directa (anotando qué obras de consulta se utilizan más y cuáles no se utilizan en absoluto) y comparación de los ítems 6,7,8 y 9 del IPT en las dos mediciones, además de los ítems 5 y 6 del ICGT.
3-Proceso de comprensión. En la segunda medición el proceso de comprensión del TO será más completo y profundo (más capacidad de análisis, síntesis y de detección de problemas; mejores hábitos de lectura)	Comparación de las respuestas de cada estudiante a los ítems 1,2,3,4,5 y 10 del IPT en las dos administraciones de este instrumento.
4-Proceso de verificación. En la segunda medición los sujetos tendrán más capacidad de autocritica y de verificación del TM.	Comparación de las respuestas de cada estudiante al ítem 13 del IPT en las dos administraciones.
5-Valoración de la dificultad del TO. En la segunda medición los sujetos considerarán el TO más difícil que en la primera.	Comparación de las respuestas de cada estudiante al ítem 14 del IPT en las dos administraciones.

Figura 11. *Hipótesis complementarias y método de contrastación.*

2. Diseño para el estudio de la generalizabilidad

Los resultados obtenidos en el estudio de generalizabilidad que hemos diseñado determinarán hasta qué punto se pueden generalizar los resultados obtenidos con los instrumentos, en qué medida existen o no factores aleatorios que puedan hacer cambiar los resultados por razones ajena a la característica que se pretende medir, ya que esperamos que las diferencias que se observen dentro de una misma medición sean debidas únicamente a los distintos niveles de competencia traductora de cada estudiante, y no a la combinación lingüística, al evaluador o al centro de formación donde estudia. Este punto es esencial si queremos que nuestros instrumentos puedan ser utilizados por otros docentes y llevar a conclusiones interesantes que después puedan aplicarse a la enseñanza de la traducción.

Según las hipótesis que hemos mencionado en el apartado anterior, las pruebas de fiabilidad, que deben determinar la variabilidad de los tres factores de confusión que pensamos que pueden influir en los resultados de los instrumentos, deben contar con un diseño que incluya (a) varios evaluadores (dándoles los mismos textos y cuestionarios para que determinen la puntuación del IET y del IPT); (b) varias facultades de traducción de distintas universidades, donde se administrarán los tres instrumentos; (c) varias combinaciones lingüísticas.

Respecto a los centros universitarios, una vez exploradas las posibilidades, decidimos realizar las pruebas en tres facultades de traducción españolas: la Facultat de Traducció i Interpretació de la Universitat de Vic, la Facultat de Ciéncies Humanes i Socials de la Universitat Jaume I de Castellón y la Facultat de Traducció i Interpretació de la Universitat Autònoma de Barcelona.

El diseño en este sentido es sencillo, ya que se trata de realizar dos mediciones en cada uno de los centros, con todos los estudiantes de primer curso matriculados en traducción del inglés al español en primer curso. Respecto a las combinaciones lingüísticas, decidimos incluir en las dos mediciones de la UAB a los estudiantes que se matricularon en traducción del francés y del alemán al español, en primer curso, cosa imposible en las otras dos facultades porque solamente ofrecen el inglés como primera lengua extranjera. De este modo, la muestra fue constituida por todos los estudiantes matriculados en primer curso el año académico 1998-99 en los tres centros mencionados, que eligieron traducir al español, independientemente de su primera lengua extranjera.

El tercer elemento del que queríamos cuantificar la variabilidad, los evaluadores, era algo más complejo, ya que para conocer el grado de variabilidad debíamos utilizar al menos dos evaluadores para cada centro, de manera que tanto el IPT como el IET de cada estudiante (en cada medición) fueran corregidos por dos evaluadores, que debían ser los mismos en las dos mediciones. De este modo, nuestro diseño cuenta con dos evaluadores por cada centro y por cada combinación lingüística para las pruebas del IPT y el IET. El diseño global se puede observar en la siguiente figura:

Universidad 1		Universidad 2		Universidad 3					
Inglés		inglés		inglés		francés		alemán	
Eval. 1	Eval.2	Eval.3	Eval.4	Eval.5	Eval.6	Eval.7	Eval.8	Eval.9	Eval.10
Grupo A		Grupo B		Grupo C		Grupo D		Grupo E	

Figura 12. *Diseño del estudio de generalizabilidad*

La figura 12 muestra que existen cinco grupos de sujetos, tres de inglés español (a, b y c), uno de francés español (d) y uno de alemán español (e); para cada grupo de sujetos existen dos evaluadores (eval.), pertenecientes al centro de donde procede la muestra cuyos datos deben corregir. En todos los casos, el evaluador al que se asigna un número impar es además el profesor del grupo, mientras que los evaluadores 2, 4, 6, 8 y 10 son otros docentes cualificados para corregir las pruebas que no impartieron clases a ese grupo de estudiantes en concreto.

Este diseño responde tanto a la primera medición (octubre 1998) como a la segunda medición (mayo 1999), de forma que los resultados de las dos mediciones permiten valorar el grado de variabilidad atribuible a cada uno de los tres elementos (universidades, evaluadores y combinaciones lingüísticas) para el IPT y para el IET.

El ICGT es diferente porque se trata de un instrumento cuya puntuación no deja opción a la subjetividad. Es un cuestionario con respuestas muy bien acotadas que no dejan margen de elección al evaluador, y si se diseñara un sistema de corrección automática podría funcionar perfectamente, de manera que es absurdo hacer que dos evaluadores corrijan cada ICGT y optamos por descargarlos de un trabajo innecesario, dado que su colaboración es totalmente desinteresada. Sin embargo, sí que probamos el ICGT en las tres universidades y con todas las combinaciones lingüísticas, para ver si existía algún grado de variabilidad atribuible a estos otros dos factores.

Por otra parte, este diseño puede permitir la detección de posibles factores de confusión que hasta ahora no hemos tenido en cuenta y que merezca la pena contemplar en lo sucesivo, como el expediente académico de los sujetos, la madurez intelectual o las distintas vías de adquisición de la competencia en lengua extranjera (padres bilingües, realización de la primaria y/o la secundaria en un colegio bilingüe, estancias en el extranjero, academias privadas, etc.).

Con este diseño, cada estudiante es evaluado por dos evaluadores, con lo que estas dos fuentes de variación están completamente cruzadas. Además, ambas están anidadas, es decir, no repetidas, dentro de los factores “Universidad” y “combinación lingüística”. Esta situación reproduce al máximo la realidad, y aunque podríamos haber cruzado “evaluador” con “Universidad” y con “combinación lingüística”, haciendo que los mismos dos evaluadores corrigieran todas las pruebas, esto se alejaría mucho de la situación real, en que los profesores nunca evalúan a alumnos que no estén matriculados en su Universidad y, menos aún, en su combinación lingüística. Lo lógico en un experimento de estas características es buscar la máxima validez externa y que los profesores corrijan a sus propios alumnos, al tiempo que todos los evaluadores evalúan alumnos de su Universidad y de su combinación lingüística.

3. Diseño para el estudio de la validez

Hemos mencionado ya en numerosas ocasiones el término “validez”, pero llegados a este punto nos gustaría aclarar que al referirnos a la validez de los instrumentos estamos hablando de la validez psicométrica, que es diferente del concepto de validez interna o externa de la metodología de la investigación.

La validez psicométrica es sinónimo de “utilidad”, es decir que indica si el instrumento construido es o no útil, si las conclusiones o inferencias que se extraerán de sus resultados estarán avaladas por datos empíricos. Si bien existen muchos tipos de validez, así como muchas pruebas para conocer esos tipos de validez, nosotros solamente mencionaremos la más importante para estos instrumentos en concreto, ya que sería absurdo, por ejemplo, intentar analizar la validez de criterio o predictiva de un instrumento que no pretende ser predictivo.

3.1. Validez de contenido

El conjunto de los tres instrumentos se ha sometido a dos pruebas de validez de contenido, ambas basadas en el criterio de los expertos en el campo.

En primer lugar, a lo largo del proceso de elaboración de los tres instrumentos, hemos utilizado un método de validación de contenido, que nos ha ayudado a ir realizando las modificaciones pertinentes en cada uno de los instrumentos. Nos referimos a las entrevistas con expertos en el campo de la traducción, algunas propiciadas por presentaciones en varios congresos internacionales y otras completamente casuales.

La presentación de comunicaciones y ponencias en diversos congresos, donde expusimos los instrumentos –la versión que manejábamos en ese momento- o donde realizamos reflexiones teóricas sobre la operacionalización del constructo, llevaron a muchas discusiones interesantes con el público asistente a la sesión y con otras personas que más tarde se acercaron a comentar el tema, de manera que obtuvimos muchos y valiosos consejos, tanto de personalidades reconocidas en el mundo de la traducción como de otros profesionales igualmente interesados por los instrumentos.

Por otra parte, también se han dado bastantes encuentros fortuitos con expertos en los que, tras un intercambio de opiniones, hemos llegado a ideas interesantes que después hemos aplicado, o simplemente nos han mostrado su interés y su apoyo en esta investigación, lo cual también ha sido de gran ayuda. En resumidas cuentas, podemos

decir que hemos ido realizando una “consulta continua” con expertos, si bien no ha sido en absoluto sistemática.

El segundo método para estudiar la validez del contenido ofrece datos concretos analizables, y se basa en las respuestas de varios expertos a un cuestionario que elaboramos, que se puede consultar en las páginas 156-163 del volumen II.

El cuestionario se entregó a diez expertos en traducción, que además son traductores, y que están interesados por la investigación sobre la adquisición de la competencia traductora, es decir que están familiarizados con el tema. Este test, que consta de 15 preguntas de elección múltiple pero también cuenta con un espacio para que los expertos anoten las razones que les han llevado a elegir una respuesta y no otra, indaga sobre la opinión de los expertos respecto a si los tres instrumentos miden realmente lo que queremos medir o no.

Para obtener los datos en cuestión, en el cuestionario se hace referencia, en primer lugar, al ICGT. El cuestionario se centra en la parte cuantitativa del ICGT, que consta de 12 ítems, y se especifica la intención de las preguntas, es decir, que los expertos nos digan hasta qué punto opinan que las preguntas del ICGT valoran lo que nosotros creemos que valoran. Respecto al IPT, y refiriéndonos también a su parte cuantitativa únicamente, hay una pregunta global (ítem 13), que explica cómo se desarrolla todo el ejercicio, el encargo de traducción, el texto y la pregunta 12, que es la que hace discernir entre un “0”, un “1” o un “2” junto con la traducción del estudiante. Para el IET también se formula un solo ítem global (pregunta 14), y finalmente hemos incluido un ítem final, el número 15, donde preguntamos por los tres instrumentos en conjunto, con lo cual también obtendremos información acerca de la adecuación de nuestra operacionalización. Por consiguiente, los resultados de estos cuestionarios aportarán datos tangibles y conclusiones respecto a la validez de contenido de los tres instrumentos.

3.2. Validez consecuencial

Si bien en nuestro caso la validez consecuencial no tiene demasiada pertinencia, puesto que los resultados de los instrumentos de medida sólo influirán en el docente de traducción, y no en los estudiantes, hemos preferido nombrarla porque nos parece conveniente que quede claro que todas estas pruebas no repercutirán de manera negativa en ningún estudiante, es decir que no solamente no se tendrán en cuenta los resultados de cada uno para la evaluación académica, sino que en todo momento se ha mantenido

la confidencialidad de su identidad. En efecto, solamente el investigador principal conoce el nombre de cada sujeto, al que se asignó un código que es el que han visto en todo momento los evaluadores. Por otra parte, no ha habido pérdida de clases de traducción para realizar las pruebas, ya que los ejercicios realizados por los estudiantes (rellenar cuestionarios y hacer traducciones cortas) durante las clases de traducción, que han ocupado en total cuatro horas lectivas del curso 1998-1999, se ajustan a los contenidos de la asignatura de traducción general directa.

El sesgo, es decir, la posibilidad de que estos instrumentos marginen a minorías por razones diversas, tampoco es pertinente en nuestro caso, porque la muestra está constituida por toda la población de las tres facultades, que son los estudiantes de primer curso que han elegido el español como lengua materna, sea cual sea la lengua extranjera, y solamente se ha eliminado del estudio a las personas en las que podían influir variables extrañas como el mayor conocimiento de la lengua extranjera (personas extranjeras, con lengua materna diferente del español) o personas con una mayor madurez intelectual (de edad superior a los 18 o 19 años) de modo que solamente se ha procedido a controlar variables con una finalidad clarificadora.

Por otra parte, estos instrumentos no están pensados para ser autoadministrados y que los estudiantes los realicen por su cuenta, ni para servir como herramientas de evaluación sumativa, sino que su finalidad es la de ayudar a los docentes de la traducción a conocer mejor el proceso de adquisición de la competencia traductora de sus estudiantes, así como para darles la oportunidad de comparar métodos docentes. Es decir que en un futuro estos instrumentos podrían repercutir de manera beneficiosa para los mismos estudiantes, redundando en una mejor enseñanza de la traducción.

4. Validación cruzada

La validación cruzada es una forma de asegurarse de que los resultados conseguidos con una serie de ítems son estables. La técnica utilizada para realizar una validación cruzada consiste en obtener, para una medición, el doble de la muestra deseada para poder repetir, con cada mitad, el cálculo de los mismos índices estadísticos. Si los resultados obtenidos son similares, este es un dato muy favorable, ya que esta prueba equivale a una réplica; en efecto, se trata de repetir las mismas pruebas con una muestra equivalente, solo que en el mismo momento, sin que haya un lapso de tiempo entre una medida y la otra.

Con el fin de realizar esta prueba, en la primera medición con los tres instrumentos realizada en la UAB, se administraron los instrumentos de medida a los dos grupos de traducción del inglés al español, en vez de administrarlos sólo a uno.

5. Sujetos

5.1. Tamaño de la muestra

Como ya hemos explicado con anterioridad, en las pruebas participaron la totalidad de los alumnos matriculados en la asignatura de traducción del inglés al español, francés al español y alemán al español de las tres facultades de traducción de las universidades UAB, UV y UJI, salvo, evidentemente, aquellos que no asistieron aquel día a clase. Sin embargo, una vez realizadas las pruebas y ayudándonos de los datos que los mismos sujetos nos brindaron, separamos los tests realizados por aquellos sujetos nacidos antes de 1979, que, por lo tanto, tenían más de 19 años en el momento de realizar las pruebas, para poder controlar las variables de madurez intelectual y experiencia previa en traducción.

Excluimos del estudio asimismo los cuestionarios realizados por personas cuya lengua materna no era el español, como por ejemplo los estudiantes de intercambio procedentes de otras universidades europeas, y finalmente quedaron también excluidos los cuestionarios incompletos, anónimos o sin fecha de nacimiento, ya que por razones obvias no eran de utilidad. Éstos últimos están señalados como “incompletos” en la figura 13, que muestra la cantidad de sujetos que participaron en las pruebas y cuántos cuestionarios fueron incluidos en el estudio y analizados finalmente. Aparecen en la misma figura los datos referidos a la primera medición, entre paréntesis, y los datos correspondientes a la segunda medición, sin paréntesis.

Centro/Idioma	Universidad 1 Inglés	Universidad 2 Inglés	Universidad 3 Inglés		Francés	Alemán
			Grupo 1	Grupo 2		
Sujetos nacidos después de 1979 (incluidos)	35 (51)	19 (26)	18 (27)	16 (25)	9 (15)	2 (10)
Sujetos nacidos antes de 1979 (excluidos)	12 (18)	5 (8)	9 (15)		5 (13)	6 (10)
Cuestionarios incompletos	2 (4)	2 (0)	2 (1)		4 (0)	0 (1)
Total sujetos	49 (73)	27 (34)	45 (68)		18 (28)	8 (21)
Total sujetos incluidos en el estudio	35 (51) 71 (70) % del total	19 (26) 70 (76) % del total	34 (52) 76 (76) % del total		9 (15) 50 (54) % del total	2 (10) 25 (48) % del total

Figura 13. *Sujetos participantes en las dos administraciones de las pruebas de fiabilidad y validez.*

Tal como muestra la figura, en la primera medición se analizaron entre un 70 % y un 76 % de los cuestionarios en inglés, un 54 % de los cuestionarios en francés y un

48 % del grupo de estudiantes de la combinación alemán español. Estos datos son interesantes porque confirman la creencia generalizada de que en nuestras licenciaturas los estudiantes más jóvenes entran mayoritariamente por inglés, mientras que el francés y el alemán van quedando cada vez más relegados a estudiantes de más edad, muchos de los cuales acceden a la Universidad por la vía de “mayores de 25 años”. Estas distinciones podrían llegar a generar variabilidad adicional en otros estudios, de modo que es importante tenerlas en cuenta.

En cualquier caso, de los 224 estudiantes participantes en las pruebas de la primera administración, en total se analizaron 154 cuestionarios, de los cuales 129 correspondían a la combinación inglés español.

Las pruebas de la segunda administración de los tres instrumentos de medida se realizaron un día de clase de las dos últimas semanas de curso, durante el mes de mayo de 1999; el día fue acordado previamente con los profesores correspondientes y tomaron parte en las pruebas los estudiantes que asistieron a esa clase ese día. Sin embargo, la cantidad de estudiantes fue menor a la de la primera medición. La explicación a esta pérdida de sujetos es bastante sencilla, ya que se trata de la cantidad de alumnos que asisten a clase regularmente, que suele disminuir a final de curso habitualmente por razones varias (cansancio de los estudiantes, acumulación de trabajos de curso que deben presentar en las últimas semanas, preparación para los exámenes, etc.).

En este caso se han cumplido los datos habituales de los procesos poco controlados en los que se toman dos medidas, es decir, una pérdida del 30 % de los sujetos en una segunda medición. Si se examina la figura 13, donde aparece entre paréntesis la cantidad de sujetos que participaron en la primera medición y fuera del paréntesis la cantidad correspondiente a la segunda medición, se puede ver cómo en la Universidad 1 se perdió un 31 % de los sujetos, mientras que en la Universidad 2 se perdió un 27 %, y en la Universidad 3 la pérdida fue del 32 % de la muestra. Si se comparan los datos se verá que el tanto por ciento de cuestionarios incluidos en el estudio respecto al total de participantes apenas varía en los grupos de inglés, mientras que varía en gran medida en los grupos de francés y alemán, en los que además la participación fue mucho menor al 70 % respecto a la primera medición. En el grupo de francés los cuestionarios válidos fueron 9, frente a 15 de la primera medición, es decir que hubo una pérdida del 40 por ciento de los sujetos, pero aún así los datos se pueden tener en cuenta, mientras que en el grupo de alemán el resultado fue peor, ya que sólo hubo dos cuestionarios válidos, es decir que hubo una pérdida del 75 % de los sujetos,

lo cual inhabilita los datos. En nuestra opinión, estos datos son indicativos de que existe una diferencia en las características generales de los sujetos que eligen las combinaciones lingüísticas alemán español y francés español respecto a los estudiantes de inglés español, y quizás estas características, que es no sólo interesante sino necesario investigar, tengan que ver con la edad, con el nivel de competencia en la lengua extranjera y con el nivel de exigencia de las asignaturas relacionadas con la traducción de la lengua extranjera.

5.2. Homogeneidad y representatividad de la muestra

Uno de los requisitos para que los resultados obtenidos mediante nuestros instrumentos puedan generalizarse a una población más extensa de estudiantes es la homogeneidad y la representatividad de las muestras. Por homogeneidad, en este caso, entendemos que los sujetos del estudio reúnan las características ya mencionadas de edad y lengua materna, es decir, que tengan entre 18 y 19 años y que su lengua materna sea el español. Estos requisitos están asegurados por el procedimiento empleado de separar todos los cuadernillos llenados por sujetos de edad o lengua materna inadecuada.

Respecto a la representatividad, pretendemos que la muestra sea representativa de todo el conjunto de estudiantes de traducción de las facultades españolas, y si bien somos conscientes de que no podemos asegurarla mediante una muestra de tres facultades, sí que nos pareció interesante conocer los perfiles académicos de los estudiantes. Aunque no existe un perfil académico concreto que defina al estudiante de traducción, sí que existen rasgos que conocemos al respecto. Por ejemplo, sabemos que el perfil académico varía entre centros, de acuerdo con su ubicación geográfica, con la concentración de centros universitarios que ofrecen la misma titulación en la misma zona o en una zona próxima y, por último, según se trate de un centro público o privado. Dadas las características de nuestra muestra, pensamos que ésta es representativa en todos los aspectos mencionados. Respecto a la ubicación geográfica, contamos con un centro muy próximo a una capital de Comunidad Autónoma (la FTI de la UAB), un centro situado en una capital de Provincia (la FTI de la UJI) y un centro situado en una capital de comarca (la FTI de la UV). En cuanto a la concentración de centros que ofrecen la misma titulación, contamos con un centro que es el único en su Comunidad Autónoma (la FTI de la UJI), otro centro que “compite” con otra Universidad en la misma ciudad (la FTI de la UAB, que tiene muy cerca la FTI de la Universidad Pompeu

Fabra) y finalmente la FTI de la UV, que por un lado se encuentra en una Comunidad Autónoma donde existen otros dos centros que ofrecen la misma titulación pero por otro lado se encuentra lo suficientemente alejada para atraer estudiantes de otras provincias de la misma Comunidad. Por último, respecto al carácter público o privado de los centros, en nuestra muestra contamos con estudiantes de dos centros públicos (las FTI de la UAB y de la UJI) y uno privado (la FTI de la UV). Así pues, nos parece que la muestra de nuestro estudio es representativa de los estudiantes de traducción en general, y tiene sentido ver cuál es el perfil académico de los estudiantes para conocer qué elementos definen, desde esta perspectiva, a la muestra. El gráfico 4 muestra el perfil académico de cada grupo de estudiantes, clasificados según la universidad en la que estudian y la nota de acceso a la Universidad de cada estudiante, que es la media entre el expediente de toda la secundaria y la nota media obtenida en el examen de selectividad. Los datos que figuran corresponden a los estudiantes de la combinación inglés español que fueron incluidos en el estudio en los tres centros, con la excepción de cuatro casos, cuyos datos no pudimos obtener por problemas administrativos.

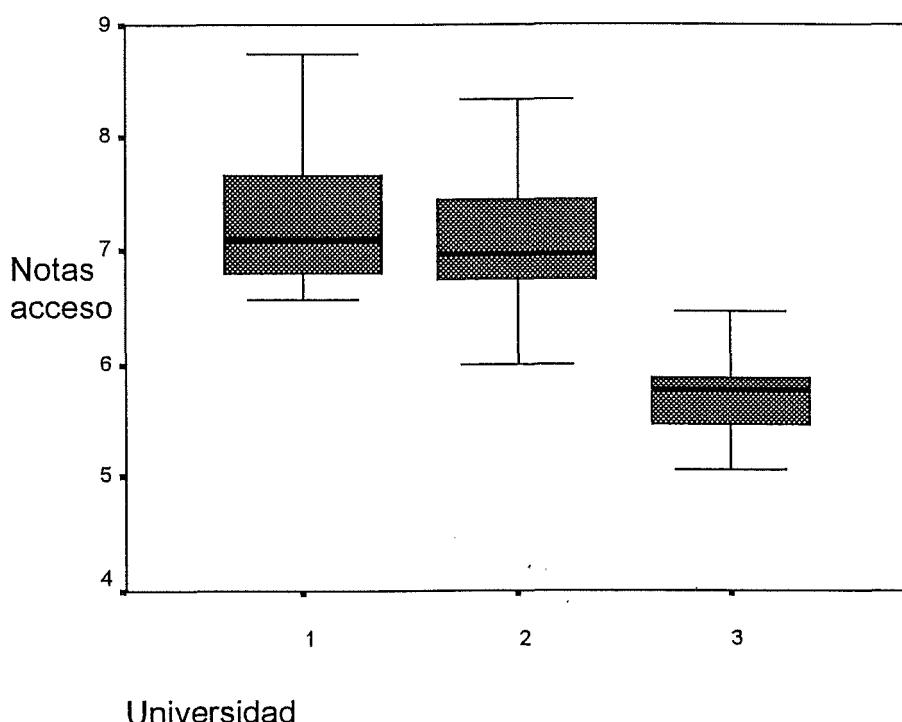


Gráfico 4. *Diagrama de caja de la comparación de las notas de acceso a la Universidad de los sujetos.*

En el gráfico 4 se puede observar cómo los grupos de sujetos de cada Universidad son homogéneos, si bien existen algunas diferencias entre las tres universidades. En este sentido, el gráfico parece confirmar nuestra idea de que cada centro responde a unos parámetros concretos respecto a la ubicación geográfica, etc. y que éstos influyen en el perfil académico de los estudiantes, ya que cada centro cuenta con un grupo de estudiantes homogéneo, con una distribución muy pareja y parece contar con sus propias características respecto al perfil académico de sus estudiantes.

El resultado de las diferencias entre los estudiantes de los tres centros, es que nuestra muestra está compuesta por sujetos con perfiles académicos variados, que van desde el 5 hasta el 9 como nota media y que están concentrados en dos zonas distintas: (a) entre 5'5 y 6; (b) entre 6'7 y 7'5. Por consiguiente, podemos afirmar que nuestra muestra es representativa respecto a los posibles perfiles académicos de los estudiantes de traducción.

6. Procedimiento

6.1. Primera administración

La primera administración de los tres instrumentos de medida se llevó a cabo el primer día de clase del curso académico en los tres centros participantes, la Facultat de Traducció i Interpretació de la Universitat de Vic, la Facultad de Ciències Humanes i Socials de la Universitat Jaume I de Castellón y la Facultat de Traducció i Interpretació de la Universitat Autònoma de Barcelona.

Todas las pruebas se realizaron durante los primeros diez días de octubre de 1998, en una sesión de dos horas de duración y con la presencia del profesor habitual de la asignatura y la de la investigadora principal, en las aulas que habitualmente se utilizan para esas clases.

El proceso de aplicación de los instrumentos fue el siguiente: el profesor de la asignatura, acompañado de la investigadora principal, se presentaron ante los estudiantes al comenzar la clase y explicaron que la primera clase del curso estaría dedicada a realizar una serie de pruebas de dos horas de duración que consistían en llenar un cuestionario y seguir las instrucciones que el mismo contenía. La única excepción fueron los grupos de traducción de francés y alemán, en los que la investigadora principal solamente se presentó al principio de la sesión para asistir a la presentación de las pruebas, dado que las clases de traducción de inglés, francés y alemán en la UAB son simultáneas, y no podía estar presente en las tres a la vez. Por lo tanto, los profesores de las asignaturas de traducción de francés y alemán permanecieron con sus respectivas clases y la investigadora se quedó en el grupo de inglés, que incluía a los dos grupos de traducción del inglés, con el fin de poder realizar la validación cruzada.

Se informó a los estudiantes de que los datos que se iban a recoger serían confidenciales, de que tenían derecho a no colaborar y de que los resultados servirían para una tesis doctoral que investigaba sobre la adquisición de la competencia traductora. Se les mostraron las obras de consulta que había sobre la mesa del profesor y se les indicó que, cuando llegaran a la parte de la traducción, podían utilizarlas siempre y cuando anotaran en el espacio señalado en el cuestionario qué términos consultaban.

Se repartieron los cuadernillos de pruebas de la combinación lingüística correspondiente, que se pueden consultar en las páginas 125-133, 134-142 y 143-151 del volumen II (versión en inglés, en francés y en alemán respectivamente) y se dieron

por comenzadas las dos horas. Los estudiantes empezaron a llenar los cuestionarios, y no hubo ningún estudiante que no quisiera colaborar, de manera que todos los presentes ese día realizaron las pruebas.

En todos los casos hubo estudiantes que levantaron la mano para preguntar, especialmente sobre el encargo de traducción, la mayoría querían saber hasta qué punto tenían que hacer caso. La respuesta en todos los casos fue la misma: sigue las instrucciones del cuestionario y utiliza tu propio criterio. También se indicó a los sujetos que si lo deseaban podían escribir notas al pie de la traducción justificando alguna de sus decisiones de traducción.

El uso de las obras de consulta fue bastante simultáneo, ya que la mayoría de estudiantes terminaron de llenar el ICGT a los 15 o 20 minutos, y en cuanto llegaron a la página 4 del cuadernillo fueron a buscar obras de consulta a la mesa del profesor. En el caso de los grupos de inglés de las tres universidades, la investigadora fue tomando nota del uso realizado por los estudiantes, fijándose especialmente en las obras que no consultó ningún alumno.

La gran mayoría de estudiantes terminaron antes de las dos horas fijadas como límite, y a medida que se recogían los cuestionarios la investigadora fue anotando el tiempo que había tardado cada sujeto en realizar la prueba completa. Al recoger las pruebas se pidió a los sujetos que entregaran también los borradores de la traducción que hubieran podido realizar, de manera que no se quedaran con nada que les recordara al texto, del que no se volvió a hablar en los ocho meses siguientes, hasta el día de la segunda medición.

6.2. Cambios realizados en el IPT

Para la segunda medición realizada con los tres instrumentos de medida, se llevaron a cabo pequeños cambios, en su mayoría prácticos, en el ICGT y el IPT, con el fin de facilitar la comprensión de los sujetos de algunas instrucciones o de aclarar las tareas. Estas pequeñas modificaciones están reflejadas en la versión definitiva de los instrumentos, unidos en un mismo cuadernillo, que se puede consultar en las páginas 125-133 del volumen II. Las modificaciones en concreto son las siguientes:

- (a) Cambios de formato: en el ICGT, que incluye las instrucciones generales para los instrumentos, se añadió la frase “cuando quieras, puedes empezar”, en negrita y detrás de una línea de puntos, para que quede más claro que en ese punto terminan las instrucciones y comienza el cuestionario; por otra parte, el texto que deben

traducir los sujetos se enmarcó en un cuadro con línea más gruesa, para que se viera mejor a primera vista; en el ítem 14 del IPT, en que se pide a los estudiantes que sitúen en una barra numerada del 0 al 10 el nivel de dificultad que presenta el texto que han traducido, se añadieron las palabras “muy fácil” y “muy difícil” a cada uno de los extremos de la barra, para que quedara claro a primera vista que el “0” era lo más fácil y el “10” el más difícil.

- (b) Contenido: en la página 4 del cuadernillo, donde se pide a los estudiantes que anoten las obras de consulta que van utilizando, se cambió la denominación de la segunda y la tercera columna, con el fin de que no hubiera posible lugar a duda. La segunda columna pasó a ser “editorial”, ya que en la primera medición hubo estudiantes que perdieron un buen rato buscando los autores de varios diccionarios, y la tercera columna pasó a indicar simplemente “término consultado”, ya que en la primera medición hubo alumnos que anotaron el número de página del diccionario consultado y otros resultaron no saber lo que era una “entrada”.
- (c) Nuevo dato: en esta nueva administración decidimos numerar los cuestionarios, de manera que si parecía obvio que dos estudiantes se habían “copiado”, se pudiera comprobar si estaban sentados al lado, para lo cual las pruebas tenían que repartirse en orden y la gente que entraba tarde debía sentarse al fondo de la clase para recibir cuestionarios con numeración posterior.

6.3. Segunda administración

La segunda medición se realizó en las dos últimas semanas de clase del curso, en una sesión de dos horas, para los tres centros y para las tres combinaciones lingüísticas. Todas las pruebas tuvieron lugar entre el 11 y el 25 de mayo de 1999. El procedimiento fue idéntico al utilizado en la primera medición, con la única salvedad de los cambios ya especificados en el apartado anterior, de manera que en las pruebas estuvieron presentes los profesores de las asignaturas y la investigadora principal, a excepción de los grupos de alemán y francés, ya que, como hemos explicado con anterioridad, las clases de traducción del inglés, del francés y del alemán son simultáneas y por lo tanto es imposible estar presente en las tres.

En esta ocasión, tal como hemos señalado al hablar de la muestra, la cantidad de estudiantes que asistieron fue menor a la de la primera medición, pero eso no impidió el transcurso normal de las pruebas, que se dio del mismo modo que en la primera

medición, y las únicas diferencias fueron la cantidad de tiempo empleada por los estudiantes y el uso de las obras de consulta.

Cuando terminaron las dos horas, se recogieron los cuestionarios de aquellos estudiantes que todavía no los habían entregado, que fueron dos en total, y se explicó a los estudiantes que durante el próximo curso se les comunicarán los resultados de esta investigación. También se les agradeció su colaboración desinteresada y se les explicó que constarían en los agradecimientos de esta tesis doctoral. Algunos estudiantes estuvieron preguntando, al finalizar la clase, sobre la utilidad de las pruebas, sobre la traducción “correcta” de algunos segmentos del texto, etc., es decir que en esta ocasión las pruebas despertaron la curiosidad de algunos sujetos.

6.4. Proceso de recogida de datos y corrección

El método para la recogida de datos fue el siguiente: una vez terminada la primera medición, se recogieron todos los cuestionarios, de los tres centros y las tres combinaciones lingüísticas, y se descartaron los cuestionarios que no cumplían las condiciones de admisión en el estudio. A continuación se procedió a asignar un código a cada estudiante, de manera que los datos reales fueran siempre confidenciales. Los códigos fueron asignados arbitrariamente, con la única precaución de que la primera letra del código coincidiera con la inicial del centro a donde pertenecían los estudiantes. De este modo, los sujetos de la Facultat de Traducció de Vic son “V-1”, “V-2”, etc., los de la Facultat de Ciéncies Humanes i Socials de la Universitat Jaume I de Castellón son “C-1”, “C-2”, etc., los alumnos de inglés de la Facultat de Traducció de la Universitat Autònoma de Barcelona son “AI-1”, “AI-2”, los de francés “AF-1”, “AF-2” y los de alemán “AA-1”, AA-2”, etc. Todos estos códigos se copiaron en todas las páginas de los cuestionarios correspondientes, de manera que al fotocopiar páginas sueltas de cada cuestionario quedara claro quién era el autor, en forma de código. Los mismos códigos se respetaron para la segunda medición, de modo que se pudiera comparar la actuación de cada sujeto en las dos mediciones.

Una vez contabilizados los sujetos de cada centro, se repartió a cada evaluador el siguiente material:

- (a) instrucciones de corrección del IPT de su combinación lingüística (cf. páginas 103, 105 y 106 del volumen II para inglés, francés y alemán, respectivamente) y del IET (cf. página 121 del volumen II);

- (b) plantillas de corrección para el IPT y el IET, es decir, tablas en blanco que debían llenar con los resultados de cada estudiante;
- (c) copia de las páginas 5 y 9 de cada cuadernillo (en las que previamente se había anotado el código correspondiente al estudiante), donde se encuentra la traducción realizada por el sujeto y sus respuestas a la pregunta 12, es decir, el material necesario para corregir el IET y la parte cuantitativa del IPT.

Evidentemente, ya se había realizado una reunión previa de los evaluadores con la investigadora principal en cada centro, de manera que todos los evaluadores sabían cómo debían proceder y estaban de acuerdo en realizar las correcciones en cuestión. El ICGT y la parte cualitativa del IPT fueron corregidos por la investigadora, siguiendo las instrucciones de corrección (cf. páginas 47-49 y 107 del volumen II para las instrucciones de corrección del ICGT y de la parte cualitativa del IPT, respectivamente).

Respecto al procedimiento tras la segunda medición, fue idéntico al ya referido, a excepción de que, dado que los códigos ya estaban asignados a los estudiantes, solamente hubo que copiar los códigos ya asignados y descartar, además de los cuestionarios incompletos o excluidos del estudio, aquellos que no estaban repetidos, es decir, los cuestionarios de estudiantes que no habían realizado las pruebas en la primera medición. A continuación se volvió a enviar a cada evaluador todo el material ya mencionado. A medida que los evaluadores terminaban sus correcciones, las iban enviando a la investigadora principal, que iba introduciendo los datos en hojas de cálculo del programa Excel, de manera que todos los datos se iban almacenado en dos bases de datos diferenciadas, una para las pruebas de la primera medición y otra para las de la segunda medición.

En el período transcurrido entre las dos mediciones se fueron recogiendo y procesando otros datos, como las notas de acceso a la Universidad de los sujetos, el tiempo utilizado para llenar los cuestionarios, etc. Asimismo, se pasaron los cuestionarios de validez de contenido (cf. páginas 156-163 del volumen II) a diez expertos y se codificaron los resultados.

Durante el proceso de corrección, algunos evaluadores realizaron comentarios de gran interés sobre dicho proceso, que serán detallados al analizar los resultados de las pruebas de validez y fiabilidad, ya que ofrecen pistas sobre las posibles razones de algunos de los resultados.

CAPÍTULO 6- PRESENTACIÓN Y DISCUSIÓN DE LOS RESULTADOS DE LA VALIDACIÓN

1. Resultados del estudio de validez de contenido

Este estudio se ha realizado mediante un cuestionario, que se puede ver en el volumen II (páginas 156-163), y que se envió a diez expertos, que son profesores de traducción en la Universidad al tiempo que traductores y que además están interesados en el estudio de la competencia traductora.

El cuestionario consta de 15 preguntas, y en cada una de ellas se pide al experto que valore hasta qué punto una de las preguntas o ejercicios de los instrumentos de medida creados responde a los objetivos planteados inicialmente por los investigadores. A cada pregunta, los expertos pueden responder (a) mucho; (b) poco o (c) nada, según piensen que la pregunta o ejercicio del instrumento en cuestión es muy adecuado, poco adecuado o nada adecuado para obtener la información que se pretende obtener.

La manera de corregir estos cuestionarios ha sido la siguiente: se han otorgado 2 puntos a todas las respuestas (a), es decir, cuando el experto ha contestado “mucho”, 1 punto si el experto ha contestado (b), “poco” y 0 puntos en caso de que el experto haya contestado (c) “nada”. Como se ha dado el caso de varios expertos que en más de una ocasión han contestado “depende” y han expuesto las razones que les llevan a dudar de la adecuación de la pregunta, en estos casos hemos otorgado 1 punto, por considerar que esta respuesta se parece más a “poco” que a las otras dos posibilidades.

Para interpretar los resultados de este estudio, es necesario mencionar que las doce primeras preguntas del cuestionario para expertos corresponden a las doce preguntas del ICGT que puntuán en la corrección cuantitativa de este instrumento, la pregunta número 13 se refiere a la valoración global del IPT (exclusivamente de su parte cuantitativa), la número 14 a la valoración global del IET y la última pregunta pide al experto que valore de forma global hasta qué punto considera que los tres instrumentos miden la competencia traductora de los estudiantes.

La siguiente figura muestra la frecuencia de respuestas de los expertos a cada una de las 15 preguntas del cuestionario:

Pregunta del cuestionario	0 (nada)	1 (poco)	2 (mucho)
1		20	80
2	20	20	60
3		40	60
4		20	80
5		20	80
6		20	80
7		40	60
8		40	60
9		40	60
10		40	60
11	20	40	40
12		80	20
13		20	80
14		20	80
15		40	60

Figura 14. *Porcentajes de respuestas sobre la adecuación de las preguntas del cuestionario a sus objetivos.*

Los resultados del estudio de validez de contenido son, como se puede apreciar a simple vista, positivos, ya que solamente en dos preguntas, la 2 y la 11, referidas respectivamente a los ítems 8 y 12 del ICGT (cf. páginas 126 y 127 del volumen II), ha habido un 20 % de los expertos que han considerado que el contenido de dichos ítems no valoraba en absoluto lo que nosotros pretendemos valorar mediante ellos. Aún así, un 60 % y un 40 % de los expertos, respectivamente, opinan que el contenido de estos mismos ítems sí que refleja lo que queremos medir.

Aparte de estas dos preguntas, en que un 20 % de los expertos ha contestado “nada”, no existe ninguna otra pregunta del ICGT, de la parte cuantitativa del IPT o del IET que uno solo de los expertos considere que no mide lo que pretendemos medir. Por otra parte, en 6 de los ítems un 80 % de los expertos han considerado que el contenido de las preguntas coincidían con el objetivo perseguido, y lo mismo ocurre en otros 7 ítems, en que un 60 % de los expertos consideran el contenido de las preguntas adecuado. Por consiguiente, podemos decir que en 13 de las 15 preguntas, los expertos consideran mayoritariamente (entre un 60 % y un 80 %) que las preguntas miden lo que pretendemos medir.

Entre los distintos ítems, conviene destacar el contenido de tres, los número 13, 14 y 15. El ítem 13 indaga sobre la adecuación de contenido de la parte cuantitativa del IPT así como sobre la forma de puntuación, y un 80 % de los expertos ha considerado que este instrumento valora “mucho” la actuación de los estudiantes frente a los problemas de traducción; el ítem 14 hace lo propio con el IET, y nuevamente el resultado ha sido de un 80 % de respuestas que opinan que el contenido del IET es

adecuado para medir lo que pretendemos medir; respecto a la pregunta número 15, que valora de forma global si los tres instrumentos miden la competencia traductora, la mayoría de los expertos (60 %), ha contestado “mucho”, mientras que un 40 % ha contestado “poco” o “depende”.

Por consiguiente, consideramos que el estudio de la validez de contenido es muy favorable, y no obstante tendremos en cuenta las opiniones de ese 20 % de los expertos para examinar nuevamente los dos ítems del ICGT sobre los que se han expresado dudas.

2. Resultados de la validación cruzada

Para realizar las pruebas de validación cruzada, se utilizaron dos grupos de clase de estudiantes de traducción de primer curso con la combinación lingüística inglés español de la FTI de la UAB que, en principio, debían ser equivalentes, dado que la distribución de los alumnos en uno u otro grupo a medida que se van matriculando se realiza a través de un ordenador que los reparte aleatoriamente en el grupo 1 o en el grupo 2. Para comprobar dicha equivalencia respecto al nivel de competencia traductora que miden nuestros instrumentos, se les administraron, el primer día de clase, los tres instrumentos de medida, y lo mismo ocurrió a final de curso, tras ocho meses de formación en traducción. En los datos que presentamos, la primera medición se identifica como "M1" y la segunda medición como "M2". El siguiente gráfico muestra las puntuaciones totales obtenidas por los dos grupos en el IPT, y están separadas las puntuaciones de las dos mediciones y de los dos evaluadores, identificados como "EV1" y "EV2".

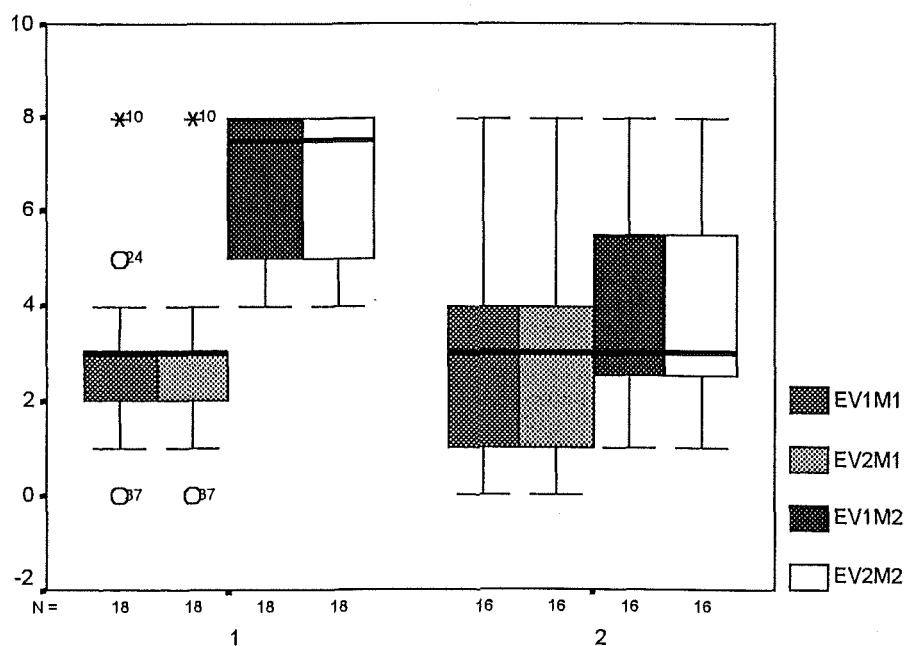


Gráfico 5. Diagrama de caja de los resultados de la validación cruzada del IPT.

Estos datos nos muestran cómo los dos grupos no son comparables, desde un principio. En la primera medida el grupo 1 es muy poco variable: todas las puntuaciones totales se encuentran entre 1 y 4 puntos, y el 50 % central de los resultados se encuentran entre los 2 y 3 puntos. Sin embargo, el grupo 2 muestra una gran

variabilidad en el IPT, ya que sus resultados ocupan toda la escala posible, es decir, desde los 0 hasta los 8 puntos. Respecto a la segunda medición, la variabilidad del primer grupo aumenta, ya que las puntuaciones totales oscilan entre los 4 y los 8 puntos, aunque el 75 % superior esté repartido entre los 5 y los 8 puntos. El segundo grupo sigue mostrando una gran variabilidad, ya que las puntuaciones totales oscilan entre 1 y 8 puntos, aunque el 50 % central se encuentra entre los 3 y los 5 puntos. La evolución del primer grupo es superior a la del segundo grupo, ya que las medias aumentan en 4 puntos (grupo 1) y 1'1 puntos (grupo 2) y esta diferencia es significativa estadísticamente ($F=11'698$; $p=0'002$). El hecho de que una diferencia sea significativa estadísticamente implica que dicha diferencia no es atribuible al azar, sino que existe algún otro factor que la causa. Sin embargo, los efectos del factor “evaluador” son estadísticamente nulos, es decir que los evaluadores no confieren variabilidad alguna. Por consiguiente, los datos nos llevan a pensar que las diferencias en la mejora de los dos grupos deben estar relacionadas con la docencia, dado que los resultados de los dos grupos son muy consecuentes dentro de cada grupo, es decir que el conjunto del grupo evoluciona en un mismo sentido.

El siguiente gráfico refleja los resultados de la validación cruzada del IET:

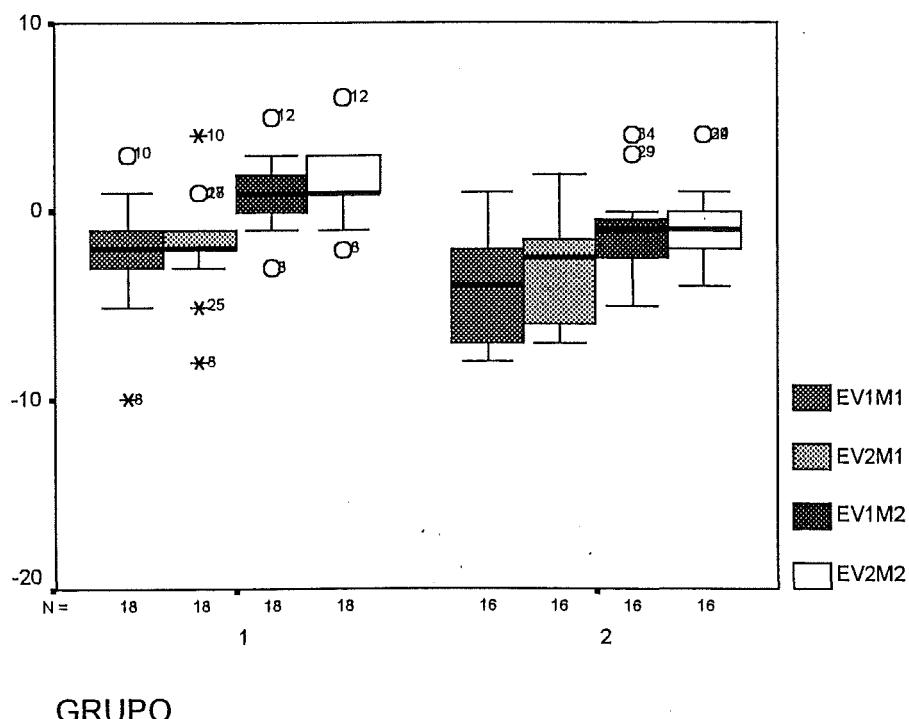


Gráfico 6. Diagrama de caja de los resultados de la validación cruzada del IET.

Las puntuaciones totales obtenidas mediante el IET muestran más variación entre los dos evaluadores, ya que en esta ocasión la diferencia entre evaluadores sí que es estadísticamente significativa ($F = 90'058$; $p < 0'001$), pero aún así se pueden definir tendencias claras de los dos grupos, y, al igual que ocurría con el IPT, en la primera medición el grupo 1 es mucho más compacto que el grupo 2, que es más variable. Respecto a esta variabilidad interna del grupo 2, si bien el rango de puntuaciones totales es de 9 puntos tanto en la primera como en la segunda medición (entre -8 y 1 punto en la primera administración y entre -5 y 4 puntos en la segunda) las puntuaciones totales del 50 % central están mucho más concentradas en la segunda medición, ya que la amplitud intercuartil, es decir, la diferencia entre el límite inferior y el límite superior de la caja, se reduce de 5 puntos en el primer evaluador y 4'8 puntos en el segundo evaluador en la primera medición, a 2 puntos y 2'5 puntos respectivamente en la segunda medición.

Esto refuerza la idea de que el grupo 1, es decir, los sujetos que conforman el grupo 1, son parecidos en lo tocante a conocimientos operativos (IPT e IET). El grupo 2, por el contrario, es variable respecto a los conocimientos operativos (IPT e IET).

Los datos de la segunda medición vuelven a mostrar que el primer grupo evoluciona de manera conjunta más que el grupo 2, y lo hace de forma significativa estadísticamente ($F = 9'097$; $p = 0'005$). Analizando las medias, observamos que el grupo 1 acaba con 2 puntos más que el grupo 2. El hecho de que un grupo mejore más que el otro, recibiendo clases en el marco de un mismo plan de estudios y en un mismo centro, nos lleva a pensar nuevamente en la diferencia atribuible al método de enseñanza, que es el único factor que, a nuestro entender, puede causar la diferencia de mejora entre los dos grupos.

En los resultados del ICGT, que presentamos a continuación, solamente aparece un evaluador, dado que en este instrumento no existe variabilidad de resultados en lo que respecta al evaluador. El siguiente gráfico muestra las puntuaciones totales obtenidas mediante el ICGT en las dos mediciones y en los dos grupos de clase:

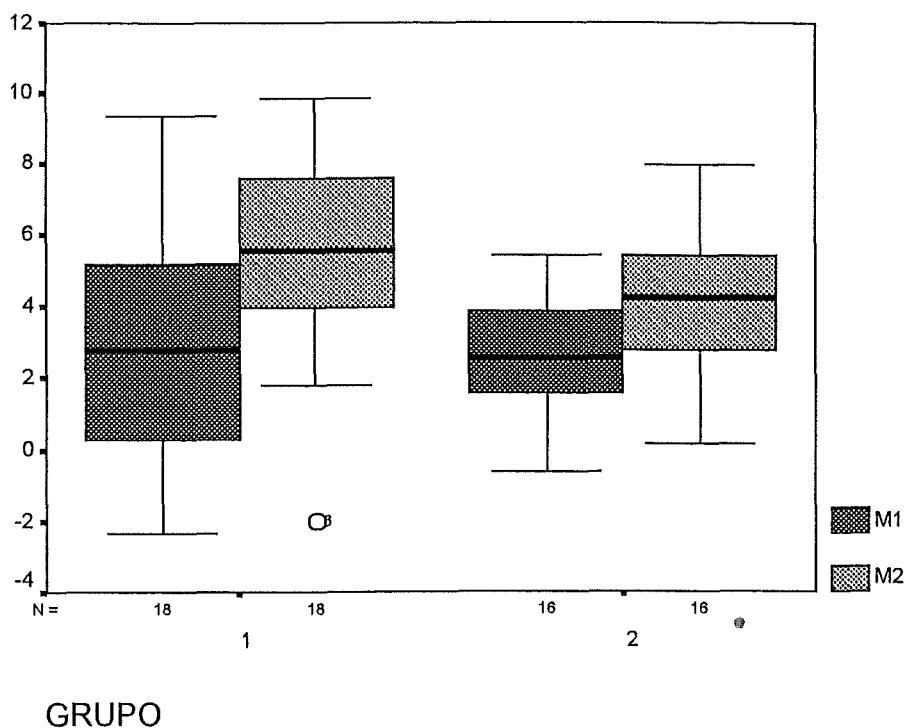


Gráfico 7. *Diagrama de caja de los resultados de la validación cruzada del ICGT.*

Los datos de este gráfico muestran cómo, en la primera medida, si bien la mediana de los dos grupos es muy similar (y también la media, que es de 2'7 puntos en el primer grupo y 2'6 en el segundo grupo) la distribución de puntuaciones totales de ambos grupos es muy diferente: el grupo 1 es mucho más variable, ya que los sujetos están muy repartidos, como indican tanto los bastones que muestran la puntuación más alta y más baja, como la amplitud intercuartil, marcada por los límites inferior y superior de las cajas. Los resultados del grupo 2 son mucho más compactos que el grupo 1, es decir que se trata de un grupo con muy poca variabilidad.

En la segunda medición, se observa un cambio en el grupo 1, cuyos resultados eran muy dispersos en la primera medición. En esta ocasión, la variedad de puntuaciones se reduce y es prácticamente equivalente al grupo dos, si bien las medias de ambos grupos difieren sensiblemente: 5'4 el primer grupo y 4'0 el segundo grupo, lo cual indica nuevamente una mejora superior del grupo 1. Los datos de los tres instrumentos nos llevan a pensar que existe un factor que causa una diferencia entre la mejora de los dos grupos, y pensamos que ese factor es el docente y el método de enseñanza empleado. En este caso, los datos sugieren que el docente del primer grupo hizo más hincapié en los conceptos declarativos que el docente del segundo grupo.

La única diferencia significativa desde el punto de vista estadístico se produce entre la medida 1 y la 2 ($F = 18,227$; $p < 0'001$), es decir que ambos grupos mejoran de forma comparable, y la diferencia se encuentra simplemente en la variabilidad dentro de los grupos.

Es curioso notar que el grupo 2, cuyos resultados en el ICGT son muy compactos, mostraban sin embargo una gran variabilidad en el IPT, donde sus resultados ocupaban toda la escala de puntuaciones totales posibles, es decir, desde los 0 puntos hasta los 8 puntos. Si comparamos los datos obtenidos en la primera administración mediante el IPT y el ICGT vemos que definitivamente, el azar ha distribuido a los sujetos de forma que los dos grupos no son equivalentes, ya que cuando se trata de conocimientos declarativos (que son los datos que se obtienen mediante el ICGT) el grupo 1 es muy variable y el grupo 2 muy compacto, mientras que en los conocimientos operativos (errores cometidos, detección y resolución de problemas) el grupo 1 es mucho menos variable que el grupo 2. Los sujetos que conforman el grupo 1 son diferentes en lo que respecta a conocimientos declarativos (ICGT) y parecidos en lo tocante a conocimientos operativos (IPT e IET). El grupo 2, por el contrario, es variable respecto a los conocimientos operativos (IPT e IET) y compacto en lo tocante a conocimientos declarativos.

Estos datos demuestran que los grupos 1 y 2 no son equivalentes respecto a las medidas de nuestros instrumentos, es decir que, en esta ocasión, el azar no ha producido una distribución de sujetos equivalente entre los dos grupos. De este modo, la validación cruzada de los tres instrumentos nos llevan a dos conclusiones importantes:

- (a) El azar, como es bien sabido, no necesariamente reparte a los sujetos en dos grupos equivalentes para nuestros propósitos. Por consiguiente, es esencial que se comparan los dos grupos de sujetos antes de iniciar un experimento, de manera que se pueda repartir a los estudiantes en dos grupos realmente equiparables; de lo contrario, sería más difícil comparar los resultados de un grupo experimental y otro de control con el fin de observar la intervención de una variable independiente.
- (b) Los datos de los dos grupos no evolucionan de forma paralela, y sin embargo los resultados de cada grupo son consecuentes, es decir que dentro de cada grupo se observa una tendencia común; dada la significación estadística nula del evaluador, concluimos que el elemento diferencial entre la evolución de los dos grupos es el profesor y el tipo de docencia impartida, lo cual refuerza la idea de que es necesario llevar a cabo investigaciones que comparan distintos tipos de docencia.

3. Contrastación de las hipótesis principales

El orden lógico de exposición de los datos referentes a la contrastación de las hipótesis principales sería, en primer lugar, los resultados de las hipótesis A, referidas a la sensibilidad al cambio de los instrumentos, y en segundo lugar, los de las hipótesis B, concernientes a la generalizabilidad de los instrumentos. Sin embargo, en términos de análisis de datos, es mucho más sencillo y natural evaluar en primer lugar la influencia de las fuentes de variabilidad en los instrumentos (evaluadores, universidades y combinaciones lingüísticas) y posteriormente analizar la sensibilidad al cambio, de manera que invertiremos el orden de las hipótesis en este apartado.

3.1. Hipótesis B: instrumentos generalizables

Las hipótesis B hacían referencia a la variabilidad atribuible a cada uno de los tres instrumentos, es decir que mediante un estudio de generalizabilidad, ya descrito, (cf. supra 5.2. *Diseño para el estudio de la generalizabilidad*) hemos obtenido datos que muestran el grado de variabilidad atribuible a los tres factores que pensamos que pueden afectar a los resultados de cada uno de los tres instrumentos.

Dado que para valorar el ICGT solamente existe un evaluador, la variabilidad atribuible a los evaluadores solamente se ha estudiado en el IPT y en el IET, y para ello hemos utilizado un **coeficiente de correlación intraclass**, que calcula la proporción de variancia atribuible a las personas (es decir, entre los sujetos) sobre la variancia total, que es la suma de las variancias existentes entre los sujetos, entre los evaluadores y la variancia residual. Recordemos que en los resultados de nuestros instrumentos existe una fuente de variancia sistemática que no podemos ni queremos eliminar, que es la diferencia en los resultados de los estudiantes entre ellos, ya que los estudiantes tienen sistemáticamente distintos niveles de competencia traductora. Por esta razón se calcula primero la variancia entre personas y después se divide entre todos los factores, que no deben en ningún caso superarla. La correlación intraclass que se mide, en este caso, es el acuerdo entre los dos evaluadores de cada centro, y el resultado de los cálculos se expresa mediante un número que puede oscilar entre 0 y 1. Un resultado de 0 mostraría el máximo desacuerdo entre los evaluadores, mientras que un 1 mostraría el máximo acuerdo entre los dos evaluadores. En general, cuando se utiliza este índice de correlación se considera que una correlación de 0'6 es aceptable, es decir que la variabilidad atribuible a otros factores que no sean las diferencias entre los mismos sujetos es muy reducida.