ESPALEX: UN CORPUS PARA EL ESTUDIO DE LA ADQUISICIÓN DEL ESPAÑOL COMO LENGUA EXTRANJERA

José M. Bustos Gisbert
Jorge J. Sánchez Iglesias
Universidad de Salamanca

RESUMEN: El corpus ESPALEX para el estudio de la adquisición del español como lengua extranjera se ha construido sobre la base de los exámenes para la obtención de los Diplomas de Español como Lengua Extranjera. Corresponde al nivel B2 de aprendizaje de nuestra lengua de acuerdo con la definición de tal nivel en el Marco de Referencia Europeo para la Enseñanza de Lenguas. En la actualidad el corpus recoge 12.576 textos distintos referidos a 50 temáticas diferentes agrupadas en torno a 5 grandes clases textuales: carta formal, carta informal, texto narrativo, texto descriptivo y texto expositivo. Tienen una longitud media entre de 175 palabras; por ello, acumula en estos momentos unas 2.200.000 palabras.

Fundamentación teórica del corpus Espalex

En el XXI Congreso de Asele presentamos una comunicación que llevaba por título «Métodos de análisis aplicados a la adquisición de segundas lenguas». En ella defendíamos, y citamos literalmente su resumen, que los trabajos en adquisición de ELE despiertan dudas que tienen que ver entre otras cosas con el número de aprendices considerados en el estudio y la valoración del grado de conocimiento de nuestra lengua por parte de los estudiantes. A partir de ahí concluíamos la necesidad de dotarnos de un procedimiento científico en el sentido estricto del término sustentado en la existencia de una base conceptual sólida y en la puesta en marcha de un análisis consistente de los datos recabados. Para ello, sugeríamos la asunción de los modelos de creación de muestras de trabajo fiables desarrollados en la estadística inferencial, basada en la realización de generalizaciones sobre una población a partir del estudio de una muestra.

Toda esa reflexión ha sido desarrollada en un artículo que lleva por título «Bases para el estudio de la adquisición del sistema verbal del español como lengua extranjera. Apuntes metodológicos» y que ha visto la luz a finales de este 2011. En dicho trabajo recordamos cómo en los estudios acerca de la adquisición del verbo por parte de los aprendices de español se ha partido de muestras insuficientes en lo referido a su tamaño y a su representatividad de un mismo estadio de aprendizaje.

La muestra de informantes

Por lo que se refiere a la primera cuestión, la estadística inferencial nos aporta datos relevantes interesantes (Véase García Ferrando, 2000): los estudios parten de

poblaciones de al menos 500 individuos como universo completo; además, en términos absolutos, como mínimo exigen casi 100 individuos de muestra.

Evidentemente, no podemos censar la población de estudiantes de ELE, y los datos con los que contamos son de una terrible imprecisión. Como muestra, un ejemplo: en 2006 el Instituto Cervantes edita la *Enciclopedia de Español en el mundo.* En ella se estima que el número de estudiantes e ELE es de 11,3 millones aunque luego se redondea la cifra en nada menos que 14 millones[1]. En esa estimación se mezclan datos desde 2003 a 2006. De ellos se atribuyen 6 millones a Estados Unidos. Pues bien, en 2008, el mismo Instituto Cervantes publica su *Enciclopedia del Español en los Estados Unidos.* Allí aporta una cifra que no llega a 5 millones de estudiantes entre la enseñanza secundaria y la universitaria. En otras palabras, el propio Instituto Cervantes solo se atreve a proponer cifras con un margen de error de ±25% para el cómputo total de aprendices de ELE. Y si ese margen del error es de ± 15% para el caso de Estados Unidos, todo ello significa que no podemos construir un universo de estudio válido sobre la base de los datos con lo que en la actualidad contamos y que se refieren a la población completa de aprendices de ELE.

Podríamos entonces por optar por realizar nuestra labor sobre la base de muestreos no aleatorios, que no necesitan de una identificación del universo de trabajo ni una selección basada en el cálculo de probabilidades[2]. El problema fundamental es que no se puede saber si en términos de investigación son representativos de la realidad. En palabras de García Ferrando (2000), no permiten estimar los márgenes de error, razón por la cual se debe evitar su uso.

En nuestro universo general de trabajo, la población teórica que deberíamos estudiar no es accesible porque no está cuantificada y censada. Pero ello no significa que no podamos realizar investigaciones estadísticamente convincentes, sino que debemos concretar ese total en lo que se denomina un universo de trabajo accesible y aprehensible. ¿Cómo? Buscando áreas donde los aprendices de ELE estén adecuadamente cuantificados. ¿Dónde? En los pruebas de evaluación de lenguas existentes en la actualidad basadas en escalas de niveles que cuenten con un alto grado de fiabilidad.

Existe abundante bibliografía en la que se establecen espacios comunes de análisis entre las investigaciones vinculadas a adquisición de segundas lenguas (ASL) y las que tienen que ver con la evaluación de lenguas. Cabe citar entre otros los trabajos de Cohen (1998), Brindley (1998), Sohamy (2000) y Douglas (2001)[3]. Brindley (1998) centra su interés en estudiar la aplicación de las escalas de niveles de conocimiento al campo de la ASL, específicamente de aquellas configuradas desde la

[1] Con el problema añadido de que esa cifra no incluye los estudiantes de ELE dentro de España, lo que introduce una variable más que no podemos tener en cuenta. Carrera y Gómez Asencio (2009) nos recuerdan que existen en la actualidad en España aproximadamente 600 centros de enseñanza de ELE.

[2] Nos estamos refiriendo a los modelos basados en muestras accidentales, muestras de cuota o en muestras intencionadas. Las primeras utilizan individuos elegidos por simple proximidad con el investigador; las segundas, muy utilizadas por los analistas de los medios de comunicación, se basan en que se construye el perfil del individuo buscado y el investigador lo elige libremente de entro los miembros del grupo; en las terceras se asume que el sentido común del investigador le permite incluir adecuadamente la muestras sobre la base normalmente de criterios subjetivos.

[3] Se sugiere al interesado la lectura del imprescindible trabajo de Bachman y Cohen (1998).

perspectiva del comportamiento lingüístico de los aprendices ya que reflejarían etapas del comportamiento lingüístico en una o varias áreas. Estarían, al fin y al cabo, indicando niveles de interlengua (en adelante, IL) desde cero hasta una expresión cuasi nativa. Son gradaciones destinadas a medir la actuación global sobre la base de una serie de competencias subyacentes. El problema, en opinión de Brindley, puede estar en cómo se construyen las escalas de niveles, en tanto que muchas de ellas, en sus palabras, carecen de fundamentación teórica. En la medida en que sean fiables, también serán aplicables a ASL, como apuntan Griffin y Nix (1991). Debemos, pues, fiarnos, de aquellas que se hayan construido de forma muy consistente y sobre la base de información perfectamente detallada. En ese sentido, y seguimos ahora a North (1993), pueden convertirse en herramientas perfectamente prácticas y operativas en los estudios de ASL. ¿De cuáles podremos fiarnos? Evidentemente, de aquellas con un alto grado de validez. Y este se puede obtener haciendo una valoración que tenga en cuenta al menos las siguientes cuestiones (Brindley, 1998): la relación entre ellas y los resultados en las pruebas de evaluación y acreditación que las reconocen; su validación psicolingüística, en tanto que demuestren que las habilidades descritas para cada nivel son las que efectivamente usan los aprendices para resolver las tareas impuestas en las pruebas de evaluación correspondientes; y la coherencia estructural de la secuencia de niveles propuesta[4].

Entendemos que contamos con una escala de niveles perfectamente fiable que emana del *Marco de referencia europeo para el aprendizaje, la enseñanza y la evaluación de lenguas*, publicado en 2001 por el Departamento de Política Lingüística del Consejo de Europa. El *Marco* propone una escala basada en tres grandes criterios: libre del contexto en tanto que la escala sea aplicable a cualquier itinerario de acceso a la lengua y en la medida en que las categorías de clasificación respondan a lo que efectivamente hacen los aprendices con la lengua; basada en una fundamentación teórica; y libre de conceptos inconsistentes y de cálculos aproximados. Para ello se ha aplicado una metodología que ha incluido procesos intuitivos, cuantitativos y cualitativos[5].

En el caso de la acreditación del español como lengua extranjera contamos con un panorama de pruebas bastante completo en el marco de ALTE. Se trata de los exámenes para la obtención de los *Diplomas de Español como lengua Extranjera* (DELE), otorgados por el Instituto Cervantes en nombre del Ministerio de Educación de España. La Universidad de Salamanca elabora los modelos de examen y evalúa las pruebas para la obtención de los diplomas de español[6]. Las correspondencias que se establecen son las siguientes[7]:

---

[4] De hecho, en el caso del inglés como lengua extranjera, Brandley duda de la validez de escalas perfectamente conocidas en los Estados Unidos como es el caso de la propuesta por la ACTFL. Dedica una importantísima parte de su trabajo a los estudios existentes para la validación de escalas de nivel.

[5] El procedimiento de fijación y definición de la escala es ciertamente complejo a la vez que muy exhaustivo. Se resume, dentro del propio documento del *Marco,* en los anejos A, B, C y D.

[6] El lector interesado encontrará más información en http://diplomas.cervantes.es/index.jsp.

[7] Las pruebas para la obtención del Diploma de Español A1 y del Diploma de Española A2 se han realizado por primera vez en noviembre de 2009 y mayo 2010 respectivamente. En el cuadro se incluyen las denominaciones actuales y sus correspondientes en el pasado.

TABLA DE CORRESPONDENCIAS						
CONSEJO DE EUROPA	A1	A2	B1	B2	C1	C2
ALTE	Acceso	Plataforma	Umbral	Avanzado	Dominio operativo eficaz	Maestría
DELE	Diploma de Español Nivel A1	Diploma de Español Nivel A2	Diploma de Español Nivel B1 (Inicial) (Antes C.I.E.)	Diploma de Español Nivel B2 (Intermedio) (Antes D.B.E.)		Diploma de Español Nivel C2 (Superior) (Antes D.S.E)

En la actualidad, los esfuerzos del Instituto Cervantes y de la Universidad de Salamanca se dirigen a que estas correspondencias sean perfectamente válidas. Por todo ello, podemos afirmar que contamos con un marco perfectamente fiable para establecer muestras de análisis orientadas a la realización de estudios de adquisición.

EL NIVEL DE LOS INFORMANTES

Es decisivo considerar el grado de dominio de ELE por parte de los individuos analizados en los estudios de adquisición. Si asumimos como cierto que en el proceso el aprendiz pasa por sucesivos estados de IL, no tiene sentido mezclar a individuos correspondientes a los diferentes estadios en el análisis o no tener en cuenta la IL en la que se encuentran en el momento en el que se lleva a cabo la investigación. Al mismo tiempo, el estudio por niveles nos va a permitir estratificar la muestra de estudio. Desde esa división sí podremos establecer y emitir juicios consistentes en relación con el proceso mismo de adquisición de ELE. Nuestra propuesta utiliza los exámenes para la obtención de los DELE, entendidos desde los diferentes niveles que se reconocen. Su estructura y distribución facilita ese estudio estratificado.

Pero además los resultados obtenidos por los individuos analizados en las pruebas para la obtención de los DELE han de ser analizados rigurosamente con el objeto de garantizar su pertenencia a un mismo grupo. Habrá que considerar no solamente si los aprendices han obtenido alguno de los DELE, sino también habrá que considerar la calificación obtenida. Solo calificaciones básicamente iguales nos garantizarán que los aprendices pertenecen efectivamente a un mismo nivel de IL. El sistema de calificación de los exámenes para la obtención de los DELE[8] prevé que la puntuación máxima que se puede obtener sea de 100 puntos, distribuidos en tres bloques. Para obtener cualquiera de los diplomas se ha de obtener al menos un 70% de acierto en cada uno de los tres grupos que componen la calificación. Solo en ese caso se suman las tres calificaciones y se obtiene el apto global. Por ello, interpretaremos como del mismo nivel a aquellos estudiantes que hayan sido considerados aptos en una prueba dada y que tengan un grado de acierto total entre el 70 y el 80% en la prueba de expresión escrita, que servirá de base para el corpus.

[8] La información está tomada de la *Descripción general del sistema de evaluación DELE* publicado por el I. Cervantes (2006). El documento se puede consultar en http://diplomas.cervantes.es/general/folletos.jsp

ESPALEX se limita inicialmente al nivel B2 por razones de accesibilidad al corpus y de representatividad del mismo: es la prueba a la que, como veremos más adelante, se han presentado más candidatos desde la creación de los DELE. No obstante, el objetivo a medio y largo plazo es incluir aprendices de A2 y C2. De ese modo, se podrá dar cuenta de resultados investigadores correspondientes a tres niveles de IL y, al mismo tiempo, nos permitirá realizar estudios referidos al proceso global de adquisición del español como lengua extranjera.

EL TIPO DE LENGUA

Ellis (1994) distingue tres tipos de datos aportados por los aprendices y que pueden usarse en la investigación sobre ASL según la naturaleza de los mismos En esencia, a las opciones que nos recuerda subyacen tres percepciones diferentes: el uso de la lengua nos informa acerca de la actuación de los aprendices, los juicios lo hacen en relación con su competencia y los auto informes nos facilitan datos sobre la conciencia de aprendizaje. Nos fijaremos exclusivamente en los dos primeros tipos de datos[9].

Los procedimientos de obtención de datos utilizados por los investigadores estudiados y vinculados a juicios metalingüísticos son básicamente de dos tipos. Una opción es la pura reflexión metalingüística acerca de determinados enunciados. La segunda posibilidad tiene que ver con la utilización de ejercicios focalizados en la forma gramatical. En relación con los juicios metalingüísticos sustentados en la pura reflexión sobre el grado de validez de los enunciados, hay que decir que nos dan datos sobre el conocimiento competencial, pero se trataría como mucho de un dominio teórico, reflexivo y por ello metaconceptual que no garantiza que el aprendiz use tales contenidos. En otras palabras, dudamos que se pueda hablar de adquisición sin detección de uso efectivo[10]. A eso se añaden las críticas que el propio Ellis, siguiendo a Birdsong (1989), reconoce y que tienen que ver con cuánto de fiables son los juicios de gramaticalidad de los hablantes nativos que normalmente funcionan como grupos de control.

En segundo término, los juicios basados en procedimientos de huecos, estén más o menos explicitadas las opciones, plantean problemas vinculados a la interpretación de los resultados; por ello, no podemos aceptar como argumento que efectivamente faciliten la investigación sobre propiedades lingüísticas que de otro modo (y estamos citando casi literalmente a Ellis) serían de difícil estudio. Los problemas son de cuádruple naturaleza: en primer término, tienen que ver con que el grado de acierto o error en la resolución de las actividades no garantiza un grado de acierto o error equivalente en el uso efectivo; por otro lado, en cualquier ejercicio de esta naturaleza será difícil evaluar el efecto de la respuesta aleatoria, sea esta acertada o incorrecta; en tercer lugar, en los ejercicios basa-

[9] Poco podemos aportar a las afirmaciones de Ellis referidas a los auto informes, y de su importancia para analizar los factores afectivos y cognitivos del proceso de adquisición, con todas las limitaciones que este tipo de datos exhiben y que muy acertadamente expone en su trabajo. En Bustos Gisbert y Sánchez Iglesias (2006) utilizamos precisamente los auto informes para estudiar la conciencia de fosilización de errores en los buenos aprendices de ELE.

[10] Es difícil medir el grado de adquisición en pruebas que no sean de aplicación de la competencia comunicativa entendida en términos globales. Todo lo demás supone una manipulación que está bien para clase en tanto que estrategia docente, pero que no refleja grado de adquisición real.

dos en textos con huecos se ha observado como un uso incorrecto puede generar una cadena de usos erróneos sin que ellos tenga que ver con otra cosa que no sea con la naturaleza discursiva del ejercicio; por último, la propia construcción de este tipo de ejercicios demuestra la dificultad de evitar las posibles dobles respuestas correctas, salvo en los casos en los que se opta por procedimientos de selección múltiple donde los distractores en el fondo no cumplen con esa función distractora.

Por todo ello, los juicios metalingüísticos son los que más dudas nos despiertan: no es tan evidente el establecimiento de una correlación automática entre la adecuación del juicio y el grado de adquisición del concepto y quedan demasiadas dudas referidas a la propia interpretación de los resultados. El juicio puede usarse pero no como fuente exclusiva de datos en tanto que no exhibe el nivel de la actuación de los aprendices; es decir, se han de validar los resultados obtenidos con su aplicación con datos extraídos del uso real de la lengua, sea este más o menos dirigido.

El uso guiado de la lengua propone un grado de libertad restringido, sí, pero libre al fin y al cabo. Permite al aprendiz optar entre opciones expresivas, entre alternativas. Le posibilita incluso obviar estratégicamente algunas opciones. Y todo esto lo encontramos en alternativas como la redacción o la entrevista oral más o menos guiadas y con independencia del grado de control que se quiera establecer en una u otra. Por poner un ejemplo en el ámbito de la expresión oral, una dramatización puede ser tan guiada o más que un ejercicio de *retelling*. Son actividades que nada tienen que ver con ejercicios de lectura en voz alta, de traducción, de reconocimiento de estructuras o de utilización de ítems discretos, que están mucho más cerca de los juicios metalingüísticos que del uso.

Defendemos el uso libre de la lengua como muestra sobre la que efectuar la investigación. Entre otras cosas, porque el texto nos va a proporcionar mucha más información sobre la adquisición en tanto que nos va a permitir cotejar aciertos y errores de uso contextualizados. Y el contexto nos va a permitir dilucidar la naturaleza de los errores detectados. En una actividad donde el aprendiz informante se limita a escribir una forma verbal no podremos nunca realizar otro tipo de interpretaciones. Proponemos para el estudio de la adquisición del sistema verbal español el uso de las pruebas de expresión escrita. En este caso las correspondientes a los exámenes para la obtención de los DELE. Son pruebas guiadas en las que sin embargo, el grado de libertad expresiva es razonablemente alto o muy alto[11].

Para el caso del estudio de la adquisición entendemos que resulta más recomendable la utilización de muestras de lengua escrita y no oral. Existen al menos tres razones a nuestro favor. En primer lugar, por la propia naturaleza del texto conversacional. En ese sentido, son aclaratorias las conclusiones de Briz (1998)[12], quien defiende que carecemos de un estudio completo de la conversación coloquial, muy especialmente en lo que afecta a gramaticalizaciones y especializaciones pragmáticas. En segundo lugar, porque la lengua conversacional se basa en cooperación y en estrategia[13]. Cooperación con el tema y con el interlocutor. Estrategia en la medida en que comunicar no es solo codificar y descodificar, sino también y sobre todo mostrar e inferir. La construcción sintáctica se hace interactiva, y los enunciados de uno influyen en los del otro. Este hecho, a la hora de

[11] Los interesados en su formato pueden acudir a la *Descripción general del sistema de evaluación DELE* publicado por el I. Cervantes en 2006.

[12] Que a su vez se inspira en Bustos Tovar (1995) y Narbona (1989).

[13] Sigo en lo esencial a Briz (1998).

valorar el grado de adquisición del sistema verbal en aprendices de L2 puede incluir variables no previstas, como es el caso del nivel de influencia de las formas de expresión usadas por un interlocutor en el otro. En tercer lugar ha de tenerse en cuenta que la descripción normativa del estándar, que es la que se utiliza habitualmente para verificar el grado de adquisición, se hace desde el registro formal escrito y especializado.

EL CORPUS ESPALEX

PUNTOS DE PARTIDA

El número total de candidatos que se presentaron a las pruebas de los DELE entre 1989 y 2006 es de 286.000, lo cual configura un universo de partida realmente amplio. No obstante, necesitamos depurar estos datos para que reflejen un universo de estudio fiable. Para ello, nos detendremos en los datos correspondientes a los candidatos presentados a las pruebas para la obtención del Diploma Básico de Español, equivalente al actual Diploma de Español Nivel B2 (Intermedio) entre 1992 y 1999, a los que hemos tenido acceso y que nos servirán como ejemplo para valorar la consistencia del corpus propuesto: el número total de candidatos presentados a las pruebas es de 55.792. Se consideraron aptos 40.491; esto es, un 72,5% de los candidatos que se presentaron. A su vez, de ellos obtuvieron una calificación entre 10,5 y 12 en la prueba correspondiente a Expresión Escrita (entre un 70 y un 80% de acierto) 10.400 candidatos, que son un 18,6% del total de los presentados y un 25,68% de los que fueron considerados aptos.

En la creación del corpus ESPALEX hemos decidido limitarnos a datos iguales o posteriores a 2009 entendiendo que a partir de esa fecha corresponden a un modelo de examen que efectivamente corresponde al nivel B2 en el que hemos decidido detenernos. No obstante, se han incluido algunos informantes correspondientes a 2007, que se usaron en su momento para la creación de una versión piloto del corpus digitalizado. Asimismo, se ha completado con un grupo de informantes que hicieron las pruebas entre 1992 y 1996; todos ellos tienen el italiano como lengua materna y son la base del estudio de Sánchez Iglesias (2003).

EL CORPUS EN SU CONFIGURACIÓN ACTUAL

Contamos en estos momentos con 6.464 de informantes distribuidos de la siguiente manera según las convocatorias a las que corresponden:

Convocatoria	Número de informantes
1992-1996 (mayo y noviembre)	147
Mayo 2007	29
Agosto 2009	251
Noviembre 2009	1631
Mayo 2010	385
Agosto 2010	4021

Este número se ha de reducir a 6.288 si contabilizamos exclusivamente los datos correspondientes a 2009 y 2010. Nuestros informantes han realizado las pruebas para la obtención de los DELE en 360 centros de examen, correspondientes a 85 países. Considerando que cada uno de los informantes redacta dos escritos diferentes, hemos de concluir que contamos con aproximadamente 12.576 textos distintos. De acuerdo con las instrucciones para la realización de la prueba, cada uno de ellos debe tener una longitud entre 150 y 200 palabras. Si asumimos una extensión media de 175 podemos suponer que el corpus ESPALEX acumula en estos momentos unas 2.200.000 palabras.

En el momento de matricularse para realizar el examen, los candidatos deben incluir su nombre, sus apellidos y su nacionalidad. La información con la que contamos nos permite deducir en la mayor parte de los casos su L1. En aquellos en los que el investigador duda acerca del origen lingüístico del informante, se ha optado por renunciar a ellos y no se han incluido en el corpus. Según los datos con los que contamos, los 12.576 textos se distribuyen de la siguiente manera según la lengua materna de sus autores:

Griego	**2.172**	**Neerlandés**	98	**Georgiano**	14
Portugués	1.922	**Checo**	96	**Azreí**	12
Italiano	1.834	**Búlgaro**	92	**Croata**	12
Francés	1.196	**Eslovaco**	62	**Lituano**	10
Alemán	742	**Sueco**	44	**Esloveno**	8
Coreano	708	**Turco**	44	**Macedonio**	8
Chino	644	**Ucraniano**	30	**Noruego**	8
Japonés	504	**Albanés**	28	**Danés**	6
Inglés	444	**Kazajo**	28	**Armenio**	2
Rumano	392	**Moldavo**	24	**Estonio**	2
Árabe	338	**Letón**	20	**Hebreo**	2
Polaco	272	**Farsi**	18	**Islandés**	2
Ruso	218	**Vietnamita**	18	**Nepalí**	2
Húngaro	122	**Serbio**	16	**Tailandés**	2

La existencia de países en los cuales hay dos o más lenguas oficiales nos ha invitado a desagregar los datos correspondientes a ellos:

PAÍS	LENGUAS OFICIALES	INFORMANTES
Bélgica	Francés/Neerlandés	82
Bosnia Herzegovina	Bosnio/Croata/Serbio	14
Burundi	Suajili/Francés	4
Camerún	Francés/Inglés	8
Filipinas	Inglés/Tagalo	20
Finlandia	Finés/Sueco	4
Guinea	Español/Francés/Portugués	4
Indonesia	Inglés/Malayo/Chino/Tamil	14
Islas Mauricio	Francés/Mauriciano	2
Kenia	Suajili/Inglés	4
Kirguistán	Kirguís/Ruso	2
Nigeria	Inglés/Yoruba	4
Pakistán	Inglés/Urdu	2
Suiza	Alemán Francés Italiano	160 26 2
Uzbekistán	Uzbeko/Ruso	8

El estado actual del corpus nos permitiría realizar estudios de adquisición consistentes relativos a 17 lenguas, que son las que alcanzan aproximadamente 100 textos diferentes:

Griego	**2.172**
Portugués	1.922
Italiano	1.834
Francés	1.196
Alemán	742
Coreano	708
Chino	644

Japonés	504
Inglés	444
Rumano	392
Árabe	338
Polaco	272
Ruso	218
Húngaro	122

Neerlandés	98
Checo	96
Búlgaro	92

DIGITALIZACIÓN DEL CORPUS

Todo nuestro corpus está disponible actualmente en soporte papel y llevamos ya tiempo trabajando en su digitalización. En la actualidad contamos con aproximada-

mente 3.100 textos ya digitalizados y organizados en una base de datos construida en el programa Microsoft Access 2007. Está distribuida de la siguiente manera de acuerdo con la L1 de los informantes:

Búlgaro	2
Húngaro	2
Letón	2
Lituano	2
Malayo	2
Noruego	2
Sueco	2
Árabe	4

Moldavo	4
Checo	6
Holandés	8
Polaco	8
Griego	10
Rumano	12
Ruso	18
Chino	22

Coreano	108
Japonés	154
Suizo Alemán	224
Italiano	321
Alemán	507
Inglés	536
Portugués	562
Francés	585

En cada ficha del corpus digitalizado se incluyen los siguientes datos: lengua materna, convocatoria en la que se redactó el texto, clase de texto (narrativo, descriptivo, expositivo, carta formal o carta informal), extensión y texto. Asimismo existe la posibilidad de incluir el texto original digitalizado. Se han incluido algunos de manera experimental y se prevé aumentar esta opción para los textos que se incorporen en adelante.

Conclusión

Los estudios de ASL deben sustentarse en una metodología consistente en lo que toca a la muestra de aprendices seleccionada para realizarla. Y lo debe ser en dos grandes sentidos: en lo que afecta al tamaño mismo de la misma y en lo que toca al nivel de los informantes seleccionados. Las elecciones muestrales en la investigación de ASL se realizan sin consistencia estadística. Se tiende a elegir a los aprendices que se tiene más a mano. En consecuencia, son grupos cuya representatividad no se puede garantizar ni en términos cuantitativos ni cualitativos. Asimismo, la consideración de la L1 de los aprendices no siempre es una variable tenida en cuenta a la hora de establecer diferencias en la selección de informantes. Sin embargo, parece que la lengua materna puede ser una herramienta perfectamente adecuada para la estratificación de la muestra, toda vez que la mayoría de los estudios de adquisición asume la importancia de los procesos de transferencia. En tercer lugar, con mucha frecuencia se mezclan aprendices correspondientes a diferentes estados de IL. El nivel de dominio de ELE es la segunda variable clave en los procesos de estratificación y, por ello, de organización de los estudios de adquisición. De hecho, así lo demuestran los avances alcanzados en el ámbito de los trabajos dedicados al análisis de los distintos niveles de IL de los aprendices. Por todo ello, los resultados obtenidos hasta ahora han de ser relativizados en la medida en que difícilmente resultan generalizables. Son aportaciones valiosas que nos indican la dirección por la que se pueden mover nuestras intuiciones, pero no pueden ser interpretados como definitivos

Los DELE nos proporcionan un material imprescindible para que los estudios de adquisición del español como lengua extranjera progresen de manera cualitativa. El convenio actual entre la Universidad de Salamanca y el Instituto Cervantes nos ha permitido llevar a cabo la tarea de construir un corpus como ESPALEX, que ha de ser una piedra angular en esa mejora en los estudios adquisicionistas. Es nuestra responsabilidad hacer todos los esfuerzos por posibilitar al investigador el acceso a esos datos imprescindibles. Y a ello nos comprometemos.

BIBLIOGRAFÍA CITADA

BACHMAN, Lyle. F. y Andrew D.. COHEN (eds.) (1998): *Interfaces between second language acquisition and language testing research*, Cambridge: Cambridge University Press.

BIRDSONG, D. (1989): *Metalinguistic performance and interlinguistic competence*. Berlin: Springer Verlag.

BRIZ, Antonio (1998): *El español coloquial en la conversación. Esbozo de pragmagramática*, Barcelona: Ariel.

BUSTOS GISBERT, José Manuel (2011)»Bases para el estudio de la adquisición del sistema verbal del español como lengua extranjera. Apuntes metodológicos», *DICENDA. Cuadernos de Filología Hispánica, 29*, 7-41.

Bustos Gisbert, J. M. (2011): Métodos de análisis aplicados a la adquisición de segundas lenguas. En AAVV: *Del texto a la lengua: la aplicación de los textos a la enseñanza-aprendizaje de español L2-LE* XXI Congreso de ASELE. Salamanca: ASELE.

BUSTOS GISBERT, José Manuel y J. SÁNCHEZ IGLESIAS (coords.) (2006): *La fosilización en la adquisición de segundas lenguas: el buen aprendiz*, Salamanca: Luso Española de Ediciones.

BUSTOS TOVAR, José Jesús (1995): «De la oralidad a la escritura», en L. Cortés (ed.): *El español coloquial. Actas del I Simposio sobre análisis del discurso oral*, Almería: Universidad de Almería, 11-28.

CARRERA, Miguel y José Jesús GÓMEZ ASENCIO (2009): La *economía de la enseñanza del español como lengua extranjera*, Barcelona: Ariel.

COHEN, Andrew (1998): «Strategies and processes in test taking and SLA». En L. F. Bachman, y A. Cohen (eds.): *Interfaces between second language acquisition and Language testing research*, Cambridge: Cambridge University Press, 90-111.

BRINDLEY, Geoff: «Describing language development? Rating scales and SLA». En L. F. Bachman y A. Cohen (eds.): *Interfaces between second language acquisition and language testing research*, Cambridge: Cambridge University Press, 112-140.

DOUGLAS, Dan 82001): «Performance consistency in second language acquisition and language testing research: a conceptual gap». *Second language research*, 17-4, 442-456.

ELLIS, Rod (1994): *The study of second language acquisition*, Oxford: Oxford University Press.

GARCÍA FERRANDO, Manuel (2000): *Socioestadística*, Madrid: Alianza.

GRIFFIN, Patrick y P. NIX (191): *Educational assessment and reporting*, Sydney: Harcourt Brace Jovanovich.

NARBONA, Antonio (1989): *Sintaxis española: nuevos y viejos enfoques*, Barcelona: Ariel.

NORTH, Brian (1993): *The development of descriptors on scales of language proficiency*, Washington DC: National Foreign Language Center.

SÁNCHEZ IGLESIAS, Jorge (2003): *Errores, corrección y fosilización en la didáctica de lenguas afines: análisis de errores en la expresión escrita de estudiantes italianos de E/LE*, Salamanca: Ediciones Universidad de Salamanca.

SOHAMY, Elana (2000): «The relationship between language testing and second language acquisition revisited». *System*, 28, 541-553.