

DEFINICIÓN DE GLOSARIOS LÉXICOS DEL ESPAÑOL: NIVELES INICIAL E INTERMEDIO

Defining lexical glossaries: beginner and intermediate level

*La définition des glossaires lexicaux: niveaux débutant
et moyen*

José M. BUSTOS GISBERT

Profesor Titular de Lengua Española

Facultad de Traducción y Documentación. Universidad de Salamanca

BIBLID [0212 - 5374 (2001) 19; 35-72]

Ref. Bibl. JOSÉ M. BUSTOS GISBERT. Definición de glosarios léxicos del español: niveles inicial e intermedio. *Enseñanza*, 19, 2001, 35-72.

RESUMEN: El artículo parte del problema existente a la hora de definir las palabras que un estudiante de español como lengua extranjera debe adquirir en las diferentes etapas de su aprendizaje. Se plantea como objetivo final ofrecer una definición consistente de los glosarios léxicos correspondientes a los dos primeros niveles de aprendizaje. En primer lugar, se revisan cuáles son los acercamientos a la cuestión realizados y los modelos de análisis utilizados hasta ahora. En segundo lugar, se realiza una presentación de un estudio destinado a la definición de los glosarios léxicos correspondientes a los niveles inicial e intermedio. Se ha partido de un análisis de corpus en el que se han utilizado como variables básicas las de frecuencia y rango, y se ha completado con un estudio de disponibilidad léxica realizado con estudiantes reales. Por último, se ofrecen en forma de apéndice ambos glosarios.

Palabras clave: Glosario léxico, corpus, frecuencia, rango, disponibilidad léxica, lengua extranjera, niveles de aprendizaje.

SUMMARY: The starting point of this paper is the problem of trying to define the lexis that should be acquired by students of Spanish as a foreign language in the different learning stages. The aim is to offer a consistent definition of the lexical glossaries corresponding to the first two levels of learning. First of all, previous approaches to the problem and models of analysis used so far are revised. Secondly, a study aimed at defining the lexical glossaries which correspond to beginner and intermediate level is presented. The study consists of a corpus analysis where range and frequency have been used as basic variables and of a lexical availability study carried out on students. Finally, the two glossaries are included in the appendix.

Key words: Lexical glossary, corpus, frequency, range, lexical availability, foreign language, learning levels.

RÉSUMÉ: L'article prend pour point de départ le problème de la délimitation du vocabulaire qu'un étudiant d'espagnol langue étrangère doit acquérir dans ses différents niveaux d'apprentissage. L'objectif final est celui d'offrir une définition complète des glossaires lexicaux correspondants aux deux premiers niveaux d'apprentissage. Il s'agit tout d'abord de réviser quels sont les approches réalisées et les modèles d'analyse utilisés jusqu'à présent, basés notamment sur une analyse de corpus avec variétés principales celles de la fréquence et de rang linguistique; puis, de présenter une étude des disponibilités lexicales réalisée avec les propres étudiants et destinée à la définition des glossaires lexicaux correspondants aux niveaux débutant et moyen; et, pour finir, d'offrir un addenda contenant les deux glossaires.

Mots clés: Glossaire lexical, corpus, fréquence, rang linguistique, disponibilité lexicale, langue étrangère, niveaux d'apprentissage.

1. INTRODUCCIÓN

Sin ningún género de dudas, el léxico es uno de los dominios más complejos en lo relativo a la enseñanza de lenguas extranjeras (en adelante, LE). La discusión al respecto es profusa, los puntos de vista adoptados son muy diversos y en todos se detecta una evidente falta de consenso a la hora de responder a las dos preguntas fundamentales: en primer lugar, ¿qué palabras hemos de enseñar en cada nivel?; en segundo término, ¿debemos proporcionar una instrucción específicamente orientada al léxico en todos los niveles de aprendizaje?

Evidentemente, resulta difícil satisfacer plenamente estas cuestiones, especialmente cuando las ponemos en relación con las actitudes metacognitivas que el vocabulario provoca tanto en docentes como en estudiantes. Éstas han sido sintetizadas por Coady (1997)¹:

¹ Al definir las, el autor pone en correlación la enseñanza del léxico con la de la comprensión de lectura.

- Los estudiantes piensan que las palabras son muy importantes y están ansiosos de aprenderlas.
- Los profesores tienden a pensar que las palabras son fáciles de aprender y que el problema radica en la gramática.
- Muchos profesores piensan que enseñar vocabulario es una actividad secundaria.
- Por ello, muchos profesores no ven el sentido a enseñar vocabulario a pesar de las peticiones de los alumnos.

El trabajo que a continuación se presenta sólo aspira a dar una respuesta parcial a la primera de las cuestiones planteadas. Para ello, en la primera parte del mismo propongo una revisión crítica de las propuestas que se han realizado a la hora de establecer diferentes glosarios. Como se verá, los estudios hasta ahora efectuados, especialmente en lo referido a la lengua inglesa nos conducen a pensar que toda construcción de un glosario debe llevarse a cabo desde las conclusiones a las que podamos llegar a partir de un estudio de corpus. Sin embargo, el corpus sólo puede ser un punto de partida que ha de completarse con un análisis de disponibilidad que confirme, rechace o matice las conclusiones a las que se pueda llegar desde variables como la frecuencia y el rango, necesarias pero no suficientes.

En su segunda parte, se dará cuenta del trabajo llevado a cabo en la Universidad de Salamanca orientado a establecer los glosarios correspondientes a los dos primeros niveles de aprendizaje: inicial e intermedio². Se trata de dar cuenta del procedimiento de trabajo realizado, de los parámetros de análisis utilizados y de las conclusiones alcanzadas en términos globales.

Por último se incluyen dos apéndices en los que se recoge una propuesta de glosario para los dos niveles antes apuntados. Obviamente, en el segundo de ellos se han excluido las entradas coincidentes con las del primero.

2. DEFINICIÓN DE GLOSARIOS LÉXICOS

2.1. *Antecedentes y primeros acercamientos*

Muchos de los esfuerzos en los trabajos sobre léxico se han orientado a definir el vocabulario mínimo correspondiente a cada nivel de aprendizaje. Ante esta cuestión se han tomado diversas posiciones de análisis.

Un primera línea de estudios se ha centrado exclusivamente en analizar el léxico dominado por estudiantes de LE. De entre ellos podemos citar el de Laufer (1991), quien ha realizado un trabajo a partir de 97 estudiantes de inglés. Desde el

² Se trata de un trabajo que he tenido el placer de dirigir desde el año 1996. En él ha intervenido de manera directa un equipo de profesores de español como lengua extranjera de Cursos Internacionales de la Universidad de Salamanca. Vaya desde aquí mi reconocimiento a todos ellos: Juan Eguiluz, Ángel Eguiluz, Marisol Martín, Alberto Buitrago, Begoña Núñez, Milagros del Castillo y Sonsoles García.

punto de vista del grado de dominio léxico, concluye que se pueden establecer cinco niveles:

Nivel 1	Dominio de menos de 2.000 palabras ³
Nivel 2	Dominio de 2.000 palabras
Nivel 3	Dominio de 3.000 palabras
Nivel 4	Dominio de 4.000 palabras
Nivel 5	Dominio de 5.000 palabras

Aun siendo una propuesta atractiva, da la sensación de que la muestra de estudiantes analizados no es suficiente como para que los resultados alcanzados sean fiables. Además, Laufer se limita a proporcionar datos puramente cuantitativos, sin indicar si los estudiantes susceptibles de ser incluidos en cada nivel dominan las mismas palabras o si simplemente conocen un número aproximadamente igual con independencia de que éstas varíen de un individuo a otro.

Una segunda línea de investigación se ha inclinado por estudiar cuál es el grado de dominio léxico de los hablantes nativos. El argumento utilizado como punto de partida se puede enunciar de la siguiente forma: si podemos definir qué palabras dominan los usuarios de una lengua materna y cuántas palabras son capaces de aprender durante cada determinado período de tiempo en su proceso de adquisición, podremos establecer las correspondencias oportunas en lo referido a los estudiantes de LE.

El problema de este tipo de trabajos es doble: de un lado, los resultados alcanzados en los diferentes proyectos realizados suelen ser muy divergentes; de otro, carecemos de estudios que nos permitan efectuar correctamente las correspondencias apuntadas. Así, por ejemplo, Goulden, Nation y Read (1990) estiman que un hablante nativo de nivel universitario de inglés domina un vocabulario de 17.000 familias de palabras. Por otro lado, D'Anna, Zechmeister y Hall (1991) afirman que un estudiante universitario medio conoce unas 16.785. Más adelante, Zechmeister, Chronis, Cull, D'Anna and Healy (1995) concluyen que los estudiantes sobrestiman su conocimiento léxico y proponen los datos siguientes:

Estudiantes de último año de secundaria	9.684 palabras
Estudiantes de primer año de universidad	16.679 palabras
Adultos mayores	21.252 palabras

Por una tercera vía transitan los trabajos en los que se ha usado el criterio de COBERTURA a la hora de definir glosarios léxicos para estudiantes de LE. El objetivo final de todos ellos sería el de seleccionar un número muy limitado de palabras que permitiera al alumno expresarse en el espectro más amplio posible de situaciones.

³ De aquí en adelante utilizo el término PALABRA aunque, en rigor, sería más preciso usar LEMA.

En esta línea, uno de los proyectos más conocidos es el BASIC (British American Scientific International Commercial)⁴. Se trata de un estudio combinado que ha permitido dar con 850 palabras (600 sustantivos, 150 adjetivos, 18 verbos y 82 operadores) y un número muy limitado de reglas morfosintácticas. Mediante estos útiles lingüísticos, el aprendiz sería capaz de transmitir cualquier información elemental. El glosario se simplifica mucho, puesto que se invita al estudiante a que exprese las palabras desconocidas mediante perífrasis. Veamos algunos ejemplos. A la izquierda, indicamos la palabra en inglés no incluida en el BASIC; y a la derecha, la forma propuesta a la hora de expresar dicho contenido:

To ask	To make a request
To eat	To take food
Enemy	One who is not a friend
Heavy	Of great weight
To like	To be pleased with
Many	A number

Obviamente, el problema de esta propuesta estriba en que, aun aceptando que el alumno fuera capaz de expresarse, difícilmente podría entender las palabras de su interlocutor, salvo que éste también se expresara mediante BASIC⁵.

Los estudios de cobertura están en desuso en el marco de los estudios de LE aunque tuvieron una enorme influencia entre 1950 y 1970. Así, podemos apuntar que es uno de los parámetros usados en la creación de uno de los trabajos más conocidos para la lengua inglesa, como es la *General Service List of English Words* (West, 1953): se trata de un glosario de 2.000 palabras, basado en los criterios de frecuencia, rango y cobertura⁶, y definido a partir de un corpus de 5 millones de palabras.

Las dificultades que implica realizar trabajos de investigación como los antes expuestos han llevado en numerosas ocasiones a que se hayan propuesto glosarios para estudiantes de LE basados exclusivamente en la introspección, la intuición y la experiencia profesional de sus autores. Tal es el caso de uno tan conocido como es el trabajo de Van Ek (1976), donde se incluye un *Lexicon for T-level English*. Es revelador que el propio Van Ek admita que se trata de un modo de actuar muy arriesgado y que sería más oportuno buscar un acercamiento más científico al problema.

2.2. La necesidad de un corpus

La mayor parte de los estudios que hemos presentado hasta este momento coincide al afirmar que es necesario partir de un corpus a la hora de intentar elaborar un glosario fiable. Sin embargo, en buena parte de los que aceptan tal

⁴ Para más información sobre este proyecto se puede consultar ODGEN (1930).

⁵ Se ha llegado a criticar el BASIC afirmando que se trata de una suerte de Esperanto.

⁶ Más adelante en este trabajo daremos cuenta de los conceptos de frecuencia y rango.

premisa se plantean numerosos problemas de procedimiento. De hecho, si nos fijamos en los trabajos de esta naturaleza realizados hasta ahora llegaremos a algunas conclusiones que resultan, cuando menos, desesperanzadoras:

- La mayor parte de los corpus se construyen exclusivamente a partir de muestras de lengua correspondientes a la lengua escrita.
- No siempre se usa el lema como unidad de recuento⁷.
- Los criterios de análisis difieren de unos a otros.
- Los corpus varían mucho en tamaño.

Las insuficiencias apuntadas se pueden observar si nos fijamos en algunos de los corpus existentes para la lengua española que nos recuerdan Van Els et al. (1984)⁸:

Autor	Número de palabras	Palabras distintas	Indicación de rango	Tipos de texto
Buchanan, 1927	1.200.000	No se indica	Sí	Lengua escrita
García Hoz, 1953	400.000	12.402	Sí	Lengua escrita
Juilland and Chang Rodríguez, 1964	500.000	5.024	Sí	Lengua escrita
Rodríguez Bou, 1952	7.000.000	10.000	No	Lengua escrita y hablada

Con independencia de que muchos de los corpus arriba citados resultan inaccesibles en la actualidad, habrá que plantearse si resultan o no adecuados para la definición de un glosario correspondiente a un nivel de aprendizaje. Efectivamente, necesitaríamos un corpus lo suficientemente representativo y adecuado en términos cuantitativos, pero también cualitativos: debe basarse en géneros textuales y discursivos muy diversos⁹. Además, debe ser consistente en lo referido a la unidad de recuento¹⁰. Finalmente, debe ser sistemático en la aplicación de criterios de análisis.

Si nos fijamos en el ámbito de la lengua española y nos preguntamos si existe un corpus que cumpla los requisitos antes apuntados, la respuesta será, necesariamente, negativa.

⁷ De hecho, el uso del LEMA tampoco está exento de problemas. Piénsese en cuestiones como los homógrafos, las lexías complejas, etc.

⁸ En este trabajo se puede encontrar una recopilación exhaustiva de los estudios de corpus realizados hasta mediados de los ochenta. Aquí sólo ofrecemos algunos de los datos correspondientes a los corpus de lengua española que aparecen en él. Aunque hoy en día se han puesto en marcha, desde numerosas instituciones de dentro y fuera de España, numerosas recopilaciones de corpus, prácticamente ninguna de ellas resulta aún accesible.

⁹ La mayor parte de corpus existentes están basados de manera casi exclusiva en la lengua literaria.

¹⁰ Parece que hay consenso a la hora de aceptar que ésta debe ser el LEMA. NATION (1983) propone que, para saber el número de palabras reales del inglés que un usuario domina, habría que multiplicar el número de lemas por 1,6.

2.3. Criterios para un análisis de corpus

Con el fin de dar respuesta a la última de las insuficiencias antes apuntadas, presentaremos a continuación los dos parámetros que la mayor parte de los investigadores coinciden en destacar como más importantes. De hecho, son los que habitualmente se utilizan a la hora de analizar un corpus léxico: frecuencia y rango.

2.3.1. Frecuencia

Con ella se analiza el número de concurrencias de una palabra en un corpus dado. Habitualmente se distinguen dos tipos: la *absoluta* maneja cifras entendida en términos totales, mientras que la *relativa* los analiza desde una perspectiva porcentual. Así, la segunda resulta mucho más interesante puesto que nos proporciona datos mucho más fiables con independencia de que aumente o se reduzca el tamaño del corpus estudiado.

En relación con ella, algunos estudios se han preocupado por descubrir qué frecuencia relativa debe mostrar una palabra antes de seleccionarla o no para un glosario. En esta línea, podemos citar, por ejemplo, el llevado a cabo por Sciarone (1979).

Otro grupo de trabajos se plantea la frecuencia en términos distintos: parte del hecho de que un número bastante reducido de palabras es la base de un alto porcentaje de cualquier texto. Así, para la lengua inglesa, numerosas investigaciones parecen confirmar que con aproximadamente 2.000 palabras se cubre el 85% de un texto¹¹. Uno de los estudios más conocidos es el de Carroll, Davies and Richman (1971), en el que trabajan con un corpus de cinco millones de palabras, de las cuales 86.741 son distintas. Analizan la relación entre el número de palabras distintas y el porcentaje de corpus cubierto. En el cuadro siguiente se da cuenta de los resultados por ellos alcanzados. Así, por ejemplo, con 10 palabras se cubre un 23,7% del corpus, y así sucesivamente:

Palabras distintas	Porcentaje de corpus cubierto
10	23,7
100	49,0
2.000	81,3
3.000	85,2
5.000	89,4
43.831	99,0
86.741 ¹²	100,0

Por una línea similar transita el trabajo de Nation (1990). En él, aplica el análisis de Carroll et al. al Diccionario Webster, de 128.000 palabras. El cotejo realizado le permite establecer la siguiente proyección:

¹¹ Desde esta percepción se ha llegado al concepto de vocabulario nuclear. Para esta cuestión puede consultarse CARTER, 1987.

¹² El 40% de las palabras sólo aparece en una ocasión.

	Palabras distintas	Porcentaje de corpus que pueden cubrir
Palabras de mucha frecuencia	2.000	85%
Vocabulario académico	800	8%
Vocabulario técnico	2.000	3%
Palabras de frecuencia baja	123.200	2%
Total	128.000	100%

A partir de estos datos, y desde el punto de vista de la enseñanza del inglés como lengua extranjera, concluye que las 2.000 primeras exigen mucho tiempo y atención tanto por parte del alumno como del docente. Añade que, una vez aprendidas, la selección del glosario para los niveles siguientes dependería de los intereses específicos del estudiante.

Pese a los resultados obtenidos, Nation reconoce que el análisis basado de manera exclusiva en la frecuencia no es del todo fiable. Llega a tal conclusión después de constatar los siguientes hechos:

- Algunas palabras muy útiles aparecen muy abajo en términos de frecuencia.
- Algunas palabras de mucha frecuencia no parecen muy adecuadas para un principiante.
- Las listas basadas en frecuencia presentan muchas diferencias.
- La frecuencia no proporciona un orden didácticamente adecuado.
- En algunos niveles de enseñanza la frecuencia deja de ser un parámetro digno de confianza.

Todo lo antes expuesto le permite concluir que es necesario usar más parámetros en el proceso de configuración del glosario. Entre ellos podrían estar los siguientes:

- Rango
- Necesidades lingüísticas
- Accesibilidad y familiaridad
- Cobertura
- Regularidad
- Facilidad de ser aprendidas

Dejando al margen el rango, en el que me detendré en la sección siguiente de este trabajo, y la cobertura, ya citada anteriormente, cabe incluir el resto de ellos en lo que se ha dado en llamar criterios psicológicos y didácticos. Efectivamente, algunas propuestas van encaminadas a justificar la inclusión de determinadas palabras desde estas dos perspectivas. Así, por ejemplo, el criterio didáctico permitiría incorporar palabras

por razones de regularidad, de necesidades lingüísticas o de facilidad para ser aprendidas¹³.

2.3.2. Rango

Con él se hace referencia a la distribución de una palabra a lo largo del corpus o de las subpartes que podamos hacer de él. Cuanto mayor es el rango de un término, es más fácil pensar que es de uso común en el lenguaje. Por tanto, se trata de un indicativo de estabilidad. Desde esta perspectiva, en algunos estudios sirve como criterio de corrección (así sucede, por ejemplo, en el corpus de Juilland and Chang-Rodríguez, 1964) y se convierte en un argumento primordial a la hora de tomar decisiones sobre palabras cuya frecuencia no sea ni muy alta ni muy baja.

Se trata de un parámetro de análisis que tradicionalmente se ha entendido como secundario en relación con el de frecuencia. No obstante, resulta esencial a la hora de relativizar los errores que puedan derivar de la construcción misma del corpus. Efectivamente, si éste resulta desequilibrado desde un punto de vista discursivo, el rango de coapariciones nos puede dar una información incluso más relevante de la que podemos deducir de la frecuencia. De hecho, en la segunda parte de este trabajo, se ha considerado el rango como primer parámetro de selección, por encima incluso de la frecuencia relativa. Lo veremos en su momento.

2.4. Estudios de disponibilidad léxica en estudiantes de LE

El punto de partida para este tipo de trabajos debe buscarse en los estudios de disponibilidad efectuados en la sociolingüística. De manera muy sucinta se puede resumir el modelo de actuación de la siguiente manera: se entrevista a un determinado número de nativos seleccionados a partir de una serie de variables (edad, sexo, trabajo y origen geográfico); se configura una lista de temas o áreas léxicas y se solicita a los informantes que escriban las palabras que consideran más útiles en ese campo. Una vez realizada la encuesta se realiza un estudio de frecuencia que toma como punto de referencia la ubicación de las palabras en la lista.

El procedimiento, no obstante, plantea problemas de difícil resolución. Así, el hecho mismo de partir de una selección de áreas léxicas condiciona las respuestas. Más aún, se ha detectado que es mucho más frecuente que en las respuestas se incluyan sustantivos frente a otras clases de palabras. Por otro lado, existen muchas dificultades a la hora de detectar la disponibilidad de términos de naturaleza abstracta.

En el contexto de la enseñanza de LE se ha planteado la cuestión del análisis de disponibilidad desde una perspectiva distinta. Uno de los modelos más conocidos es el de Paribakht and Wesche (1993 y 1997). Estos autores centran su interés en si, a

¹³ Para profundizar en la cuestión del criterio didáctico y psicológico, puede consultarse T. VAN ELS et al. (1984).

la hora de ampliar el glosario léxico de estudiantes en un nivel intermedio, la enseñanza sistemática del vocabulario a partir de ejercicios descontextualizados tiene más éxito que la lectura extensiva. Así, establecen un protocolo con el cual evaluar el dominio de un glosario léxico¹⁴. Para cada palabra se le plantea al estudiante una serie de opciones entre las que debe decidir. A continuación se ofrece una transcripción del protocolo utilizado:

Self-report categories	
I	I don't remember having seen this word before.
II	I have seen this word before, but I don't know what it means.
III	I have seen this word before, and I think it means ____ (synonym or translation).
IV	I know this word. It means ____ (synonym or translation).
V	I can use this word in a sentence: _____.
(If you do this section, please also do Section IV)	

Cada respuesta se valora numéricamente de acuerdo con la siguiente escala:

I	1	The word is not familiar at all.
→		
II	2	The word is familiar but its meaning is not known.
→		
III	3	A correct synonym or translation is given.
→		
IV	4	The word is used with semantic appropriateness in a sentence.
→		
V	5	The word is used with semantic appropriateness and grammatical accuracy in a sentence.
→		

La asignación de puntuación para cada respuesta depende de si el informante ha sido capaz de contestar correcta o incorrectamente¹⁵.

El modelo de Paribakht y Wesche ha sido adaptado por Slagter (2000) para el castellano. Ésta es la versión utilizada por el profesor holandés:

¹⁴ Este protocolo habría de ser cumplimentado por dos grupos paralelos de estudiantes que habrían recibido una formación metodológicamente distinta, según lo apuntado en la hipótesis que les sirve de punto de partida. Para justificar la validez de su escala efectuaron en 1993 un prexamen en un grupo de 93 estudiantes correspondientes a 6 niveles distintos de enseñanza (Escuela de Verano del Instituto de Segundas Lenguas de la Universidad de Ottawa), a quienes se les aplicó una lista de 32 palabras.

¹⁵ El lector interesado puede encontrar una descripción muy completa del experimento en PARI-BAKHT and WESCHE (1997). Por razones de espacio tengo que renunciar ahora a darle todo el espacio que, en rigor, se merece ese excelente trabajo.

I	No recuerdo haber visto esta palabra.
II	He encontrado esta palabra alguna vez, pero no recuerdo su significado.
III	He visto esta palabra antes y creo que significa ____ (Sinónimo o traducción).
IV	Conozco esta palabra. Significa _____ (Sinónimo o traducción).
V	Puedo utilizar esta palabra en una oración.

Renuncio ahora a transcribir los criterios de valoración ponderada de cada respuesta porque Slagter mantiene básicamente los mismos que existían para el inglés¹⁶.

Es indudable que se hace necesario efectuar estudios de disponibilidad que nos permitan cotejar los resultados alcanzados desde los análisis de rango y frecuencia. Sin embargo, todavía quedan problemas por resolver. En mi opinión, éstos han sido acertadamente apuntados por Nation (1990). Así, este autor recuerda las dificultades inherentes a la puesta en marcha de este tipo de pruebas.

De un lado, apunta obstáculos de naturaleza cuantitativa. Efectivamente, la realización de cuestionarios muy extensos incide en el nivel de motivación de los informantes y, por ello, en sus respuestas.

De otro, afirma que es complicado decidir la forma misma de evaluar. Se han propuesto diferentes alternativas para que los estudiantes puedan responder a una encuesta de disponibilidad: traducción, redacción de una frase, selección múltiple o simple reconocimiento. Sin embargo, todas ellas plantean problemas. La primera de las posibilidades resulta muy adecuada en lo que tiene que ver con la administración de la prueba; sin embargo, plantea grandes dificultades a la hora de evaluar los resultados, especialmente cuando se trata de grupos multilingües. La segunda opción es muy poco operativa, en la medida en que exige una extraordinaria cantidad de tiempo tanto para la administración como para la evaluación. El uso de ítems de selección múltiple conduce a graves problemas de fiabilidad. Para soslayarlos exigiría un nivel de preexamen tal que convertiría el mecanismo en impracticable. Por otro lado, resultaría excepcionalmente difícil construir ítems para palabras de determinados dominios léxicos. La última de las hipótesis implica problemas derivados del hecho de que hay una tendencia en los aprendices a sobrestimar su conocimiento de vocabulario; no obstante, si ésta es una tendencia general en todos los estudiantes, se podrían buscar mecanismos de corrección que nos ofrecieran resultados más fiables.

¹⁶ En la segunda parte de este trabajo, presentaré los protocolos de disponibilidad que se han utilizado para el estudio realizado en la Universidad de Salamanca.

3. DEFINICIÓN DE UN GLOSARIO LÉXICO PARA UN NIVEL INICIAL DE ESTUDIO DEL ESPAÑOL COMO LENGUA EXTRANJERA

3.1. *Introducción*

La necesidad de incorporar un diccionario a un método multimedia de aprendizaje de nivel inicial nos ha movido a efectuar un estudio sobre si el glosario inicial que en principio parece gozar con el consenso de la mayor parte de los investigadores efectivamente coincide con el léxico disponible del grupo de estudiantes que deben dominarlo. Para ello, durante un plazo de aproximadamente seis meses se ha efectuado un estudio de disponibilidad de dicho glosario.

La selección de las palabras evaluadas se ha realizado en dos fases. En la primera de ellas se han comparado varias de las listas de palabras existentes en la actualidad para este nivel¹⁷. Hemos trabajado sobre la base de tres de ellas: la definida para el Nivel Umbral por el Consejo de Europa, la utilizada en la construcción del método *Viaje al Español* y la correspondiente al examen para la obtención del Certificado Inicial de Español, expedido por el Ministerio de Educación y Cultura. Las elegidas se han distribuido en los siguientes campos léxicos:

- Origen y procedencia: la nacionalidad.
- Origen y procedencia: el país.
- El ser humano: el cuerpo humano.
- El ser humano: la identificación personal.
- La percepción física: la vista.
- La percepción física: el oído.
- La percepción física: el tacto y el olfato.
- El ser humano: la higiene.
- El ser humano: la salud.
- El ser humano: la apariencia física.
- El ser humano: la actividad intelectual.
- El comportamiento humano: los gustos y las preferencias.
- El comportamiento humano: opiniones, comparaciones y consejos.
- El comportamiento humano: argumentos, causas y consecuencias.
- El comportamiento humano: formas de ser y estados de ánimo.
- El comportamiento humano: posibilidad, duda, fines y posesión.
- El entorno privado: la vivienda.
- El entorno privado: el vestuario y los complementos.
- El entorno privado: la alimentación.
- Las relaciones de parentesco y amistad.
- La comunicación personal.
- La relación personal.

¹⁷ Este estudio comparativo nos ha permitido comprobar que existe acuerdo entre las diferentes listas en aproximadamente un 90% de las palabras incluidas, por lo que cabe afirmar, a priori, que se trata de una lista bastante fiable.

La actividad profesional: profesiones y actividades laborales.
La actividad profesional: relaciones laborales.
Conocimientos y creencias: el arte, la ciencia la religión.
Conocimientos y creencias: el aprendizaje y el estudio.
El entorno público: lugares de residencia y alojamiento.
El entorno público: establecimientos públicos.
El entorno público: servicios públicos.
El entorno público: medios de comunicación.
El entorno público: medios de transporte.
El ocio: espectáculos, juegos y aficiones.
El ocio: espectáculos, lugares, fiestas y celebraciones.
La ciudad.
El mundo: el universo.
El mundo: la materia.
La tierra: los accidentes geográficos.
La tierra: la flora.
La tierra: la fauna.
La tierra: el clima y el medio ambiente.
Los objetos.
El tiempo: la división del tiempo.
El tiempo: relaciones temporales.
El espacio: relaciones espaciales.
El espacio: la división del espacio.
El espacio: el movimiento en el espacio.
El peso y el volumen.
La forma.
El dinero y las relaciones económicas.
Cantidades y frecuencias.
El tamaño.

3.2. *Estudio de disponibilidad*

Una vez distribuidas, se ha corregido de nuevo la lista, incorporando y eliminando algunas entradas con el fin de equilibrar los diferentes campos. Como resultado de todo ello se ha llegado a un glosario final de 1.211 palabras. Se han establecido dos tipos de grupos meta para completarlo. En primer lugar estarían los estudiantes que inician un curso de nivel intermedio; en este caso, la encuesta se les hace durante los tres primeros días de clase. En segundo lugar, estarían los estudiantes que finalizan un curso de nivel inicial¹⁸; a estos alumnos se les administra la prueba durante los tres últimos días de su curso. En términos cuantitativos, éstos son los datos más relevantes del estudio:

¹⁸ Estos estudiantes de nivel inicial o intermedio participaban en cursos de 80, 100, 120 ó 150 horas de clase.

Número de palabras evaluadas:	1.211
Número de grupos de estudiantes analizados:	50
Número total de encuestas realizadas:	25.107
Media de encuestas por palabras:	21

Para efectuar la encuesta se ha diseñado el siguiente protocolo:

Ancho	
1. No conozco esta palabra:	<input type="checkbox"/>
2. Creo conocer esta palabra pero no estoy seguro de su significado:	<input type="checkbox"/>
3. Conozco esta palabra y soy capaz de usarla en una frase: ¹⁹	<input type="checkbox"/>

Tras completar el trabajo de campo se ha efectuado la informatización de los datos. A continuación se muestra un ejemplo del programa utilizado²⁰:

Palabra			
Ancho			
Respuesta 1	Respuesta 2	Respuesta 3	Grupos
9	2	3	Grupo 2
8	3	3	Grupo 38

Después de ejecutar los recuentos oportunos, se han analizado los datos. Para ello han sido sometidos a dos hipótesis de trabajo independientes.

Hipótesis 1. Se asigna una valoración numérica a cada una de las respuestas posibles, de acuerdo con el cuadro siguiente:

Variable	«Desconozco»	-2,5 para el 100%
Variable	«Creo conocer»	5 para el 100%
Variable	«Uso»	10 para el 100%

¹⁹ A efectos de cómputo, si un estudiante respondía con la opción 3, pero construía una frase incorrecta en lo referido al uso de la palabra, se interpretaba la respuesta como correspondiente al grupo 2.

²⁰ El programa informático de Microsoft Access utilizado fue diseñado por don José Luis Alonso Berrocal, profesor del Departamento de Informática y Automática de la Universidad de Salamanca.

Veámoslo en la palabra ANCHO:

Total respuestas:	28	
VARIABLE	RESPUESTAS	VALORACIÓN PARCIAL
Variable «Desconozco»	17	-1,51
Variable «Creo conocer»	5	0,89
Variable «Uso»	6	2,14
Valoración total:	1,52	

Se decide que deben eliminarse todas aquellas palabras que alcancen una valoración total inferior a 2,77. Según esta hipótesis, de las 1.211 punto de partida, se aceptan 961 y se desdeñan 250.

Hipótesis 2. A las tres variables anteriores se incorpora otra, combinación de la segunda y la tercera:

Variable 1	«Desconozco»
Variable 2	«Creo conocer»
Variable 3	«Uso»
Variable 4 (=2+3)	«Conozco»

Se efectúa un proceso de selección graduada de las palabras, merced a la cual se han alcanzado los resultados que se presentan en el cuadro siguiente:

Se aceptan aquellas que «Conoce» es mayor del 60%	848
De las restantes, se acepta aquellas donde «Conoce» está entre el 55% y el 60%	33
De las restantes, se acepta aquellas donde «Uso» es mayor del 40%	64
De las restantes, se acepta aquellas donde «Uso» está entre el 35% y el 40%	17

De acuerdo con este procedimiento, de las 1.211 punto de partida, se aceptan 962 y se eliminan 249.

Una vez ejecutadas las dos hipótesis, se cotejan las dos listas de palabras rechazadas. Se decide suprimir aquellas que aparecen en ambas, y mantener las que sólo resultan excluidas desde una de las hipótesis (esto es, 27 palabras). Finalmente, se ha tomado la decisión de recuperar algunas palabras²¹. Resultado de todo este proceso ha sido un glosario final compuesto de 1.012 entradas. En el Anexo 1 a este trabajo se incluyen listadas alfabéticamente.

²¹ Esto se debe a diversas razones: en algunos casos se trataba de palabras difícilmente interpretables descontextualizadas. En otros, porque eran esenciales en términos de configuración de algunos campos léxicos. Finalmente, se recuperaron algunas en las que se detectaron erratas en la presentación de las mismas. En cualquier caso estamos hablando de un número de palabras que no superaba las quince.

4. DEFINICIÓN DE UN GLOSARIO LÉXICO PARA UN NIVEL INTERMEDIO DE ESTUDIO DEL ESPAÑOL COMO LENGUA EXTRANJERA

4.1. *Introducción*

En el origen de la investigación que a continuación se presenta está la responsabilidad adquirida por la Universidad de Salamanca ante el Ministerio de Educación y Cultura de hacerse cargo, desde el año 1991, de la elaboración y corrección de los Diplomas de Español como Lengua Extranjera. A partir de ese momento, una de las mayores preocupaciones de dicha institución ha sido la de garantizar la fiabilidad de las pruebas para conseguir cada uno de los tres diplomas en la actualidad existentes. Cubrir este objetivo pasa, entre otras cosas, por contar con una exhaustiva descripción del nivel de dificultad de tales exámenes. Con el fin de satisfacer esta necesidad, durante 1992 se realizó un notable esfuerzo colectivo en el marco de la Universidad de Salamanca. Fruto de él fue la edición, por parte del entonces Ministerio de Educación y Ciencia, de la *Guía para la obtención de los Diplomas de Español como Lengua Extranjera (DELE)*²². En ella se da cuenta del grado de exigencia de las pruebas para la obtención de los exámenes correspondientes a los dos diplomas de nivel intermedio y superior: el Diploma Básico de Español (DBE) y el Diploma Superior de Español (DSE), respectivamente. Sin embargo, en cierto sentido el trabajo realizado quedó incompleto, puesto que carece de una definición del nivel léxico.

En el caso del Diploma Superior de Español (DSE) los problemas son objetivamente menores, puesto que el propio nivel de exigencia permite poder incluir virtualmente cualquier palabra de la lengua española.

Mucho más espinosa, sin embargo, resulta la cuestión del examen para la obtención del Diploma Básico de Español (DBE), que reconoce un «nivel intermedio» de conocimiento de nuestra lengua. El entrecomillado es voluntario: ¿qué es un nivel intermedio? De la misma manera que parece haber cierta unanimidad a la hora de establecer una correlación entre el nivel inicial y el nivel umbral²³, no se detecta tanto acuerdo en el momento de fijar los contenidos «intermedios». Mayores divergencias se observan cuando se trata de efectuar tal labor en el plano léxico.

Sobre la base de estas premisas, la pregunta que nos planteamos los responsables de la creación de las pruebas para la obtención del DBE es si debemos fiar la selección léxica de las mismas a nuestra intuición, a nuestra experiencia como docentes y como evaluadores de español LE. Los resultados alcanzados por los candidatos que se han presentado a las diferentes convocatorias parecen avalar que

²² Desgraciadamente, se trata de una edición muy limitada y poco accesible. No obstante, a muy corto plazo los interesados podrán consultarla en la siguiente dirección electrónica: <http://www.usal.es/curespus>.

²³ Con todas las matizaciones que se quieran hacer a esta afirmación, claro está.

nuestra labor no va muy desencaminada y que la selección léxica propuesta para cada examen es, en términos generales, aceptable.

Sin embargo, se hace necesario dar un paso adelante, y plantear una propuesta de actuación científicamente consolidada que sustituya un modelo de comportamiento dirigido por la intuición y la subjetividad. Lógicamente, ese trabajo sólo ha sido posible cuando se ha contado con un corpus, en forma de pruebas de examen, mínimamente amplio como para poder llevar a cabo un análisis que nos pueda proporcionar datos fiables. Esta labor se inició en el año 1996, a partir de 14 pruebas de examen para la obtención del DBE efectuadas entre 1991 y 1995. El trabajo de análisis informático nos ha permitido resumir las características del corpus de la siguiente manera²⁴:

Corpus total	57.617	
Glosario	4.100	
Frecuencia máxima	5.434 ²⁵	9,4312%
Frecuencia mínima	1 ²⁶	0,0017%

Relación frecuencia/Cantidad de texto cubierto:		
	Palabras (%)	Total de texto cubierto (%)
Frecuencia inferior a 4:	2.472 (60%)	3.832 (6,6%)
Frecuencia igual o superior a 4:	1.628 (40%)	53.785 (93,4%)

Los datos resultan cuando menos tranquilizadores: un glosario de 1.628 palabras cubre más de un 93% del total del corpus. Dicho de otro modo, parece lícito deducir que cualquier candidato que conozca esa lista estará preparado para comprender prácticamente la totalidad de cualquiera de los textos incluidos en los exámenes. Este primer análisis, en principio, nos indica que el proceso de creación de las pruebas no va por mal camino.

4.2. *La necesidad de definir un glosario de nivel intermedio*

El paso siguiente nace de la pregunta que inevitablemente nos hemos realizado: ¿se puede colegir de los datos anteriores que ese glosario se corresponde con el del nivel intermedio? Responder a esa pregunta de modo afirmativo resulta, cuando menos, apresurado. Se hace necesario contar con un corpus mucho más amplio. Desgraciadamente ese material no existe o no está disponible, ya lo hemos

²⁴ Para la cobertura informática fue decisiva la colaboración de profesorado adscrito al Departamento de Informática y Automática de la Universidad de Salamanca, especialmente del profesor José Luis Alonso Berrocal.

²⁵ Se trata del artículo determinado EL.

²⁶ Un total de 1.369 palabras distintas.

dicho antes, en la actualidad. Sin embargo, de lo que sí disfrutamos en nuestro país es de una floreciente actividad editorial en materia de enseñanza de español LE. Fruto de ella ha sido la publicación de numerosos manuales y libros de texto para los diferentes estadios del aprendizaje. Podemos, entonces, plantear una hipótesis atractiva: si completamos el corpus extraído a partir de los exámenes para la obtención del DBE con el que podamos deducir de diversos métodos de enseñanza que «afirman» dirigirse a un nivel intermedio, podremos detectar cuáles son los puntos de acuerdo entre unos y otros. Dicho de otro modo, podremos dejar de hablar de un nivel intermedio para pasar a proponer el nivel intermedio en el que parece haber consenso.

Así pues, para ampliar el corpus se han seleccionado aleatoriamente los siguientes métodos de enseñanza:

- EQUIPO PRAGMA (1997): *Esto funciona B*. 9ª edición, 2ª reimpr. Madrid, Edelsa.
 EQUIPO PRAGMA (1998): *Esto funciona A*. 9ª edición, 3ª reimpr. Madrid, Edelsa.
 GÓMEZ ASENCIO, J.; PRIETO, E. y BORREGO, J. (1986): *Progresos*. Salamanca, Universidad de Salamanca.
 MILLARES, S. y CENTELLAS, A. (1996): *Método de español para extranjeros. Nivel intermedio*. Madrid, Edinumen.
 MIQUEL, L. y SANS, N. (1995): *Rápido*. Madrid, Difusión.
 SÁNCHEZ, A. y PASCUAL, M.ª T. (1995): *Cumbre. Nivel medio*. Madrid, SGEL.

Tres de ellos, *Esto Funciona*, *Rápido* y *Cumbre* incluyen un apéndice en el que se listan las palabras que, en opinión de los autores configuran el glosario correspondiente al nivel de aprendizaje hacia el que se orientan. Tal hecho ha facilitado la labor de informatización del material. Los tres restantes se digitalizaron íntegramente, por lo que también se cuenta para ellos con un cálculo de frecuencia relativa.

Finalmente, el corpus se ha completado con el glosario activo y pasivo que se exige para la obtención del Certificado Inicial de Español (CIE)²⁷: parece legítimo aceptar que *a priori* todo él debería formar parte del glosario intermedio. Una vez realizado todo el proceso informático, el corpus ha quedado configurado de la siguiente manera:

²⁷ Se trata del tercer título oficial del MEC. En su momento, un grupo de expertos convocados por el Instituto Cervantes y la Universidad de Salamanca redactó el nivel de exigencia del mismo. En él se incluía el vocabulario activo y pasivo exigible.

FUENTE	CORPUS	GLOSARIO	GLOSARIO REFINADO ²⁸	OBSERVACIONES
CUMBRE	N ²⁹	1.944	1.912	Incluye vacías
ELE	18.814	1.906	1.906	
MILLARES	13.255	1.728	1.728	
ESTO FUNCIONA	N	1.775	1.258	Excluye <i>Para Empezar</i> ³⁰
PROGRESOS	26.678	3.002	3.002	
RÁPIDO	N	3.229	2.713	Incluye vacías
DBE	57.617	4.100	3.831	
CIE ACTIVO Y PASIVO	N	1.876	1.805 ³¹	Excluye vacías

Total palabras corpus: 125.238

Palabras distintas: 6.499

4.3. Selección de un primer glosario

A continuación se ha decidido establecer un primer glosario sobre la base del rango de aparición de las palabras en relación con las ocho fuentes de las que se alimenta el corpus. En su totalidad, la distribución de rango es como sigue:

Rango	Número de palabras	%
1	3.027	46,57
2	1.062	16,34
3	663	10,20
4	465	7,15
5	395	6,07
6	359	5,52
7	365	5,61
8	163	2,50

En principio, se han aceptado las palabras cuyo rango fuera igual o mayor a 4. Sin embargo, tras efectuar un revisión de la lista resultante, se ha optado por eliminar 46 palabras con un rango entre 4 y 7. A continuación, se ofrecen algunos ejemplos:

²⁸ Se alcanza este glosario una vez revisadas las repeticiones, los errores ortográficos, etc.

²⁹ La N indica que no se digitalizó el texto completo sino solamente el glosario incluido al final del mismo.

³⁰ Incluye alguna palabra vacía, aunque no todas, porque estaban integradas en alguna lexía compleja del glosario final. *Para empezar* es el libro de texto elaborado por los mismos autores para el nivel inicial. Por ello no incluyeron en el glosario de *Esto funciona* las palabras que se enseñaban en él.

³¹ El vocabulario activo exigido en el CIE es de 1.182 palabras. El pasivo, incluidas las anteriores es de 1876.

Rango 4				
ajustar	capa	portavoz	refugio	rellenar
Rango 5				
avance	dolor	operación		
Rango 6				
estupendo	serie			
Rango 7				
arma	empleado	imagen		

La fase inicial del análisis ha permitido incluir 1.691³² palabras. Esta selección se ha incrementado con otras 393³³. A la hora de incluirlas se han tomado en consideración esencialmente dos factores: si formaban parte de la lista CIE o si estaban dentro de la lista DBE con una frecuencia absoluta superior a 3. Las características de este grupo se presenta en el cuadro siguiente:

Palabras de rango menor de 4	392	19% del total
En el CIE	145 (37%)	7% del total
Fuera del CIE	247 (63%)	12% del total
En el DBE	155:	
Frecuencia > 3:	58 (37,45%)	2,8% del total
Frecuencia < 4:	97 (62,55%)	4,65% del total
Rango 3:	52	
Rango 2:	23	
Rango 1:	22	
Fuera del DBE	92	
Rango 3:	34 (8,6%)	1,63% del total
Rango 2:	26 (6,6%)	1,24% del total
Rango 1:	32 (8% de 392)	1,5% del total ³⁴

Como se puede observar, las palabras de rango inferior a 4 y que no estarían plenamente justificadas en términos de rango y frecuencia apenas suman un 9% del total de palabras aceptadas para la siguiente fase del procedimiento.

³² De éstas, 108 no aparecían en el DBE.

³³ Se aceptaron palabras con un rango inferior a 4 por dos razones: a) que fueran palabras vacías, y b) que fueran necesarias para equilibrar determinados campos léxicos.

³⁴ Por poner sólo un ejemplo, en este grupo de palabras de rango 1 y que no aparecían en el DBE ni en el CIE se incluyeron palabras como las siguientes: austriaco, boliviano, brasileño, irlandés, marroquí, panameño, puertorriqueño, suizo o turco.

4.4. Estudio de disponibilidad

A continuación se ha realizado un estudio de disponibilidad de las 2.083 palabras que en un principio se han seleccionado. Considerando las dificultades, ya apuntadas más arriba en este trabajo, inherentes a la realización de este tipo de pruebas, se ha optado por reducir al máximo el número de palabras evaluadas. Para ello se parte del glosario existente para el nivel inicial que ya he descrito en el capítulo anterior.

De las 1.012 que componen el glosario inicial ya evaluado en clases de español LE³⁵, 886 aparecen exactamente iguales en el glosario intermedio, mientras que 126 aparecen en el glosario inicial pero no en el glosario intermedio. De estas últimas se detecta que aunque 102 muestran diferencias ortográficas, realmente son las mismas³⁶. Así pues, hay un total de 24 palabras que deben agregarse, aunque sin necesidad de evaluarlas, al glosario intermedio.

Por tanto, de las 2.083 palabras del glosario intermedio, 988 ya han sido explícitamente confirmadas merced al estudio de disponibilidad efectuado para el nivel inicial. Asimismo, se renuncia a evaluar un bloque de 151 palabras de naturaleza vacía o deíctica³⁷. Por tanto, han de ser examinadas con un nuevo test un total de 944 palabras³⁸.

Con el fin de agilizar el procedimiento se ha desarrollado un protocolo que se presenta a continuación:

Abandonar	
1. No recuerdo haber visto esta palabra o expresión antes	<input type="checkbox"/>
2. He encontrado esta palabra o expresión alguna vez, pero no recuerdo su significado	<input type="checkbox"/>
3. He visto esta palabra o expresión antes y creo conocer su significado	<input type="checkbox"/>
4. Conozco esta palabra o expresión y estoy seguro de su significado	<input type="checkbox"/>

El procedimiento se inició en el mes de marzo de 1999 y terminó en diciembre de ese mismo año. Para efectuarla se ha seleccionado el siguiente público meta. En primer lugar estarían los estudiantes que inician un curso de nivel avanzado; en este caso, la encuesta se les efectuaría durante los tres primeros días de clase. En segundo lugar, estarían los estudiantes que finalizan un curso de nivel

³⁵ Del que se da cuenta al final de este documento en un anexo

³⁶ Por ejemplo, se trataría de casos en los que en el glosario inicial aparece la palabra ABURRIRSE, y en el glosario intermedio, ABURRIR.

³⁷ Se han eliminado del glosario las preposiciones, conjunciones, artículos, pronombres personales, demostrativos, indefinidos, numerales, etc.

³⁸ El número real de encuestas fue de 972.

intermedio³⁹; a estos alumnos se les administra la prueba durante los tres últimos días de su curso.

En términos cuantitativos, éstos son los datos más relevantes del estudio:

Número de palabras evaluadas:	972
Número de grupos de estudiantes analizados:	50
Número total de encuestas realizadas:	35.211
Media de encuestas por palabras:	37

Tras completar el trabajo de campo se ha efectuado la informatización de los datos. A continuación se muestra un ejemplo del programa utilizado⁴⁰:

Palabra				
Abandonar				
Grupos	Respuesta 1	Respuesta 2	Respuesta 3	Respuesta 4
Grupo 1	1	1	4	4
Grupo 12	2	0	4	3
Grupo 13	2	2	2	6

Después de ejecutar los recuentos oportunos, se han analizado los datos. Para ello han sido sometidos a dos hipótesis de trabajo independientes.

Hipótesis 1. Se asigna una valoración numérica a cada una de las respuestas posibles, de acuerdo con el cuadro siguiente:

Variable 1	-2,5 para el 100%
Variable 2	1 para el 100%
Variable 3	5 para el 100%
Variable 4	10 para el 100%

³⁹ Estos estudiantes de nivel intermedio o avanzado participan en cursos de 80, 100, 120 ó 150 horas de clase.

⁴⁰ El programa informático de Microsoft Access utilizado fue diseñado por don José Luis Alonso Berrocal, profesor del Departamento de Informática y Automática de la Universidad de Salamanca.

Veámoslo en la palabra CIERTO:

Total respuestas:	39	
Variable	Respuestas	Valoración parcial
Variable 1	1	-0,06
Variable 2	1	0,03
Variable 3	4	0,51
Variable 4	33	8,46
Valoración total: 8,94		

Se decide que deben eliminarse todas aquellas palabras que alcancen una valoración total inferior o igual a 2,3. Según esta hipótesis, de las 972 punto de partida, se aceptan 811 palabras.

Hipótesis 2. Se efectúa un proceso de selección graduada de las palabras, merced a la cual se han alcanzado los resultados que se presentan en el cuadro siguiente:

Se aceptan aquellas donde la suma de las Variables 3+4 es mayor del 60%	717
De las restantes, se acepta aquellas donde la suma de las Variables 2+3+4 es mayor del 65%	95

De acuerdo con este procedimiento, de las 972 punto de partida, se aceptan 812.

Una vez ejecutadas las dos hipótesis, se cotejan las dos listas de palabras rechazadas. Se decide incluir aquellas que aparecen en ambas, y suprimir las que resultan excluidas desde una de las hipótesis (esto es, 37 palabras más 5 incluidas erróneamente en el glosario). Finalmente, se ha tomado la decisión de recuperar algunas palabras⁴¹. Resultado de todo este proceso ha sido un glosario final compuesto de 799 entradas que se debe sumar a las 1.012 palabras del nivel inicial, lo que suma 1.811 a la que se deben incorporar 199 palabras vacías. Todo ello arroja un total de 2.010 palabras. En el Anexo 2 a este trabajo se incluyen listadas alfabéticamente, excluidas las coincidentes con el glosario correspondientes al nivel inicial y las palabras vacías.

⁴¹ Esto se debe a diversas razones: en algunos casos se trataba de palabras difícilmente interpretables descontextualizadas. En otros, porque eran esenciales en términos de configuración de algunos campos léxicos. Finalmente, se recuperaron algunas en las que se detectaron erratas en la presentación de las mismas. En cualquier caso estamos hablando de un número de palabras que no superaba las quince.

5. CONCLUSIÓN

El objetivo último de este trabajo es el de proporcionar una herramienta consistente al profesional de la enseñanza de español como lengua extranjera. Ya es momento de que dejemos de hacer depender de nuestra intuición una cuestión tan delicada como la selección del léxico que debemos enseñar o sobre la cual construir ejercicios y actividades para los estudiantes: no olvidemos que muchos de ellos fracasan entre otras cosas porque los estudiantes son incapaces de entender el contexto léxico en el que se proponen.

Obviamente, los glosarios que proponemos pueden ser discutibles, pero entendemos que gozan de dos virtudes: la primera de ellas es que nacen con una voluntad de consenso; la segunda, que son el resultado de un análisis científico de los datos. Por ello, creemos que son extraordinariamente novedosas en el ámbito de la enseñanza de ELE en nuestro país. Quedan, pues, al servicio de los interesados.

En la actualidad trabajamos en la creación de una herramienta informática merced a la cual los creadores de materiales didácticos puedan comprobar si éstos se corresponden en términos léxicos con el nivel de enseñanza intermedio o inicial. Con esto daremos por concluida la investigación puesta en marcha. Renunciamos a tratar de definir niveles por encima de los ya propuestos, Entre otras cosas, porque, siguiendo a Nation, dudamos que pueda llevarse a cabo.

ANEXO 1:
Lista de palabras correspondientes al nivel inicial

abajo	americano	ayudar
abogado	amigo	ayuntamiento
abrazo	amor	azúcar
abrigo	andar	azul
abril	animal	bailar
abrir	anoche	bajar
abuelo	anterior	bajo
aburrido	antes	balcón
aburrirse	antipático	balón
acabar	anuncio	baloncesto
accidente	año	banco
aceite	apagar	bañarse
aceituna	aparato	baño
aconsejar	apartamento	bar
acordarse	apellido	barato
acostarse	aprender	barba
actor	aprobar	barco
adecuado	aproximadamente	barrio
adelante	aquí	bastante
adelgazar	árabe	basura
adiós	árbol	bebé
aeropuerto	arena	beber
afeitarse	argentino	bebida
africano	armario	belga
agencia	arriba	beso
agenda	arroz	biblioteca
agosto	arte	bicicleta
agradable	artesanía	bien
agua	artículo	bienvenido
ahí	artista	bigote
ahora	ascensor	billete
aire	así	blanco
ajo	asiático	blando
al lado de	aspecto	boca
alcohol	aspirina	bocadillo
alegrarse	atención	boda
alegría	aún	bolígrafo
alemán	australiano	boliviano
alfombra	austriaco	bolsa
alquilar	autobús	bolsillo
alto	autonomía	bolso
alumno	autopista	bonito
allí	avenida	bosque
amable	avión	bota
amarillo	aviso	botella
ambulancia	ayer	botón

brasileño	casarse	contento
brazo	casete	contestador automático
broma	casi	contestar
bueno	castillo	conveniente
burro	casualidad	conversación
buscar	catarro	corazón
buzón	catedral	corbata
caballo	causa	coreano
cabeza	cava	correcto
cabina de teléfonos	cebolla	correo
caerse	celebrar	correr
café	cena	cortar
cafetería	cenar	corto
caja	centímetro	cosa
cajero automático	centro	costar
calcetín	centro comercial	crear
calefacción	centroamericano	crema
caliente	cepillo	crystal
calor	cerca (de)	cuaderno
calvo	cerdo	cuadrado
calle	cerrar	cuarto
cama	cerveza	cubano
cámara (de fotos)	cigarrillo	cuchara
camarero	cine	cuchillo
cambiar	cita	cuello
cambio	ciudad	cuenta
camino	clase	cuero
camisa	clásico	cuerpo
camiseta	cliente	cuidado
cámping	clima	cultura
campo	cocina	cumpleaños
canadiense	cocinar	curioso
canal de televisión	cocinero	curso
canción	coche	champú
cansarse	coger	chaqueta
cantante	colegio	chico
cantar	colombiano	chileno
capital	color	chino
cara	comedor	chocolate
carne	comer	danés
carné	comida	dar
carnicería	compañero	de acuerdo
caro	comprar	debajo (de)
carretera	comprender	deber
carta	concierto	decir
cartel	conducir	dedo
cartera	conocer	defender
cartero	consejo	dejar
casa	contaminación	delante (de)

delgado	dulce	estudiar
demasiado	duro	estupendo
demostrar	ecuatoriano	estúpido
dentista	edad	etcétera
dentro (de)	edificio	européo
depende	egipcio	exacto
dependiente	ejemplo	examen
deporte	eléctrico	excursión
desayunar	elefante	explicar
desayuno	elegante	exposición
descanso	empezar	exterior
describir	empresa	fábrica
descripción	en punto	fácil
desea	enamorarse	falda
despacio	encantar	familia
después	encima (de)	familiar
detrás (de)	encontrar(se)	famoso
día	enero	fantástico
diario	enfermedad	farmacia
dibujo	enfermero	favorito
diccionario	enfermo	fax
diciembre	enfrente (de)	febrero
diente	engordar	fecha
diferencia	enorme	felicidades
diferente	ensalada	feliz
difícil	enseguida	feo
dinero	enseñar	fiebre
dios	entender	fiesta
dirección	entonces	filete
dirección	entrada	fin
director	entrar	firma
disco	época	flor
discoteca	error	fondo
discusión	escalera(s)	forma
disfrutar	escribir	formal
distinto	escritor	fotografía
divertido	escuchar	fotógrafo
divertirse	escuela	francés
divorciarse	español	frío
docena	especial	fruta
documento	esperar	frutería
domicilio	esquiar	fuego
domingo	esquina	fuera (de)
dominicano	estación	fuerte
don	estadounidense	fumar
dormir	estanco	funcionar
dormitorio	este	fútbol
ducha	estómago	futuro
ducharse	estudiante	gafas

garaje	idea	latinoamericano
gas	idioma	lavar
gasolina	iglesia	lección
gato	igual	lectura
gente	imaginar	leche
gimnasio	importante	leer
gobierno	importar	lejos (de)
gordo	imposible	lengua
gracias	increíble	león
grande (gran)	indio	letra
gratis	individual	levantar(se)
griego	información	libre
gripe	informático	librería
gris	inglés	libro
grupo	inmediatamente	limón
guapo	instrucciones	limpiar
guatemalteco	inteligente	limpio
guerra	interesante	lista
guía	interesarse	listo
guitarra	interior	litro
gustar	internacional	loco
habitación	invierno	luego
habitante	invitar	luna
habitual	ir(se)	lunes
hablar	irlandés	luxemburgués
hacer	isla	luz
hambre	israelí	llamar(se)
harto	italiano	llave
helado	izquierdo	llegada
hermano	jabón	llegar
hermoso	japonés	lleno
hielo	jardín	llevar
hijo	jefe	llorar
historia	jersey	llover
hoja	joven	madre
hola	juego	maestro
holandés	jueves	maleta
hombre	jugador	malo (mal)
hondureño	jugar	mamá
hora	julio	mano
horario	junio	manta
horrible	juntos	manzana
hospital	kilogramo	mañana
hostal	kilómetro	mapa
hotel	labio	máquina
hoy	lago	mar
huevo	lámpara	maravilloso
húmedo	lápiz	marcar un teléfono
humor	largo	marido

marrón	mueble	oferta
marroquí	muerte	oficial
martes	mujer	oficina
marzo	mundo	ofrecer
más	museo	oír
matrimonio	música	ojo
mayo	músico	olvidar(se)
mayor	muy	opinión
media	nacer	orden
medicina	nacionalidad	ordenador
médico	nada	oreja
medio	nadar	organización
medio ambiente	naranja	organizar
mediodía	nariz	oro
mejor	natural	otoño
memoria	necesario	oveja
menor	necesidad	padre
menos	necesitar	paella
mensaje	negar	pagar
mentira	negro	página
menú	nervioso	país
mercado	nevar	paisaje
mes	nicaragüense	pájaro
mesa	nieto	palabra
metal	niño	palacio
metro	noche	pan
mexicano	nombre	panadería
miedo	normal	panameño
mientras	norte	pantalones
miércoles	norteamericano	papá
mínimo	noruego	papel
ministro	noticia	paquete
minuto	novela	paraguayo
mirar	noviembre	parar
mismo	novio	parecerse
moda	nube	pareja
moderno	nuevo	parque
modo	número	parte
molestar	nunca	particular
momento	objeto	partido
moneda	obligación	pasado
mono	obligatorio	pasajero
montaña	obra	pasaporte
monumento	obrero	pasar
morir(se)	observar	paseo
motocicleta	octubre	paso de peatones
mover	ocupar	pasta de dientes
mucho	odiar	pastel
mucho gusto	oeste	pastelería

pastilla	político	receta
patata	poner	recibir
paz	ponerse	recomendar
pedir	portugués	recordar
peinar	posibilidad	recto
película	posible	redondo
pelirrojo	postal	regalar
pelo	postre	regalo
pelota	precio	región
pensar	precioso	regular
pensión	preferir	reír(se)
pequeño	pregunta	religión
perdón	preguntar	reloj
perfecto	premio	repetir
periódico	preocuparse	residencia
periodista	presentar	responder
permiso	presente	respuesta
perro	presidente	reunión
persona	primavera	reunirse
personal	primo	revista
personalidad	principio	rey
peruano	prisa	rico
pesado	probable	río
pesar	probar	rojo
pescadería	probarse	romper(se)
pescado	problema	ropa
peseta	producto	rosa
petición	profesión	rubio
pez	profesor	ruido
pie	programa	rumano
piel	prohibir	ruso
pierna	pronto	sábado
pijama	propina	saber
pintor	provincia	sacar
piscina	próximo	sal
piso	público	salado
planchar	pueblo	salida
planeta	puente	salir
plano	puerta	salud
planta	puertorriqueño	salvadoreño
plástico	querer	sangre
plato	queso	secar(se)
playa	quitarse	seco
plaza	quizá	secretario
pobre	radio	secreto
poco	rápido	sed
poder	rato	seguir
polaco	rebajas	seguro
policía	repcionista	sello

semáforo	té	venir
semana	teatro	ventana
sentarse	teléfono	ver
sentirse	telegrama	verano
señor	televisión	verdad
señorita	temperatura	verde
septiembre	temprano	verdura
serio	tener	vestido
servicio	tenis	vestirse
servir	terminar(se)	vez
siempre	termómetro	viajar
siesta	terrazza	viaje
siglo	tiempo	vida
significar	tienda	vídeo
siguiente	tigre	viejo
silencio	tímido	viernes
silla	tío	vino
sillón	típico	visitar
simpático	tipo	vivir
situación	título	voluntario
sobrino	toalla	volver
socorro	todo	zapatería
sofá	tomar	zapatilla
sol	tomate	zapato
solo	torero	zoológico
soltero	tormenta	zumo
sopa	tortilla	
sorprender	trabajar	
suave	tráfico	
subir(se)	traje	
sucio	tranquilo	
sudafricano	tren	
sudamericano	triste	
sueco	turco	
suerte	turismo	
suficiente	turista	
suizo	último	
supermercado	universidad	
sur	urgente	
tabaco	uruguayo	
también	usar	
tampoco	vaca	
tapa	vacaciones	
tarde	vacío	
tarjeta de crédito	vale	
tarta	valer	
taxi	vaso	
taxista	vender	
taza	venezolano	

ANEXO 2:

Lista de palabras correspondientes al nivel intermedio
(excluidas las que aparecen en el nivel inicial)

abandonar	asiento	caramelo
absoluto	asignatura	cárcel
absurdo	asistir	cargo
acción	asociación	cariño
acento	atmósfera	carpetas
aceptar	atractivo	carrera
acercar	atrás	caso
acompañar	atravesar	castaño
acondicionado	auténtico	castellano
acostumbrar	automático	castigar
actividad	autor	catalán
acto	avanzar	católico
actual	aventura	cazar
actualidad	avisar	celebración
actuar	ayuda	cenicero
adaptar	ayudante	central
además	azafata	cercano
admirar	baile	certificado
admitir	barcelonés	charlar
adolescente	barón	chicle
adulto	básico	chiste
afectar	bastar	chocar
afición	beca	cielo
aficionado	belleza	ciencia
afirmación	besar	científico
afirmar	biografía	cierto
alarma	bomba	cinturón
alimento	bombero	circo
alrededor	borracho	círculo
amar	brillante	ciudadano
amoroso	británico	claro
analizar	caber	clasificar
andaluz	cable	club
animar	cabo	coincidir
anteayer	cajón	cola
antiguo	calcular	colaborar
anunciar	calendario	colección
aparcamiento	calidad	colgar
aparcar	callar	colonia
aparecer	caminar	columna
argumento	candidato	comentar
arquitecto	cantidad	comenzar
arreglar	carácter	comercio
artificial	característico	comienzo

comité	controlar	deportivo
cómodo	convencer	deprimir
compañía	convenir	deprisa
comparar	convertir	derecho
completar	convivir	desaparecer
completo	corregir	desaparición
complicado	correspondencia	desarrollar
compra	corresponder	desarrollo
compuesto	correspondiente	desastre
común	corrida	descansar
comunicación	corriente	desconocer
comunicar	costa	descubrir
comunidad	costumbre	descuento
comunista	crear	deseo
conclusión	crecer	desierto
concreto	crimen	desnudo
concurso	crisis	desordenar
condición	crítica	despacho
conductor	criticar	despedir
conferencia	cruzar	despertador
confianza	cuadro	despertar
confundir	cualidad	despierto
conocimiento	cuento	destino
consecuencia	cuestión	destrucción
conseguir	cuidar	destruir
conservador	culpa	detalle
conservar	culpable	detective
considerar	culto	detener
consistir	cultural	determinado
constantemente	cuñado	devolver
construcción	curar	diálogo
construir	curva	dibujar
consultar	dato	dificultad
consumidor	de repente	directamente
consumir	debate	directo
consumo	deberes	dirigir
contaminar	decidir	disponer
contar	decisión	dispuesto
contexto	declarar	distancia
continente	deducir	diverso
continuación	definir	doble
continuamente	definitivo	doctor
continuar	demás	dogmático
contrario	democracia	droga
contraste	democrático	droguería
contratar	departamento	duda
contrato	deportista	dudar

dueño	evidente	ganador
durar	evitar	ganar
echar	excepto	garganta
ecología	exclusivamente	gastar
ecologista	exclusivo	gasto
economía	existencia	general
económico	existir	generoso
editorial	éxito	gesto
educación	experiencia	gimnasia
efecto	explicación	girar
egoísta	explosión	golpe
ejercicio	explotar	grado
ejército	expresar	gramática
elaborar	expresión	grave
elección	extender	gritar
electricidad	extrañar	guardar
electrodoméstico	extranjero	hábito
elegir	extraño	histórico
elemento	extraordinario	horno
empleo	fabricar	humanidad
empresario	factura	humano
encender	falso	ideal
encerrar	falta	idéntico
enemigo	faltar	identificar
energía	fatal	ilegal
enfadar	favor	ilógico
engañar	ficción	ilustración
enterarse	filosofía	ilustrado
entrevista	firmar	ilustrar
entrevistar	física	ilustre
enviar	físico	imaginable
equipaje	flamenco	imaginario
equipo	formación	importancia
equivalente	formar	impreso
equivocar	fórmula	incendio
escapar	fragmento	incluir
esencial	frecuencia	incompleto
espacio	frecuente	inconveniente
espalda	frente	indicar
espectáculo	fresa	industrial
espectador	fresco	infancia
espejo	frontera	infantil
esposo	fuerza	inferior
esquí	función	informar
estado	función	informativo
estatura	fundamental	informe
estilo	gallego	iniciar

inocente	manifestación	notar
insistir	manifestar	noticiero
instalar	mantener	objetivo
instituto	maravilla	obligar
instrumento	marca	obtener
insultar	marchar	ocasión
intención	marco	occidental
intentar	masculino	océano
interés	matar	ocurrir
interpretar	matemáticas	oído
intervenir	material	olímpico
introducir	materna	opinar
inútil	mayoría	oportunidad
invencible	mecánico	oposición
inventar	medir	ordenar
inventor	mejorar	oriental
investigación	mente	origen
joya	mentir	original
jubilado	menudo	orquesta
justificar	meter	oscuro
justo	miembro	par
juventud	militar	parada
ladrón	modelo	parado
lamentar	monarquía	paraguas
lástima	montar	pared
lata	montón	pariente
lento	moral	participar
ley	moreno	partir
libertad	mostrar	pasear
licenciado	motivo	peligro
ligero	motor	peligroso
línea	muchacho	pena
literatura	muerto	pendiente
llamada	mundial	península
llenar	municipal	peor
lluvia	musical	perder
local	nacimiento	perfeccionar
localizar	nacional	perfumería
lógico	naturaleza	período
lotería	navidad	permitir
luchar	negativo	perseguir
lugar	negocio	personaje
madera	nervio	personalmente
madrileño	nieve	pertenecer
madrugada	nombrar	pesca
mandar	norma	pescador
manera	nota	pescar

peso	propio	resultado
petróleo	proponer	resultar
piedra	protagonista	resumen
pintar	proteger	resumir
pintura	provocar	retirar
pisar	proyecto	retraso
plan	psicólogo	reunir
plancha	publicar	revelar
plata	publicidad	revisar
plátano	puerto	robar
plazo	puntual	robo
pluma	quedar	rodilla
plural	raro	romano
población	rayo	romántico
poderoso	razón	rueda
poema	reacción	sala
poesía	reaccionar	salón
poeta	real	saltar
política	realidad	saludar
póliza	realizar	saludo
popular	reciente	santo
porcentaje	recoger	satélite
portero	recomendación	satisfecho
posiblemente	reconocer	sección
posición	recuerdo	seguridad
positivo	reducir	selección
posterior	referencia	seleccionar
practicar	referir	sensación
práctico	reflejar	sentimiento
precisamente	régimen	separar
preciso	regresar	serpiente
preferencia	regreso	sexo
preocupación	relación	similar
preparar	relacionar	simple
presencia	reparar	sistema
presentación	repartir	sitio
prestar	representar	situar
pretender	república	sobrevivir
principal	reservar	social
príncipe	resolver	sociedad
privado	respeto	soldado
procurar	respetar	soledad
producir	respirar	solicitar
profesional	responsabilidad	solicitud
progreso	responsable	solitario
prometer	restaurante	sólo
pronunciar	resto	solución

solucionar	terrible	valle
sombrero	texto	valor
sonar	tierra	valoración
soñar	timbre	vaquero
sonido	tirar	varios
sonreír	titular	vasco
sonrisa	tocar	vecino
soportar	todavía	vegetación
sorpresa	tonto	vegetariano
sucedecr	toro	velocidad
suceso	total	vendedor
suegro	trabajador	venta
sueldo	trabajo	ventaja
suelo	tradicción	verdadero
sueño	tradicional	víctima
sufrir	traducción	viento
sugerir	traer	vigilar
suponer	tranquilizar	violencia
supuesto	transportar	violento
suspender	transporte	visita
sustituir	tras	vista
taco	tratar	viudo
tardar	tropical	vivo
tarea	trozo	volar
taza	turístico	volumen
técnica	único	votar
técnico	unidad	voto
tecnología	unión	voz
telefónico	unir	vuelo
televisor	universitario	vuelta
tema	urgencia	ya
temer	uso	yugoslavo
temporada	útil	zona
tenedor	utilizar	
tener ganas	valenciano	

BIBLIOGRAFÍA

- CARTER, R. (1987): *Vocabulary*. London, Allen & Unwin.
- CARROLL, J.; DAVIES, B. and RICHMAN, B. (1971): *The American Heritage Frequency Book*. Boston, American Heritage.
- COADY, J. and HUCKIN, T. (1997): *Second Language Vocabulary Acquisition*. Cambridge, C.U.P.
- D'ANNA, C. A.; ZECHMEISTER, E. B. and HALL, J. W. (1991): Toward a meaningful definition of vocabulary size, *Journal of Reading Behaviour*, 23 (1), 109-122.
- GOULDEN, R.; NATION, P. and READ, J. (1990): How large can a receptive vocabulary be, *Applied Linguistics*, 11 (4), 341-363.
- LAUFER, B. (1991): How much lexis is necessary for reading comprehension. En P. J. L. ARNAUD and H. BÉJOINT (eds.). *Vocabulary and applied linguistics*. Basingstoke, Macmillan, pp. 126-132.
- NATION, I. S. P. (1983): *Learning and teaching vocabulary*. Wellington, Victoria University.
- NATION, P. (1990): *Teaching and learning vocabulary*. Boston, MA, Heine & Heine Pub.
- ODGEN, C. K. (1930): *Basic English: A general introduction with rules and grammar*. London, Kogan, Paul, Rench and Trubner.
- PARIBAKHT, T. and WESCHE, M. (1993): The relationship between reading comprehension and second language development in a comprehension-based ESL program, *TESL Canada Journal*, 11, 9-29.
- (1997): Vocabulary enhancement activities and reading for meaning in second language vocabulary acquisition. En J. COADY and T. HUCKIN (eds.): *Second Language Vocabulary Acquisition*. Cambridge, C.U.P.
- SLAGTER, P. J. (1999): ¿Interactuar y negociar el significado con un PC. En J. FERNÁNDEZ GONZÁLEZ et al.: *Lingüística para el siglo XXI*. Salamanca: Ediciones Universidad de Salamanca, pp. 1511-1520.
- VAN EK, J. A. (1976): *The threshold level for modern language in schools*. London, Longman.
- VAN ELS, T. et al. (1984): *Applied Linguistics and the Learning and Teaching of Foreign Languages*. Baltimore, MA, Edward Arnold.
- WEST, M. (1953): *A general service list of English words*. London: Longman, Green and Co.
- ZECHMEISTER, E. B.; CHRONIS, A. M.; CULL, W. L.; D'ANNA, C. A. and HEALY, N. A. (1995): Growth of a functionally important lexicon, *Journal of Reading Behaviour*, 27 (2), 201-212.