

Tesis doctoral

El verbo *dar* en el español escrito de aprendientes de L1 inglés: estudio comparativo entre hablantes no nativos y hablantes nativos basado en corpus



Anna Sánchez Rufat

DEPARTAMENTO DE FILOLOGÍA HISPÁNICA Y LINGÜÍSTICA GENERAL
2015



Tesis doctoral

El verbo dar en el español escrito de aprendientes de L1 inglés: estudio comparativo entre hablantes no nativos y hablantes nativos basado en corpus

Anna Sánchez Rufat

DEPARTAMENTO DE FILOLOGÍA HISPÁNICA
Y LINGÜÍSTICA GENERAL

Conformidad de las directoras

Dra. Carmen Galán Rodríguez

Dra. Fiona MacArthur Purdon

2015

ÍNDICE DE CONTENIDOS

1. INTRODUCCIÓN	1
1.1. Tema y objetivo de la investigación	1
1.2. La investigación de corpus (informatizados) de aprendientes	1
1.3. Antecedentes y motivación de la investigación	3
1.4. El objeto de la investigación y la metodología empleada	5
1.5. Estructura de la tesis	11
2. ESTADO DE LA CUESTIÓN SOBRE LOS ESTUDIOS DE LA LENGUA DEL APRENDIENTE	15
2.1. Estudios sobre la lengua del aprendiente antes de los corpus informatizados	18
2.1.1. El Análisis Contrastivo	18
2.1.2. El Análisis de Errores	20
2.1.3. La Adquisición de Segundas Lenguas. Del prejuicio sobre los corpus de aprendientes a su apreciación en el análisis de la interlengua	25
2.2. La investigación de corpus de aprendientes	28
2.2.1. Revisión de los corpus de aprendientes	32
2.2.2. Algunos aspectos metodológicos y limitaciones de la investigación basada en corpus	40
3. LA REVALORIZACIÓN DEL LÉXICO Y LA COMPETENCIA LÉXICA DEL VERBO	45
3.1. La revalorización del léxico	47
3.1.1. Algunas contribuciones de la teoría lingüística	48
3.1.2. Algunas contribuciones de la teoría psicolingüística	52
3.1.3. Repercusiones pedagógicas	59
3.2. El conocimiento del verbo: la competencia léxica	69
3.2.1. Los rasgos morfológicos	71
3.2.2. La naturaleza semántica	73
3.2.3. El significado y sus relaciones paradigmáticas y sintagmáticas	78
3.2.4. La interfaz léxico-sintaxis	80
3.2.5. El factor sociolingüístico	86
3.2.6. Rasgos de la competencia léxica del verbo	89

4. EL VERBO <i>DAR</i> : NATURALEZA LÉXICO-SEMÁNTICA Y PROYECCIÓN SINTÁCTICA	95
4.1. Consideraciones sobre la naturaleza léxico-semántica de <i>dar</i>	96
4.1.1. Hacia una definición de dar que se visualice en la sintaxis: problemas encontrados y soluciones para el análisis de los datos	96
4.1.1.1. Sobre los contenidos aspectuales y las idiosincrasias léxicas en el significado unitario de <i>dar</i> : la trayectoria y la meta en la transferencia y en la creación	97
4.1.1.2. Sobre la existencia de rasgos mínimos de <i>dar</i> que determinan la selección semántica de sus argumentos. La concordancia léxica y la redundancia	103
4.1.2. Descripción de los sentidos de dar: problemas encontrados y soluciones para el análisis de los datos	111
4.2. Tablas	117
5. LA CUESTIÓN DE LA NORMA EN ELE	125
5.1. La norma descriptiva	128
5.2. Debilidades de la norma descriptiva	131
5.3. Norma y estándar	137
5.4. Conclusiones	145
6. METODOLOGÍA	149
6.1. Los datos y herramientas	149
6.1.1. El CEDEL2	149
6.1.2. Antconc	156
6.2. Los procedimientos	161
6.2.1. Cuestiones preliminares sobre la metodología del análisis de interlengua basado en corpus	162
6.2.2. Estructuras consideradas en el análisis	164
6.2.3. Fuentes seleccionadas para contrastar los datos	166
6.2.3.1. Introspección	167
6.2.3.2. Otros corpus textuales	169
6.2.3.3. Diccionarios combinatorios	170
6.2.3.4. Obras de consulta generales de carácter panhispánico	171
6.2.4. Criterios para determinar el grado de aceptabilidad de los datos	172
6.2.5. Criterios para determinar la inaceptabilidad de los datos	180
6.2.5.1. Localización del error	181
6.2.5.1.1. Error relevante frente a error no relevante	184

6.2.5.2. Descripción del error	186
6.2.5.2.1. El modelo de funcionamiento del lexicón utilizado en la descripción del error: el lexicón y el sistema computacional	187
6.2.5.2.2. Descripción del error: error léxico y error computacional	189
6.2.5.3. Explicación del error	196
7. EL USO DEL VERBO <i>DAR</i>: ANÁLISIS DE CORPUS CONTRASTIVO	201
7.1. Relaciones del índice de frecuencia entre los dos corpus. Análisis cuantitativo y cualitativo	202
7.1.1. Frecuencia total del lema <i>dar</i> : ¿tienden los HNN a utilizar este lema más o menos que los HN?	205
7.1.1.1. Análisis cuantitativo	205
7.1.1.2. Análisis cualitativo	206
7.1.2. Construcciones sintácticas del verbo y su frecuencia	207
7.1.2.1. Análisis cuantitativo	209
7.1.2.2. Análisis cualitativo	211
7.1.2.3. Interpretación de las relaciones de frecuencia en cada una de las estructuras sintácticas entre el corpus nativo y el no nativo	214
7.1.2.3.1. Alta frecuencia en HNN <i>vs.</i> alta frecuencia en HN	214
7.1.2.3.2. Baja frecuencia en los HNN <i>vs.</i> baja frecuencia en los HN	222
7.1.2.3.3. No frecuencia en HNN <i>vs.</i> baja frecuencia en HN	232
7.1.2.3.4. Síntesis del análisis de las relaciones de frecuencias	233
7.1.3. Palabras y tipos de palabras con los que se combina <i>dar</i>	235
7.1.3.1. El uso de <i>dar</i> como verbo de transferencia	237
7.1.3.1.1. Análisis cuantitativo	237
7.1.3.1.2. Análisis cualitativo	240
7.1.3.2. El uso de <i>dar</i> como verbo de creación	241
7.1.3.2.1. Análisis cuantitativo	241
7.1.3.2.2. Análisis cualitativo	244
7.2. Aceptabilidad e inaceptabilidad de los datos no nativos. Análisis cuantitativo y cualitativo	247
7.2.1. Frecuencia de los datos aceptable-naturales, aceptable-restringidos y de los inaceptables	247
7.2.2. Localización del error	249

7.2.3. Descripción y explicación del error en el núcleo verbal	249
7.2.4. El error y el acierto en las estructuras analizadas	258
7.2.5. Descripción y explicación del error en el complemento	261
7.2.6. Descripción y explicación del error en el sujeto	264
7.2.7. El error en el sintagma	265
7.3. Conclusiones sobre el uso de dar	267
7.3.1. Diferencias y semejanzas por estructuras	267
7.3.2. Diferencias y semejanzas por usos de dar en sus dos significados	269
7.3.3. El papel de la transferencia de la L1	270
8. DISCUSIÓN	273
8.1. Revisión de la metodología desarrollada en esta investigación	273
8.1.1. El enfoque cuantitativo y cualitativo. Hipótesis de partida contrastadas y conclusiones sobre los resultados obtenidos	273
8.1.2. La identificación de las necesidades de los HNN a partir de las semejanzas y diferencias en el español nativo y el español no nativo. La necesidad de una norma prescriptiva y de una matización de los sobregeneralizadores conceptos de infrautilización y sobreutilización	277
8.2. Limitaciones de la investigación	279
8.3. Futuras investigaciones	282
8.4. Conclusión	283
9. CONCLUSIÓN	285
9.1. Implicaciones en los estudios de interlengua del español	285
9.1.1. La función de diagnóstico	285
9.1.2. La función evaluativa	290
9.2. Algunas posibles aplicaciones didácticas de esta investigación	292
9.2.1. Materiales didácticos	292
9.2.2. El aula de ELE	295
9.2.3. Diccionarios	304
9.3. Perspectivas de futuro de los estudios de la interlengua del español basados en corpus	305
REFERENCIAS BIBLIOGRÁFICAS	309
ANEXO	329

ÍNDICE DE FIGURAS, TABLAS Y GRÁFICOS

Figura 1. Relación de los significados del verbo <i>dar</i>	77
Tabla 1. Conocimientos necesarios para la producción escrita del verbo	91
Tabla 2. Combinatoria composicional de <i>dar</i> en la construcción V + N	120
Tabla 3. Combinatoria de <i>dar</i> en otras construcciones sintácticas diferentes a la de V + N.	122
Tabla 4. Combinatoria de <i>dar</i> en construcciones fraseológicas	123
Figura 2. Diseño del corpus CEDEL2	151
Gráfico 1. Proporción de palabras en el análisis del CEDEL2	153
Figura 3. Concordancias de <i>dar</i> (del CEDEL2) en Antconc	159
Figura 4. Tabla de Excel facilitada por Cristóbal Lozano con las redacciones, información adicional sobre sus condiciones de producción, y datos personales y lingüísticos de los participantes	160
Figura 5. Concordancias importadas en una tabla de Excel una vez iniciado el análisis de los datos	161
Tabla 5. Juicios de aceptabilidad	177
Tabla 6. Dimensión de la localización en la tipología del error	183
Tabla 7. Dimensión descriptiva de la tipología del error	196
Tabla 8. Dimensión explicativa de la tipología del error	199
Tabla 9. Relación entre los índices de frecuencia nativos y no nativos	203
Tabla 10. Frecuencia de <i>dar</i> en HN y HNN	206
Tabla 11. Tipos de construcciones sintácticas en las que participa <i>dar</i>	208
Tabla 12. Frecuencia de <i>dar</i> en las diferentes estructuras	209
Gráfico 2. Frecuencias de <i>dar</i> de cada estructura en el corpus no nativo	210
Gráfico 3. Frecuencias de <i>dar</i> de cada estructura en el corpus nativo	210
Tabla 13. Frecuencia de <i>dar</i> en las estructuras ditransitivas y transitivas	215
Tabla 14. Frecuencia de <i>dar</i> en las estructuras impersonales	217
Tabla 15. Frecuencia de cada tipo de impersonalidad	218
Tabla 16. Frecuencia de <i>dar</i> en las locuciones	220
Tabla 17. Frecuencia de cada tipo de locución	220
Tabla 18. Frecuencia de <i>dar</i> en las estructuras intransitivas	222
Tabla 19. Frecuencia de <i>dar</i> en las estructuras pasivas	225

Tabla 20. Frecuencia de <i>dar</i> en las construcción absolutas	226
Tabla 21. Frecuencia de <i>dar</i> en las pronominales	228
Tabla 22. Frecuencia de <i>dar</i> en las construcciones preposicionales	228
Tabla 23. Frecuencia de <i>dar</i> en los esquemas fraseológicos semiproductivos	230
Tabla 24. Frecuencia de <i>dar</i> en la construcción introducida por <i>como</i> y en la construcción con adjetivo	232
Tabla 25. Clases léxicas y sustantivos con los que se combina <i>dar</i> con el sentido de “proporcionar” (en CVP) en el CEDEL2	239
Tabla 26. Las clases léxicas con las que se combina <i>dar</i> con el sentido de “proporcionar” (en las CVP) en las que difieren los porcentajes de uso nativo y no nativo en más de un 5 %	240
Tabla 27. Clases léxicas y sustantivos con los que se combina <i>dar</i> con el sentido de “producir(se)” (en CVA) en el CEDEL2	242
Tabla 28. Las clases léxicas con las que se combina <i>dar</i> con el sentido de “producir(se)” (CVA) en las que difieren los porcentajes de uso nativo y no nativo en más de un 5%.	244
Gráfico 4. Índices de frecuencia de los usos aceptable-naturales, aceptable-restringidos e inaceptables	248
Gráfico 5. Localización del error	249
Gráfico 6. Errores léxicos <i>vs.</i> errores computacionales en el uso de <i>dar</i>	250
Gráfico 7. Frecuencia de los diferentes tipos de error léxico del verbo <i>dar</i>	251
Gráfico 8. Frecuencia en los diferentes tipos de error computacional del verbo <i>dar</i>	253
Gráfico 9. Porcentaje de error de cada estructura	259
Gráfico 10. Porcentaje de error en el núcleo verbal de cada estructura	259

LISTA DE ABREVIATURAS

AC	Análisis Contrastivo
ACI	Análisis Contrastivo de la Interlengua
AE	Análisis de Errores
ASL	Adquisición de Segundas Lenguas
CAI	Corpus de Aprendientes Informatizado
CLCI	Centro de Lingüística de Corpus del Inglés
CEDEL2	Corpus Escrito del Español como Segunda Lengua
CD	Complemento Directo
CI	Complemento Indirecto
CVA	Construcción con verbo de apoyo
CVP	Construcción con verbo que actualiza su sentido pleno
<i>DPD</i>	<i>Diccionario Panhispánico de Dudas</i>
ELE	Español como Lengua Extranjera
HN	Hablante Nativo
HNN	Hablante No Nativo
ICLE	International Corpus of Learner English
L1	Lengua materna
L2	Lengua meta
LE	Lengua meta
LC	Lingüística de Corpus
LINDSEI	International Database of Spoken English
<i>MCER</i>	<i>Marco Común Europeo de Referencia para las Lenguas</i>
N.	Nota
N	Nombre
<i>NGLE</i>	<i>Nueva Gramática de la Lengua Española</i>
SN	Sintagma Nominal
SPLLOC	Spanish Learner Language Oral Corpus
V	Verbo
<i>V.s.</i>	<i>Versus</i>

1. INTRODUCCIÓN

1.1. TEMA Y OBJETIVO DE LA INVESTIGACIÓN

Esta tesis constituye un estudio del uso del verbo *dar* en el español de un grupo de aprendientes anglófonos. El objetivo de la investigación es mostrar la manera en la que un enfoque lingüístico basado en corpus aplicado a los estudios de interlengua¹ del español contribuye a encontrar las semejanzas y diferencias entre el español escrito de un grupo de hablantes no nativos (HNN) y un grupo equivalente de hablantes nativos (HN). Las diferencias y similitudes entre el español de los HNN y el español de los HN permite identificar las necesidades de los HNN referidas al verbo *dar* en la modalidad escrita, pues ese es el carácter de los corpus utilizados.

1.2. LA INVESTIGACIÓN DE CORPUS (INFORMATIZADOS) DE APRENDIENTES²

El principal interés de la investigación en el ámbito de estudio de la Adquisición de Segundas Lenguas (ASL) –al que pertenece este análisis– es encontrar los principios subyacentes que regulan la construcción e interpretación de las estructuras lingüísticas y comunicativas en los diferentes estadios de adquisición de la segunda lengua (L2)³. De esta manera, la ASL se plantea como objetivo construir modelos subyacentes de representación de la lengua de los HNN en estadios particulares de la adquisición, atendiendo a los factores que limitan o favorecen el desarrollo de dicha adquisición (Larsen-Freeman y Long, 1991; Lozano y Mendikoetxea, 2013). La principal fuente de datos para describir el proceso de adquisición y los factores que afectan a este proceso es la propia lengua de los HNN, ya sea producida de manera natural o a través de procedimientos experimentales o de juicios metalingüísticos (Granger, 2002; Lozano y Mendikoetxea, 2013: 2). Así pues, como señalan Lozano y Mendikoetxea (2013: 1), el

¹ *Interlengua* es el término introducido por Selinker (1972) para referirse al sistema estructurado que construye el aprendiente de la lengua extranjera (LE) en un estadio determinado del desarrollo del aprendizaje. Es un sistema lingüístico dinámico que no se concibe como defectuoso o desviado de la norma; se altera continua pero sistemáticamente –ya que se mantiene cierto orden en las secuencias de desarrollo–, pese a la influencia que ejerce sobre ella la lengua materna del aprendiente (Larsen-Freeman y Long, 1991). Este concepto es también denominado *sistema aproximado* (Nemser, 1971) o *competencia transitoria* (Corder, 1967); de aquí en adelante utilizaremos los términos *interlengua* y *lengua de los aprendientes* indistintamente para aludir al sistema lingüístico no nativo, por ser los más extendidos para referirse al sistema lingüístico no nativo.

² En este apartado presentamos la investigación de corpus de aprendientes, que será desarrollada en el capítulo 2.

³ Los términos *segunda lengua* y *lengua extranjera* se emplean en esta investigación indistintamente para referirnos a la lengua meta del aprendiente.

éxito de la investigación en ASL depende de la validez y fiabilidad de estos procedimientos de obtención de datos.

Uno de estos procedimientos desarrollados para la obtención de datos naturales constituye la base del área de investigación lingüística conocida como *investigación de corpus de aprendientes*, que es el resultado de la interrelación de dos disciplinas, hasta hace pocos años desvinculadas entre sí: la Lingüística de Corpus (LC) y la ASL (Granger, 2002: 4).

El desarrollo de los corpus de aprendientes informatizados y el de la investigación de la interlengua basada en corpus ha supuesto un gran avance en los estudios de ASL. Mientras que los trabajos anteriores –basados en otros procedimientos de obtención de datos– eran limitados en cuanto al número de sujetos estudiados para poder controlar las variables que afectan a la producción del aprendiente, los nuevos corpus informatizados permiten trabajar con más y mejor calidad de datos de lengua natural. Así pues, poder analizar amplios corpus de aprendientes que estén bien diseñados, de acuerdo con los criterios establecidos (Sinclair, 2005), proporciona una base empírica sólida para la descripción de la interlengua. Al mismo tiempo, estas descripciones mejoradas de la lengua del aprendiente pueden ser utilizadas en la enseñanza de las L2 (Granger, 2002: 4; Pastor Cesteros, 2004: 98), lo cual enlaza directamente con nuestros intereses en el ámbito de la actuación en el aula; cuando nos referimos a las posibilidades de aplicación de los resultados de este estudio de adquisición a la enseñanza lingüística como nuestro objetivo último, no lo hacemos simplemente desde una perspectiva investigadora, sino también –y fundamentalmente– sobre la base de nuestra trayectoria en la enseñanza de ELE⁴ (véase el capítulo 9 para estas posibles aplicaciones didácticas).

Sylviane Granger ha sido, sin duda, uno de los principales motores de la investigación de la interlengua basada en corpus. En los años 90 inició, en la Universidad Católica de Lovaina, dos proyectos pioneros que han marcado el desarrollo de esta área de estudios: el International Corpus of Learner English (ICLE) y el International Database of Spoken English (LINDSEI). La segunda versión del ICLE y la primera del LINDSEI fueron lanzadas en 2009 y 2010, respectivamente. Las cuestiones metodológicas han sido también una de las principales preocupaciones de Granger; al mismo tiempo que se desarrollaba el corpus ICLE, se diseñaba un marco de investigación del corpus de aprendientes, conocido como Análisis Contrastivo de la

⁴ Dicha trayectoria comienza en 2001, y continúa en la actualidad. Impartí mis primeras clases en la Universidad de Salamanca, tras realizar en dicha Universidad el Máster de Enseñanza del Español como Lengua Extranjera (ELE); entre 2002 y 2007 me dediqué a la enseñanza de ELE como Profesora Asociada en la Universidad de Cambridge (Reino Unido). Entre 2007 y 2010 coordiné e impartí cursos de ELE en la Universidad de Extremadura. Actualmente, compagino la docencia en la Universidad de Córdoba con la impartición de cursos de formación de profesores de ELE en dicha Universidad y en otros centros.

Interlengua (ACI). Este enfoque, a diferencia de los anteriores análisis de la lengua del aprendiente –análisis contrastivo (AC) y análisis de errores (AE)–, trata la lengua del aprendiente, esto es, la interlengua como una entidad autónoma e independiente y permite realizar comparaciones sistemáticas cuantitativas y cualitativas entre la lengua nativa y la interlengua, de las que se obtienen las semejanzas y las diferencias entre el español de ambos grupos y, por lo tanto, los rasgos lingüísticos que los HNN habrán de adquirir en otros estadios de la adquisición si desean alcanzar la naturalidad nativa.

Este enfoque del ACI ha sido muy fructífero en los estudios de la interlengua del inglés, como lo demuestra la gran cantidad de estudios de interlengua contrastivos originados a partir del ICLE⁵, motivo por el que este corpus ha acabado inspirando la creación de corpus similares en otras lenguas, como es el caso del corpus de aprendientes de español utilizado en esta investigación, el Corpus Escrito del Español como Segunda Lengua (CEDEL2).

1.3. ANTECEDENTES Y MOTIVACIÓN DE LA INVESTIGACIÓN

Aunque ha habido un importante desarrollo en el terreno de los estudios de la lengua del aprendiente desde los primeros análisis contrastivos –entre la lengua materna (L1) del aprendiente y la L2– hasta el análisis de la interlengua en la investigación de la ASL, el aspecto colectivo de la naturaleza de la interlengua del español no ha sido suficientemente estudiado (véase el capítulo 2 para un recorrido por los diferentes modelos de análisis de la lengua del aprendiente).

Dentro del campo de la ASL, el objeto de la investigación ha sido más la producción lingüística de HNN individuales que la de grupos de HNN con el mismo perfil (Guo, 2006: 3, 9); no obstante, el interés de investigadores y profesores debiera consistir en conocer los rasgos de la interlengua de un grupo homogéneo de HNN: esos rasgos son generalizables a todos los HNN con un mismo perfil, no solo son aplicables al grupo analizado, por lo que las descripciones y explicaciones lingüísticas obtenidas por los investigadores son muy abarcadoras; y los profesores, si conocen los rasgos de todo un grupo, pueden asegurarse de que las necesidades cubiertas en el aula son las sentidas y compartidas por la mayoría de los aprendientes del grupo. Por todo ello, es necesario examinar el español producido por un grupo más que por el de una sola mente, por lo que, como reivindica Leech (1998: xix), el aspecto colectivo debería ser una faceta de la investigación de la ASL. Sin lugar a dudas, la investigación de la

⁵ La página *web* del proyecto ICLE contiene un apartado que recoge las publicaciones que se sirven de esta base de datos: <<http://www.uclouvain.be/en-169926.html>>.

interlengua basada en corpus es muy útil para descubrir los rasgos que caracterizan la lengua de un grupo determinado.

De todos los aspectos que conforman la interlengua del español, se ha seleccionado para este análisis de corpus un elemento léxico por varias razones. Por un lado, son muy pocos los estudios de interlengua del español que tienen como objeto de estudio el léxico. Este nivel lingüístico es el que más se resiste a la estructuración y, por tanto, el más difícil de analizar sistemáticamente (Terrádez Gurrea, 2001: 14). Por otro lado, cada vez son más las teorías que plantean el vocabulario como el verdadero esqueleto de una lengua (Hunston y Francis, 2000; Sinclair y Mauranen, 2006) y que las palabras son, por lo tanto, las que determinan la forma sintáctica final de las oraciones (Hudson, 2007). Al mismo tiempo, se considera que la competencia léxica está en la base del desarrollo de las destrezas y habilidades comunicativas en una lengua extranjera (Baralo, 2007: 384)⁶, idea comúnmente aceptada desde que ya observara Wilkins hace más de cuarenta años que “Without grammar very little can be conveyed; without words, nothing can be conveyed” (1972: 111). En relación con ello, son los propios aprendientes los que a menudo señalan en el aula su frustración por los bloqueos y malentendidos que les provoca la falta de dominio activo de vocabulario (Cervero y Pichardo, 2000: 93), lo que les lleva a expresar su intención de expandir su vocabulario; pero lo cierto es que estas carencias suelen ser fruto no tanto de la falta de palabras almacenadas en el lexicón mental, sino de su incapacidad de usar las palabras que creen conocer. Es decir, los profesores, antes de disponerse a trabajar nuevas palabras en un nivel avanzado (nivel de dominio que se corresponde con las muestras de lengua analizadas en esta investigación), debieran determinar primero si su grupo de aprendientes ha alcanzado un conocimiento profundo de cada una de las palabras que creen conocer, especialmente de las palabras “simples” comúnmente usadas. Esto permitiría a los profesores conocer el nivel de complejidad en el que los aprendientes pueden operar entre los diferentes sentidos de las palabras polisémicas (Guo, 2006: 4). Un enfoque de análisis de la interlengua basado en corpus puede ayudar a determinar cuál es el grado de profundidad en el conocimiento de una palabra por parte de un grupo de aprendientes.

En suma, la consideración de las anteriores premisas –esto es, 1) el aspecto colectivo de la naturaleza de la interlengua del español no ha sido suficientemente

⁶ Al respecto de la correlación entre el léxico y la comunicación, Linnarud (1986: 120) expresa que “Writing a composition is not primarily an exercise in using correct language, but a means of expressing ideas and communicating a message where lexical choice plays a crucial role”. Así pues, el hecho de que la selección léxica tiene un papel destacado en la redacción –que es precisamente el tipo de texto al que pertenecen los datos analizados en esta investigación– explica la necesidad de iniciar el desarrollo de los estudios de la interlengua del español basados en corpus analizando aspectos de la competencia léxica de los aprendientes.

estudiado y 2) existe una correlación entre el léxico y la forma final de las oraciones y entre el léxico y la comunicación— explica la motivación de este proyecto de investigación, que selecciona un aspecto léxico como objeto de estudio y que aplica para su análisis un enfoque basado en corpus.

1.4. EL OBJETO DE LA INVESTIGACIÓN Y LA METODOLOGÍA EMPLEADA

Dentro del conocimiento léxico, esta investigación se centra en el verbo, concretamente en *dar*. Por un lado, reconocemos la capacidad del verbo para contener concreciones argumentales que provocan que el significado de la oración se asimile al del verbo que la proyecta, lo que lo convierte en un área muy importante de la estructura de cualquier lengua susceptible de generar errores, como observa Housen (2002: 78): “The choice of the verb system as the focus of study in second language acquisition (SLA) is based on the assumption that this is a centrally important area for the structure of any language which is moreover likely to pose major learning problems of any age”. Además, como señala Ringbom (1998a: 192), “Verbs are less topic-sensitive than nouns, and *the most frequently used verbs* may thus provide a good starting point for an assessment of linguistic features characteristic of one group of learners”⁷.

En relación con ello, las palabras simples o “pequeñas” (del inglés *smallwords*) desempeñan un papel central en el éxito comunicativo (Sinclair, 1991; Hasselgren, 2002), lo que las convierte en un importante objeto de análisis. De hecho, *dar* es uno de los verbos usados con más frecuencia en español y forma parte del grupo de esas palabras pequeñas, lo que lo sitúa como un buen punto de partida para evaluar la lengua de un grupo de HNN: ocupa el puesto 39 en el *ranking* de las 5000 palabras más frecuentes del español recogidas en el *Frequency Dictionary of Spanish* (Davies, 2006), que incluye artículos, preposiciones, conjunciones y demás categorías gramaticales. De esas 39 palabras más frecuentes, 25 son monosílabas y el resto bisílabas, lo que apoya la anterior idea de que muchas de las palabras monosílabas, pequeñas, del español se utilizan con frecuencia en la comunicación diaria.

En el corpus nativo del CEDEL2 el verbo *dar* aparece un promedio de una vez cada 546 palabras, lo cual supone casi una aparición por cada página de texto escrito. El uso tan extendido de este verbo significa que los HNN deben adquirirlo. Parece evidente que el hecho de que un elemento sea frecuente en la lengua implica que el HNN está más expuesto a él y, por lo tanto, resulta más fácil de recordar (Nation, 2001; Nagini, 2008; Schmitt, 2010; Schmitt y Redwood, 2011). No obstante, las características

⁷Las cursivas en la cita son nuestras.

atribuidas frecuentemente a los verbos frecuentes (Viberg, 1996) y nuestra experiencia en el aula nos inclinan a pensar que estos se perciben por los profesores y por los alumnos como elementos difíciles de manejar.

Puede haber varias razones que expliquen esta dificultad sentida. En primer lugar, estos verbos polisémicos participan en construcciones sintácticas muy variadas; esta variación sintáctica es intralingüística, ya que el verbo *dar* se localiza en numerosas expresiones idiomáticas del español y selecciona para determinados significados diferentes preposiciones, por lo que HNN debe adquirir el uso que los HN hacemos no solo del verbo *dar* en su construcción básica *dar* + nombre (N), sino también de unidades multipalabra (del inglés *multiword units*) del tipo a *dar con* (+ sintagma nominal [SN]), *dar en* (+ SN), *dar a* (+ infinitivo o + SN), *dar contra* (+ SN) o *dar por* (+ infinitivo o + participio), así como *darse cuenta de* o *dar a luz* (entre otras muchas combinaciones); y la variación sintáctica de *dar* es también interlingüística, pues aunque exista un verbo frecuente equivalente en otra lengua –como *to give* en el caso del inglés– es improbable que se dé una correspondencia completa entre las unidades multipalabra en ambas lenguas, entre las unidades fraseológicas e incluso entre las construcciones verbo (V) + N⁸. Y es que la construcción V + N, que es –de todas las estructuras en las que puede aparecer el verbo *dar*– la más frecuente (presente en los usos ditransitivos, transitivos, pasivos, impersonales e intransitivos), y, aparentemente, la más sencilla por su estructura formal y por su equivalencia formal en la L1, puede resultar compleja desde una perspectiva semántica. El contenido intensional de este verbo en esta construcción (V + N) es tan bajo que su capacidad extensional es máxima, esto es, puede coaparecer con una gran cantidad de sustantivos (véase el capítulo 4 para una reflexión sobre la naturaleza léxico-semántica y la proyección sintáctica del verbo *dar*). En relación con ello, las restricciones que ejerce este verbo sobre sus argumentos son, muy a menudo, intralingüísticas, por lo que están basadas en la selección léxica del verbo en función de su estructura interna y de los rasgos subléxicos contenidos en las palabras con las que se combina y con los que concuerda; y estos rasgos fácilmente difieren de una lengua a otra. Por ello, aunque en español tenemos *dar un libro a alguien* y en inglés *to give a book to somebody* (la restricción extralingüística del verbo –basada en nociones físicas– es la misma⁹), y aunque también tenemos *dar cariño a alguien* y *to give love to somebody* (ejemplos de restricción intralingüística compartida), la combinatoria de ambos verbos no es, en absoluto, equivalente en muchos casos.

⁸ Al final del capítulo 4 se incluyen tres tablas que reflejan de manera esquematizada las diferentes estructuras en las que participa *dar*: construcción V + N (tabla 2); construcciones preposicionales, pronominales, con participio y con *como* (tabla 3), y la tabla con construcciones (semi)fraseológicas (tabla 4).

⁹ Para una explicación más detallada de los conceptos *restricción intralingüística* y *extralingüística*, véase el apartado 3.2.

El verbo *dar* y el verbo *to give* difieren de manera importante en las construcciones con verbo de apoyo (CVA), como en *dar un salto* –cuya característica semántica distintiva es el significado abstracto del verbo, próximo a “hacer” (Bosque, 2001b: 23)–, por lo que estas estructuras pueden ser un foco de desviaciones para los HNN. Trabajos como el de Wanner *et al.* (2013b) refuerzan esta hipótesis en tanto en cuanto sus conclusiones sobre las colocaciones muestran que los HNN cometen más errores en las CVA, que usan relativamente a menudo, que en otras combinaciones V + N.

Las CVA son opacas –comparadas con las construcciones en las que *dar* actualiza su significado pleno (CVP), próximo a “proporcionar”, como en *dar cariño*–, ya que el vínculo entre el significado figurado del verbo de apoyo –“hacer”– y el significado básico o físico a partir del cual se proyecta –“entregar una entidad material”, como en *dar un lápiz*– es difícil de rastrear en ellas. Así, los sentidos figurados se proyectan a partir del significado básico, pero este no está tan visible en *dar un salto* o *un paseo* (CVA) como en *dar cariño* o *una opinión* (CVP).

Consideramos que tras el significado básico y los significados figurados (ya sea el que actualiza el verbo en las CVA o en las CVP) se oculta un mismo significado *sistémico* (al que se refería Rhul [1989] con el término *inherente*) a partir del cual todos los usos del verbo se proyectan (básico y figurados), que en el caso de *dar* sería: (*A*) causa que *B* pase de no existir a existir (en *C*, de manera que los seres *C* disfruten o padezcan *B*); pudiendo ser *B un lápiz, un salto* o *una opinión*. Así pues, todos los significados del verbo están contemplados en la definición sistémica; no obstante, el significado que se percibe en las CVA es fundamentalmente el aspectual, dado que la gramática en estas construcciones prescinde de la idiosincrasia léxica del verbo –que en el caso de *dar* es la transferencia– y selecciona fundamentalmente los rasgos aspectuales del significado del verbo –próximo a “hacer, ejecutar o producir”– (Bosque 2001b). Esto provoca ese distanciamiento entre el significado figurado del verbo de apoyo y el básico, físico, a partir del cual se proyecta; lo que revierte en la dificultad a la hora de descodificar y, principalmente, codificar estos usos lingüísticos. Asimismo, al ser compartido dicho significado aspectual por otros verbos de apoyo, esto contribuye, de nuevo, a esa percepción de arbitrariedad y a que no haya necesariamente una coincidencia en las gramáticas de las diferentes lenguas a la hora de seleccionar un verbo de apoyo para un sustantivo eventivo, como son *salto* o *paseo*¹⁰.

Todo esto hace que posiblemente la L1 sea una fuente de interferencias en estas CVA –mientras que en español *damos un paseo* o *un paso*, en inglés se *toman* (*take a walk* y

¹⁰ En el capítulo 4 se explica detenidamente la relación entre el sentido básico o físico del verbo, los sentidos figurados así como la relación existente entre ambos tipos de significado y el significado sistémico, a partir del cual los anteriores se proyectan.

take a step)– o que el verbo de apoyo sea elegido al azar de entre varios verbos frecuentes (Chi Man-lai *et al.*, 1994), dado que, a la hora de producir un mensaje, el HNN no interpreta con la misma facilidad la extensión natural del significado básico al significado figurado en las CVP (*dar cariño* o *una opinión*) que en las CVA (*dar un salto* o *paseo*).

En suma, pese a las semejanzas semánticas entre las lenguas, los verbos frecuentes han desarrollado en cada lengua muchas especificidades; así lo sintetizan Altenberg y Granger (2001: 174): “Despite fundamental semantic similarities across languages, high-frequency verbs have developed many language-specific differences, which make them treacherous for foreign language learners”.

Ténganse en cuenta, por tanto, estas diferencias interlingüísticas y complejidades intralingüísticas –sintácticas y semánticas– que, a nuestro juicio, pueden dificultar la adquisición del verbo *dar*. Dado que en el panorama de ELE no existen estudios que atiendan este aspecto de la interlengua –no contamos hasta la fecha con trabajos empíricos que describan el uso de estos verbos por parte de los HNN en diferentes niveles de su interlengua–, y considerando que estos verbos están muy extendidos, lo que hace que los HNN tengan que adquirirlos obligatoriamente si desean comunicarse eficazmente y con precisión, un análisis como el que aquí se realiza se hace necesario.

Consideramos que las hipótesis formuladas acerca de los posibles focos de desviaciones en el uso de este verbo constituyen un certero punto de partida para este análisis, motivado no solo por nuestra experiencia en el aula y la falta de trabajos que arrojen algo de luz sobre este aspecto de la interlengua del español, sino también por las discrepancias observadas en la bibliografía revisada en torno al uso de los verbos frecuentes en inglés por parte de grupos de aprendientes de diferente L1. Hay estudios que se refieren a la sobreutilización de estos verbos (Hasselgren, 1994; Granger, 1996b; Ringbom, 1998b; Källvist, 1999; Guo, 2006), mientras que otros hablan de infrautilización, como Altenberg y Granger (2001) o Sinclair (1991: 79): “many learners avoid the common verbs as much as possible, and especially where they make up idiomatic phrases. Instead of using them, they rely on larger, rarer, and clumsier words which make their language sound stilted and awkward”.

Ahora bien, para identificar qué aspectos de la interlengua del español, en lo referido al verbo *dar*, son producidos en consonancia con los usos nativos, cuáles se distancian un tanto y cuáles son los que se desvían claramente del español nativo, no nos basta con conocer las semejanzas y diferencias en cuanto a las frecuencias de uso, sino que también se requiere un análisis detallado de la actuación, que va a permitir, además, determinar la adecuación de cada uso de acuerdo con el concepto de norma prefijado (a esta cuestión le dedicamos el capítulo 5). Y a este respecto tampoco parece haber acuerdo entre los investigadores de la interlengua del inglés. Mientras que algunos

establecen que estamos ante verbos “seguros” para el HNN –como señala Hasselgren (1994: 250) en relación con su propuesta *teddy bear principle*, según la cual “core words, learnt early, widely usable, and above all safe (because they do not show up as errors) are hugely overused, even among learners sufficiently advanced to have been weaned off them”–, otros consideran que son una fuente de errores (Chi *et al.* 1994; Howarth, 1996; Källvist, 1999; Granger, 2001; Yan, 2006; Guo, 2006). Esta falta de acuerdo entre los investigadores de la interlengua del inglés acerca de si estos verbos se sobreutilizan o infrautilizan, o si son seguros o una fuente de errores, unido, a su vez, a la escasez de trabajos que se centran en este aspecto de la interlengua del español, refuerzan, por lo tanto, la necesidad del presente análisis.

Teniendo en cuenta estas consideraciones, esta investigación arranca de las siguientes preguntas, cuyas respuestas permitirán confirmar o descartar las hipótesis anteriores:

1. ¿Cuáles son las semejanzas y diferencias más destacadas entre el español de los HNN y el español de los HN con respecto al uso del verbo *dar*?

1.1. ¿Tienden los HNN a utilizar este lema más o menos que los HN?

1.2. ¿Qué estructuras sintácticas utiliza cada grupo, y cuál es la relación del índice de frecuencias?

1.3. ¿Con qué palabras o tipos de palabras combinan los HN y los HNN el verbo *dar*?

2. ¿Es un verbo *seguro* (su uso no muestra errores perceptibles) para el HNN o, por el contrario, es una fuente de problemas fruto de las diferencias interlingüísticas y complejidades intralingüísticas?

3. ¿Qué papel desempeña la lengua materna (L1) en el uso de este verbo? ¿y qué presencia tiene el factor intralingüístico en la adquisición de este verbo?

Esta investigación pertenece al campo de estudios de la ASL; en concreto, forma parte de la rama de estudios de la lengua del aprendiente. Se aplica para el análisis un enfoque basado en corpus a través de la metodología específica del ACI, por medio de la cual se estudia el español escrito de un grupo concreto de aprendientes de nivel avanzado de L1 inglés –el español de los HNN del CEDEL2– y se compara con el español nativo de un corpus de control, también integrado en el corpus CEDEL2; esta comparación permite descubrir las semejanzas y las diferencias entre los usos nativos y los no nativos. Dado que no podemos asumir que el español de los HN puede erigirse en la única norma para los HNN, es necesaria la aplicación de una norma reguladora que precise el grado de adecuación de los usos no nativos (el capítulo 5 constituye una discusión sobre dicha norma). Para un corpus como el CEDEL2, en el que tanto los HN como los HNN manejan referentes diatópicos distintos, la unión de los estándares peninsular y americano en un estándar panhispánico constituye un referente lingüístico

operativo para categorizar los datos de la interlengua. En los casos en los que se alberguen dudas sobre la estandarización de algún elemento y se sienta la necesidad de contrastar las intuiciones, se recurre a herramientas como un corpus más amplio (CREA), diccionarios apropiados (*Redes, Diccionario panhispánico de dudas [DPD]*, etcétera) o la *Nueva gramática de la lengua española (NGLE)* (el capítulo 6 se centra en los procedimientos seguidos para clasificar los usos).

El *software* utilizado para la recuperación de los usos de *dar* es Antconc3.2.3m (a las herramientas utilizadas para el análisis nos referimos en el capítulo 6) y el método para analizar los datos obtenidos a partir de las concordancias es cuantitativo y cualitativo. A este respecto, Nesselhauf (2004: 136) señala lo siguiente:

Many studies are exclusively or primarily quantitative...While such studies can be interesting starting points for further quantitative analyses, they do not usually in themselves contribute much to language learner analysis, let alone to language teaching. If progress is to be made, it is imperative that this current stage is left behind and that more qualitative analyses are carried out.

En relación con esta distinción, Bosque (2005: CLVIII-CLIX [2004]) diferencia la Lingüística de Corpus de la Lingüística con Corpus. La primera se caracteriza por reducir las distinciones teóricas, evitar las consideraciones que pueden parecer conjeturas y especulaciones y confiar ciegamente en los cálculos estadísticos que se realizan sobre los corpus textuales; a la Lingüística de Corpus interpretada de este modo le corresponde un enfoque de análisis de los datos cuantitativo. Por el contrario, en la Lingüística con Corpus –que se corresponde con la investigación cualitativa de corpus– no se reduce el aporte teórico; el investigador se sirve del corpus como fuente de datos, de manera que “los datos encontrados se filtran por su introspección, se evalúan y se completan con otros que el corpus no proporciona, pero que la introspección considera naturales” (Bosque, 2005: CLIX [2004]). Esta introspección, en nuestro caso, está motivada por estudios de otras áreas de ASL, como la psicolingüística, y por nuestra experiencia docente en el ámbito de ELE.

Como bien observa Guo (2006: 5), solo por medio de un enfoque cuantitativo y cualitativo de los datos es posible aprovechar las ventajas de la tecnología de la informática actual y la práctica y las teorías de la LC y de otras áreas, como la de la enseñanza de ELE (una discusión más profunda sobre el enfoque cuantitativo y el cualitativo se encuentra en el apartado 2.2.2). Por ello, consideramos que esta investigación constituye una muestra de la mencionada Lingüística con Corpus; las reflexiones teóricas surgen de los resultados estadísticos obtenidos de una investigación cuantitativa (como se desprende del análisis realizado en el capítulo 7).

En cuanto al marco teórico de esta investigación, conviene añadir que se entiende el lenguaje como un prisma; cada corriente lingüística se ha aproximado al estudio del léxico de manera diferente, con enfoques distintos, adoptados en función del núcleo de interés y de las finalidades perseguidas. Como se concibe que hay diferentes maneras de aproximarse al lenguaje, en general, y al léxico, en particular, resultaría sesgado tomar solo una perspectiva de análisis del léxico, pues eliminaría otras que serían muy enriquecedoras. Esto explica la línea de trabajo que hemos seguido en el tratamiento del verbo *dar*, en el que optamos por combinar varias perspectivas de análisis. De este modo, nos servimos de ciertos presupuestos de la semántica cognitiva (véanse los capítulos 3 y 4), concretamente de los modelos cognitivos de Lakoff y Johnson (1999), porque consideramos que ofrecen una explicación plausible sobre la variedad de significados del verbo, esto es, sobre la relación entre el significado básico o físico de *dar* y los usos figurados del verbo a partir de ciertas proyecciones metafóricas que provienen de un dominio de origen físico; también recurrimos a ciertos planteamientos semánticos (véanse capítulos los 4 y 6) que proponen la existencia de una definición sistémica del verbo (Ruhl, 1989) que explicaría que los nuevos significados que surgen de la combinación de *dar* (tanto los físicos como los figurados) están contenidos como posibilidad en esa definición de la palabra en el léxico. De esta manera, la combinatoria del verbo –que pone de manifiesto la relación natural que existe entre el significado físico y los sentidos figurados– se regiría a través de mecanismos de concordancia de rasgos léxicos (Bosque, 2005 [2004]; véase el capítulo 4). Por otro lado, los principios generativos del lenguaje fundamentan la propuesta de Baralo (2001) de organización y funcionamiento del lexicón mental no nativo, que da cuenta de los diferentes elementos que se activan en el procesamiento productivo de la L2; de su capacidad descriptiva nos servimos para configurar nuestra tipología del error léxico, originada a partir del marco del AE propuesto por Corder (1971).

1.5. ESTRUCTURA DE LA TESIS

La exploración que realizamos del grado de conocimiento del verbo *dar* por parte de un grupo concreto de HNN tiene como fin contribuir a la mejora de la descripción de la competencia léxica de este grupo en un estadio de interlengua específico. A continuación describimos brevemente los capítulos que estructuran esta investigación.

El primer capítulo (“Introducción”) presenta el tema y el objetivo de la investigación: un estudio contrastivo de interlengua basado en corpus (sobre el verbo *dar*). Esta metodología (ACI) se presenta como una evolución a partir de otros enfoques y métodos de análisis de la lengua del aprendiente.

Partiendo de estas consideraciones, el capítulo segundo (“Estado de la cuestión sobre los estudios de la lengua del aprendiente”) recoge los avances y cambios que ha habido en la investigación de la lengua del aprendiente para centrarse en el uso de corpus; en este capítulo se revisan los corpus de HNN más destacados y se subrayan algunos de los aspectos metodológicos más característicos. Tras revisarse los escasos estudios del ACI que versan sobre un aspecto del léxico del español, la elección de nuestro tema queda, así, justificada.

En el tercer capítulo (“Revalorización del léxico y competencia léxica del verbo”) se recoge una propuesta formalizada de los aspectos que integran la competencia léxica del verbo. Para acometer esta tarea, se parte de una sucinta revisión historiográfica sobre el papel del léxico en relación con la lengua, fundamentalmente durante las últimas cinco décadas –etapa decisiva en la revalorización de este componente–: desde una perspectiva más bien teórica, desde otra “adquisicionista” y, por último, desde una perspectiva pedagógica.

En el cuarto capítulo (“El verbo *dar*: la naturaleza léxico-semántica y proyección sintáctica”), con la intención de desarrollar herramientas eficaces para la categorización de los datos analizados del corpus y para facilitar la explicación de los datos, se propone una descripción de los diferentes significados que actualiza el verbo *dar* y de las diferentes estructuras sintácticas en las que participa.

El capítulo quinto (“La cuestión de la norma en ELE”) plantea la problemática cuestión (Altenberg, 2011: xv) de la selección de la norma en los estudios de interlengua y delimita el estándar de referencia para el análisis: se propone la aplicación de un concepto de norma que integra un enfoque descriptivo y otro más regulador (prescriptivo).

En el capítulo sexto (“Metodología”) se diseña la metodología de la investigación; tras referirnos a las herramientas empleadas en la investigación y describir en profundidad los corpus comparados y analizados en este trabajo, se recuperan algunas cuestiones metodológicas avanzadas o esbozadas en los capítulos anteriores y se presentan los procedimientos seguidos en el análisis.

Partiendo de estas consideraciones, en el capítulo séptimo (“El uso del verbo *dar*. Análisis de corpus contrastivo”) se analiza la producción de *dar* en el corpus nativo y en el no nativo y se comparan desde un enfoque cualitativo y cuantitativo.

En el capítulo octavo (“Discusión”) se discuten los resultados del análisis, las limitaciones y se apuntan varios estudios que sería interesante realizar para completar el presente.

Por último, en el capítulo noveno (“Conclusión”) nos referimos a las aportaciones de esta investigación en los estudios de corpus de aprendientes de ELE y se recogen las principales aplicaciones del análisis; de esta manera, se proponen posibles

aplicaciones didácticas que pueden tenerse en cuenta en los materiales de ELE y en el aula de ELE, partiendo de la asunción –comúnmente aceptada en el campo de estudio de la lengua del aprendiente– de que el español nativo representa el estado de lengua ideal que los HNN aspiran alcanzar. Por último, se indican las perspectivas de futuro próximo de los estudios de la lengua del aprendiente.

2. ESTADO DE LA CUESTIÓN SOBRE LOS ESTUDIOS DE LA LENGUA DEL APRENDIENTE

Aprender una lengua extranjera es un proceso lento y costoso, que en raras ocasiones culmina en un dominio completo de la lengua meta. Los problemas pueden estar relacionados con la interferencia de la L1, con los rasgos de la LE o con el propio proceso de aprendizaje. Revelar los rasgos de la interlengua se ha convertido en un medio importante de medir las diferencias entre la interlengua y la actuación del HN, lo que puede llevar potencialmente a una mejora en la enseñanza de la lengua y en el conocimiento del proceso de aprendizaje de la lengua (Hasselgard y Johansson, 2011: 33).

La investigación basada en corpus de aprendientes es una rama de estudios sobre la lengua del aprendiente relativamente nueva, cuyo desarrollo supuso un paso significativo en los estudios de interlengua. En este ámbito específico, *corpus* tiene un significado restringido: “a collection of pieces of language text in electronic form, selected according to external criteria to represent, as far as possible, a language or language variety as a source of data for linguistic research” (Sinclair, 2005: 16).

Esta rama de estudios nace en la década de los noventa de la mano de Sylviane Granger, quien –tras fundar en 1990 el Centro de Lingüística de Corpus del Inglés (CLCI¹¹) de la Universidad Católica de Lovaina– inicia un exitoso proyecto de compilación del primer corpus internacional de aprendientes de inglés denominado International Corpus of Learner English (ICLE, Granger *et al.*, 2002; 2009 [segunda versión]), que ha inspirado trabajos similares en muchos otros países, como es el caso del CEDEL2 (Lozano, 2009a; Lozano y Mendikoetxea, 2013) en España, el corpus de aprendientes de español utilizado en esta investigación (como se señalaba en el capítulo anterior).

Estos corpus surgen más de treinta años después de que comenzaran a compilarse los primeros corpus de HN. El objetivo principal de los lingüistas y los lexicógrafos era entonces construir una representación de la producción escrita y hablada de los HN de una lengua, lo que supuso el inicio del desarrollo de los corpus informatizados. En rigor, puede decirse que la revolución de los corpus en lingüística – *revolución* en el sentido de que ha originado toda una relación de nuevas preguntas de investigación– comenzó con la finalización y la distribución del Brown Corpus, en 1964, publicado por Kučera y Francis, de la Universidad de Brown, con el título *A Standard Sample of Present-Day Edited American English, for Use with Digital Computers*

¹¹ Para una información completa sobre el CLCI, véase <www.uclouvain.be/encecl.html> (10-06-14).

(Leech, 2011: 10). Poco después, Kučera y Francis (1967) lo usaron para crear la primera lista de frecuencias del inglés basada en datos de un corpus; en 1982 publicaron listas de frecuencias lematizadas, basadas en la versión del corpus anotada con las categorías de palabras (Leech, 2011: 10). En ese mismo año, los lingüistas escandinavos Hofland y Johansson compararon el trabajo de Brown con el que ellos acababan de concluir, el Corpus of British English Lancaster-Oslo/Bergen (LOB), y publicaron el trabajo titulado *Word Frequencies in British and American English*. Por primera vez, listas de palabras frecuentes –del inglés americano y el británico– con información gramatical derivadas automáticamente de un corpus informatizado estaban disponibles para el investigador y para el profesor de lenguas. Pese a su innovación, estas técnicas ya habían sido empleadas en un trabajo más revolucionario –*The OSTI Report*– que implicaba un cambio en la concepción de la lengua, sobre todo en la relación entre la gramática y el léxico. Fue dirigido por Sinclair desde la Universidad de Birmingham y concluido en 1970 pero no llegó a ver la luz hasta el año 2004¹². Este trabajo usaba un pequeño corpus de inglés hablado de 135000 palabras, que en los años ochenta y noventa fue ampliado bajo la dirección de este mismo lingüista en lo que se denominó *Birmingham Collection of English Texts*, conocido después como el *Bank of English* (1997).

Un análisis como el nuestro, que parte de la interrelación entre los niveles gramatical y léxico –centrado precisamente en la exploración de la interfaz léxico-sintaxis para dar cuenta de la estructura de la lengua–, que analiza el vínculo entre la combinatoria léxica de una palabra y la estructura de su significado, es heredero, en cierta medida, de estos trabajos de Sinclair (1966, 1970) –surgidos a raíz de las ideas desarrolladas por Firth (1957)– y de Halliday (1961, 1966), su colega entonces en Edimburgo, quien declaró (1966: 273-277) que el nivel del léxico (incluyendo la colocación) merecía constituir un nivel de descripción lingüística diferenciado de los demás, interrelacionado en un continuo con el nivel gramatical; los corpus daban muestra de ello¹³. En este enfoque, las palabras y la gramática están ineluctablemente unidas en lo que se conoce como *lexicogrammar* (Halliday, 1985; Sinclair, 1991) o *lexical grammar*, lo que supuso un impacto en el campo de la lexicografía: “Corpus data forced lexicographers to shift away from a focus upon regularities which were easily attested by a few examples towards what Hanks (2009: 216) describes as ‘idiosyncratic conventions that are associated with each word’” (McEnery y Hardie, 2012: 80). En este sentido, puede decirse que

¹² Sobre los orígenes, avances y progresos en la descripción de la lengua basada en los análisis de corpus, véase Sinclair *et al.* (2005).

¹³ Las principales contribuciones sistemáticas de la LC a la descripción del léxico y la gramática se hicieron dentro de la LC del inglés (McEnery y Hardie, 2012: 72), donde se desarrollaron y refinaron conceptos como los de colocación y anotación de corpus.

the corpus revolution has introduced a new theoretical perspective on linguistic structuring: one in bold contrast to the mainstream paradigm of Chomsky (e.g. Chomsky 1965: 84-88) whereby grammar and lexicon are two clearly distinct components. It also challenges a tradition long established in language study, whereby grammars and dictionaries provide distinct kinds of information about a language, and are published in separate covers (Leech, 2011: 12).

En relación con esto, han sido publicadas recientemente obras lexicográficas que contienen información de cierta naturaleza lingüística que no aparece en los diccionarios habituales. Nos referimos a los diccionarios combinatorios *Redes* (2004) y *Práctico* (2006) –basados en un corpus de 250 millones de palabras–, que supusieron una innovación en el terreno lexicográfico al proporcionar la combinatoria de cada palabra basada en la restricción semántica que los predicados ejercen sobre sus argumentos, al tiempo que incluyen información sobre la frecuencia de uso de las combinaciones. Estos diccionarios de restricciones léxicas no tienen equivalentes directos en ningún diccionario del español o de cualquier otro idioma (Bosque, 2005: LXXXVI [2004]); y unidos a los diccionarios de colocaciones –basados también en análisis de corpus– ponen de manifiesto la estrecha relación que existe entre el trabajo lexicográfico, el lexicológico y el gramatical, como también puede verse en el recientemente publicado diccionario de frecuencias del inglés americano (Davies y Gardner, 2010), basado en un corpus de 385 millones palabras, que presenta, junto a la frecuencia individual de cada palabra, una lista de colocaciones comunes para cada palabra (estas cifras muestran el aumento del tamaño de los corpus utilizados desde las primeras compilaciones de textos hasta la actualidad).

En suma, se puede decir que la LC se ha convertido en los últimos veinte años en un componente indispensable del aparato metodológico de la lingüística, en general. Esto es, en rigor, se trata más bien de una metodología que de una teoría del lenguaje; una metodología que ha facilitado nuevos accesos para el estudio de la lengua –su naturaleza y la manera en la que la procesamos– y nuevas maneras de unir teoría y datos (Altenberg, 2011: XIII). Así lo señalaba Lewis (2000: 126) “In recent years, since the widespread availability of large computer-based corpora –collections of natural written and spoken text– which have been statistically analysed, we have better descriptions of English available to us than ever before”.

Así pues, desde hace ya varios años la LC y otras ramas de la lingüística tienden a converger (McEneary y Hardy, 2012: 226); es decir, la metodología de corpus forma parte de la práctica diaria de lingüistas funcionalistas, sociolingüistas, analistas del discurso o de estudiosos de la lengua del aprendiente, entre otros. En relación con ello,

Granger (1998a: XXI) resume de la siguiente manera esta última rama de estudios centrada en la lengua del aprendiente y su análisis a través de los corpus: “With roots both in corpus linguistics and second language acquisition studies (SLA), it uses methods and tools of corpus linguistics to gain better insights into authentic learner language”. A los orígenes de los estudios de corpus de aprendientes (informatizados) (CAI) y la LC nos acabamos de referir brevemente; volveremos sobre ello con mayor profundidad en el apartado 2.2. En cuanto a la relación de los estudios de CAI con la ASL, dado que han sido varios los procedimientos en la investigación del proceso de adquisición anteriores a los estudios de CAI de los que este se nutre –como el Análisis Contrastivo (AC), el Análisis de Errores (AE) y el Análisis de la Interlengua o de la Actuación (AI)–, se incluye a continuación una revisión de estos procedimientos. Después, se señala la relación entre estos y la ASL, y se plantea la conexión entre esta disciplina y la investigación de CAI (2.1). La segunda parte de este capítulo (2.2) se refiere a la investigación de CAI y se presentan algunos de los corpus de aprendientes más destacados. El capítulo finaliza con algunas cuestiones fundamentales de esta metodología.

2.1. ESTUDIOS SOBRE LA LENGUA DEL APRENDIENTE ANTES DE LOS CORPUS INFORMATIZADOS

El desarrollo progresivo de los diferentes modelos de análisis de datos aplicados a la investigación en el ámbito de la ASL suele explicarse de un modo similar (Larsen-Freeman y Long, 1991; Santos Gargallo, 1993; Baralo, 2004 [1999]; Pastor Cesteros, 2004; Guo, 2006; De Alba, 2009). Lo sintetizamos a continuación.

2.1.1. *El Análisis Contrastivo*

La lengua del aprendiente fue investigada antes de la aparición de los corpus de aprendientes; se pueden datar los orígenes del análisis de este objeto a finales de los años sesenta y en los años setenta, con el AE. No obstante, antes del periodo del AE, entre los años 40 y 60, surgen los primeros estudios de adquisición –a medio camino entre los estudios lingüísticos y psicolingüísticos– en un momento en el que se está produciendo el debate sobre la teoría del aprendizaje lingüístico y el consiguiente enfrentamiento entre conductistas y cognitivistas (Pastor Cesteros, 2004: 99). Estos estudios orientan por primera vez el proceso de enseñanza-aprendizaje desde la perspectiva de los aprendientes; por ello, se considera el AC como el primer modelo de

análisis centrado en la adquisición de la L2, aunque el acercamiento científico fue prácticamente teórico en su totalidad, pues apenas produjo resultados prácticos concretos (De Alba, 2009). En plena vigencia de la teoría lingüística estructuralista y del modelo de aprendizaje conductista, se consideraba que el error es señal de no adquisición –por lo que debía ser penalizado– y de interferencia de los hábitos de la L1. Desde que los investigadores del AC establecieron una relación entre los errores de los HNN y la diferencia entre la L1 y la L2, trataron de localizar la fuente del error por medio de la comparación entre las dos lenguas, “in the comparison between native and foreign language lies the key to ease or difficulty in foreign language teaching” (Lado, 1957: 1); véase como muestra del AC aplicado al aprendizaje de ELE, *The grammatical structures of English and Spanish*, de Stockwell *et al.* (1965), una de las obras más representativas de este modelo de análisis. Por consiguiente, el objetivo de la comparación era identificar los rasgos fáciles y difíciles de la L2; la L1 podía ayudar a los HNN o interferir negativamente durante la producción de las estructuras gramaticales y léxicas. De esta manera, se presuponía que había una relación directamente proporcional entre la distancia lingüística y la dificultad del aprendizaje: a más diferencia mayor dificultad (Pastor Cesteros, 2004: 101).

Este método de análisis conllevó una serie de problemas o carencias que los profesores de lenguas detectaron pronto; así lo señala Corder (1967: 161):

Teachers have not always been very impressed by [the contributions of the Contrastive Analysis Research] for the reason that their practical experience has usually already shown them where these difficulties lie and they have not felt that the contribution of [this research] has provided them with any significantly new information.

Los profesores percibían que dos lenguas más distantes no generaban más dificultades que dos cercanas, como el AC defendía. Además, no se demostró que por medio de este método se pudieran predecir todos los errores producidos en el aula; asimismo, no se logró probar que todos los errores que se predecían se materializaban.

No obstante, conviene reconocer los logros del AC: inicia la investigación centrada en el aprendiente y su proceso de aprendizaje, y supone el punto de partida de los actuales estudios de ASL (Pastor Cesteros, 2004: 103); de hecho, su desarrollo se vincula a la comprobación o negación de las propuestas del AC. Por todo ello, este tipo de análisis debe aparecer en cualquier recorrido sobre la evolución en la metodología de investigación del fenómeno de la adquisición y aprendizaje de las L2, como también argumenta De Alba (2009).

Con respecto al desarrollo del AC en relación con el aprendizaje del español, la *Bibliografía de Lingüística general y española (1964-1990)* (Báez, 1995), que abarca un periodo

extenso de tiempo, muestra la existencia de un conjunto amplio de trabajos en los que se analizan temas de fonología, morfología y sintaxis del español y el inglés, o el español y el francés. Estos trabajos representan lo que se conoce como la *versión débil* del AC, que destaca el carácter explicativo –pero no predictivo– de la interferencia lingüística para entender las dificultades en el aprendizaje. En Penadés (1999: 9) se hace referencia también a la existencia de algunos trabajos, menos numerosos que los anteriores, que contrastan el español con otras lenguas, como el italiano, el rumano, el alemán, el portugués, el checo, el ruso, el japonés o el sueco. Asimismo, los números 51 y 52 de la revista *Carabela*, del año 2002, que están dedicados a la lingüística contrastiva, presentan trabajos muy completos con este enfoque. Así, aunque no han dejado de realizarse estudios contrastivos –de hecho, a finales del siglo XX se hablaba de una revalorización de la lingüística contrastiva (Fernández González, 1995: 14-19)–, las críticas a este modelo de análisis conllevaron la aparición de un nuevo modelo: el Análisis de Errores, que supuso un gran avance en los estudios de adquisición al tener como objeto de estudio la lengua del aprendiente, y no la L1 ni la L2, como observa Guo (2006: 2-3) “it was a significant advance when EA [error analysis] researchers to have placed the learner language (rather than L1 and L2) under examination”.

2.1.2. *El Análisis de Errores*

A finales de los años sesenta y durante los años setenta, la investigación de la lengua del aprendiente, conocida entonces como el AE, resultó en una actividad muy popular. Este método de análisis se erigió como el nuevo modelo de investigación de la adquisición de la LE, tras constatarse empíricamente la incapacidad del AC para alcanzar los objetivos propuestos por Lado (1957) y tras la caída de sus bases teóricas –representadas lingüísticamente por el estructuralismo y psicolingüísticamente por el conductismo–, provocada por la irrupción de los planteamientos teóricos de la lingüística chomskiana y de las teorías cognitivas del aprendizaje. Como se ha referido anteriormente, lo que ahora comienza a interesar a los estudiosos son las producciones concretas de los HNN:

el cambio metodológico es crucial: se pasa de las predicciones desde el plano de la abstracción —en el que hasta ahora se habían desarrollado las investigaciones en L2— al espacio real y concreto de las producciones de los discentes, con el objeto de obtener datos empíricos que favorezcan la explicación de los errores en el proceso de adquisición y aprendizaje de las lenguas extranjeras (De Alba, 2009: s/p).

Desde el AE se concibe la adquisición de una segunda lengua como un proceso cognitivo, interiorizado de formación de reglas de la L2 –lo que supone un traslado de la hipótesis innatista al respecto de la adquisición de la L1 al ámbito de la adquisición de segundas lenguas–, en el que se establece una línea de continuidad entre la L1 y la L2. El sistema lingüístico que se desarrolla a través de ese proceso se conoce como interlengua, concepto al que ya nos hemos referido (véase nota [n.] 1), y que nace, precisamente, en el seno del AE (Pastor Cesteros, 2004: 105). Corder (1967) consideraba la lengua del aprendiente una especie de *dialecto idiosincrásico*, con sus propios rasgos, diferentes de la L1 y de la L2, en el que el error constituye un aspecto fundamental, pues comporta un gran valor para el estudio del proceso de aprendizaje e informa del nivel de adquisición; esto es, entre los investigadores del AE existe la opinión compartida de que los errores de los HNN, lejos de ser considerados negativos, son inevitables y necesarios en el desarrollo del aprendizaje, pues constituyen evidencias positivas de que el aprendizaje está teniendo lugar:

A learner's errors, then, provide evidence of the system of the language that he is using (i.e. has learned) at a particular point in the course (and it must be repeated that he is using some system, although it is not yet the right system). They are significant in three ways. First to the teacher, in that they tell him, if he undertakes a systematic analysis, how far towards the goal the learner has progressed and consequently, what remains for him to learn. Second, they provide to the researcher evidence of how language is learned or acquired, what strategies or procedures the learner is employing in his discovery of the language. Thirdly (and in a sense this is their most important aspect) they are indispensable to the learner himself, because we can regard the making of errors as a device the learner uses in order to learn. It is a way the learner has of testing his hypothesis about the nature of the language he is learning (Corder, 1967: 169).

De esta manera, el error se convierte en objeto de estudio y es utilizado con fines didácticos.

El tratamiento del error planteado por Corder (1971) se puede resumir en cuatro etapas (Ellis, 1994: 68-69), a saber: la recogida de errores a partir de la recopilación de muestras de lengua de aprendientes, su identificación, su descripción y su explicación:

The first step in carrying out an EA was to collect a massive, specific, or incidental sample of learner language. The sample could consist of natural language use or be elicited either clinically or experimentally. It could also be collected cross-sectionally or longitudinally. The second stage involved identifying the errors in the sample. Corder distinguished errors of competence from mistakes in performance and argued that EA should investigate only errors. [...] The third stage consisted of description. Two types of

descriptive taxonomies have been used: linguistic and surface strategy. The former provides an indication of the number and proportion of errors in either different levels of language (i.e. lexis, morphology, and syntax) or in specific grammatical categories (for example, articles, propositions, or word order). The latter classifies errors according to whether they involve omission, additions, misinformations, or misordering. The fourth stage involves an attempt to explain the errors psycholinguistically.

Así pues, en cuanto a la primera etapa –la recopilación de muestras–, el análisis del error llevado a cabo por los investigadores del AE puede estar basado en datos procedentes de test de elicitación (en Mairal Usón *et al.*, 2010: 55, se enumeran algunos de estos procedimientos) y datos de corpus preinformático¹⁴; puede centrarse en un solo individuo (estudios longitudinales, que involucran a HNN que son seguidos durante un periodo de tiempo) o en un grupo representativo de una población (estudios longitudinales o transversales, estos últimos basados en distintos niveles de competencia); no obstante, la capacidad de generalización de los resultados de estos estudios está vinculada al número de representantes y al número de muestras, aunque, como bien observa de Alba (2009), no existe ningún criterio normalizador en este ámbito.

En lo que se refiere a la segunda y tercera etapa en el tratamiento del error (identificación y descripción), Corder ya señaló que no se pueden considerar errores todos los fallos que comete el HNN, sentando las bases para diferenciar error (fallo sistemático, relacionado con la competencia) y falta (fallo asistemático, relacionado con la actuación); tras la identificación del error, se suele establecer una taxonomía del error para clasificar los errores hallados, aunque esta también puede realizarse con anterioridad a la identificación o con posterioridad.

En la cuarta etapa, se analizan las causas de los errores, esto es, por qué se transgreden las normas.

Aunque sin duda este proceso de análisis aporta información muy valiosa sobre el modo en el que se produce la adquisición de la L2 (Baralo, 2004 [1999]; Guo, 2006, entre muchos otros), no son pocas las limitaciones de este modelo. Schachter y Celce-Murcia (1977: 41, citado en Guo, 2006: 9), en pleno apogeo del AE, expusieron sus reservas con respecto a este método de análisis, que evidencian las limitaciones del método a la hora de describir el proceso de adquisición de la L2. A nuestro juicio, de entre todas las limitaciones sobresalen tres por su vinculación directa con la caída del

¹⁴ El término *corpus* en este periodo se refiere a una colección de muestras de lengua natural que ha sido compilada para un estudio lingüístico (Hunston, 2002: 2). Esta definición está relacionada con la información de *corpus* obtenida del DRAE (2001): “conjunto lo más extenso y ordenado posible de datos o textos científicos, literarios, etc., que pueden servir de base a una investigación”. En la actualidad, el término implica que el corpus se almacena y que se accede a él electrónicamente.

AE y la llegada de nuevos modelos de análisis de la lengua del aprendiente: el análisis del los errores aislados, la clasificación de los errores identificados y la adscripción de las causas a los errores. En primer lugar, el error se analiza aislado; los investigadores extraían de los datos los errores de los HNN; una vez que eran identificados, los datos se descartaban, por lo que no existía la posibilidad de recuperarlos para confirmar resultados o reevaluar el contexto. En segundo lugar, está la dificultad para identificar los errores –pues no siempre es fácil determinar qué es un error y qué es una falta, ni si un determinado uso es verdaderamente una desviación de la L2 o no lo es– y la dificultad para clasificar los errores, esto es, establecer con qué tipo de estructura se corresponde un determinado error, pues, como señala Penadés (2003), existen numerosas taxonomías en relación con la clasificación de los errores, lo que ha generado mucha controversia debido a la falta de criterios normalizadores¹⁵. Por último, otro de los déficits asociados al AE es la adscripción de las causas a los errores. A diferencia del AC, en el AE las causas no se restringen a la transferencia interlingüística (Hammarberg, 1973: 29); estas pueden ser numerosas entre las interlinguales y las intralinguales, pero es una práctica común entre los investigadores de AE hacer un análisis de errores aislados dentro de un alcance muy limitado y luego etiquetarlos como intralinguales o interlinguales (Guo, 2006: 10)¹⁶. Por otro lado, pero en relación con este punto, Dulay *et al.* (1982) manifiestan que uno de los principales problemas del AE es que en los estudios se mezclan la descripción del error y la explicación, esto es, qué es lo que está incorrecto en un determinado uso y por qué. Se podría evitar esta situación si –como señala De Alba (2009)– en cada taxonomía existiera una breve explicación, por cada uno de los apartados en los que se ha dividido la clasificación, de las reglas que han sido trasgredidas antes de abordar las explicaciones de estos conflictos¹⁷.

Es evidente que el tratamiento de los errores aislados y la complejidad y la falta de sistematicidad para identificar y clasificar los errores son debilidades importantes de este modelo; pero no menos problemático es el hecho de tener como objeto de estudio únicamente los errores, lo cual no revierte en un conocimiento completo sobre la lengua producida por el HNN, como pronto reivindicaron Hammarberg y Enkvist en

¹⁵ Este aspecto es importante para esta investigación, en tanto que conduce, por un lado, a reflexionar en un capítulo (el 5) sobre la norma con la que contrastar los datos –para así poder fijar qué es un error y qué no lo es– y, por otro lado, a establecer en otro capítulo (el 6) unas pautas y unos parámetros en la categorización o catalogación de los datos, en general, y del error en particular, una vez que los errores son identificados, y no antes, para no influir en el análisis con los tipos de errores que previsiblemente nos podríamos encontrar.

¹⁶ A este respecto, y en un estudio como el nuestro, en el que no se analiza el uso del verbo en HNN de diferentes L1 con un mismo nivel de dominio, sino que todos son hablantes anglófonos, las recomendaciones de Schachter y Celce-Murcia (1977: 41, citado en Guo, 2006: 9) recobran su relevancia: “It would be wise, then, for investigators to suggest causes of error only very cautiously. What we see happening, however, is just the reverse”.

¹⁷ Esta cuestión metodológica es atendida en esta investigación.

sendos trabajos: “The insufficiency of error analysis” (1973) y “Should we count errors or measure success?” (1973). Svartvik sugiere ya en ese mismo año (1973) que la expresión *análisis del error* debería ser reemplazada por la de *análisis de actuación*, aunque finalmente la que se estableció fue la de *estudios de interlengua*, en relación con el término *interlengua*, de Selinker (1972) (Hasselgard y Johansson, 2011: 35): “Although the study of errors is a natural starting-point, the final analysis should include linguistic performance as a whole, not just deviation” (Svartvik, 1973: 8)¹⁸.

Ellis (1994: 67) se refiere a esta misma idea cuando años después señala: “A frequently mentioned limitation is that EA fails to provide a complete picture of learner language. We need to know what learners do correctly as well as what they do wrongly”. Así pues, queda constatado que el objetivo del estudio de la lengua del aprendiente no se aplica solo a los errores, sino que debe representar el nivel de dominio del HNN.

Debido a los problemas metodológicos señalados, el AE fue gradualmente absorbido por un campo de estudio de la adquisición de la L2 más general, el conocido hoy como ASL (Guo, 2006: 3, 10-11).

Pese a estas limitaciones, el AE, orientado hacia la descripción de la interlengua, ha sido el modelo de investigación de la adquisición de ELE más productivo hasta ahora, aunque, como señala Baralo (2004: 38 [1999]), la fase explicativa del análisis del error no se haya desarrollado totalmente. Algunos de los trabajos más exhaustivos han sido los de Vázquez (tesis doctoral defendida en 1987 y publicada en 1991) –centrado en los errores principalmente morfosintácticos producidos en la interlengua de estudiantes alemanes–, Fernández (tesis defendida en 1991 y parcialmente publicada en 1997) –que examina los errores en todo el sistema de la lengua en cuatro grupos de L1 diferentes, en tres estadios de evolución de su interlengua– y Santos Gargallo (tesis defendida en 1992) –que examina los errores de la producción escrita en alumnos serbocroatas–. Disponemos también de numerosos estudios menos abarcadores de análisis de errores de HNN de español, como los de Penadés *et al.* (1999)¹⁹, que incluyen memorias de maestría de alumnos de la Universidad de Alcalá. Estos se suman a otras tesinas, tesis doctorales y otros trabajos de investigación que utilizan esta metodología a finales de los años 90 y durante los años transcurridos en el siglo XXI – las actas de los últimos congresos de ASELE (véanse las de 2005, 2006 o 2007, por

¹⁸ La investigación de Linnarud, de 1986 –iniciada en el contexto del proyecto de estudios contrastivos sueco-inglés, dirigido por Svartvik (1973)–, es un análisis de la actuación léxica, no solo de los errores. Este estudio usa material textual combinado con material procedente de técnicas de elicitación.

¹⁹ Para una idea más completa sobre la situación del AC, AE y AI del español en el último cuarto del siglo XX, véase el anexo al número 43 de la revista *Carabela*, donde se incluyen las referencias bibliográficas extraídas de publicaciones periódicas entre 1983 y 1997.

poner algunos ejemplos), la base de datos Teso²⁰ de tesis doctorales producidas en España y la red de didáctica del español como lengua extranjera del Ministerio²¹ son buena prueba de ello, pues incluyen muchos trabajos de investigación centrados en el AE-. La evolución de estos estudios muestra que a partir de los análisis únicamente lingüísticos se ha ido dando paso a los análisis que integran también los elementos discursivos, aunque los primeros en ningún momento han dejado de producirse. Un análisis de errores puede centrarse en cualquiera de las cuatro destrezas, aunque las dos productivas (expresión oral y escrita) son las más estudiadas, dado que en ellas es más fácil cuantificar datos (De Alba, 2009).

El tratamiento del error que se lleva a cabo en el AE –propuesto por Corder (1967, 1971)– constituye una contribución que sigue teniendo vigencia hasta ahora en los análisis de interlengua o de la actuación, ya estén basados en corpus o no lo estén, en tanto en cuanto los errores forman parte de la lengua de los HNN, es decir, contribuyen a determinar el nivel de dominio lingüístico en el que se encuentran los HNN. Así pues, investigaciones como la que aquí se desarrolla, no pertenecen al AE pero sí que adoptan y adaptan la metodología del AE para analizar el error.

2.1.3. La Adquisición de Segundas Lenguas. Del prejuicio sobre los corpus de aprendientes a su apreciación en el análisis de la interlengua

Tras varias décadas de desarrollo desde finales de los años sesenta, la investigación de ASL “has become a rather amorphous field of study with elastic boundaries” (Ellis, 1994: 15). Larsen-Freeman y Long (1991) sostienen que el ámbito de la ASL es fundamentalmente la naturaleza del proceso de adquisición de la L2 y los factores que afectan a la lengua de los HNN²².

La principal fuente de datos para analizar este proceso de adquisición es la lengua producida por los HNN, ya sea espontáneamente o a través de los procedimientos de elicitación o de introspección (juicios de gramaticalidad, tareas de comprensión, etcétera). El éxito de la investigación en ASL depende de la validez y fiabilidad de estos medios de obtención de datos (como se ha señalado en el apartado 1.2). Precisamente,

²⁰ Para hacer una consulta de esta base de datos, véase <<https://www.educacion.gob.es/teseo/irGestionarConsulta.do>> (12-06-14).

²¹ Véase la página del Ministerio de Educación <<http://www.mecd.gob.es/redele/Biblioteca-Virtual/Presentacion.html>> (12-06-14).

²² Trabajos como los de Corder (1969), Selinker (1972), Schumann (1976) y Krashen (1977) son representativos del nacimiento de la ASL; en ellos se separa la enseñanza del aprendizaje y “se sientan las bases de lo que luego se va a consolidar como acercamientos psicolingüísticos, lingüísticos y sociolingüísticos al estudio de la adquisición” (Muñoz Licerias, 2009).

una de las principales limitaciones de la actual investigación en ASL está relacionada con la recopilación de datos. La mayoría de los estudios favorece los datos experimentales y los procedentes de la introspección, y tiende a rechazar los datos de usos lingüísticos naturales, que suelen estar representados en los corpus. Granger (2002: 5-6) explica esta preferencia en la investigación de ASL refiriéndose a la dificultad que existe en los contextos no experimentales para controlar las variables que afectan a la producción de los HNN. A esto hay que añadirle la falta de formación de los lingüistas aplicados en el uso de las metodologías informatizadas, que son las que permiten trabajar con datos naturales a gran escala, como señala Tono (2003: 806).

Como es difícil someter a una gran cantidad de informantes a la experimentación, la investigación de la ASL tiende a emplear una base empírica relativamente estrecha, con el foco puesto en la lengua de un número muy limitado de individuos, lo que provoca cuestionamientos acerca de la generalización de los resultados (Granger, 2002: 6).

Es cierto que existen razones para valorar las técnicas de elicitación, pues por medio de estos procedimientos no solo resulta más sencillo considerar las variables que afectan a la producción del HNN, sino que el investigador se asegura de que la estructura que desea investigar está presente en el material analizado, y que lo que está presente lo está porque el HNN lo conoce.

No obstante, como señala Granger, pese a estos beneficios, uno de los problemas principales es que la base empírica de la investigación en la ASL es pequeña, lo que pone en tela de juicio la generalización de los resultados, como en el estudio de Lardiere (1998), que usa los datos de un solo aprendiente de inglés recopilados durante varios años. Esto, por consiguiente, demuestra que la lengua colectiva del aprendiente no es foco de interés de la investigación actual de la ASL (Guo, 2006: 3): “When the learner’s output is considered, the focus of the research is rather more on the output of individual learners than on the output of a group of learners with the same background”. Esto resulta llamativo cuando precisamente las dificultades típicas de un grupo concreto son las que deben atenderse en la enseñanza de lenguas, y por lo tanto deben interesar en las investigaciones de la interlengua.

Por ello, Granger (1998b: 5), Guo (2006: 11) o Lozano y Mendikoetxea (2013: 1-3), entre otros muchos, reivindican el uso de los corpus de aprendientes informatizados como una fuente de datos naturales muy valiosa: “There is clearly a need for more, and better quality, data and this is particularly acute in the case of natural language data [...] learner corpora are a valuable addition to current SLA data sources” (Granger, 1998b: 5). Lozano y Mendikoetxea (2013: 1-2) argumentan con acierto el hecho de que los estudios sobre la adquisición de la L1 se han servido del uso de corpus extensos informatizados durante los últimos 25 años, como el Child

Language Data Exchange System o CHILDES²³ (MacWhinney, 2000), que ha sido la fuente de datos de más de 3200 trabajos, lo que ha supuesto un gran paso en el conocimiento sobre la manera en la que los niños adquieren y desarrollan su L1. Estos corpus en L2 son todavía escasos –aunque la situación está cambiando (véase a este respecto el apartado 2.2)–. Consecuentemente, la investigación de la L2 se ha podido beneficiar poco hasta ahora de esta fuente de datos naturales a gran escala que tiene mucho que aportar a nuestro conocimiento sobre el modo en que se desarrolla la lengua del aprendiente.

En general, la contribución de la investigación de CAI ha sido más substancial en la descripción que en la interpretación de los datos referidos a la ASL (Granger, 2004)²⁴, más centrada en un acercamiento pedagógico a la ASL, con pocas referencias a los debates, hipótesis y teorías actuales sobre la ASL y sus implicaciones en el desarrollo de la lengua del aprendiente (Myles, 2005, citado en Lozano y Mendikoetxea, 2013: 1-2) –en su lugar, los estudios de corpus, al igual que los estudios experimentales, han servido en el intento de construir hipótesis en la investigación en la ASL–²⁵.

Así, queda constatado, como se ha señalado en el comienzo de este apartado, que el campo de investigación lingüística conocido como investigación de CAI, al que pertenece este trabajo, ha surgido como resultado de la confluencia de dos campos hasta ahora discrepantes: LC y ASL.

En suma, se ha visto a lo largo de este apartado (2.1) que el AE no puede proporcionar una descripción completa de la lengua del aprendiente (recuérdese la reivindicación de Svartvik [1973: 8]), aunque sí intentó presentar el panorama de los errores de los HNN con claros propósitos pedagógicos; por otro lado, se ha puesto de manifiesto que el aspecto colectivo de la lengua del aprendiente ha recibido poca atención en la investigación actual sobre la ASL. Los corpus de aprendientes informatizados pueden suplir estas carencias, y sumar otras ventajas, como sugieren Leech (1998: xix), Nesselhauf (2005: 43) o Guo (2006: 12), entre otros. Estas se presentan en el siguiente apartado.

²³ CHILDES abarca tanto la dimensión monolingüe del lenguaje como la bilingüe, la no nativa y la patológica.

²⁴ Para un compendio de publicaciones recientes –basadas en corpus– más descriptivas que interpretativas, véase Lozano y Medikoetxea (2013: 3).

²⁵ Esta investigación también pretende contribuir en esta nueva dirección, pues se busca probar varias hipótesis en torno a la interfaz léxico-sintaxis en una base de datos más amplia y bien construida que facilite la generalización de los resultados; esta interfaz es “one of the much debated issues in second language research [...], as a potential source of observed deficits [together with the syntax-discourse interface] in the development of learners’ interlanguage grammars” (Lozano y Mendikoetxea, 2013: 6).

2.2. LA INVESTIGACIÓN DE CORPUS DE APRENDIENTES

Los comienzos de la LC se remontan a los años sesenta –como se ha referido al comienzo de este capítulo–, cuando se compilan los primeros corpus que permiten mejorar las descripciones del inglés. En el ámbito específico de la LC, el significado de *corpus* es mucho más restringido que el que se utiliza en la lengua general o en el periodo anterior al nacimiento de esta rama de la lingüística, en los años sesenta, dado que se concibe necesariamente en soporte electrónico²⁶.

En los últimos 20 años se han confeccionado grandes corpus, como son, en el caso del inglés, el British National Corpus (BNC) –con 100 millones de palabras– y el Corpus of Contemporary American English (COCA) –con 410 millones de palabras–; y, en el caso del español, el Corpus del Español –con 100 millones de palabras– y el Corpus de la Real Academia Española (CREA) –con 154 millones de palabras–. La RAE ha puesto en marcha, con la colaboración de la Asociación de Academias de la Lengua Española (ASALE), el Corpus del Español del Siglo XXI (CORPES XXI), cuya primera fase de elaboración finalizará en diciembre de 2014, con 300 millones de palabras. Este corpus de referencia y herramienta de investigación se presentó en el VI Congreso Internacional de la Lengua Española, en Panamá, y reunirá textos reales de todos los países hispanohablantes, al igual que el CREA y el Corpus Diacrónico del Español (CORDE), pero con muestras de lengua recientes –desde el año 2001 hasta el 2012– que nos permitirán obtener mejores descripciones sobre el uso que se hace de la lengua durante el siglo XXI.

Ya se ha señalado que el uso de corpus en la investigación sobre la adquisición de la L1 tampoco es nuevo; desde los años 70 es una práctica común en los estudios de la lengua del niño. El corpus más extenso es el ya mencionado CHILDES, que cuenta con 44 millones de palabras en más de 30 lenguas diferentes, y constituye un referente en el estudio de la adquisición de la L1 y del bilingüismo (Lozano y Mendikoetxea, 2013: 3)

Ahora bien, los corpus de aprendientes informatizados no aparecieron hasta comienzos de los 90²⁷, cuando la tecnología y el análisis de corpus nativos se

²⁶ Véase n. 14 al respecto y la definición de *corpus* de Sinclair (2005: 16) recogida al comienzo del capítulo.

²⁷ Dentro del proyecto danés conocido como Project in foreign language pedagogy (PIF) se llevaron a cabo determinados estudios que se publicaron en el libro *Learner Language and Language Learning* (Færch *et al.*, 1984, citado en Hasselgard y Johansson, 2011: 34), en el que se discuten muchos aspectos de la lengua del aprendiente a partir de un corpus de HNN único tanto en el tamaño como en el tipo. El corpus escrito contaba con 100000 palabras y el oral con 250000 palabras. Færch presentó en 1979 planes para informatizarlo, pero su muerte en 1987 impidió que el desarrollo del proyecto PIF fuera más allá y que se conformara el primer corpus de aprendientes informatizado (Hasselgard y Johansson, 2011: 34).

encontraban relativamente desarrollados²⁸ –tras originarse el interés, a partir de finales de los 80, por la creación de recursos de corpus útiles de acuerdo con las necesidades de los aprendientes de lenguas extranjeras (Tribble, 2011: 85)–.

En relación lógica con la ya expuesta definición de corpus de Sinclair (2005: 16), se entiende por *corpus de aprendientes*

Electronic collections of authentic²⁹ FL/SL textual data according to explicit design criteria for a particular SLA/FLT purpose. They are encoded in a standardised and homogeneous way and are documented as to their origin of provenance (Granger, 2002: 7).

En esta definición de Granger se sintetizan las características fundamentales de un corpus de aprendientes: es una recopilación de textos en soporte electrónico que requiere un diseño particular (de acuerdo con unos criterios estándar que son definidos a propósito del CEDEL2 en el capítulo 6). A los principales objetivos de este tipo de corpus se refiere la misma autora: dispone de unas herramientas y métodos procedentes de la LC que ayudan a aportar descripciones mejoradas de la lengua del aprendiente que pueden ser utilizadas en la investigación de la adquisición y en la enseñanza de las L2:

Using the main principles, tools and methods from corpus linguistics, it aims to provide improved descriptions of learner language, which can be used for a wide range of purposes in foreign/second language acquisition research and also to improve foreign language teaching (Granger, 2002: 4).

Así pues, surgen en esta época los corpus de aprendientes porque, por un lado, la LC ya está lo suficientemente consolidada desde un punto de vista tecnológico-metodológico, como ya se ha señalado, y, por otro lado, por las necesidades planteadas en el ámbito de la ASL que desembocaron en la confluencia de ambas ramas, por lo que los análisis de la interlengua basados en extensos corpus disfrutaron a partir de entonces de una buena acogida en este campo de la ASL. A estas necesidades nos referimos a continuación.

En primer lugar, los estudios basados en el análisis del comportamiento de los HN no informan de la dificultad de las estructuras específicas que deben ser enseñadas

²⁸ El University College de Londres, la Universidad de Lancaster, la de Birmingham y la de Nottingham, en el Reino Unido; y las de Northern Arizona y Brigham Young, en EE.UU., han tenido mucho que ver en este desarrollo, mediante las aportaciones de lingüistas como Randolph Quirk, Geoffrey Leech, John Sinclair, Ron Carter, Mike McCarthy, Douglas Biber o Mark Davies.

²⁹ La autora especifica más adelante (Granger, 2002: 8) lo que se entiende por datos auténticos: los datos que resultan de actividades reales de clase.

en el aula de LE, ni de su proceso de aprendizaje; por ello, Granger (1998b: 7) reprueba que los materiales de enseñanza de inglés como lengua extranjera se diseñen “with a very fuzzy, intuitive, non-corpus-based view of the needs of an archetypal learner”. Es imprescindible conocer las verdaderas dificultades de los HNN en el aprendizaje de la L2 para que puedan ser trabajadas adecuadamente en los materiales didácticos, y estas pueden ser fácilmente identificadas en los corpus de HNN. A este respecto, Leech (2001: 339) defiende que “corpus based interlanguage analysis enables us to identify areas of difficulty which are not derivable from Ns corpora alone”. Estas ideas son recogidas por Nesselhauf (2004: 125-126) en la siguiente afirmación, en la que destaca para el desarrollo de la enseñanza de una L2 la importancia de disponer tanto de corpus nativos para conocer lo que estos verdaderamente utilizan como la de contar con corpus de HNN para averiguar sus dificultades:

Hardly anyone will doubt any longer that native speaker corpora are indeed useful for the improvement of language teaching. They are useful mainly because they can reveal –better than native speaker intuition– what native speakers of the language in question typically write or say (either in general or in a situation/in a certain text type). For language teaching, however, it is not only essential to know what native speakers typically say, but also what the typical difficulties of the learners of a certain language, or rather of certain groups of learners of this language, are.

En segundo lugar, se hacen necesarios los corpus de aprendientes informatizados por las ventajas que ofrecen con respecto a los procedimientos adoptados normalmente en el AE tradicional (Altenberg y Granger, 2001: 189; Nesselhauf, 2005: 41), como son los corpus textuales preinformatizados y las pruebas de elicitación –de donde se obtienen datos seminaturales, ya que las tareas están diseñadas para controlar la lengua que los HNN deben producir–. Además, los corpus informatizados permiten aumentar el tamaño del material analizado –una investigación a gran escala puede revelar rasgos de uso que hayan escapado a lingüistas que emplearan la intuición o una pequeña cantidad de muestras– y la variedad, normalmente producto de una gran cantidad de participantes:

Unless they are very large, learner corpora with a great number of words from individual learners have the disadvantage of being less balanced than those that incorporate a smaller number of words from more learners, as individual usage may easily distort the overall results. For this reason, the contribution of individual learners is usually kept small in learner corpora (Nesselhauf, 2005: 43).

Las principales ventajas son bien conocidas. En primer lugar, la recogida de datos en las que se basan el AE no suelen ser sistemáticas –frente a la recopilación de los datos de un corpus que sí lo son–: los detalles de los HNN y las circunstancias de la producción no se recogen o no lo suficientemente, y esta información es necesaria para realizar una adecuada interpretación de los datos, esto es, un buen análisis de la interlengua (Nesselhauf, 2005: 41-42). En segundo lugar, como ya señalaron Schachter y Celce-Murcia en 1977, la colección de textos se concibe como un depósito de errores que se desecha una vez que estos son extraídos, por lo que no se cuenta con el contexto para verificar resultados. Tampoco se atiende, como observa Nesselhauf (2005: 41-42), a la lengua que los HNN producen correctamente, hecho denunciado por Svartvik en 1973; los errores son contabilizados como absolutos y no suelen compararse con los aciertos en esa misma estructura o elemento. Asimismo, ni la sobreutilización ni la infrautilización pueden analizarse con este método. Con las siguientes palabras lo expone Leech (1998: xvii): el corpus de aprendientes

enables us to investigate the non-native speaking learner's language (in relation to the native speakers) not only from a negative point of view (what did the learner get wrong?) but from a positive one (what did the learner get right?). For the first time it also allows a systematic and detailed study of the learner's linguistic behaviour from the point of view of "overuse" (what linguistic features does the learner use more than a native speaker?) and "underuse" (what features does the learner use less than a native speaker?).

Asimismo, la consolidación de la LC, la evolución tecnológica y los métodos desarrollados en la investigación permitieron que los nuevos corpus informatizados pudieran ser desarrollados como herramientas de investigación para ser usadas por muchos especialistas en este campo, y no solo por un investigador individual como en el análisis del material usado en el periodo anterior; este es un aspecto clave del éxito de la investigación en ASL basada en corpus. Además, por medio del ordenador, se pueden realizar nuevos tipos de estudios, como obtener la frecuencia de coaparición –o estudios cuantitativos en general–, y descubrir patrones de uso lingüístico en un grupo concreto de HNN. Este es, al fin y al cabo, el objetivo de esta investigación, en la que el ordenador facilita ciertas búsquedas, como la de extraer las concordancias del verbo, tarea que hubiera resultado muy costosa de haber tenido que realizarse manualmente –para lo que habría sido necesario trabajar con un corpus mucho más limitado, con el consiguiente problema de generalización de los resultados–. Hay diferentes tipos de *software* de recuperación de los datos en función de los propósitos del investigador. Micro Concord, Word Smith Tools o AntConc son algunas de las herramientas de

recuperación más usadas (para detalles del *software* empleado en esta investigación, véase el capítulo 6).

No obstante, pese a estos beneficios en el análisis por la informatización de los corpus, los investigadores no deberían limitar sus investigaciones a lo que el ordenador puede hacer, pues, como señala Granger (1998b: 16), “a computerized approach has linguistic limitations”. Por un lado, las diferencias o las semejanzas superficiales entre los aspectos de la lengua nativa y no nativa siempre requieren investigación cualitativa (Meunier, 1998: 36). Dado que los hallazgos cuantitativos han de ser considerados cuidadosamente y comparados con análisis cualitativos, se plantea en Hasselgard y Johansson (2011: 55-56) la necesidad de utilizar corpus alternativos de referencia para contrastar los datos y buscar causas externas para algunos de los resultados, recursos que se utilizan en este trabajo. De este modo, hay que considerar los datos cuantitativos como punto de partida para un análisis posterior, por lo que filtrar los datos obtenidos por medios informáticos debería ser la esencia de la investigación de corpus de aprendientes (Guo, 2006: 23).

El siguiente apartado presenta algunos de los corpus de HNN más destacados entre los que se incluye el corpus del que nos hemos servido en esta investigación.

2.2.1. Revisión de los corpus de aprendientes

El interés por los corpus de aprendientes se desarrolló rápidamente en los años noventa y primera década del siglo XXI, inspirado por el trabajo iniciado por Granger y su equipo en el CECL, el ICLE, al que ya nos hemos referido –en la introducción de este capítulo– como el primer corpus de aprendientes reconocido internacionalmente³⁰. Este proyecto ha estimulado la creación de corpus similares en otras interlenguas (como la del español con el CEDEL2), ha generado una gran cantidad de estudios académicos y ha influido enormemente en el campo de las publicaciones en enseñanza del inglés (como las de Longman y Cambridge University Press que, pese a usar sus propios corpus de HNN, reciben la influencia del ICLE directa o indirectamente) y en el de la ASL (McEnery y Hardie, 2012: 82).

La aparición de la investigación de CAI puede decirse que se produce en 1998 con la publicación de *Learner English on Computer*, de Granger (De Cock *et al.*, 2011: 2;

³⁰ Con anterioridad al trabajo realizado en el proyecto ICLE, se llevaron a cabo estudios basados en datos de corpus de aprendientes –entendiendo este concepto desde una perspectiva muy amplia– como el de Krashen *et al.* (1978), pero el alcance y la naturaleza sistemática del trabajo de Granger se concibe como el primer compromiso a gran escala con la lengua del aprendiente por medio de un enfoque basado en corpus (McEnery y Hardie, 2012: 82).

Nesselhauf, 2005: 40), una colección de artículos pioneros sobre la lengua del aprendiente, basada en el ICLE. El volumen coeditado por Granger (Granger *et al.*, 2002) *Computer Learner Corpora, Second Language Acquisition and Foreign Language Teaching* es una continuación del anterior con más desarrollos en el campo. El ICLE es un corpus informatizado disponible comercialmente³¹ de 3,3 millones de palabras procedentes de ensayos argumentativos escritos por aprendientes de inglés con dieciséis L1 diferentes (3,7 millones de palabras en la versión del corpus publicada en 2009). Pronto se completó con un corpus hablado (LINDSEI, Gilquin *et al.* 2010), constituido por 800000 palabras de entrevistas a aprendientes de inglés de diferentes L1. Ambos corpus –anotados según la categoría de palabra y error– se corresponden únicamente con un diseño transversal, esto es, basado en distintos niveles de competencia. Para suplir esta limitación, se inició en 2008 un nuevo proyecto, el Longitudinal Database of Learner English (LONGDALE), para compilar un corpus longitudinal, que integra HNN de nivel avanzado con diferentes L1 que son seguidos durante un periodo de tres años. Este corpus puede revelar información sobre el propio proceso de aprendizaje.

Estos corpus de aprendientes desarrollados en el CECL permiten mostrar los errores y las desviaciones cuantitativas (sobreutilización e infrautilización) en los usos de los HNN a partir de una norma o estándar comparativo seleccionado. Como señala Guo (2006: 22), “The characteristics of learner IL [interlengua] will become obvious only when learner output is put into a comparison with some kind of norm (even though is impossible to establish a norm acceptable to all the researchers in this field” (véase el capítulo 5 para una discusión y propuesta sobre el tema de la norma). Estos errores y desviaciones difieren en tipo y frecuencia dependiendo de la L1 del aprendiente (Altenberg, 2011: XIV), por lo que para poder determinar el papel que desempeña la transferencia de la L1 en la producción del aprendiente lo ideal es poder contar con un corpus multilingüe, como los anteriores, para desde una perspectiva contrastiva poder demostrar empíricamente las intuiciones en la teoría de ASL. Por ello, Granger propone en 1996 el Análisis Contrastivo de Interlengua (ACI) para analizar los corpus de aprendientes, y lo coloca en el centro del proyecto ICLE (De Cock *et al.*, 2011: 2; Hasselgard y Johanson, 2011: 38)³². A diferencia del estudio contrastivo, que implica normalmente la comparación de dos lenguas, el ACI involucra diferentes variedades de la misma lengua: “involves quantitative and qualitative comparisons between native language and learner language (L1 *vs.* [*versus*] L2) and between different

³¹ Se puede acceder al CD-Rom en <<http://www.i6doc.com/en/collections/cdicle/>> (18-06-14).

³² La utilidad del ACI ha sido demostrada en una gran cantidad de estudios, como se muestra en Granger *et al.* (2002). Esta metodología ha sido adoptada también en otros proyectos en los que se utilizan corpus de interlengua de francés (FRIDA [Granger, 2003]), alemán (el corpus FALKO [Lüdeling *et al.*, 2005]), italiano (el corpus VALICO [Corino, 2008]) y noruego (el corpus ASK [Tenfjord *et al.* 2006]). Para una información sobre estos corpus véase <www.uclouvain.be/en-ccel-lcWorld.html> (19-06-14).

varieties of interlanguage (L2 *vs.* L2)” (Granger, 2009: 18; véase también Granger, 1996). No obstante, Hasselgard y Johansson (2011: 43) subrayan que la mayoría de los estudios de ACI realizados hasta ahora comparan la lengua nativa con una sola variedad no nativa, si acaso con dos; no en vano, Guo (2006: 22) mantiene que el ACI no solo se refiere a la comparación entre interlenguas diferentes, sino también a la comparación entre una interlengua particular y la lengua meta. De hecho, como reconoce Nesselhauf (2005: 9), en función del aspecto de la lengua analizado, puede que sea necesario restringir el análisis a un grupo de L1 en lugar de analizar más datos de muchos grupos diferentes, como es el caso de la combinatoria léxica, ya que muchos estudios sugieren que la L1 desempeña un papel importante en la producción de las combinaciones de palabras³³.

Los estudios de interlengua basados en corpus requieren, por lo tanto, un corpus nativo comparable, esto es, diseñado para permitir comparaciones: “Only those components of corpora which have been designed to be independently contrastive should be contrasted” (Sinclair, 2005: 3). Este corpus de control sirve para medir las desviaciones de la lengua de los HNN con respecto a la nativa, aunque es posible realizar un análisis de interlengua basado en corpus no contrastivo; este se integra en los estudios de ASL, pero no en los de ACI; no obstante, esta no es una práctica estándar en la investigación de la L2 (Lozano y Mendikoetxea, 2013: 10). El corpus de aprendientes suecos de Linnarud (1986), al que ya nos hemos referido, aunque no estaba informatizado, contaba ya con un corpus comparable de HN de inglés, por lo que puede ser considerado germen del ACI.

El corpus nativo de comparación con el que cuenta el ICLE es el Louvain Corpus of Native English Essays (LOCNESS), que contiene unas 324000 palabras procedentes de ensayos argumentativos escritos por estudiantes británicos y norteamericanos. Por su parte, el LINDSEI también dispone de un corpus nativo de control, el Louvain Corpus of Native English Conversation (LOCNEC). Todos estos corpus están suficientemente documentados, esto es, contienen información sobre el aprendiente (edad, sexo, L1, otras lenguas extranjeras conocidas, región, nivel de dominio, etcétera), datos considerados necesarios en una investigación de interlengua basada en un corpus.

La influencia del ICLE se puede observar en numerosos corpus de aprendientes³⁴, como el Written Corpus of Learner English o WriCLE (Rollinson y

³³ De acuerdo con este sentido menos restrictivo del concepto, la presente investigación se encuadra en el ACI.

³⁴ Véase <<http://www.uclouvain.be/en-cecl-lcworld.html>> (20-06-14) para un extenso listado de diferentes corpus de aprendientes, integrado por varias interlenguas –en su mayoría inglés– y diferentes L1,

Mendikoetxea, 2010), un corpus disponible gratuitamente³⁵ de unas 750000 palabras de aprendientes de inglés con L1 español, creado en la Universidad Autónoma de Madrid con el mismo diseño que el ICLE. En la actualidad se está compilando un subcorpus de escritura no académica (WriCLEinf), en su mayoría blogs, para permitir la comparación entre los diferentes registros, así como el estudio de estructuras que no se encuentran normalmente en escritos académicos más formales (Lozano y Mendikoetxea, 2013: 5).

Otros importantes corpus de aprendientes fueron compilados para satisfacer las necesidades de los creadores de materiales didácticos de consulta. Destaca el Longman Learners' Corpus (compilado por Longman Publishers), que pretende servir en la creación de diccionarios de enseñanza de lengua inglesa, como el *Longman Dictionary of Contemporary English* (1995) y para otros recursos de enseñanza y materiales didácticos que están diseñados para contrarrestar los errores hechos por estudiantes de determinadas L1. La recopilación de muestras escritas de aprendientes se inició en 1987, por Longman. Este corpus cuenta con 10 millones de palabras escritas por HNN de 117 nacionalidades con diferentes niveles de dominio. El corpus está ahora disponible comercialmente, aunque ha originado muchas menos investigaciones que el ICLE (Guo, 2006: 14).

El Cambridge Learners Corpus (Nicholls, 2003) es el mayor corpus de aprendientes de inglés –cuenta con 30 millones de palabras, y está en continuo desarrollo–, y también fue compilado para desarrollar material didáctico de inglés como lengua extranjera, en este caso, en la editorial Cambridge University Press; pero también está disponible comercialmente. Está creado a partir de más de 200000 exámenes escritos por aprendientes de inglés de 217 países diferentes. Contiene información sobre el aprendiente y las condiciones de redacción. Resulta de particular interés el hecho de que está anotado con la localización y tipos de errores hechos por los HNN en su producción escrita. Las ventajas de disponer de un corpus anotado con los errores son, a menudo, destacadas (en el apartado 2.2.2 retomamos esta cuestión):

We can see which errors are typical of different learner levels or of particular language groups because all the scripts have information about the first language and English level of the writer. This means that when we produce a book designed for a particular level, e.g. Upper Intermediate, we can look at all scripts written by Upper Intermediate learners

con información sobre el tipo de datos que conforman el corpus, número de palabras y nivel de dominio del aprendiente.

³⁵ El corpus se encuentra disponible en <<http://www.uam.es/proyectosinv/woslac/Wricle/>> (20-06-14).

and very easily see exactly what mistakes they make. In this way we can make sure the book contains appropriate help for an Upper intermediate student³⁶.

Otros corpus de aprendientes, como The Hong Kong University of Science and Technology Learner Corpus o The Chinese Learner English Corpus, por poner algún ejemplo más, han sido también utilizados para la investigación en estudios de la interlengua del inglés de sinohablantes (Chi *et al.*, 1994, o Guo, 2006, entre muchos otros); el primero consta de 25 millones de palabras escritas por aprendientes de inglés en bachillerato y universidad. Al segundo, formado por un millón de palabras, solo pueden acceder los usuarios del departamento de inglés de HKPU.

Desde no hace mucho tiempo, contamos también con datos de la interlengua del español; lo que no es de extrañar, dado el tremendo desarrollo que esta lengua ha experimentado en esta última década: “se ha producido un auge notable en lo que al estudio del español y a todo lo relacionado con las culturas hispánicas respecta” (Muñoz-Basols *et al.*, 2014: 1)³⁷. Este hecho ha revertido en un creciente interés por la adquisición del español como segunda lengua (Pastor Cesteros, 2001; Pérez Leroux y Muñoz Licerias, 2002 o Montrul, 2004 –centrados en el análisis de aspectos morfosintácticos–; o los más abarcadores Lafford y Salaberry, 2003 o Geeslin, 2013³⁸), que ha coincidido con la creación de varios corpus de aprendientes de español. No es nuestro objetivo ser exhaustivos proporcionando un listado con todos los corpus de aprendientes de español, sino presentar los más destacados por su calidad en el diseño y, por lo tanto, por su potencial para servir como buena base de datos en las investigaciones de la interlengua del español. Además, como bien observan Lozano y Mendikoetxea (2013: 21), los investigadores están continuamente creando corpus para satisfacer sus necesidades, lo que dificulta un posible recuento de todos los corpus existentes en un momento dado.

Uno de los corpus del español más representativos por su rigor metodológico y tamaño es el CEDEL2³⁹ (Lozano, 2009a; Lozano y Mendikoetxea, 2013), que nace en el proyecto WOSLAC, de la Universidad Autónoma de Madrid y la Universidad de

³⁶ Cita de la página web de Cambridge University Press recogida por McEnery y Hardie (2012: 83) aunque ya no se encuentra localizable.

³⁷ En este trabajo de Muñoz-Basols *et al.* (2014) queda constatada la importancia del español no solamente como lengua de estudio sino también como un idioma unido a lo económico.

³⁸ En esta obra, el español como L2 y el proceso de adquisición del español como L2 son analizados desde diferentes perspectivas de investigación y se atiende a diversos rasgos de los niveles lingüísticos e interfaces; no obstante, los elementos más trabajados pertenecen al nivel fonológico y gramatical.

³⁹ En la siguiente página web se presenta el corpus y la dirección de contacto de Cristóbal Lozano, compilador y creador del corpus, a quien se le puede solicitar una muestra del CEDEL2, dado que todavía no ha sido publicado, <<http://www.uam.es/proyectosinv/woslac/collaborating.htm>> (23-06-14).

Granada, cuyo objetivo principal es el estudio del papel de las interfaces (léxico-sintaxis y sintaxis-discurso) como fuentes de error en el desarrollo de la interlengua del HNN. Es un corpus escrito de aprendientes anglófonos de español clasificados en un nivel de dominio lingüístico inicial, intermedio o avanzado, según los resultados obtenidos en un test de diagnóstico estandarizado. Este hecho es un dato importante para el análisis que, sin embargo, no es tenido en cuenta para la configuración del ICLE, donde se clasifican los participantes por el nivel de dominio que deberían tener de acuerdo con la edad y curso en el que se encuentran; Tono (2003: 801) también recoge estas reservas a la hora de medir la lengua de los HNN por medio de factores externos:

Selection based upon external criteria such as school year or age does not necessarily ensure that the subjects selected are comparable in terms of language proficiency. This happens to be the case for the Japanese-speaking EFL learner group. Although their learner profile fulfilled all the criteria, their proficiency levels are so markedly lower than those from other European countries that the inclusion of the Japanese data seems to skew the overall results.

El CEDEL2 cuenta, asimismo, con un subcorpus nativo de control diseñado con las mismas características que el no nativo con el objetivo de servir como base de datos con la que comparar los usos de la interlengua. En marzo de 2011 (fecha del último recuento publicado) el corpus contaba con 750000 palabras, aunque se encuentra todavía en la fase de recopilación de datos; se pretende alcanzar el millón de palabras. Hasta ahora, los estudios realizados a partir del CEDEL2 han analizado algunos aspectos del español aislados, pertenecientes a diferentes niveles lingüísticos e interfaces. Así, Pino (2012) estudia el uso de *lo que*, *de que*, *algo que* y *dice que* en un grupo de aprendientes suecos de español comparado con el uso que hacen los HN del CEDEL2, y Lozano (2009b, 2009c, 2011) se ha centrado en el análisis de la adquisición de sujetos gramaticales. Algunas muestras del CEDEL2 han sido etiquetadas de acuerdo con la estructura sintáctica y colocaciones –por medio de UAM Corpus Tool (O’ Donnell, 2009)–. De hecho, se ha iniciado el estudio de las colocaciones en el CEDEL2 (Prieto *et al.*, 2009; Pérez Serrano, 2012; Orol González y Alonso Ramos, 2013) con vistas a diseñar aplicaciones en línea, asistentes informáticos que sirvan de herramienta de ayuda a la redacción en español, que detecten el error colocacional y aporten estrategias de corrección (Ferraro *et al.*, 2011; Vincze *et al.*, 2011; Wanner *et al.*, 2013a; Ferraro *et al.*, 2013;), lo cual requiere la creación de una tipología del error colocacional que permita etiquetar el corpus de aprendientes (Alonso Ramos *et al.*, 2010a y 2010b; Wanner *et al.*, 2013b). Por otro lado, Escutia (2010 y 2012) analiza el uso de predicados inacusativos y del *se* por aprendientes de nivel avanzado.

El diseño del CEDEL2 atiende a los 10 principios estándar recomendados por Sinclair (2005). El hecho de que un corpus esté bien diseñado permite poder contestar a cualquier pregunta de investigación sobre la adquisición del español; además, la redacción –que es el tipo de texto que configura este corpus– es una actividad voluntaria para los HNN que carece de restricciones temáticas, ya que el participante elige de entre una gran variedad de títulos. Esto hace que todas las estructuras y elementos léxicos estén suficientemente representados. Por eso, el CEDEL2 se considera una base de datos naturales (una información más detallada sobre este corpus se halla en el capítulo 6).

El CEDEL2 es un buen complemento del Spanish Learner Language Oral Corpus (SPLLOC 1 y SPLLOC 2⁴⁰) (Mitchell *et al.*, 2008), que lanzaron poco antes la Universidad de Southampton, la Universidad de Newcastle y la Universidad de York con financiación del Economic and Social Research Council (ESRC) del Reino Unido. Este, a diferencia del CEDEL2, contiene datos orales, y no escritos, de anglohablantes clasificados en un nivel de dominio lingüístico en función de la edad y los años de estudio del idioma –y no a través de un test de diagnóstico estandarizado–, que aprenden español en contextos institucionales (secundaria y universidad). Este corpus también contiene datos de HN de español obtenidos a partir de las mismas entrevistas con objeto de poder comparar el español nativo y la interlengua española de los hablantes de inglés. Los datos se obtienen a partir de unas tareas de elicitación diseñadas para este proyecto, entre las que se incluye la narración, la descripción de imágenes, el debate sobre un tema determinado y la entrevista individual. Con los datos obtenidos de estas tareas se pretende contestar a las preguntas de investigación del proyecto, relacionadas con el uso de los pronombres clíticos y el orden de palabras en el caso de SPLLOC1, y con el tiempo y aspecto en el de SPLLOC2. Por consiguiente, a diferencia del CEDEL2, este corpus no contiene datos de producción naturales, sino *semi-naturales*, pues está diseñado para obtener determinados usos o estructuras que permiten al investigador dar respuesta a sus preguntas de investigación. Pese a que estas técnicas se desaconsejan según los criterios estándar de diseño de corpus, el SPLLOC supone un hito para la investigación de corpus de aprendientes de español, por ser el primero que está compilado de acuerdo con ciertos criterios estándar de diseño, lo que lo convierte en un corpus muy útil para la investigación en ASL.

Existen otros corpus de aprendientes de español, la mayoría de orientación pedagógica que a menudo se basan en análisis de errores. Destacan el Corpus de Aprendices Taiwaneses de Español (CATE; Lu, 2010), todavía en desarrollo⁴¹ y que ya

⁴⁰ Véase <<http://www.sploc.soton.ac.uk/>> (20-06-14).

⁴¹ Véase <<http://corpora.fild.ncku.edu.tw/#>> (20-06-14).

cuenta con casi medio millón de palabras, o The Japanese Learner Corpus of Spanish (Kamakura), con 83400 palabras escritas por aprendientes japoneses; de reciente creación es el Corpus Español de Aprendientes Italianos (SCIL; Bailini, 2013) y en el corpus The Anglian Polytechnic University Learner Spanish Corpus, de la Universidad Anglia Ruskin, en el Reino Unido (Ife, 2004) los participantes son aprendientes de varias L1. Por otro lado, el corpus Aprender a Escribir en Lovaina, de la Universidad Católica de Lovaina, es un corpus escrito de 1 millón de palabras de aprendientes de L1 alemán de todos los niveles de dominio; su acceso en línea está restringido. En España contamos también con el Corpus para el Análisis de Errores de Aprendices de E/LE (CORANE) (Cestero Mancera *et al.*, 2001), creado con fines pedagógicos, que desde 2009 se distribuye en CD (Cestero Mancera y Penadés, 2009). Reúne más de mil escritos, etiquetados, de informantes de diferentes niveles de dominio y muy variadas L1. Asimismo, ha sido compilado recientemente el corpus oral CORELE (Campillos Llanos, 2014), que incluye muestras de lenguas de aprendientes de español con varias L1 y que contribuye al análisis del error ayudado por las técnicas informatizadas (del inglés *Computer-aided Error Analysis*, Granger [2008]). Este enfoque emplea las técnicas de la LC y hunde sus raíces en el marco del AE propuesto por Corder (1971), al que ya nos hemos referido (identificación, descripción y explicación del error), pero requiere dar un paso más –la etiquetación de los errores de acuerdo con una descripción y clasificación del error previamente establecida (a las ventajas y limitaciones de un corpus etiquetado con los errores nos referimos en el apartado 2.2.2)–; este corpus (CORELE) también cuenta con un corpus de control nativo, por lo que permite la comparación entre el español de los HNN y el de los HN⁴².

De todo esto se deduce que ya existe material para poder investigar la lengua de los aprendientes de español, lo que permite enriquecer el panorama de los estudios de esta interlengua. Una vez que se diseñan corpus que reúnen las condiciones de fiabilidad y de estandarización de los datos, como es el caso del CEDEL2, los cimientos para estrechar la relación entre la LC y la investigación de ASL ya están establecidos; si sobre ambos pilares se construye la investigación de CAI, este es el momento, por lo tanto, de realizar un sólido avance en esta rama de estudio de la lengua del aprendiente de español.

⁴² Para más información sobre diferentes enfoques basados en corpus para investigar la interlengua del español véase Campillos Llanos (2014) y Mendikoetxea (2013).

2.2.2. Algunos aspectos metodológicos y limitaciones de la investigación basada en corpus

La investigación en la LC puede clasificarse en función de unos criterios que distinguen los tipos de corpus con los que puede trabajarse. A continuación, se presentan brevemente algunos de los criterios que consideramos más relevantes (véase McEnery y Hardy, 2012, para una detallada clasificación sobre la tipología de corpus):

1) El uso de corpus anotado *versus* no anotado. Cuando el análisis lingüístico está codificado en los propios datos del corpus, se trata de un corpus anotado. En lugar de editar el texto directamente, es posible almacenar las anotaciones separadas del texto – para que el texto no se sobrecargue de etiquetas (Sinclair, 2004: 191)–, de tal manera que son los programas informáticos los que combinan e integran el texto y las anotaciones en función de lo que requiera el analista. Un corpus se puede anotar de modo que cada palabra tenga, por poner algún ejemplo, una etiqueta con la categoría gramatical a la que pertenece; se puede anotar el número y la persona de los verbos o pronombres, si el sustantivo tiene un referente animado o inanimado, o los errores. Al respecto de etiquetación de estos últimos, Granger (1998b: 15) defiende las cualidades de un corpus anotado; así, argumenta que una vez que se diseña una taxonomía del error y se insertan las etiquetas en los archivos de texto, el corpus puede ser consultado de manera automática, y pueden obtenerse listas de tipos de errores específicos o las frecuencias de cada tipo de error (Campillos Llanos, 2014).

Para automatizar la etiquetación del error se utiliza un *software* especial, como el Error Editor, usado en el CLCI, el Knowator, aplicado al CEDEL2 en los errores colocacionales o el Exchanger XML Lite 3.2. (eXtensible Markup Language). De este modo, la explotación de los datos puede ser rápida si se cuenta con herramientas de búsqueda adecuadas que mejoren la precisión de las búsquedas –si el CEDEL2 tuviera los errores ortográficos anotados, se podría garantizar que la recuperación de los datos de la búsqueda del lema *dar* incluyera todos los usos de *dar* de los HNN⁴³–. Para Leech (1997: 4-6), la anotación añade valor al corpus original porque permite la extracción directa de la información requerida por el analista y porque permite la reutilización del corpus por otros usuarios sin que estos tengan que reanotar el mismo material lingüístico.

Estas ventajas no implican que el proceso de etiquetación de errores esté exento de limitaciones: ni existe un estándar uniforme sobre lo que es correcto e incorrecto – los nativos tampoco parecen estar siempre de acuerdo en lo que es un error y lo que es aceptable en su lengua– ni, por tanto, son incuestionables los hallazgos obtenidos a

⁴³ En Rayson y Baron (2011) se presenta una nueva aplicación, el llamado *Variant Detector*, para detectar los errores ortográficos en corpus escritos de HNN.

partir de un corpus de HNN etiquetado con los errores (acerca de la selección de un estándar estable, véase el capítulo 5). Al mismo tiempo, no conviene perder de vista que un corpus etiquetado con los errores no permite descubrir otros aspectos de la interlengua que también contribuyen a caracterizar la lengua del aprendiente como lengua no nativa –por lo que adolece de las mismas debilidades del AE, que se centraba en los usos incorrectos y no consideraba los usos correctos ni las estrategias de evitación–, como señala Granger (2003: 11):

It is also important to bear in mind that error tagging, in spite of its numerous advantages, is only concerned with learner misuse. It fails to uncover other aspects of interlanguage such as the under –and overuse of words and phrases, which together with downright errors contribute to the nonnativeness of learner productions.

En suma, la etiquetación del error, aunque puede resultar útil, no es una panacea para el problema de lo no nativo en el uso de la L2. Es un tipo de información, entre toda la que puede proporcionar el corpus de HNN, que puede servir de ayuda al investigador de la interlengua, al estudiante y al profesor.

Como se ha adelantado, en lugar de editar el texto directamente, otra opción es tener el texto libre de anotaciones externas –las cuales aparecen aisladas fuera del texto–, y parece que esta dicotomía enfrenta a los especialistas. Los defensores del texto no anotado consideran que las etiquetas pueden imponer un análisis a los usuarios de los datos, o hasta pueden ser imprecisas o inconsistentes, críticas que Sinclair (2005: 5) tiene en cuenta cuando declara que “Any information about a text other than the alphanumeric string of its words and punctuation should be stored separately from the plain text and merged when required in applications”. Separar el texto de las etiquetas está en consonancia con la propuesta de Hunston (2002: 94) de no mezclar el texto original y la anotación, dado que debe haber un movimiento constante entre el uso de técnicas avanzadas de búsqueda en un corpus anotado y la visión de los datos lingüísticos originales, sin procesar.

En línea con las anteriores argumentaciones de Sinclair y Hunston, en esta investigación –que usa un corpus no procesado– habría sido útil poder contar con el corpus anotado con los lemas verbales, para poder saber con exactitud que todas las formas de *dar* son recuperadas para el análisis (teniendo en cuenta el margen de error de la búsqueda debido a fallos ortográficos al que nos acabamos de referir). O, si el corpus hubiera estado etiquetado según las categorías gramaticales de las palabras, también se habría evitado el trabajo de filtración de los datos no relevantes en los que pueda aparecer *dar*, como conjunciones del tipo *dado que*, la preposición *de* o el sustantivo *dado*. En cambio, para el estudio de las diferentes construcciones verbales en las que aparece

dar y de sus coapariciones léxicas, no se requiere el texto anotado ni los errores etiquetados, por las limitaciones señaladas *supra*.

2) Uso de un corpus longitudinal *versus* transversal. Los corpus se han utilizado para estudiar las diferencias en la lengua a lo largo del tiempo; son los corpus diacrónicos los que, aplicados a la investigación de corpus de aprendientes, se conocen como *corpus longitudinales* (Granger, 2002: 11), usados para rastrear el desarrollo lingüístico de los HNN durante un periodo de tiempo. Estos son muy útiles para describir el progreso y la evolución de la interlengua. Debido a la dificultad a la hora de compilar un corpus de este tipo –que requiere un seguimiento de un grupo de HNN a lo largo de unos años–, existen pocos de esta naturaleza, como el LONGDALE, de la Universidad de Lovaina, o el LANGSNAP (Tracey-Ventura *et al.*, 2013) –este último para la interlengua del español–, que es una fuente reciente de datos longitudinales de entrevistas a aprendientes que estudiaron español en el extranjero. Los corpus longitudinales, al poder aportar información valiosísima acerca de la adquisición de la L2, son de gran interés para los investigadores, que intentan cubrir la carencia de corpus longitudinales por medio de la creación de corpus de aprendientes de diferentes edades o niveles (de inicial a avanzado), de manera que la estructura del corpus se asemeje a la de un corpus longitudinal (Guo, 2006: 17). Este tipo de corpus se conoce como *quasi-longitudinal* (Granger, 2002: 11), y el CEDEL2 es un buen ejemplo al contener muestras de lengua de HNN de español de nivel inicial, intermedio y avanzado. No obstante, la mayoría de estudios realizados hasta ahora están basados en corpus de aprendientes sincrónicos, también llamados *transversales* –como es el caso de nuestro estudio, centrado únicamente en el subcorpus del CEDEL2 perteneciente a los hablantes de nivel avanzado–; estos corpus permiten estudiar tendencias en un grupo de HNN, aunque son los longitudinales los que guardan una relación más estrecha con la investigación en ASL, ya que la principal ocupación de esta disciplina hasta la fecha ha sido la naturaleza del proceso de adquisición de la segunda lengua y los factores que afectan a la lengua de los aprendientes (Larsen-Freeman y Long, 1991).

3) Análisis de corpus cuantitativo y cualitativo. Los corpus son una fuente de datos cuantitativos para los lingüistas, por lo que a menudo sintetizan sus hallazgos cuantitativos a través de estadísticas (McEnery y Hardie, 2012: 48). Es posible usar un corpus y no recurrir a análisis estadísticos, como cuando queremos establecer si un fenómeno particular existe en la lengua. Nos basta con recuperar una sola aparición de uso para probar que existe, aunque la no aparición en un corpus no implica la no existencia. En cambio, si lo que nos interesa es determinar si el fenómeno en cuestión es frecuente en la lengua, hay que emplear estadísticas. La frecuencia en los datos se produce con tanta regularidad en los análisis de corpus que es raro dar con un estudio en LC en el que no se lleve a cabo algún análisis estadístico (McEnery y Hardie, 2012:

49), aunque este sea relativamente básico y descriptivo, como es el caso de nuestra investigación cuando queremos descubrir la frecuencia con la que los HN y los HNN usan el verbo *dar* o cada una de sus estructuras y significados; o averiguar el índice de errores que cometen, y en qué estructuras son más frecuentes. No obstante, cualquier estudio sobre la lengua del aprendiente resultaría superficial si el único análisis realizado consistiera en la recuperación de datos de manera mecanizada. Como observa Meunier (1998: 36), las diferencias o semejanzas superficiales entre la lengua nativa y no nativa siempre requieren una investigación posterior, por lo que los datos obtenidos de las búsquedas informatizadas suponen el punto de partida para el análisis ulterior, que ha de ser cualitativo (aspecto señalado en el apartado 2.2) Esto no siempre sucede: como señala Muñoz Liceras (2009), en muchos casos se ha querido primar una investigación cuantitativa en el sentido más literal y menos útil del término.

La investigación de la adquisición del español L2 basada en corpus está aún en una fase embrionaria (Lozano, *en prensa*), y el léxico es de los niveles menos analizados en la interlengua, pese a que está muy extendida la idea de que la interlengua se caracteriza por un uso menos profundo y extenso del vocabulario.

En relación con ello, en inglés existen algunos trabajos, como los de Chi *et al.* (1994) o Ringbom (1998a), que informan del uso y la distribución de las palabras, y de los diferentes sentidos de las palabras polisémicas; en cambio, la mayoría de los estudios sobre el léxico se limitan a investigar la cantidad o extensión de vocabulario, lo que resulta peligroso dado que ello esconde el problema de si ese vocabulario se conoce en profundidad o no, especialmente las palabras polisémicas. Estas investigaciones –aunque nos informan de si los HNN deben ampliar el tamaño de su vocabulario para mejorar su competencia lingüística en un nivel de dominio dado– no hacen referencia a si los HNN deben ampliar o mejorar el uso de las formas que parecen o creen conocer. Altenberg y Granger (2001) o Guo (2006) marcan una diferencia –con respecto a la mayoría de estos trabajos– en los estudios de interlengua del inglés, pues se dedican al análisis profundo de los verbos *make*, *take* y *keep*. Sobre la interlengua del español no existe este tipo de estudios, está prácticamente todo por hacer. En este aspecto es en el que nuestra investigación cobra valor, pues realiza un examen sobre la profundidad del conocimiento de un aspecto del vocabulario no estudiado hasta ahora en la interlengua del español, el del verbo frecuente *dar*. Guo (2006: 42) afirma que “it is essential for researches to know how much of a word has become a part of the learners’ English and how much is yet to be learned and used by the learners”. De esta manera, los resultados serán significativos desde una perspectiva teórica –en el conocimiento sobre el proceso de adquisición del vocabulario– y aplicada o didáctica.

Otros enfoques que distinguen los estudios de CAI están relacionados con la posibilidad de contar o no con un número de variedades de la misma interlengua

representadas en el corpus o con la posibilidad de disponer de textos comparables en dos interlenguas diferentes. Los primeros corpus constituyen la base de datos de los estudios contrastivos de interlengua, como ya se ha señalado; el ICLE es uno de los más representativos, ya que ofrece comparaciones entre 16 interlenguas del inglés, pues contiene textos escritos en inglés por hablantes de 16 L1 diferentes. Este enfoque permite describir la interlengua en un nivel de dominio determinado atendiendo a los factores interlinguales y a los intralinguales del aprendizaje, lo que posibilita un desarrollo enorme de los estudios sobre el proceso de adquisición de la L2 (en este caso, del inglés). El CEDEL2 no cuenta con un subcorpus comparable de esta misma interlengua producida por hablantes de otra L1 que no sea el inglés, ni tampoco el SPLLOC, por lo que no se pueden realizar análisis de esta naturaleza en ninguno de los corpus más representativos de HNN de español. En cambio, el CEDEL2 sí que cuenta con un corpus comparable de una interlengua diferente, como es el caso del WriCLE. La comparación entre el CEDEL2 y el WriCLE permite determinar si las particularidades o déficits en una interlengua concreta (del español o del inglés) resultan de la influencia de la L1, si es consecuencia del *input* o de patrones de desarrollo universales (Lozano y Mendikoetxea, 2013: 6). El WriCle se encuentra todavía en fase de compilación de datos, por lo que queda pendiente su utilización en futuras investigaciones.

En suma, el estudio de CAI es ahora un campo establecido de la lingüística aplicada, un campo que está en continua evolución. Aunque el ACI se ha convertido en un paradigma fructífero, en el caso de la interlengua del español se está iniciando ahora su desarrollo. No obstante, todavía quedan muchas cuestiones metodológicas por resolver para que los investigadores de ASL se beneficien plenamente de los estudios de CAI.

3. LA REVALORIZACIÓN DEL LÉXICO Y LA COMPETENCIA LÉXICA DEL VERBO

La idea que en la actualidad se tiene sobre el conocimiento léxico es la de una representación verbal compleja que, por un lado, comprende una amplia gama de ítems léxicos que abarcaría tanto las unidades monoverbales, las pluriverbales (compuestos) y las secuencias formulaicas, esto es, “word strings that can be retrieved from memory more quickly than they can be composed word by word on the basis on one’s knowledge of grammar” (Boers y Lindstromberg, 2009: 10); Schmitt (2000: 2) y Nation (2001: 47) también consideran las familias de palabras como unidades psicológicas que se integran en el conocimiento léxico. Por otro lado, el conocimiento léxico –igualmente complejo (Nation, 2001: 26)– implica varios aspectos, tales como el dominio receptivo y productivo de la forma, el significado, la función gramatical, las asociaciones y colocaciones, así como el registro y la frecuencia, aspectos o dimensiones que completan de alguna manera la definición de *competencia léxica* recogida en el *Marco común europeo de referencia para las lenguas: aprendizaje, enseñanza, evaluación (MCER)*⁴⁴: “el conocimiento del vocabulario de una lengua y la capacidad para utilizarlo” (2002: 108)⁴⁵. De todo ello surgen tres características fundamentales del ítem léxico: la forma, el significado y el uso de esta entidad de doble cara (con forma y significado). La última característica es la más abarcadora, dado que presupone o engloba las dos anteriores, a las que habrían de sumarse además las señaladas por Nation.

Desde que en los años 80 Canale y Swain aplicaran a la enseñanza de LE el concepto de *competencia comunicativa*, por medio del cual Hymes (1972) reformulaba la competencia lingüística de Chomsky, el léxico –componente hasta entonces desatendido en las aulas de lenguas– se convierte en un elemento imprescindible para lograr el objetivo de enseñanza de la entonces recién nacida metodología comunicativa: que el aprendiente, como agente social (concepto fundamental del *MCER*), sea capaz de desenvolverse con éxito en las situaciones

⁴⁴ Tomamos este documento del Consejo de Europa como referencia en nuestro trabajo porque está concebido como orientación y modelo para todos los aspectos involucrados en el proceso de enseñanza y aprendizaje de L2.

⁴⁵ Por razones estilísticas, los términos *léxico* y *vocabulario* son empleados en este trabajo como equivalentes, coincidiendo con las definiciones del *Diccionario de términos clave de ELE* (12-06-13). No obstante, somos conscientes de que, en sentido riguroso, no denotan lo mismo: cada uno se relaciona con un plano diferente, el de la lengua en su sistema (*langue*) y el del uso de la lengua (*parole*) (Tréville, 2000: 27). En sentido estricto, en nuestro análisis de corpus nos ocuparemos más del vocabulario, ya que trataremos de establecer y delimitar el conocimiento que del verbo *dar* tiene un grupo determinado.

habituales de comunicación que se dan en la interacción social. Por consiguiente, el léxico terminará disponiendo de un dominio de conocimiento propio, susceptible de ser acotado y caracterizado de manera precisa, como muestran los trabajos de Richards (1976), Nation (2001) y Baralo (2006) que intentan delimitar el saber implicado en el conocimiento de una palabra. Nuestra propuesta pretende contribuir en esa dirección: adopta los principios y contenidos del *MCEER* y adapta los presupuestos de Richards (1976), Nation (2001) y Baralo (2006) con el fin último de servir de herramienta útil a los estudios de interlengua, en los que tanto el conocimiento léxico en general como el del verbo en particular ocupan un lugar central, pues constituyen la base del desarrollo comunicativo.

Para acometer esta tarea, se parte de una sucinta revisión historiográfica (3.1) sobre el papel del léxico en las teorías lingüísticas, psicolingüísticas y pedagógicas, fundamentalmente durante las últimas cinco décadas –etapa decisiva en la revalorización de este componente en la que el modelo de gramática centrado en las reglas y en la jerarquía estructural cede paulatinamente el paso a otros modelos basados en la selección léxica fruto, a su vez, del giro lingüístico hacia el funcionalismo comunicativo–.

Así las cosas, si durante las tres cuartas partes del siglo XX saber y aprender vocabulario equivalía únicamente a conocer la forma y el significado de la palabra para poder construir oraciones gramaticalmente correctas, en el último cuarto de siglo –encontrado en un marco comunicativo– este conocimiento es absorbido por un concepto más abarcador, el de uso, que se convierte en el nuevo foco de análisis tanto en las teorías lingüísticas como en las de adquisición. Gracias al conocimiento sobre el uso del léxico, los enunciados producidos son ya no solo correctos desde una perspectiva gramatical, sino también adecuados al contexto, a la situación comunicativa.

A lo largo del breve recorrido historiográfico que presentamos a continuación van recogiendo los rasgos que desde un primer momento integraron el conocimiento léxico, así como algunos de los más destacados que progresivamente –gracias a los desarrollos en semántica léxica, lingüística del texto, análisis del discurso y lingüística de corpus– se han ido incorporando a esta competencia. Recuperamos los que consideramos más relevantes en la conformación del panorama actual. Estos aspectos son abordados por separado en la segunda parte de este capítulo (3.2.), que desemboca en el desglose de los contenidos que conforman la competencia léxica del verbo en la producción escrita.

3.1. LA REVALORIZACIÓN DEL LÉXICO

Como se ha señalado, entrado el último tercio del siglo XX el léxico todavía era un componente lingüístico menospreciado en el campo de la ASL (Richards, 1985: 176). Este hecho coincide con el escaso papel que, a su vez, las teorías lingüísticas habían concedido a este componente; en consecuencia, esta situación se trasladaba al aula de lenguas, donde la enseñanza del léxico quedaba relegada a un segundo plano y escaseaban los manuales que incorporaran de manera sistemática su tratamiento; y en el caso de que se ejercitara, su enseñanza se limitaba a los listados de palabras aisladas que habían de ser memorizadas para que sirvieran al fin último de la construcción de oraciones bien formadas gramaticalmente. Ciertamente, este es el caso del Método Gramática-Traducción (1840-1940), en el que el único modo de acercamiento explícito al léxico lo constituyen glosarios acoplados al texto como ayuda para la traducción. En general, no se facilita ayuda a los aprendientes para que retengan el vocabulario y puedan recurrir a él durante el uso activo de la lengua. Asimismo, bajo los presupuestos del Método Audio-oral (1940-1970), el léxico no sale mejor parado. Dado que se concede prioridad a la fluidez y precisión, se trabaja con la memorización de diálogos seguidos de una ejercitación de estructuras intensiva y prolongada; estas prácticas reflejan la infravaloración de la comprensión y del procesamiento cognitivo, por lo que se ofrecen escasas técnicas y estrategias que promuevan la retención del léxico en LE. De hecho, la causa del fracaso de este método radica en su dependencia de los ejercicios orales basados en la sustitución, lo que impide una asociación constante entre las formas léxicas y los significados particulares (Segalowitz, 2000: 212, citado en Boers y Lindstromberg, 2008a: 3).

A partir de los años 70 el léxico comienza a pasar de ser el elemento marginado hasta llegar a ocupar a finales de siglo XX y comienzos del siglo XXI un lugar central en las teorías pedagógicas gracias al enfoque comunicativo. La conocida cita de Wilkins (1972: 111) –a la que ya nos hemos referido en el capítulo 1–, “Without grammar very little can be conveyed; without words, nothing can be conveyed” –de la que destacan por su representatividad las palabras *word* y *convey*–, marca el comienzo de un cambio de tendencia. Y apenas una década después, la idea manifestada por Allen (1983) de que no se puede desdeñar el valor del vocabulario como elemento vertebrador del pensamiento (función simbólica) ni como requisito imprescindible para la interacción social (función comunicativa) era comúnmente aceptada.

¿Qué está sucediendo en el panorama lingüístico durante estos años para que se produzca esta transformación en el valor del léxico que acabará revolucionando el panorama de la enseñanza de lenguas? En las secciones 3.1.1. y 3.1.2 se presentan algunos de los factores determinantes.

3.1.1. *Algunas contribuciones de la teoría lingüística*

Apuntábamos en el apartado anterior que hasta no hace mucho tiempo la gramática constituía el elemento vertebrador del conocimiento lingüístico; era esencial ensamblar el esqueleto primero, de modo que las palabras que hubieran de ser aprendidas pudieran ser encajadas en el lugar adecuado. Esta manera de proceder refleja la tradición chomskyana, que fundamentaba el dominio de la lengua en una dicotomía: el control de reglas gramaticales, por un lado, y el dominio de unidades léxicas⁴⁶, por otro.

Pese a que esta concepción del lenguaje estaba muy extendida en los años setenta, se asientan en esta época las aportaciones de lingüistas funcionalistas británicos como Halliday o Firth, las de sociolingüistas americanos como Hymes, así como las del ámbito de la filosofía del lenguaje, con Austin y Searle. Estos autores, entre otros, configuran un nuevo marco teórico –que supone una ruptura con el modelo anterior– en el que se destaca la dimensión pragmática de la lengua, y se reorganiza entonces el conocimiento lingüístico en torno a criterios semánticos y nociofuncionales.

Así, en este nuevo contexto teórico y con la ayuda de las nuevas herramientas informatizadas de análisis de datos (los corpus), surgen nuevos planteamientos funcionalistas que ponen en cuestión los presupuestos formales del generativismo tan extendidos entonces.

En primer lugar, se defiende que para poder constituir enunciados correctos (adecuados) y naturales, esto es, comunicativos, la sintaxis no puede ser independiente ni del léxico ni de la semántica y, además, tampoco puede obviar el contexto situacional (Dik, 1997)⁴⁷, puesto que este último factor determina la adecuación al registro, la frecuencia y, en definitiva, la corrección y la naturalidad

⁴⁶ Conviene aclarar que en este trabajo se emplea el término *unidad léxica* para referirnos tanto a las palabras gráficamente simples como a las expresiones complejas formadas por varias palabras que, sin embargo, funcionan de forma unitaria a efectos léxicos, por lo que respetan el principio de composicionalidad semántica.

⁴⁷ El germen de esta idea se encuentra en la gramática sistémico-funcional de Halliday (1977).

de un uso particular de una palabra⁴⁸. En este nuevo paradigma de análisis el sociolingüista norteamericano Hymes (1972) acuña la expresión *competencia comunicativa*, el nuevo objeto de la descripción lingüístico-teórica, entendida como la capacidad del usuario natural del lenguaje no solo para construir e interpretar expresiones lingüísticas (tal y como se define la competencia lingüística de Chomsky), sino también para usar esas expresiones de forma apropiada y efectiva según las convenciones de interacción verbal predominantes en una comunidad lingüística. Este nuevo concepto se ha aplicado a la enseñanza de LE por Canale y Swain (1980), en lo que constituye los cimientos del enfoque comunicativo en la enseñanza de lenguas⁴⁹, vigente todavía hoy.

Los postulados anteriores encuentran confirmación en las evidencias que proporcionan los análisis de corpus (Sinclair, 1991).

En este sentido, Biber *et al.*, autores de la obra *Longman Grammar of Spoken and Written English* (1999), observan en su introducción:

Syntax and lexicon are often treated as independent components of English. Analysis of real texts shows however, that most syntactic structures tend to have an associated set of words or phrases that are frequently used with them.

Pese a la novedad de los conceptos de ‘coaparición’ y ‘frecuencia’ en la cita, el origen del interés por las coapariciones frecuentes puede rastrearse incluso muchos años antes, hacia finales del siglo XIX, cuando el lingüista polaco Kruszewsky (1887: 172, citado en Koerner, 1982: 233) defendía que las palabras estaban relacionadas en la mente humana por medio de dos leyes de asociación, a saber: la de similitud y la de continuidad; la primera conecta palabras por medio de similitud formal y semántica, y la segunda por la coocurrencia frecuente en una serie sintáctica. Igualmente, las propuestas de Palmer no estaban tan alejadas de estos presupuestos cuando en 1933 planteó que las *coapariciones frecuentes* de palabras debían ser enseñadas a los aprendientes de lenguas extranjeras como “an

⁴⁸ Estas consideraciones dejan entrever la necesidad de ampliar los límites del análisis lingüístico, relegado hasta el momento a la oración; como es bien sabido, esto, junto con la constatación de la importancia de los elementos contextuales e interactivos en el estudio de determinadas unidades, provoca el nacimiento de la Lingüística del Texto, la Pragmática, el Análisis del Discurso o el Análisis de la Conversación.

⁴⁹ Como mantiene Pastor Cesteros (2005: 381), no hay que ver el enfoque comunicativo únicamente como un producto de este cambio teórico lingüístico, pues en él también influyeron otros factores pedagógicos y socioculturales.

integral whole or independent entity, rather than by the process of piecing together their component parts” (citado en Stengers, 2009: 6)⁵⁰.

En los años cincuenta reaparecen parcialmente estos conceptos cuando la semántica léxica estructural europea analiza las palabras a través de las complejas, pero perfectamente imbricadas, relaciones sintagmáticas y paradigmáticas. Son estas últimas las que conforman una red de relaciones de oposición y equivalencia que permite definir el significado de los términos, aspecto que culminará en la teoría de los campos léxicos (Pottier, 1963; Coseriu, 1977). Pese a estos avances acerca de las relaciones léxicas, ni el comportamiento colocacional de las palabras ni la selección léxica despertaron el interés de los generativistas; no obstante, estos trabajos, y otros similares como el de Firth (1957: 11), de quien proviene la conocida cita “you shall know a word by the company it keeps”, recobran interés años después. Este trabajo confiere un papel clave en su teoría del significado a la colocación como fenómeno lingüístico, pues la coaparición habitual de lexemas individuales o “the distribution of common words” (Firth, 1957: 196) permite analizar qué determina que una palabra en un contexto dado actualice una acepción u otra; dicho de otro modo, el significado de una palabra está determinado por las otras palabras con las que se combina⁵¹.

⁵⁰ La expresión *coaparición frecuente* se asimila normalmente al concepto de *colocación*, considerado inestable ya que son múltiples las clasificaciones y definiciones en torno a esta cuestión, lo que genera “una gran confusión en la bibliografía” (Alonso Ramos, 1993: 165). Las referencias a este concepto suelen incluir las combinaciones frecuentes y clichés, que no se sustentan necesariamente en bases semánticas restringidas. Sobre este concepto, véase la n. 51.

⁵¹ Halliday (1966) y Sinclair (1966), dentro del marco de la escuela sistémica británica, desarrollaron la idea de Firth y trataron de probar que la colocación era una parte importante en la formación de significado. Precisamente esta escuela parte del papel fundamental que Firth otorga a la colocación como fenómeno lingüístico en su teoría del significado. Hasta ahora, todos los acercamientos para demostrar esta hipótesis y aplicar el concepto de colocación a la descripción de la lengua se habían quedado en la mera especulación; así que, en 1963, en el proyecto descrito en *The OSTI Report* se inició la primera investigación empírica sobre la colocación, la primera aplicación de este concepto a la lexicografía computacional. La definición del término tenida en cuenta en estos primeros trabajos era la de Jones y Sinclair (1974: 19): “‘Collocation’ is the co-occurrence of two items in a text within a specified environment. ‘Significant collocation’ is regular collocation between items, such that they co-occur more often than their respective frequencies, and the length of text in which they appear, would predict”. En los trabajos realizados en esta escuela británica se tuvieron en cuenta las características colocacionales de una palabra para la desambiguación de homógrafos y términos polisémicos, así como para explicar la diferencia entre una palabra empleada en una locución y la misma palabra usada fuera de esta. No obstante, pese a que de los programas de concordancias se obtiene una información valiosísima, este enfoque ha recibido numerosas críticas (véase Alonso Ramos, 1994-1995; Corpas Pastor, 2001; Koike, 2001; y Bosque, 2001, entre otros trabajos), precisamente por su carácter exclusivamente estadístico. Los resultados de los programas informáticos de concordancias no constituyen en sí mismos estudios científicos del lenguaje si a aquellos no les sigue un análisis cualitativo de los datos. Si entendemos que todo término como objeto de estudio de la lingüística ha de designar un concepto propiamente lingüístico, el criterio de la coocurrencia frecuente –al que se refiere la anterior definición– resulta insuficiente para establecer la existencia de una colocación, ya que no es posible dar un contenido

Estos aspectos sintagmáticos, que toman en consideración la asociación de significado y construcción o sentido y sintaxis, son fundamentales en algunos de los modelos de descripción lingüística que se han desarrollado en las últimas décadas como *Construction Grammar* (Fillmore *et al.*, 1988), *Pattern Grammar* (Hunston y Francis, 2000) y *Linear Unit Grammar* (Sinclair y Mauranen, 2006). Incluso Chomsky, que en sus primeras etapas tenía como centro de su descripción lingüística la sintaxis, ha adoptado recientemente una postura más lexicalista en su programa minimalista, por lo que presta más atención al léxico y a la manera en la que este se formatea, codifica y organiza (Chomsky, 1995).

Desde una perspectiva similar, otros generativistas, como Jackendoff (2003), proponen una relación simétrica entre sintaxis y léxico. Y es que, como recoge Lezcano (1994-95), a partir de Tesnière y su gramática de valencias (1959), poco a poco ha ido calando la idea de que la arquitectura del verbo es el elemento central que determina la estructura básica de toda oración. Asimismo, la propuesta de Fillmore en su teoría de la Gramática de Casos (1976) se mueve en ese ámbito de revalorización del léxico, en concreto, de la semántica del verbo a la hora de motivar el fenómeno sintáctico y morfológico.

En la lingüística cognitiva se va más allá en el valor concedido al léxico (Langacker, 1987; y en relación con la adquisición de L1 Tomasello, 2003), al proponer que la lengua en uso (del inglés *language-in-use*) consiste en la memorización de ítems simbólicos que varían en forma –desde el morfema a expresiones idiomáticas– y en sustantividad –desde formas plenamente especificadas hasta esquemas más abiertos–; aquí, la sintaxis es un fenómeno emergente que se va desarrollando a partir de la analogía y la generalización.

homogéneo comprensible conceptualmente en términos teóricos al variado grupo de fenómenos que comparten este criterio con el que se identifica el término. Es llamativo que en la enseñanza de LE se adoptara el término colocación con la definición original de la escuela sistémica (véase Lewis, 1993, 1997, 2000), cuando en el momento en que aparecieron los enfoques que introdujeron la enseñanza de este concepto ya se estaba cuestionando y debatiendo la ausencia de criterios semánticos en la caracterización del concepto. Precisamente la aceptación de estos criterios es lo que transformó la escuela sistémica en el nuevo contextualismo británico. Estos trabajos destacan que no todas las coapariciones frecuentes son de la misma naturaleza (véase Sinclair *et al.*, 2005): mientras que en unas apenas se da restricción semántica (colocaciones ascendentes), en otras sí se observa una determinación semántica (colocaciones descendentes). Aun así, la frecuencia continúa siendo el criterio del que parte este análisis. Bosque, en un intento de definir y categorizar el concepto de colocación respondiendo a criterios propiamente lingüísticos, lo incluye como una subclase de la noción de *selección léxica* (véase Bosque, 2001 y 2005 [2004]), que es con la que nosotros trabajaremos en el apartado 3.2 al establecer una distinción entre los diferentes tipos de combinaciones léxicas en los que participa el verbo. Para una revisión sobre el concepto de colocación, véase Sánchez Rufat (2010).

3.1.2. Algunas contribuciones de la teoría psicolingüística

En relación con estos planteamientos teóricos basados en el papel esencial del vocabulario en la descripción lingüística, la importancia de las unidades léxicas también ha sido reivindicada en la investigación sobre la adquisición tanto de la L1 como de la L2; véase Cruttenden (1981), Peters (1983), Ellis (1994) o Wray (2002). Acorde con la transformación señalada en el panorama lingüístico general, las siguientes palabras de Krashen y Terrell (1983: 155) anuncian este cambio en los estudios de adquisición:

Vocabulary is also very important for the acquisition process. The popular belief is that one uses form and grammar to understand meaning. The truth is probably closer to the opposite: we acquire morphology and syntax because we understand the meaning of utterances.

Cuando los niños adquieren su L1, están expuestos a un continuo fónico en el que tratan de discriminar unidades recurrentes –que no son necesariamente siempre palabras, dado que no distinguen los límites entre estas– y determinar su significado a partir del contexto en el que aparecen (Boers y Lindstromberg, 2009: 29). Wray (2002) distingue entre la adquisición infantil del léxico de una L1 –que se produce mediante un almacenamiento holístico de secuencias de palabras (o secuencias formulaicas)– de la adquisición de una L2 en jóvenes o adultos, que se corresponde con un modelo más bien analítico, por lo que tiende a procesar esos bloques de palabras frecuentes (incluso los que tienen un sentido relativamente unitario) como secuencias de palabras aisladas, bloqueando o dificultando su adquisición como unidad (multipalabra). Ello no impide que el adulto pueda guardar almacenados desde su niñez gran cantidad de bloques formulaicos en su L1, y que incluso pueda aprender con posterioridad algunos nuevos por medio de este procedimiento, aunque no es lo más usual. No obstante, hay quienes, como Boers y Lindstromberg (2009: 31), no ven tanta diferencia entre la adquisición de los bloques prefabricados en L1 y en L2, por lo que, aunque el procesamiento sea más analítico en los adultos, no se detiene en absoluto el aprendizaje de estos bloques⁵².

⁵² El trabajo de Prendergast titulado *The mastery of languages; or, The art of speaking foreign tongues idiomatically* (1864) destaca –de entre toda la bibliografía consultada– como el primer intento de elaborar una teoría psicolingüística sobre el modo en que los niños adquieren la L1 y su aplicación a la enseñanza de L2 que concede al léxico un estatus extraordinario. En su propuesta se esboza el concepto de secuencias *prefabricadas* que evolucionará hasta dar lugar a lo que hoy conocemos como *lengua formulaica*.

Estas diferencias en cuanto a los modelos de adquisición del vocabulario entre niños y adultos ponen de manifiesto el desconocimiento que todavía existe en lo que se refiere al modo en que funcionan nuestras capacidades cognitivas, pese al reciente avance en el ámbito de las ciencias neurológicas. No obstante, sí que es comúnmente aceptado que tanto niños como adultos tratan de segmentar en unidades discretas, en lo que comprende la primera fase de la adquisición lingüística, el continuo fónico al que están expuestos para así determinar el significado de estas a partir del contexto, por lo que el cerebro las percibe en términos de información semántica al tiempo que aprecia su forma fonológica. De acuerdo con esto, el vocabulario desempeña un papel importante en el proceso de adquisición de una lengua; es el significado que se adquiere a partir del contexto.

Baralo (2001) –sirviéndose de los presupuestos que la teoría generativa en su etapa más reciente presenta para la descripción de la competencia nativa– lleva a cabo un interesante trabajo sobre la organización del lexicón mental⁵³ en el que, a partir de la hipótesis de que la organización y funcionamiento del lexicón es cualitativamente semejante en L1 y en L2⁵⁴, demuestra que existe un vínculo muy directo también en la interlengua entre saber vocabulario y saber construir oraciones en las que aparecen las palabras que lo integran.

En lo que se refiere a la mencionada primera fase en la adquisición de la lengua, Baralo (2001: 14-15) aclara que el reconocimiento de una palabra en el continuo fónico y el dominio de su pronunciación constituyen un proceso de aprendizaje diferente al de la apropiación del concepto que expresa. Para lo primero, se requiere escucharla varias veces y reconocerla como unidad diferente de otras; pero para apropiarse de su significado se pasa al menos por dos etapas, una rápida y otra mucho más lenta. Como señala la autora, la primera vez que el niño reconoce una nueva realidad y la palabra que la nombra –por ejemplo, un nuevo objeto y su nombre–, asigna esa palabra a una clase semántica amplia –por ejemplo, a la categoría fruta–. La etapa siguiente implica desarrollar la

⁵³ En este trabajo adoptamos precisamente la definición que esta autora ofrece del término *lexicón*: “la parte de la competencia lingüística que contiene las piezas léxicas o formantes, es decir, las raíces y temas, los afijos flexivos y derivativos, y las reglas que regulan su combinación. El concepto es cognitivo, dinámico, procesual, a diferencia del concepto de “léxico”, como sinónimo de ‘vocabulario’, entendido como un simple listado de palabras, o como una organización de campos semánticos” (Baralo, 2001: 12). Es, en definitiva, un modelo abstracto que representa la manera en que se almacenan y recuperan las unidades léxicas; contiene las unidades y las reglas que disponen la manera en la que dichas unidades se combinan, y las asociaciones entre ellas.

⁵⁴ Aunque aclara que “obviamente, en el caso de apropiación de las lenguas no nativas, estará siempre presente el conocimiento de la LM [lengua materna] y de las otras lenguas extranjeras aprendidas, disponible para transferir sus bases léxicas”.

especificidad del término nuevo dentro de su clase semántica; se corresponde con la elaboración de las diferencias de rasgos y relaciones que existen entre esos términos que pertenecen a la misma clase. Así, en un proceso lento y gradual, las entradas léxicas se van especificando en sus rasgos semánticos y van ajustando sus redes relacionales con las otras entradas léxicas semánticamente próximas.

Ahora bien, una vez que el niño ha aprendido una palabra –lo que en este estadio implica saber cómo suena (competencia fónica), qué significado señala la forma de la palabra y qué está incluido en el concepto (competencia semántica), la almacena en el lexicón –tejiendo redes léxicas semánticas y formales que facilitan tanto el almacenamiento de las palabras como la recuperación– con la información categorial adecuada para aplicar las reglas correspondientes –supone saber qué partes se reconocen en la palabra, en qué estructuras podría aparecer y qué otras palabras aparecen con ella, se corresponde con una competencia gramatical, en concreto morfológica, sintáctica y léxico-sintáctica–. Estas fases se corresponden con las tres tareas interconectadas –etiquetado, empaquetado y construcción de redes que ya había señalado Aitchison (1994)–: en el proceso de etiquetación se crea una conexión entre el concepto, la imagen acústica del signo y el referente; a continuación se comprueba qué conceptos se agrupan (empaquetan) dentro de la misma etiqueta y si la pieza léxica tiene matices dialectales, pragmáticos, sociolingüísticos y metafóricos; este segundo proceso requiere exposición a la lengua e interacción con hablantes nativos. Mientras estos dos procesos se desarrollan, comienza, de manera paralela, una tercera etapa: la construcción de redes que va estableciendo las relaciones de sentido entre las palabras. Esta manera de organizarse el lexicón se conoce como *la metáfora o modelo de la red* (de Collins y Quillian, citado en Aitchison, 1994): las palabras son los nudos que se intercomunican por medio de numerosos hilos, estableciendo redes de relaciones extremadamente complejas. A las redes léxicas dedica Baralo varios trabajos (2006 y 2007), y define el concepto como “el conjunto de elementos interrelacionados y de procesos combinatorios, de manera que el resultado de tales procesos es algo más complejo y diferente que la mera suma de sus componentes” (Baralo, 2007: 384). Muchos investigadores defienden la idea de que nuestro lexicón mental se encuentra altamente organizado de formas muy diferentes en redes interrelacionadas (y, de hecho, uno de los mejores diccionarios combinatorios del español, el de Ignacio Bosque, se titula *Redes*): existen asociaciones fonológicas, morfológicas (derivación y composición), discursivas, semánticas (hiponimia e hiperonimia, sinonimia y antonimia, metonimia, polisemia, campos semánticos y asociativos), categoriales, léxicas y léxico-sintácticas (cognados, expresiones idiomáticas, coapariciones

frecuentes), pragmáticas (elección léxica determinada por las variables de la situación comunicativa), psicológicas (elementos que se organizan en torno al prototipo), personales o afectivas, socioculturales, intralingüísticas e interlingüísticas. Evidentemente, ciertas asociaciones son más fuertes que otras. En este sentido, diversos test de asociación de palabras han demostrado que las redes colocaciones y de coordinación (léxicas y léxico-sintácticas) son las más fuertes (Nagini, 2008: 184).

Por lo tanto, como decíamos al abrir este capítulo, saber una palabra es conocer, por un lado, su forma y el significado, y, por otro lado, saber usarla. Así, si el niño aprende un verbo, como en el caso de *dar*, podrá aplicar las reglas de flexión propias de los verbos de la primera conjugación (aunque en el proceso de adquisición de este verbo intervienen ciertas irregularidades); de esta manera, el niño asigna el mismo significado a formas con rasgos propios de persona, número, tiempo, modo y aspecto, como *daría, daremos, dabais, dieron, habías dado*, etcétera. De igual modo, al categorizar el verbo como una acción, se adquiere la estructura léxico conceptual correspondiente, que tiene toda la información necesaria para proyectar la estructura sintáctica de la oración. Si se trata del verbo *dar*, será capaz de reconocer el verbo como una acción transitiva que contiene tres argumentos en su estructura léxico conceptual: el agente, el tema y el destinatario. Saber usar este verbo implica saber proyectarlo en el eje sintagmático y construir oraciones en las que el agente se proyecte como el sujeto de la oración, el tema como el complemento directo (CD) y el destinatario como el complemento indirecto (CI) (Baralo, 2001: 11).

De acuerdo con estos datos, Baralo (2001: 18-23) ofrece una hipótesis de funcionamiento de lexicón en L1 y en L2 en un marco psicolingüístico⁵⁵. Las propiedades del lexicón vienen impuestas por una dotación genética especificada para el lenguaje humano, conocida como Gramática Universal (GU), que funciona de una manera computacional con dos tipos de elementos: las unidades léxicas y sus reglas combinatorias. Disponemos de un procesador léxico que interrelaciona dos dominios distintos: el representacional y el computacional (Marslen-Wilson, 1989, citado en Baralo, 2001: 18). Esto quiere decir que vincula, por un lado, el continuo fónico que percibimos por los sentidos con el sistema de procesamiento léxico (lo que implica competencia fónica y semántica), y, por otro lado, conecta este sistema con la representación de la información semántica y sintáctica asociada con la palabra analizada (lo que

⁵⁵ El tratamiento del error en nuestro análisis (apartado 6.2.4) está fundamentado en la capacidad descriptiva de este modelo.

supone una competencia léxico-sintáctica). El hablante reconoce en el lexicón relaciones entre la palabra y otras unidades léxicas por su parecido fónico y semántico, por la regularidad en los patrones de formación. El conocimiento intuitivo del nativo, tácito y operativo, le permite distinguir el primitivo de la palabra derivada. Y más aún, la competencia morfológica le permite percibir las relaciones jerárquicas que se dan en el interior de una palabra, esto es, las relaciones de dominio y de precedencia, el orden en la incorporación de afijos en el proceso de derivación de una palabra, y el tipo de relación que mantienen con la base. Este mismo conocimiento morfológico le permite saber qué palabras son posibles y cuáles no podrían serlo en su lengua.

Parece ser, de acuerdo con los resultados de investigaciones realizadas por Baralo (2001: 20), que el procesamiento léxico depende fundamentalmente de señales e interconexiones semánticas, y que en el proceso de construcción del lexicón en la interlengua la L1 influye en el procesamiento de la L2 (todos los subsistemas lingüísticos –fónico, morfológico, etcétera– de la L2 se interpretan a través de los parámetros ya fijados)⁵⁶. La transferencia⁵⁷, aún así, está condicionada por otros procesos: la preferencia por el uso de la forma menos marcada, por lo que los rasgos marcados de la L1 no se transfieren a la L2, la sobregeneralización de las reglas que tienen más capacidad generativa en la L1, la hipercorrección o la neutralización. Pese a todos estos procesos que intervienen en el procesamiento léxico de la interlengua, Baralo (2001: 23-34) demuestra que el conocimiento léxico –que permite la creatividad léxica– en L2 se construye por medio de las mismas reglas de formación de palabras y de buena formación conceptual⁵⁸ que la teoría lingüística generativa ha propuesto para la descripción de la competencia nativa.

La visión que se tiene del conocimiento del vocabulario es la de un constructo complejo que implica un número de dimensiones diferentes que

⁵⁶ Los trabajos de Laufer (1994) o Singleton (2000) ponen en evidencia estos mismos resultados.

⁵⁷ La transferencia se caracteriza por construir hipótesis sobre el funcionamiento de la L2 a partir del conocimiento nativo de la L1 y sus conocimientos previos del mundo.

⁵⁸ La autora se refiere a una “GU de conceptos” que nos permite construir conceptos léxicos a partir de un conjunto finito de primitivos mentales guiados por las Reglas de Buena Formación Conceptual (conjunto finito de principios combinatorios), sobre las que se basa el aprendizaje. Frente a esta idea, la semántica cognitiva postula que creamos los conceptos a partir de las categorías perceptivas (esquemas de imágenes que abstraemos de nuestra experiencia corporal cotidiana), un proceso que llamamos *abstracción*, y mediante proyecciones metafóricas y metonímicas que toman como dominio de origen alguna de las imágenes esquemáticas; por lo tanto, un dominio físico. Los efectos prototípicos (Rosch, 1973) y categorías radiales (Lakoff, 1987) son el resultado de la manera en la que organizamos el conocimiento a partir de las anteriores imágenes esquemáticas.

conlleven representaciones mentales explícitas e implícitas (Nation, 2001; Schmitt, 2010), así lo observa Baralo (2001) a juzgar por la permanencia de los errores idiosincrásicos en la interlengua; el conocimiento explícito es tal vez almacenado en la memoria a corto plazo, y conlleva un proceso de recuperación diferente al conocimiento inconsciente, esto es, aquel que no se manifiesta de manera expresa y que se almacena en la memoria a largo plazo. A este respecto, Nagini (2008: 182) afirma que “The LTM [long term memory] is said to contain the mental lexicon”; en la memoria a largo plazo, añade, una persona tiene un almacén de palabras que están organizadas y representadas de tal manera que el reconocimiento y la recuperación de la palabra resulta fácil, lo que demuestra una relación estrecha entre lenguaje y memoria. De estas aportaciones se concluye que todo el conocimiento que se retenga en la memoria a corto plazo no forma parte del léxico mental. Además, la competencia nativa general se distingue por un equilibrio entre un conocimiento generativo, basado en reglas, que contribuye en la elaboración de mensajes planificados –como *dar* cuando se combina con sustantivos que designan “eventos que se interpretan como unidades informativas”, como *clase*, *lección*, *discurso*, *rueda de prensa*, etcétera–; y uno formulaico, o memorístico, que se manifiesta cuando toda una secuencia de palabras está disponible en el habla y en la escucha como una sola unidad, sin ninguna ayuda mnemotécnica, por lo que aporta fluidez y precisión en la lengua espontánea (como podría ser *da clase*, *le dieron gato por liebre*, *voy a darle el pésame* o *porque me da la gana*), y, al mismo tiempo, también contribuye en la planificada (Sinclair, 1991).

En definitiva, este modelo psicolingüístico (Baralo, 2001) ofrece una hipótesis acerca de la organización del léxico y presenta la relación de este con las reglas de la gramática “como una interfaz entre dos subsistemas cognitivos: el conocimiento del mundo y la información contenida en los actos de habla, de un lado, y el sistema computacional de la lengua, sujeto a principios generales y abstractos, responsable de la forma y el orden de las palabras para su procesamiento” (Baralo, 2001: 9). Este planteamiento es contrastado con un análisis de corpus no nativo, en el que se demuestra, como decíamos, que el léxico constituye el motor generativo de la gramática, en el sentido de que la información léxica contenida en las diferentes palabras que participan en una oración constituye el germen de lo que será su forma final.

Otro modelo psicolingüístico que destaca el valor del léxico –en este caso la hipótesis se aplica al uso lingüístico en el proceso de producción oral– es el de Levelt (1989). De acuerdo con este modelo, el léxico contiene un conocimiento declarativo, es decir, accesible a través de la introspección, y al

igual que en el modelo propuesto por Baralo, la elección de ciertas palabras es lo que determina la forma gramatical y fonológica de la oración en las que aquellas participan, por lo que la gramática es uno de los componentes integrados en el conocimiento de una palabra. Levelt divide el lexicón en dos partes: una que contiene lemas y otra que contiene formas. Por cada lema se requiere un conocimiento semántico y gramatical, esto es, supone el conocimiento de los significados que componen la palabra –es posible que cada lema contenga además información sobre adecuación, estilo o frecuencia que nos advierte de las restricciones de uso– y la categoría sintáctica a la que pertenece, junto con sus funciones gramaticales y otras restricciones gramaticales. Estos aspectos determinan el uso de la palabra. La información –semántica y gramatical– contenida en el lema está conectada a la forma morfofonológica de la palabra, lo que implica que el significado y la forma están unidos. En la propuesta de producción lingüística (oral) se distinguen tres etapas fundamentales: conceptualización, formulación y articulación, que reflejan la estrecha relación que existe entre los elementos que componen el conocimiento de una palabra. En el proceso de producción se desencadena la siguiente secuencia: 1) se produce a través del conceptualizador, que es de donde parte el mensaje, un mensaje preverbal que contiene la información que el hablante desea transmitir (primera etapa). 2) El formulador, transforma el mensaje preverbal en un plan fonético: accede al lexicón para encontrar el lema que contenga los componentes significativos apropiados (segunda etapa). 3) Esta transformación implica, por un lado, codificar el mensaje gramaticalmente; algunos de estos componentes estarán conectados directamente con los rasgos gramaticales concretos. Se activarán otros componentes gramaticales que forman parte del lema (segunda etapa). 4) Dicha transformación supone, por otro lado, codificar fonológicamente el mensaje. Los componentes significativos y gramaticales del lema están unidos a los rasgos morfológicos y fonológicos de la palabra, lo cual indica que la gramática, la morfología y la fonología están determinadas por las palabras escogidas (segunda etapa). 5) Se elige la morfología apropiada para que codifique las funciones significativas y gramaticales de la palabra (segunda etapa). 6) Se producen los rasgos fonológicos que corresponden a la forma de la palabra (etapa segunda). 7) El articulador produce finalmente la palabra (tercera etapa).

Nation (2001: 38) aporta el ejemplo de *painter* para ilustrar la materialización de las 7 fases en la producción de esta palabra según Levelt. Tras la primera fase, en la que se lleva a cabo el mensaje preverbal, se establece una relación –segunda y tercera fases– entre el significado “una persona que pinta” y el sustantivo singular contable que se refiere al agente que lleva a cabo la acción

expresada por la raíz del verbo; la cuarta fase une la anterior con la forma *painter*; en la fase quinta se elige la forma *paint* + *er*, ya que el sufijo *-er* se conecta con quien lleva a cabo la acción de un verbo y es un afijo que al unirse a una raíz da como resultado un sustantivo; durante la sexta fase se elige el modelo apropiado de acentuación; el sufijo *-er* no modifica la forma fonológica de la raíz *paint*.

Todas estas fases que se suceden en la producción lingüística muestran la manera en la que los distintos aspectos que intervienen en el conocimiento de una palabra se influyen o interrelacionan (Nation, 2001: 39)⁵⁹, lo que pone de manifiesto la importancia del léxico en la producción de los mensajes. Por consiguiente, Levelt –al igual que otros autores citados, como Baralo, Wray, Krashen y Terrell, o Tomasello– atribuyen al vocabulario un papel primordial en el proceso de adquisición de una lengua, pues es el significado vertebrador del pensamiento que se adquiere a partir del contexto.

3.1.3. Repercusiones pedagógicas

Como cabe esperar, las consideraciones teóricas descritas subyacen en los cambios que tendrán lugar en las prácticas pedagógicas:

The role of vocabulary teaching, which in the light of the findings concerning a deeper understanding of the nature of vocabulary in both the fields of linguistics and applied linguistics, raised the issue that the teaching of vocabulary should be redefined to fit into the way we design our syllabi (Richards, 1976: 77).

Para ilustrar esta transformación, resulta útil revisar el estatus del componente léxico en los principales paradigmas de enseñanza del último siglo⁶⁰.

Como su propio nombre indica, el Método Gramática-Traducción (o Tradicional), que dominó la enseñanza de LE desde aproximadamente 1840

⁵⁹ Levelt también se refiere a las redes entre palabras. Distingue entre las relaciones que se dan dentro de una misma entrada en el lexicón (como las que se acaban de señalar en el ejemplo anterior, o las relaciones en la flexión de una palabra) y las relaciones entre diferente entradas. Estas últimas las divide en intrínsecas y asociativas. Las intrínsecas están basadas en rasgos semánticos (sinonimia, antonimia; días de la semana), gramaticales (pertenencia a la misma categoría léxica, o desempeño de la misma función gramatical), morfológicos (derivación) y fonológicos (mismos sonidos iniciales o finales). Las asociativas dependen de las colocaciones frecuentes de la palabra y no de los aspectos significativos, como *verde* y *césped* o *trueno* y *relámpago*. Bogaards (1994) también se refiere a ellas en los mismos términos.

⁶⁰ Véase más sobre ello en Boers y Lindstromberg (2008: 1-4), Richards y Rodgers (2009 [1998]), Pastor Cesteros (2004) y Morante Vallejo (2005).

hasta 1940 –aunque con la oposición de movimientos reformadores–, se centra en la morfología y en la sintaxis –la naturaleza de la L2 determina el peso que se concede a estos dos componentes–, en textos escritos. En este método el tratamiento explícito del vocabulario se realiza en forma de glosarios unidos a los textos que sirven de ayuda para la traducción, normalmente de L2 a L1. Los aprendientes deben memorizar esas listas de palabras bilingües que se consideran necesarias para poder formar oraciones, que constituyen la unidad básica de aprendizaje, mediante la aplicación de las reglas aprendidas. Por lo general, no se ayuda a los aprendientes a retener el vocabulario nuevo para un uso activo.

El Método Directo –introducido en Alemania y Francia a principios del siglo XX, y difundido en EEUU a través de las escuelas fundadas por Sauveur y Berlitz– se construye a partir de la observación de cómo aprenden a hablar los niños. Se concede prioridad a la lengua oral. El vocabulario aparece ahora contextualizado dentro de la situación cotidiana en la que el alumno se debe desenvolver; el significado se transmite, en el caso del vocabulario concreto, a través de la demostración y asociación de objetos o imágenes con la palabra; en el caso de las palabras abstractas, por medio de la ejemplificación de acciones y de la asociación de ideas. La práctica de este método demostró que se obtenían buenos resultados únicamente en un contexto de inmersión lingüística. Sus propuestas metodológicas eran de difícil aplicación en el aula de lenguas, ya que el esfuerzo y el tiempo empleados para la explicación de muchas palabras lo hacían poco rentable y poco viable. Más adelante, para solventar estos inconvenientes, se comienza a trabajar la lectura de textos en lengua no literaria, es decir, textos periodísticos y diálogos principalmente que requerían un conocimiento léxico específico. Es entonces cuando aparece por primera vez el estudio del vocabulario como herramienta importante para acceder a una de las destrezas lingüísticas: la comprensión lectora. Y como consecuencia, se elaboran las primeras listas con las dos mil palabras más frecuentes en inglés; estos listados se convirtieron en una referencia obligada en la creación de materiales didácticos (Richards y Rodgers, 1998: 38).

Las críticas a este método se concretaron en los años veinte y treinta, y están en la base del Método Situacional británico y del Método Audio-oral, procedente de EE.UU., ambos de orientación estructuralista.

Estos métodos intentan dotar de una fundamentación teórica al método anterior. El dominio de la oralidad sigue siendo prioritario; la fluidez y el rigor se alcanzan por medio de la memorización de diálogos seguidos de ejercicios prolongados e intensivos; de aquí se desprende su base psicológica conductista según la cual el aprendizaje es fruto de la formación de hábitos. La repetición de

estructuras para practicar la lengua, la preeminencia de la destreza oral y el criterio del uso para la selección del léxico ya se encontraban en el método anterior, pero factores lingüísticos y sociopolíticos consagran al Método Audio-oral hasta los años 70 (Pastor Cesteros, 2004). Con respecto a los primeros, la escuela estructuralista americana, por medio de Bloomfield y Fries, proporciona un nuevo modo de análisis lingüístico. Conciben la lengua como un sistema formado por estructuras; la morfosintaxis y la fonología, al ser los únicos niveles susceptibles de estructuración desde los supuestos estructurales, se convierten en el foco de interés de la enseñanza y aprendizaje. El léxico, sin embargo, “muestra un comportamiento arbitrario no regido por reglas y que no participa en la configuración de la estructura lingüística” (Morante Vallejo, 2005: 14), por lo que no existe ningún interés por analizar la manera en que se adquiere el vocabulario: basta con memorizar algunas listas de palabras para poder ejercitar las estructuras sintácticas. Únicamente se podrá proceder a la ampliación de vocabulario una vez que se dominen las estructuras. Además, ya señalábamos (apartado 3.1) que, dado que la consideración por la comprensión y el procesamiento cognitivo es tan baja, se ofrecen escasas técnicas y estrategias que promuevan la retención del léxico en L2.

Hacia los años ochenta, la enseñanza de lengua comunicativa se convierte en el paradigma dominante en la enseñanza de L2, metodología que, pese haber sufrido transformaciones a lo largo de estos años, continúa vigente. En ello influyen factores socioculturales y pedagógicos, así como las innovaciones lingüísticas y psicolingüísticas descritas en las dos anteriores secciones. A continuación, nos detenemos en los principales factores –algunos ya sugeridos– que desencadenan la implantación de este nuevo paradigma de enseñanza.

A finales de los años sesenta se produce la crisis del Método Audio-oral por varias razones. En primer lugar, se produce un rechazo de la teoría psicolingüística conductista; Chomsky ya había demostrado que las teorías estructurales no podían explicar las características fundamentales de la lengua, como la creatividad y la singularidad de cada una de las oraciones. Como manifiesta Howatt (1984), a finales de los años sesenta parecía claro que el enfoque situacional estructural había terminado su recorrido. No había futuro en continuar tratando de predecir la lengua a partir de las situaciones. Lo que se necesitaba era un estudio más minucioso de la lengua y promover un retorno al concepto tradicional de que los enunciados tienen significado por sí mismos y expresan los significados y las intenciones de los hablantes y de los autores que los crean.

La nueva psicología cognitiva a la que se adhiere el generativismo postula que la L1 se adquiere a partir de la formación de reglas –hipótesis que se construyen y destruyen–, no de hábitos –ya se ha señalado que en Baralo (2001) se defiende que este proceso es también aplicable a la L2–. Estas reglas se interiorizan, son el producto de la GU y de la exposición al *input* lingüístico; de ahí que en los métodos comunicativos se promueva un aprendizaje basado en la experiencia, en el que se ofrecen muchas muestras de *input*, pues se conciben como oportunidades para que el aprendiente formule sus hipótesis y las pueda contrastar (véase el Enfoque Natural, de Krashen y Terrell, o el Enfoque por Tareas, de Candlin)⁶¹.

Por otro lado, poco después, a finales de los sesenta y principios de los setenta, los lingüistas británicos –como ya hemos señalado– destacaron otra dimensión fundamental de la lengua: el potencial funcional y comunicativo. El objeto de la descripción de la teoría lingüística es ahora la competencia comunicativa, frente a la competencia lingüística de Chomsky. El objetivo de la enseñanza-aprendizaje de L2 es asimismo la competencia comunicativa. Se puede decir que el primer programa de orientación comunicativa es el Nociofuncional. Surge de la mano de Wilkins (1976), encargado por el organismo de la cooperación cultural y educativa del Consejo Europeo de diseñar una nueva programación de enseñanza de idiomas basada en las necesidades reales de los aprendientes de lenguas en Europa. Esta se organiza en torno a nociones y funciones, y no sobre las estructuras, lo que implica concebir la lengua más allá de la gramática y del vocabulario para entenderla como un discurso conformado por funciones y nociones para cuya adquisición se requiere el componente lingüístico, esto es, el léxico, la fonología, la morfología y la sintaxis. Este componente, junto al discursivo, el sociocultural y el estratégico, integran la competencia comunicativa. Queda superada, por tanto, la oración como unidad máxima del análisis lingüístico estructuralista, por lo que en el aula se trabajará con actividades de tipología textual y, de acuerdo con los nuevos criterios pragmáticos y sociolingüísticos, se incidirá en la consecución del objetivo comunicativo perseguido y en la adecuación contextual en el uso de la lengua, y no tanto en la corrección gramatical. Y esto repercute directamente sobre la consideración del léxico; las palabras tienen una carga cultural, por lo que no se puede prescindir de la dimensión pragmática y sociolingüística, ni de la adquisición ni de la enseñanza del léxico; por lo tanto, como señala Brieno

⁶¹ No obstante, en la actualidad no se rechaza que parte del aprendizaje sea tanto producto de la formación de hipótesis como imitativo. En ambos procesos la cantidad de *input* es determinante.

García (2013: 11), toda teoría lingüística debe incluir el léxico, no solo con sus particularidades formales, fonológicas, sintácticas y semánticas, sino también el vocabulario como depositario de todas las propiedades idiosincrásicas de la lengua. Estas mismas particularidades se han de adquirir en la L2, y por ello son tenidas en cuenta en nuestra propuesta del apartado 3.2.

Así, al igual que Wilkins (1972: 111) cuando afirmó que “Without grammar very little can be conveyed; without words, nothing can be conveyed”, muchos otros investigadores (como Meara, 1996 o Gómez Molina, 2000) a los que nos sumamos, consideran que el léxico, como se ha destacado en las teorías de adquisición y en las más recientes teorías lingüísticas, es más que un componente situado al mismo nivel que el resto, pues es la clave para el desarrollo comunicativo: si comunicar es transmitir significado, como dice Lewis (1993: 38) “words carry more meaning than grammar”, por lo que desarrollar vocabulario es desarrollar comunicación. Por consiguiente, el léxico puede considerarse el verdadero esqueleto de la lengua y, por lo tanto, protagonista en la clase de LE: “lexical knowledge is known to be an absolutely crucial factor across the whole spectrum of L2 activities” (Singleton, 1999: 3). Estas ideas que revalorizan el léxico han contribuido a que en algunos de los materiales más recientes de ELE se trabaje el léxico de forma sistemática, en todas las unidades didácticas y a lo largo de toda la unidad, como es el caso de *Bitácora* (2011), de la Editorial Difusión, aunque podría decirse que este es un caso aislado.

Así las cosas, en los años setenta se abandona, finalmente, la asunción general tan extendida hasta los años sesenta de que “most learners could get by adequately with a very limited vocabulary” (Meara, 1996: 1); de hecho, como señala Fernández (1997), un error en la selección de una palabra o en el entendimiento puede producir una situación de completa incompreensión.

Podemos afirmar que, una vez que la dicotomía entre gramática y léxico se pone en entredicho desde el marco teórico funcionalista, se produce su anulación también en las aulas de LE; el léxico deja de estar al servicio de aquella, y ahora ambos componentes se complementan, lo que refuerza la idea de que la gramática de una lengua solo se puede adquirir unida al desarrollo del conocimiento léxico (Lewis, 1993; Bogaards, 1996).

De acuerdo con esta nueva interdependencia entre gramática y léxico, el Enfoque Léxico de Lewis, de base comunicativa, se construye sobre la asunción general señalada anteriormente de que la separación entre vocabulario y gramática es artificial. Según Lewis (2000: 137), mantener este punto de vista crea confusión y conlleva que muchos rasgos útiles acerca de cómo se usan las palabras (en lugar de lo que significan) sean pasados por alto. De esta manera, da

un paso más y concede primacía al léxico al reconocer que la gramática sobregeneraliza y que el sistema es más restrictivo de lo esperado. Desde esta perspectiva postula que la lengua consiste en léxico gramaticalizado y no en gramática lexicalizada, lo que constituye el principio fundamental del Enfoque Léxico: “Recognising that every word has its own grammar, however, means that any approach based on the central role of lexis is in many ways more grammatical than any traditional grammar syllabus” (Lewis, 2000: 8). Este autor manifiesta (2000: 137) que en su enfoque recibe la influencia de los avances en lingüística teórica –como los de la semántica léxica– y de los del análisis de corpus, pues es este marco de léxico gramaticalizado en el que los investigadores están tratando de encontrar descripciones del inglés más precisas.

En la enseñanza de orientación comunicativa, en la que –como se ha visto– aprender vocabulario es adquirir significados, se reclama que una palabra no se puede aprender de forma aislada, sino en contexto (Wilkins, 1972), en línea con las aportaciones de Firth y, posteriormente, de la Escuela sistémica británica, que le confieren un papel fundamental a la colocación como fenómeno lingüístico que participa en la determinación del significado que actualiza una palabra. En este nuevo paradigma de investigación léxica, Brown (1974), siguiendo a Firth, introdujo la noción de colocación en el área de la didáctica de la lengua para destacar la importancia del contexto a la hora de determinar el significado de una palabra. Lewis (1993: 80-81), a este mismo respecto, señala que tanto el contexto situacional como el cotexto no pueden considerarse como algo optativo, sino como parte intrínseca del significado de las palabras; son, por consiguiente, esenciales para la concreción de significados, y, por ende, para la comprensión y elección de las unidades léxicas. De esta manera, se produce un rechazo a la memorización de listados de palabras frecuentes, tan habitual en los métodos anteriores, en las que únicamente se atendía a la forma de las palabras y a su significado aislado de todo contexto⁶². En el polo opuesto a estas listas se

⁶² Resulta sorprendente que todavía hoy en el aula y en los manuales de LE la enseñanza de las palabras aisladas de su uso sea una práctica tan extendida. Pese a que normalmente se parte de un *input* en el que las palabras están contextualizadas, se termina con un tratamiento de la palabra descontextualizado, por lo que se produce el error cuando el alumno trata de usarla en un nuevo contexto. El siguiente ejemplo está tomado de Sánchez Rufat (2011). En él se muestra que el aprendiente detecta un problema a partir de un texto con la palabra *dissolver*, en *El azúcar se disuelve en el café*; se siguen aportando en el aula definiciones con sinónimos que son tratados como si fueran absolutos, o se siguen a pie juntillas las definiciones de los diccionarios. Si buscamos *dissolver*, encontramos “deshacer, destruir, aniquilar” y “separar, desunir lo que estaba unido”. No obstante, un *edificio se destruye*, el *camino* y la *cama* se *deshacen*, y los *siameses* pueden llegar a *separarse* pero ninguno puede *dissolve*. Por otro lado, el *azúcar se disuelve* o *se deshace* en el café, pero no *se destruye*. Al mismo tiempo, estas definiciones no nos permiten codificar automáticamente que una *manifestación*, *concentración*, *disturbio*, *pacto* o *rumor se disuelven*. Estos problemas se solucionarían si nos atuviéramos

encuentran los nuevos diccionarios combinatorios como *Redes* (2004) y *Práctico* (2006) –fruto de más de treinta años de trabajo en equipo–, centrados en los criterios semánticos mediante los que los predicados restringen a sus argumentos –a menudo las combinaciones de una palabra no se eligen individualmente, sino que la elección obedece a criterios semánticos restrictivos–, o los diccionarios de colocaciones que existen en otras lenguas⁶³. Su utilidad pedagógica resulta evidente.

Desde el momento en el que se concibe la lengua como un sistema de comunicación, de transmisión de significados, y dado que el significado de una secuencia compleja –de acuerdo con el modelo de adquisición analítico de los adultos– es el resultado de los significados de las unidades menores que contiene y del modo en que se combinan (principio de composicionalidad), hemos de aceptar que ni es posible comunicar con pocas palabras, ni estas se deben aprender descontextualizadas, pues el contexto y la combinatoria proporcionan la clave para interpretar el significado (recuérdense las propuestas asociacionistas de Wilkins [1972] o Aitchinson [1994], ya comentadas, a las que se suman otras como las de Thornbury [2002], Pérez Basanta [2003], Higuera [2006] o Baralo [2006]).

Y a esta misma dirección apunta el *MCER* (2002: 13):

[la competencia comunicativa] se relaciona no sólo con el alcance y la calidad de los conocimientos (por ejemplo, las distinciones fonéticas realizadas o la extensión y la precisión del vocabulario), sino también con la organización cognitiva y la forma en que se almacenan estos conocimientos (por ejemplo, las distintas redes asociativas en que el hablante coloca un elemento léxico) y con su accesibilidad (activación, recuperación y disponibilidad).

Si se acepta la hipótesis de que las redes combinatorias son un poderoso principio para organizar y almacenar el léxico, su enseñanza ayudará indudablemente tanto a la memorización de unidades léxicas como a su posterior recuperación. De hecho, las pruebas llevadas a cabo por Pawley y Syder (1983), Erma y Warren (2000) y Wray (2002) confirman que las asociaciones prefabricadas de palabras pueden llegar a ocupar un 50 % del discurso nativo.

al contexto y se le explicara a los aprendientes que *dissolver* significa, entre otras cosas, mezclarse de forma homogénea algo sólido o líquido en otro líquido, y en ese sentido físico se construye con sustantivos que designan sustancias constituidas por partículas (azúcar, sal, café, cacao, arena). Para ello, la dificultad no está en conocer el significado de *dissolver*, sino en saber usar la palabra.

⁶³ Resultaría inadecuado denominar a *Redes* y *Práctico* diccionarios de colocaciones si se entiende por ellas combinaciones frecuentes, clichés que no están integrados en el sistema lingüístico, que no se sustentan necesariamente en bases semánticas restringidas.

Incluso en el paradigma neurocognitivo se defiende la hipótesis de que estos bloques formulaicos tengan capacidad –ya sea aplicando la analogía o la extensión– para generar parte de la gramática de una lengua en los primeros estadios de la adquisición de L1 (Carter y McCarthy, 1988; Tomasello, 2003).

Todos estos precedentes sentaron la base para que una proporción significativa de los teóricos de la enseñanza de LE viesen en la enseñanza del vocabulario, en el sentido amplio de palabras y elementos multipalabra, un factor clave para alcanzar un dominio lingüístico alto (Boers y Lindstromberg, 2008: 4). Las teorías pedagógicas de Willis (1990) expuestas en *The Lexical Syllabus*, y las de Lewis en su Enfoque Léxico (1993, 1997, 2000), así como la obra de Nattinger y De Carrico (1992), son claras muestras de la especial relevancia que se otorga al léxico en el proceso de enseñanza-aprendizaje de una lengua.

A partir de mediados de los ochenta, la importancia de aprender vocabulario había cuajado, pero en el paradigma pedagógico se concede prioridad al dominio de las funciones del lenguaje, así que, pese a que las palabras contienen más significado que la gramática y las funciones, la metodología comunicativa no indica las pautas de tratamiento del aprendizaje del léxico, que tiene como único papel servir de apoyo a la lengua funcional (Schmitt, 2000). En este mismo sentido, aunque han ido aumentando los estudios que ponen de manifiesto la necesidad de tomar como punto de partida en el aula las unidades léxicas superiores a la palabra (véase Álvarez Cavanillas, 2008; Higuera, 2006; y Castillo Carballo, 2001, entre otros), no acaban de conformarse estrategias didácticas consolidadas ni existe tampoco un total acuerdo entre los teóricos acerca de si el componente léxico debe ser objeto directo de la enseñanza o si debe enseñarse de manera implícita (recuérdense las propuestas del Enfoque Natural de Krashen y Terrell, 1983; y las apreciaciones de Baralo [2001] y Nagini [2008]).

En la adquisición, ya sea consciente o inconsciente, participa tanto un procesamiento analítico como holístico; este último predomina en las situaciones de interacción rutinarias, y aquel permite explotar los recursos lingüísticos de cada uno de forma más creativa (Wray, 2002: 278). En esta misma línea, en el modelo de Skehan (1998) se distingue entre un modo de procesamiento que facilita la lengua a tiempo real y otro que se encarga del uso de lengua planeada; esta distinción se corresponde con la establecida entre el conocimiento generativo (o basado en reglas, según el principio de composicionalidad) y el memorístico (basado en la memoria, según el principio de idiomática), de Sinclair (1991), a la que ya nos hemos referido anteriormente.

A finales de los ochenta y principios de los noventa, surgen las primeras voces que reclaman que, pese a la abundancia de contextos significativos y a las numerosas oportunidades de interacción, los aprendientes no logran adquirir algunos elementos lingüísticos (Long, 1997); y es que, como señala Schmidt (1990) “picking up target language forms from input when they do not carry information crucial to the task, appears unlikely for adults. Paying attention to language form is hypothesized to be facilitative in all cases”.

En esta época surge el enfoque denominado *Focus on form* (1991), de Long, en el que se plantea una propuesta que intenta conciliar la enseñanza del contenido y los aspectos lingüísticos de la forma, pues considera que ambos elementos son esenciales en el proceso de enseñanza-aprendizaje de una L2.

Así se establece la necesidad de incluir estrategias pedagógicas destinadas al aprendizaje explícito de la forma; dentro de una tarea comunicativa se incorporan actividades esporádicas de “toma de conciencia” (hipótesis del *noticing*, de Schmidt, 1990) de los fenómenos lingüísticos, pues el *input* que recibe el aprendiente no siempre ofrece información suficiente para que se construya el conocimiento o para corregir la producción equivocada; en otras palabras, a partir del procesamiento del contenido, es preciso llamar la atención de los aprendientes hacia las características de la lengua que están aprendiendo (palabras, estructuras gramaticales, coapariciones frecuentes, patrones pragmáticos, etcétera), a fin de que se produzca una adquisición de la forma y de los significados que alcanza la lengua en la comunicación.

Aunque el concepto de “toma de conciencia” se ha empleado principalmente para discutir asuntos de la enseñanza de la gramática, también más recientemente se ha utilizado en el debate en torno al asunto de la instrucción de vocabulario explícita *vs.* implícita que comentábamos en páginas anteriores. Laufer (2005), defiende con muy buenos argumentos la necesidad de una enseñanza explícita, con foco en la forma. Craik y Lockhart (1972) afirman que la verdadera adquisición del vocabulario se produce solo si se alcanza un nivel de procesamiento mental profundo, lo que equivale a establecer asociaciones cognitivas. La hipótesis de la Profundidad de Procesamiento sostiene que las actividades mentales que requieren más elaboración de pensamiento, manipulación o procesamiento de una nueva palabra ayudarán en el aprendizaje de aquella palabra (Craik y Lockhart, 1972: 673; Cervero y Pichardo, 2000: 96-104).

En suma, desde que Schmidt (1990) formulara su hipótesis de la toma de conciencia o *noticing*, en la que proponía que el esfuerzo cognitivo consciente – que conlleva la experiencia subjetiva de la toma de conciencia– es una condición

necesaria para la transformación de *input* en entrada, se considera que la adquisición por parte de los adultos de este conocimiento en L2 es diferente al modo en el que lo adquieren los niños en su L1: lo que en L2 se pueda adquirir de forma implícita a partir de contextos comunicativos es muy limitado (en parte por transferencia de la L1; la L2 se percibe por medio de los mecanismos de la percepción óptima de la L1), y el logro de la precisión exige fuentes adicionales de consciencia y aprendizaje explícito –lo cual no impide que lo aprendido de este modo no pueda pasar en algún momento a ser conocimiento inconsciente–. Esta postura se conoce en ASL como *Weak interface position* y se basa en los estudios que demuestran que diferentes tipos de instrucción explícita son más efectivos que los implícitos, y que esta eficacia es duradera (véase Ellis [2011] para una revisión de estos trabajos). Nation (2001) o Gómez Molina (1997) se refieren también a la utilidad de las actividades de léxico descontextualizadas que tienen como único objetivo el desarrollo de la competencia léxica, y que permiten lograr una mejora en el uso de la LE. No obstante, Ellis (2011: 39) afirma:

In the latter part of the twentieth century, research (...) converged on the conclusion that explicit and implicit knowledge of language are distinct and dissociated, they involve different types of representations, they are substantiated in separate parts of the brain, and yet they can come into mutual influence in processing.

En la actualidad se suele trabajar el léxico en los métodos comunicativos de manera dual, tanto implícitamente como de forma explícita, lo que es coherente con las ideas expuestas: “El conocimiento de una palabra es una representación mental de gran complejidad, que integra diferentes aspectos y componentes cognitivos, algunos más automáticos e inconscientes otros más conscientes, reflexivos y experienciales” (Baralo, 2007: 41).

Así, en el proceso de enseñanza-aprendizaje en ELE, al tiempo que se ofrece al alumno gran cantidad de *input* para que desarrolle una adquisición inconsciente del vocabulario, por medio de la imitación, la analogía y la formulación de hipótesis, se está comenzando a trabajar este componente de forma organizada, a partir de la reflexión sobre su estructura en muestras contextualizadas y considerando los campos léxicos, así como los diferentes tipos de relaciones que se establecen entre las palabras –atendiendo, por lo tanto a las teorías de adquisición de Aitchison, Levelt y Baralo, entre otros–; el manual *Bitácora* (2011) y los más específicos *Vocabulario* (2009) y *Uso interactivo del*

vocabulario B2-C2 (2013)⁶⁴ son una buena muestra de ello; proponen actividades que favorecen la comprensión, la retención y la utilización de las unidades léxicas. Para la enseñanza de inglés como L2, la obra *Teaching Chunks of Language. From noticing to remembering*, de Lindstromberg y Boers (2008), es una buena propuesta de técnicas y ejercicios que ayudan al aprendiente a recordar, a almacenar a largo plazo cientos de expresiones frecuentes multipalabra, o *chunks of language*, como el título de la obra indica.

Como síntesis de todo lo expuesto en esta parte dedicada a la revalorización del léxico en las últimas décadas, concluimos que las teorías lingüísticas y psicolingüísticas, por un lado, centran el análisis lingüístico en el léxico gramaticalizado y en la relación entre sentido y sintaxis; y, por otro lado, postulan la adquisición de los ítems léxicos a través de un procesamiento en parte memorístico –adquisición de varios ítems de manera holística (secuencia formulaica)– y en parte generativo y analítico, en parte consciente y en parte inconsciente; y explican su almacenamiento en el lexicón por medio de redes formales y semánticas. Precisamente la fórmula o secuencia formulaica ha sido reconocida como elemento clave en el proceso de adquisición de la L1 (Peters, 1983; Wray, 2002) y en la adquisición de una producción casi nativa en el caso de la L2 (Pawley y Syder, 1983; Wray, 2002). Por todo ello no es de extrañar que surjan –como hemos recogido en este último apartado– enfoques metodológicos comunicativos que promuevan la enseñanza del léxico, y fundamentalmente las fórmulas, para lograr la adquisición de la L2 (*The Lexical Syllabus*, Willis, 1990; *Lexical Phrases and Language Teaching*, Nattinger y DeCarrico, 1992; *El enfoque léxico*, Lewis, 1993, 1997, 2000).

3.2. EL CONOCIMIENTO DEL VERBO: LA COMPETENCIA LÉXICA

Desde los años setenta, las aportaciones de las teorías lingüísticas a la revalorización del léxico en el proceso de adquisición de una lengua conllevaron la necesidad de determinar qué es lo que sabe un hablante acerca de una palabra para que pueda decirse que la conoce. Dado que la Didáctica de Lenguas requiere en todo momento un cuerpo de conocimientos facilitados por –aparte

⁶⁴ Para una descripción de su fundamentación teórica basada en estos nuevos principios y enfoques que revalorizan el léxico, véase el análisis realizado de este manual en Sánchez Rufat (2014).

de la Pedagogía— la Lingüística Teórica y la Psicolingüística⁶⁵, los nuevos tiempos demandaban una enseñanza del vocabulario acorde con estos cambios que estaban teniendo lugar. La descripción propuesta por Richards (1976) sobre el conocimiento de la palabra, enfocada a la enseñanza del componente léxico, configura el marco general para los estudios posteriores (Nation, 1990 y 2001; Schmitt y Meara, 1997), entre los que se incluye, de hecho, el nuestro. Según Richards (1976: 83), los aspectos integrados en el conocimiento léxico son estos: 1) conocer una palabra significa conocer el grado de probabilidad de encontrarse la palabra en el discurso; en el caso de muchas palabras, también conocemos el tipo de palabras que pueden encontrarse con mayor probabilidad asociadas a ellas; 2) conocer una palabra implica conocer las limitaciones en el uso de la palabra de acuerdo con las variaciones en la función y en la situación; 3) conocer una palabra significa conocer el comportamiento sintáctico asociado a la palabra; 4) conocer una palabra conlleva un conocimiento de la forma y de sus palabras derivadas; 5) conocer una palabra conlleva un conocimiento de la red de asociaciones entre esa palabra y otras palabras en la lengua; 6) conocer una palabra significa conocer el valor semántico de esa palabra; y, por último, 7) conocer una palabra significa conocer muchos de los diferentes significados asociados a la palabra.

Como puede verse, saber una palabra abarca una variedad de conocimientos. Algunos de ellos suponen la incorporación de conceptos nuevos que se estaban empezando a investigar a mediados de los años setenta, como las colocaciones —sugeridas en la primera afirmación—, entendidas como las coapariciones frecuentes de palabras obtenidas de los análisis de los extensos corpus informatizados. Asimismo, se mencionan las variaciones diafásicas —o registros— y diatópicas, y al discurso como nueva unidad de análisis —como se infiere de los puntos primero y segundo—. También se toma en consideración el almacenamiento de las palabras por medio de redes o asociaciones —como se indica en el quinto enunciado—.

En relación con ello, y como base para un estudio de interlengua, que en el trabajo que nos ocupa consiste en descubrir la profundidad⁶⁶ del conocimiento

⁶⁵ Entendemos que esta colaboración entre disciplinas no es unilateral: la Didáctica de Lenguas puede ser, asimismo, una fuente de datos para otros ámbitos como la Lingüística Teórica o la Psicolingüística; las evidencias empíricas proporcionadas por esta rama de la Lingüística Aplicada (la Didáctica de LE), como lo referente a los procesos de aprendizaje, pueden servir a aquellas disciplinas (Lingüística Teórica y Psicolingüística) para que enriquezcan a su vez sus propios planteamientos.

⁶⁶ La profundidad y la extensión del vocabulario constituyen las dos dimensiones tradicionales de la competencia léxica. Entendemos por profundidad el conjunto de aspectos que se conocen de una palabra y que indican el nivel de conocimiento alcanzado.

productivo⁶⁷ del verbo por parte de un grupo concreto de HNN –en otras palabras, analizar los aciertos y las desviaciones a la hora de usar el verbo *dar*–, se hace necesario fijar previamente los diferentes contenidos que integran dicho conocimiento, es decir, qué implica con exactitud conocer un verbo. Para ello nos serviremos de las propuestas de Richards (1976), Nation (2001) y Baralo (2006)⁶⁸, y las adaptamos –al igual que en este último trabajo– a los contenidos del *MCER*, dado que el documento del Consejo de Europa sirve de orientación y modelo para establecer una homogeneidad en la enseñanza-aprendizaje de lenguas en los distintos países y una obligada coherencia pedagógica.

3.2.1. *Los rasgos morfológicos*

El *MCER* (2002: 108), como se recoge al comienzo de este capítulo, se refiere a la competencia léxica como “el conocimiento del vocabulario de una lengua y la capacidad para utilizarlo”. Ahora bien, para poder determinar esa capacidad de utilizar un verbo por parte de un grupo de HNN, se ha de examinar la compleja interfaz en la que se extiende este conocimiento, y que interfiere con otros dominios de la competencia lingüística y sociolingüística (Nation, 2001: 23; Baralo, 2006: 1).

El *MCER* (2002: 107) distingue seis competencias que integran la competencia lingüística, a saber: léxica, gramatical, semántica, fonológica, ortográfica y ortoépica. Tomemos la perspectiva de un análisis del conocimiento de un HNN de ELE que actúa como emisor (y no como receptor) de un mensaje que requiere como medio de transmisión la modalidad escrita⁶⁹. Se descartan, entonces, por su irrelevancia las competencias ortoépica y la fonológica, aunque reconocemos que para poder escribir una palabra suele ser necesario disponer previamente de su representación fonológica. El *MCER* (2002: 108-116) presenta las restantes como si fueran compartimentos de

⁶⁷ Entendemos por *conocimiento productivo* la capacidad de un hablante de “wanting to express a meaning through speaking or writing and retrieving and producing the appropriate spoken or written word form” (Nation, 2001: 25).

⁶⁸ Tréville (2000), en un interesante trabajo, adapta la descripción de la competencia comunicativa de Canale (1983) a la competencia léxica; considera que las unidades léxicas son representantes de fenómenos gramaticales, referenciales, discursivos, socioculturales y de procesos estratégicos y, por lo tanto, instrumentos de desarrollo de la competencia comunicativa. Nosotros, en relación con ello, nos referimos a este conocimiento léxico como una competencia multidimensional.

⁶⁹ En este apartado 3.2. nos referimos, por tanto, al término competencia léxica en el sentido de la capacidad de recuperar esta palabra únicamente en el escrito.

conocimientos estancos, bien delimitados. Sin embargo, el conocimiento que atañe a esta competencia, el léxico, no puede representarse de esta manera. La competencia léxica se extiende de manera transversal: si queremos determinar la extensión que abarca este conocimiento, se han de examinar, como decimos, las competencias vecinas.

En primer lugar, saber usar un verbo por escrito requiere competencia ortográfica, lo que supone el conocimiento de la correcta escritura del verbo –en cuanto a las letras que lo forman y la ausencia o presencia de tilde–; esto conlleva conocer aquellos casos en los que a un fonema le corresponden distintas grafías, o una grafía más de un fonema o ninguno, como en el caso de la *h*. Pero más compleja es la consideración de la morfología del verbo, que responde al reconocimiento de su organización interna: el verbo tiene morfemas flexivos de tiempo, modo, número, persona y aspecto. Además, el verbo puede derivar de un sustantivo, de un adjetivo o de un adverbio, y esta información es de orden léxico. Por el contrario, la información de modo, tiempo y aspecto corresponde al componente gramatical –aunque a menudo se vincule con el léxico– ya sea porque depende de un rector, como en el caso del modo, o porque se construya composicionalmente, como ocurre con el aspecto⁷⁰. Sí debe atenderse, en cambio, a errores relativos a la construcción morfológica del paradigma verbal, esto es, al conocimiento de las formas flexivas. Los relativos al componente léxico son las formas irregulares, en los aspectos que afectan a la base como a los elementos flexivos. El HNN no puede deducir que *enredar* es regular y *cerrar*, irregular. Aitchison (1987) tras revisar los resultados de unos análisis psicolingüísticos empíricos, concluye que los morfemas flexivos se añaden en línea, puesto que este proceso está basado en reglas, y al ser relativamente consistente puede realizarse con poco esfuerzo cognitivo. En cambio, el proceso de añadir prefijos y sufijos –esto es, la derivación– es menos transparente –por ejemplo, la mayor parte de los verbos transitivos forman adjetivos en *-ble*; la

⁷⁰ El uso de las nociones *modo*, *tiempo* y *aspecto gramatical*, pese a que se pueden expresar mediante morfemas, implica un conocimiento gramatical procedimental que va más allá del conocimiento del propio verbo, por lo que la selección del tiempo verbal, del modo o del aspecto gramatical no tendrían por qué ser tenidos en cuenta en el análisis de un verbo de la interlengua. Puede dejarse fuera, asimismo, el aspecto léxico, pues esta noción no es léxica en sentido estricto (De Miguel, 1999); es decir, aunque el aspecto léxico forma parte del contenido semántico del verbo, tiene más bien un carácter sintáctico, por lo que intervienen en él factores de índole léxico-semántica (información semántica contenida en la raíz relacionada con el modo en que tiene lugar el evento que describe el verbo) pero también léxico-sintáctica (características gramaticales de los participantes en el evento: función semántica y sintáctica, número, determinación y cuantificación). Así, para analizar el aspecto léxico han de abordarse todos estos factores de manera conjunta y tomando en consideración configuraciones oracionales, por lo que el aspecto léxico del verbo puede ser modificado por la información que aportan otros elementos que lo acompañan; esto hace que la tarea de sistematizar el análisis del aspecto léxico del verbo pueda quedar al margen.

alternancia entre *-ez* y *-eza* depende en gran medida de la cantidad de sílabas del adjetivo base; tenemos *almacenamiento* y no *almacenación*, y, sin embargo, decimos *rotación* y no *rotamiento*—, por lo que las palabras derivadas se recuperan de la memoria como un todo polimorfemático, y ya no hablaríamos, por lo tanto, del conocimiento de una misma o una sola palabra. No obstante, a la hora de juzgar la competencia morfológica del HNN, sí que destacamos la importancia de que reconozca que la raíz del verbo puede aparecer en otras palabras, y conviene, además, conocer los mecanismos de formación de palabras, lo que incluye a los afijos de derivación. Por ello, en nuestro estudio consideramos también si el HNN conoce la forma de las palabras derivadas que usa del verbo *dar*, pese a que, como decimos, del proceso de derivación se obtengan palabras diferentes. Aun así, ya se ha señalado (al comienzo de este capítulo) que hay estudios que evidencian que las familias de palabras, que comparten la raíz, son entidades reales desde una perspectiva psicológica, y que cuando nos referimos a conocer una palabra debemos referirnos a conocer una familia de palabras.

3.2.2. *La naturaleza semántica*

Además de reconocer su forma escrita (ortografía) y las partes que lo integran (morfología), conocer un verbo implica estar familiarizado con su significado. Este tercer aspecto es fundamental, pues, dejando por el momento a un lado la situación comunicativa, se podría afirmar que para hacer un buen uso escrito del verbo basta con conocer su representación gráfica y su *naturaleza semántica* (si se entiende que esta determina la estructura temática y aspectual del verbo, factores ambos que interactúan con la sintaxis). Tradicionalmente la palabra se ha asociado siempre al significado; es indudable la estrecha interrelación que existe entre la competencia léxica y la semántica. Esta última “comprende la conciencia y el control de la organización del significado” con que cuenta el hablante (MCER, 2002: 112), y abarca cuestiones relacionadas con el significado lingüístico expresado por medio de las unidades simples y de sus combinaciones⁷¹. Ya hemos señalado también en el apartado 3.1.2 que, en esta misma línea, en el modelo de funcionamiento del lexicón de Baralo (2001: 20), si

⁷¹ Precisamente, el que la competencia léxica sea considerada como un conocimiento que se extiende por otros dominios lingüísticos se explica, en parte, porque se encuentra indisolublemente unida a la competencia semántica, cuyo contenido se integra en todo el sistema – significado poseen, además de la palabra, las unidades que conforman los demás niveles lingüísticos: el morfema, el sintagma, la oración y el texto— en tanto que la lengua se basa en una organización de la forma o expresión y en una organización del significado o contenido.

algo caracteriza el procesamiento léxico es el hecho de que depende fundamentalmente de señales e interconexiones semánticas.

Si reconocemos la complejidad del significado, habremos de asumir que el aspecto semántico de la palabra es el más complejo de todos los rasgos que integran el conocimiento léxico⁷². Por un lado, se debe conocer qué significado señala la forma de la palabra y cuáles son los límites que separan la palabra de otras con las que está relacionada, y, por lo tanto, qué está incluido en el concepto. Asimismo, deben reconocerse las entidades y actividades que pueden referir el concepto al que remite la palabra. Téngase presente, sin embargo, que las palabras no suelen tener un significado preciso, o único, con lo que la competencia léxica implica conocer los diferentes significados que estas señalan, si bien es cierto que “los límites entre los sentidos figurados de las palabras son escurridizos, y la separación que se establece entre los usos físicos y los figurados es unas veces mínima y otras máxima” (Bosque, 2005: CXIII [2004]). Cervero y Pichardo (2000: 58) distinguen los denotativos –referenciales o literales–, aquellos “significados objetivos de las palabras o el contenido que la lengua da a la palabra”, de los connotativos⁷³ –culturales o figurados–, por los que entienden los matices o notas significativas que el hablante añade por motivos culturales o personales.

La semántica cognitiva, que se desarrolla también a partir de los años setenta, aporta una explicación –menos vaga– de esta variedad de significados a partir del modelo cognitivo de Lakoff y Johnson (1999). Si aplicamos las hipótesis sobre el modelo cognitivo a nuestro objeto de estudio, podemos establecer que *dar* forma parte de un modelo cognitivo, pues constituye una estructura significativa que nos permite reconocer acciones que se presentan en nuestro campo visual (*dar un lápiz a alguien, un abrazo, una clase o un susto*) y que está formada a partir de la regularidad (patrón recurrente) de nuestros movimientos corporales y manipulación de objetos. Ahora bien, por medio de *dar* podemos conceptualizar fragmentos específicos de la realidad directamente o a través de proyecciones metafóricas. Precisamente en esta distinción radica la polisemia de este verbo: en *dar* confluyen dos tipos de modelos cognitivos.

A) Uno más básico, que se corresponde con una estructura de esquema de imágenes: se construye únicamente por medio de una estructura

⁷² Así lo señala también Higuera (1996: 111), que se refiere a la intervención de múltiples factores en el ámbito semántico, desde la representación mental hasta las relaciones combinatorias de unas palabras con otras.

⁷³ En este trabajo no entramos en la consideración de si el concepto *connotativo* pertenece al significado o al sentido; de hecho utilizamos estos dos términos indistintamente.

“preconceptual”, a partir de imágenes que constantemente aparecen en nuestra experiencia corporal cotidiana; en el caso de *dar*, el modelo cognitivo se corresponde con el Esquema de Imagen del CAMINO (Lakoff y Johnson, 1999: 33), una experiencia física que tiene que ver con dominios básicos —en este caso, el espacio—. El CAMINO surge de nuestra percepción de entidades que se mueven a lo largo de un trayecto en el que hay un punto inicial (la salida), una meta y un recorrido: *ir de un sitio a otro, lanzar una pelota o dar un lápiz a un niño* tienen en común que hay una entidad X que está en un lugar de origen, se desplaza siguiendo una trayectoria y alcanza una meta. Por lo tanto, la idea de *dar algo* (entendiendo por *algo* cualquier sustantivo que designa una entidad material) *a alguien* está en nuestra experiencia corporal cotidiana, por lo que a través de este modelo conceptualizamos de manera directa recurriendo únicamente a las categorías perceptivas.

B) Otro más complejo, que se forma mediante proyecciones metafóricas que toman como dominio de origen las estructuras preconceptuales, esas categorías básicas e imágenes esquemáticas —a las que ya nos hemos referido— que provienen de un dominio físico, y que permiten conformar un dominio abstracto —que es lo que se quiere conceptualizar: hablamos de *estructuras preconceptuales* como previas a la conceptualización abstracta—. Este modelo nos permite ampliar nuestro conocimiento del mundo a partir de lo conocido; nos facilita la tarea de comprender aspectos complejos de la realidad en función de otros más básicos y mejor conocidos.

Así, el anterior Esquema de Imagen del CAMINO, que se encuentra en un dominio del espacio físico, se proyecta hacia una estructura del “espacio conceptual”; una vez que hemos asumido la estructura del esquema origen-camino-destino, podemos derivarla a casos abstractos, y así se producen las metáforas conceptuales.

En el caso del verbo *dar*, a partir del Esquema de Imagen del CAMINO que procede de nuestra experiencia corporal —en el cual una entidad X se desplaza— se construyen diferentes proyecciones metafóricas: 1) *A da* [hacer llegar, proporcionar] un consejo, permiso, una orden, un ejemplo, una explicación (a C); 2) *A da* [hacer efectivo, ejecutar] un abrazo, un salto, un golpe, una clase, una respuesta (a C); 3) *A da* [suscitar] miedo, tristeza, sed, felicidad, asco (a C), donde *A* suele ser un sujeto no agentivo; o sin posibilidad alguna de agente: *darle a C* [suscitar(se) en C] impresión, sensación. Por lo tanto, el modelo de Esquema de Imagen caracteriza la estructura; y los modelos metafóricos, las proyecciones que parten del modelo estructural. El enlace entre la cognición y la experiencia está, por tanto, motivado.

A esta distinción cognitiva podemos añadirle todavía una variable de significado más a partir del trabajo de Ruhl (1989), en el que se defiende que las palabras polisémicas, como *dar*, sean tratadas como una misma palabra. Siguiendo su propuesta, distinguimos la existencia de dos fuentes de significado en una palabra contextualizada.

Por un lado, se halla el significado *sistémico*⁷⁴: el hablante tiene un concepto subyacente de la palabra que es compartido por toda la gama de significados con la que esa palabra se usa –obtener este significado a partir de las distintas acepciones de un vocablo es uno de los principales problemas de la lexicografía–. Hay, por tanto, algún rasgo intensional que está presente en todos los sentidos del verbo, tanto en *dar un lápiz* o *un consejo* como en *dar un abrazo* o *miedo*; dicho de otro modo, ese rasgo es lo que significa *dar* como palabra aislada: a modo de hipótesis (nuestra), *A da B (a C)* significaría “*A* causa que *B* pase de no existir a existir (en *C*, de manera que los seres *C* disfruten o padezcan *B*)”. Obsérvese que en ambos casos –es decir, tanto si *B* “al ser dado” pasa de no existir a existir (el miedo, el abrazo) como si pasa de existir en manos de *A* a existir en manos de *C* (el lápiz, el consejo)– se manifiesta el mismo concepto de *trayectoria* con un punto final o estado resultativo que marca el término del evento. En el capítulo 4 profundizamos en estas consideraciones.

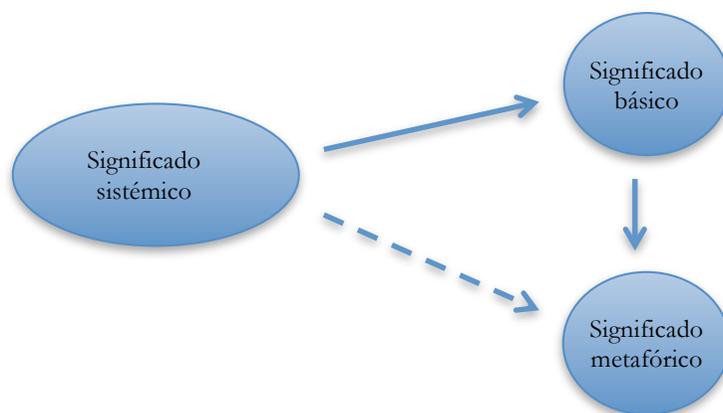
Otros rasgos intensionales, en cambio, son activados por el contexto. Es decir, que, por otro lado, está el significado que inferimos a partir de las palabras que se encuentran en el contexto inmediato y por nuestro conocimiento del mundo. Puede verse que existen diferencias en los *dar* de *dar un lápiz*, *dar un consejo*, *dar un abrazo* y *dar miedo* –pese a que todos compartan la idea de *trayectoria* con un estado final–. Para explicar esta diferencia nos servimos de los anteriores presupuestos de la semántica cognitiva. Si estamos ante una estructura conceptual que no ha sido originada a partir de proyecciones metafóricas, sino únicamente a partir de imágenes que están en nuestra experiencia corporal cotidiana, nos estamos refiriendo a una estructura preconceptual, del tipo a la que se produce en “*A da [hace llegar, proporciona] una entidad material a C*”, o, lo que es lo mismo, “*A causa que una entidad material pase de no existir en C a existir en C*”; esto es, *alguien da el lápiz a alguien*. Así, las palabras que se encuentran en el contexto inmediato del verbo *dar* serán sustantivos que designen realidades materiales.

⁷⁴ Ruhl (1987) se refiere a este concepto con el término *inherente* y Alonso (1997) con el de *puente semántico*. Mel’čuk (1988: 29 citado en Alonso, 1997: 6) también alude a una lexía que está semánticamente vinculada de manera directa a la mayor parte de las otras lexías del vocablo.

El grupo Pragglejaz (2007: 3) presenta una serie de criterios para distinguir –con sistematicidad y rigor– en el discurso estos sentidos, que son los considerados *básicos* –esto es, los literales o físicos–, de los metafóricos o figurados. Así, el sentido más básico de una palabra sería el más concreto –“lo que evoca es más fácil de imaginar, de ver, o sentir”–, que está relacionado con acciones corporales –la correspondencia de este sentido básico con el concepto de Esquema de Imagen de Lakoff y Johnson es clara–, es más preciso –en oposición a vago– e históricamente anterior, y no es necesariamente el más frecuente de todos. Por otro lado, el significado que se proyecta a partir del anterior, a través de las categorías básicas o imágenes esquemáticas del significado básico que se da en el dominio físico para conformar un dominio abstracto, sería el figurado, el que encontramos en *dar un consejo, un abrazo* y *miedo*.

Esta relación entre ambos significados (literal o preconceptual y figurado o metafórico) es demostrable, pues, aunque el significado figurado contrasta con el significado básico, aquel puede ser entendido al ser comparado con este. Si el segundo no se hubiera construido a partir del primero, los dos significados no podrían compartir el significado sistémico: *A* causa que *B* pase de no existir a existir (en *C*, de manera que los seres *C* disfruten o padezcan *B*), lo cual se corresponde con aquellos rasgos del significado básico o preconceptual que permanecen en el figurado o metafórico. Se ejemplifica en la siguiente figura:

Figura 1. Relación de los significados del verbo *dar*



3.2.3. El significado y sus relaciones paradigmáticas y sintagmáticas

Desde la perspectiva de la semántica léxica, acabamos de señalar que cada palabra actualiza una acepción –un significado, ya sea el básico o uno metafórico– en el discurso en función de las piezas léxicas que la acompañan (recuérdense, asimismo, las aportaciones de la teoría del significado de Firth, a las que nos hemos referido en secciones anteriores). Así, la acepción que se actualiza en *dar un abrazo*, *clase* o *un salto* no es la misma que en *dar cariño*, *alojamiento* o *confianza*: mientras que en los primeros ejemplos el sentido es el de “ejecutar”, en los últimos se refiere a “proporcionar” –aunque los dos son significados figurados entendidos como proyecciones naturales del significado básico–. Esta diferencia significativa radica en la estructura interna del significado de una palabra, conformada por una serie de rasgos menores abstractos que intervienen en una gran variedad de conexiones significativas con otras palabras; los rasgos que se activan en una combinación no se activan en otra.

Así, las informaciones léxicas contenidas en una palabra justifican su selección por parte de otras palabras, por lo que determinan la manera en que la usamos o la combinamos dentro del discurso. Aprender una nueva palabra implica saber utilizarla en sus diferentes contextos combinatorios, lo que requiere un conocimiento de sus acepciones⁷⁵. En relación con ello, Lyons (1980) se refiere a la significación de una palabra como una serie de relaciones paradigmáticas que la unidad contrae con otras unidades de la lengua en los diferentes contextos de aparición. Dado que el significado puede descomponerse en rasgos menores, las relaciones que se dan entre los significados de las palabras tienen que ver con el hecho de que compartan algunos de estos rasgos. Así, sabemos que el sistema nos permite sustituir *dar* en *dar consejo* por *ofrecer*, y eso se explica porque tanto *dar* como *ofrecer* comparten ciertos rasgos semánticos contenidos en el significado de *consejo*, lo que determina su selección léxica. Hay autores (Bosque, 2005 [2004]; De Miguel, 2011) que se refieren a esto con la expresión *concordancia léxica*⁷⁶. Lo mismo sucede con *dar* e *impartir*, que son intercambiables cuando seleccionan *clase* o *curso*; sin embargo, no se puede **impartir un consejo* porque en este contexto los rasgos que comparte *dar* con *consejo* no los contiene *impartir*. Así, el contexto en que se encuentre la palabra o, lo que

⁷⁵ Este dominio de conocimiento forma parte del abarcado por la interfaz léxico-sintaxis, a la que nos referimos en el siguiente apartado 3.2.4, al tratar la competencia que lleva el mismo nombre.

⁷⁶ Al respecto de la concordancia de rasgos semánticos nos detenemos en mayor profundidad en el capítulo 4.

es lo mismo, las relaciones sintagmáticas en las que participe determinarán las relaciones paradigmáticas de esa acepción, pues cada significado entra en relaciones paradigmáticas y sintagmáticas propias.

La relación entre los rasgos que se da en el eje paradigmático puede ser, por un lado, de identidad, esto es, de sinonimia, como acabamos de ver. No obstante, convendría matizar que, como bien señalan Cervero y Pichardo (2000: 86), no hay sinónimos absolutos, ya que estos se diferencian desde una perspectiva estilística y también de combinatoria sintagmática, como es el caso de los ejemplos anteriores. En ellos veíamos que *dar* puede sustituirse por *impartir* cuando va acompañado de *clase*, *lección*, *conferencia*, *curso*..., es decir, de sustantivos que designan “eventos que se interpretan como unidades informativas, muy frecuentemente relativas a la presentación de contenidos”; sin embargo, se utiliza *dar* pero no *impartir* con otros sustantivos que, por extensión a los anteriores, expresan “eventos que se ofrecen a la contemplación pública”, como *concierto*, *espectáculo* o *mitin*. Por otro lado, la relación entre rasgos puede ser de inclusión, es decir, de hiperonimia e hiponimia o cohiponimia, o de holonimia y meronimia; y puede ser de oposición o exclusión, por lo que hablaríamos de antonimia.

Así las cosas, como observa Richards (1976), conocer una palabra conlleva un conocimiento de esta red de asociaciones paradigmáticas entre esa palabra y otras palabras en la lengua. Este conocimiento también se integra en la competencia léxica del hablante. No obstante, hemos de tener en cuenta que “Las asociaciones suscitadas por una unidad léxica [...] son hasta cierto punto individuales de cada persona, puesto que las vamos configurando de acuerdo con nuestra experiencia lingüística y vital (Oster, 2009: 35); y, además:

Con respecto al significado conceptual de una palabra deben añadirse algunas relaciones que participan en la construcción del significado, pero que se configuran a partir de la experiencia lingüística y vital de cada individuo, lo que provoca que el significado de una palabra adquiera aspectos únicos y variables para cada persona (Briño García, 2013: 17-18).

Estas asociaciones ayudan a recuperar y fijar las palabras, por lo que son tenidas en cuenta en el proceso de enseñanza. Pero, como bien observa Briño García, precisamente por su carácter individual y porque son relaciones en ausencia, y no en presencia, como las sintagmáticas, no podemos averiguar si los alumnos las tienen, ni, en tal caso, cuáles son. Por estas mismas razones, es imposible conocer sistemáticamente el dominio que los HNN tienen sobre este contenido –las asociaciones paradigmáticas– de la competencia semántica a partir

del análisis de un corpus. Vemos, por lo tanto, cómo el conocimiento léxico sobrepasa el conocimiento lingüístico –en correspondencia con la competencia lingüística a la que se refiere el *MCER*– pues también depende de la experiencia cognitiva, afectiva y sensorial del hablante con la realidad.

3.2.4. La interfaz léxico-sintaxis

Por otro lado, en la estructura semántica y conceptual del verbo el hablante reconoce una serie de argumentos: en el caso de *dar*, dos son internos, correspondientes a lo dado (tema) y a quien se da (destinatario); y otro es externo, correspondiente al dador (agente). Como señalaba Baralo (2001) en su modelo de organización del lexicón, esta red argumental léxico-semántica contiene toda la información requerida para proyectar la estructura sintáctica de la oración; en este caso, organiza una estructura ditransitiva en la que el hablante puede asignar función de sujeto sintáctico al agente, función de CD al tema y función de CI al destinatario. La selección léxica del verbo, como sugiere el estudio basado en corpus de Sinclair (1987), determina la construcción gramatical de la oración, para lo que se requiere una competencia léxico-sintáctica. Al mismo tiempo, el hablante cuenta con una competencia sintáctica, es decir, conoce los vínculos que establece el verbo con las demás piezas léxicas por medio de sus marcas formales, mediante relaciones de concordancia de rasgos; así, la flexión verbal refleja la concordancia del verbo en número y persona con la flexión nominal del sujeto, pero no con la del CD, con el que no tiene ninguna relación de dependencia estructural⁷⁷.

Si profundizamos en la mencionada competencia léxico-sintáctica, a un HNN, para poder usar un verbo transitivo no le basta con saber que la oración requiere un CD, sino que además ha de conocer el contenido semántico que deben tener las piezas que desempeñan esa función en la oración. El verbo selecciona rasgos (+/-animado, +/-concreto, +/-contenedor [rasgo señalado por De Miguel, 2011: 140]); en el caso de *dar*, el verbo selecciona, entre muchos otros, el rasgo subléxico [+desplazamiento breve], lo que explica su coaparición con *paseo*, *salto*, *paso* o *vuelta*; esto justifica a su vez por qué *dar* no se combina con

⁷⁷ El número o la persona, en el caso del sustantivo que desempeña la función de CD (de persona), expresan significados *inherentes*; es decir, su uso no tiene más restricción que la de adaptarse a las propiedades del estado de cosas que se pretende comunicar; sin embargo, el número y la persona en el verbo expresan significados *dependientes* desde una perspectiva estructural; la concordancia del verbo con el sujeto del que se predica muestra la dependencia estructural entre los dos constituyentes.

viaje, recorrido o excursión, sustantivos que se interpretan como desplazamientos largos (en el capítulo 4 nos detenemos en esta cuestión). De esta manera, el HNN tiene que aprender con qué palabras o tipos de palabra puede aparecer el verbo⁷⁸; ha de saber si el verbo ejerce algún tipo de restricción semántica sobre las palabras que lo acompañan en el eje sintagmático. Si se considera que no ejerce ninguna, se trata de combinaciones léxicas favorecidas por los hábitos de la comunidad de HN; y, en el caso de que el verbo sí ejerza restricción, se interpreta que la combinación está forzada por la naturaleza interna del sistema. Aun así, hay que distinguir cuál es el tipo y el grado de dicha restricción para sistematizar las combinaciones léxicas atendiendo a criterios puramente lingüísticos; pues, además de las restricciones semánticas, los verbos imponen restricciones categoriales (tipos de sintagmas, oraciones de infinitivo o flexionadas en indicativo o subjuntivo) –y este conocimiento también parece formar parte del conocimiento léxico del verbo–.

Otro aspecto que debe considerarse es el vinculado con el proceso de gramaticalización, involucrados en los casos en que *dar* se convierte en auxiliar, como en *dar por* (le da por llorar), o en verbo de apoyo (dar un abrazo) –en este último caso, la combinación constituye un predicado complejo en el que el verbo no predica en exclusiva ni de manera fundamental⁷⁹⁸⁰–.

En relación con esto último, y con el fin de facilitar el análisis de este conocimiento de la interfaz léxico-sintaxis en la interlengua, hemos realizado una propuesta de organización conceptual y terminológica de las combinaciones léxicas del verbo –apoyada en la coocurrencia de *dar* con el nombre que funciona como tema (normalmente el CD)– partiendo del concepto de selección léxica de Bosque (2001 y 2005 [2004]), basado en la motivación de la restricción de las combinaciones.

A este último tipo de conocimiento sintagmático –que va más allá de reconocer las funciones semánticas y sintácticas requeridas por el verbo– incluido en el dominio de la interfaz léxico-sintaxis lo hemos denominado

⁷⁸ Algunos autores se refieren a este dominio de conocimiento como el nivel de las redes léxicas (véase Baralo, 2006).

⁷⁹ En el capítulo 4 abordamos la cuestión de la naturaleza léxico-semántica de *dar*. Dejando a un lado las expresiones idiomáticas o fraseológicas, las combinaciones en las que participa este verbo no se forman, en su mayoría, aplicando libremente las reglas de la gramática, como sucede en *dar un caramelo*, sino que están determinadas, por un lado, por las características particulares de las piezas léxicas involucradas –como es el caso de *dar una explicación* o *dar un salto*, ejemplos de uso del verbo en sentido pleno y de apoyo, respectivamente– y por las restricciones categoriales que impone el verbo.

⁸⁰ A los casos de lexicalización, nos referimos a continuación.

subcompetencia léxico-combinatoria⁸¹, y en ella, a su vez, distinguimos la subcompetencia léxico-combinatoria sistemática y la subcompetencia léxico-combinatoria rutinaria (véase también Sánchez Rufat, 2010 y 2011, para una discusión sobre estos conceptos). Esta última es la que permite al hablante conocer la combinatoria asistemática y construir unidades fraseológicas en las que el verbo participa⁸², como *darse cuenta (de)* o *abí me las den todas*, en las que la restricción no viene impuesta por el sistema lingüístico, sino por factores en ocasiones difíciles de sistematizar, ya sean culturales, sociales o históricos⁸³. En las combinaciones asistemáticas es posible que se desconozcan de qué son producto estas agrupaciones de palabras, cuál es su motivación. Tal es el caso de *hacer una llave*, entendiéndose aquí por *llave* el encuentro físico que “consiste en hacer presa en el cuerpo del adversario, o en alguna parte de él, para inmovilizarlo o derribarlo” (DRAE, 2001). El hablante conoce que *dar*, en el

⁸¹ Hemos desestimado la expresión *subcompetencia colocacional* (introducida por Hill, 1999) porque el término *colocación* suele ir vinculado al de *frecuencia* y, como exponemos en anteriores trabajos (Sánchez Rufat, 2010), la frecuencia no es un criterio que aporte información sobre la naturaleza lingüística de una combinación (Bosque, 2001); frecuentes pueden ser estados, acciones o eventos que sean repetidos o habituales sin tener por qué serlo desde el punto de vista del sistema lingüístico, como sucede con las acciones diarias del tipo *lavar los platos*, *sacar la basura* o *ver la televisión*. Es, además, un concepto relativo, pues la combinación puede ser frecuente dependiendo de los tipos de textos que constituyen el corpus de análisis. Por todo esto, si el término *colocación* no restringe –en términos léxicos y sintácticos– los dominios que necesitan ser establecidos en estudios del lenguaje, consideramos que el empleo de este término y el concepto que designa no es útil como unidad funcional (diferenciada) de descripción en la interlengua ni en la didáctica de lenguas. Por lo aquí expuesto, a la hora de elaborar nuestra propuesta taxonómica no tomamos en consideración la clasificación de combinaciones léxicas señalada por Bally en 1909 (1951), en la que recurre precisamente al criterio de la frecuencia, aunque sin lugar a dudas supone un importante punto de partida para los estudios sobre fraseología y sobre colocaciones.

⁸² Pese a que no existe unanimidad entre los especialistas a la hora de clasificar el conjunto de unidades de las que se ocupa la fraseología, en este trabajo se incluye bajo el término *unidad fraseológica* un primer grupo formado por las unidades que son conmutables por un enunciado, como las paremias (refranes, proverbios, aforismos, citas, máximas, apotegmas, sentencias, dichos) y las fórmulas rutinarias y pragmáticas (que acogen las tradicionales locuciones interjectivas); y un segundo grupo constituido por unidades fraseológicas que no conforman un enunciado completo, esto es, las locuciones con función sintáctica de elemento oracional y los esquemas fraseológicos semiproductivos. A diferencia de las demás combinaciones léxicas, estas unidades pluriverbales institucionalizadas no son composicionales desde un punto de vista semántico (ni tampoco formal), de ahí su idiomatidad. En términos de no composicionalidad define Casares (1950: 170) la idiomatidad como la “combinación estable de dos o más términos, que funciona como elemento oracional y cuyo sentido unitario consabido no se justifica, sin más, como una suma del significado normal de los componentes”; en este mismo sentido es en el que Bolinger (1976: 100) define *idiom* como “groups of words with set meanings that cannot be calculated by adding up separate meanings of the parts”, o al que Makkai se refiere en su introducción al *A Dictionary of American Idioms* (1975: IV): “the assigning of new meaning to a group of words which already have their own meaning”.

⁸³ Para facilitar nuestra clasificación de las competencias involucradas en la competencia léxica, hemos decidido integrar los conocimientos fraseológicos en la competencia léxico-sintáctica, aunque nada o poco tengan que ver con el conocimiento de los recursos gramaticales de una lengua y la capacidad de utilizarlos.

sentido de “hacer efectivo o ejecutar”, se combina con sustantivos que designan “ciertos movimientos, reales o figurados, con frecuencia bruscos, vehementes, impulsivos o resultantes de algún contacto” (Redes, 2005: 639 [2004]), como *abrazo, apretón, espaldarazo, empujón, estirón, toque, manotazo, bofetón* o *codazo*, a los que podemos añadir otros muchos. Sin embargo, debe asimilar que es el verbo *hacer* el que se combina con *llave*, pese a que este sustantivo pertenezca también a la clase léxica descrita. Y es que, como señala Alonso (2004: 94), no es posible controlar semánticamente todos los nombres que aparecen con el verbo. Si se considera que esta combinación –hacer una llave– es ajena a la descripción del sistema lingüístico, al igual que la combinatoria en las unidades fraseológicas, hay que hablar de norma o rutina, de ahí el nombre que le hemos asignado a esta subcompetencia. La adquisición y uso de las combinaciones rutinarias se producen en bloque –con esto no queremos decir que todas las secuencias que adquirimos y usamos en bloque sean fraseológicas, como ha quedado demostrado a partir del ejemplo de combinatoria asistemática, la cual es composicional–; dichas combinaciones se almacenan y recuperan de manera holística (recuérdese a propósito de esto el principio de idiomaticidad, de Sinclair, 1991), lo cual depende únicamente de la capacidad memorística, y no del conocimiento del sistema lingüístico del español.

Por el contrario, en el conocimiento que implica tener subcompetencia léxico-combinatoria sistemática entran en juego las propiedades combinatorias de las palabras⁸⁴; es decir, las combinaciones sistemáticas son producto del carácter restrictivo del sistema lingüístico.

Los predicados seleccionan sus argumentos restringiendo el conjunto de piezas léxicas que pueden denotar de acuerdo con sus rasgos semánticos, por lo que las unidades seleccionadas no son piezas aisladas –como en el caso de las combinaciones asistemáticas (recuérdese *hacer una llave*)–, sino nociones muy diversas que pueden ser denotadas por grupos de sustantivos que son seleccionados por el verbo (lugares, objetos físicos, movimientos bruscos o impulsivos resultantes de algún contacto físico). La teoría de los predicados de Bosque, de la que nos servimos, explica la capacidad de generación de estas restricciones sistemáticas; atender a estas restricciones no como combinaciones en las que la relación entre los elementos es biunívoca, sino como casos de selección léxica, implica adquirir –de entrada– un conocimiento de la lengua

⁸⁴ Los dos diccionarios combinatorios que existen en español, *Redes* (2005 [2004]) y *Práctico* (2006) (*Dice* es un diccionario de colocaciones que está en proceso de creación; véase <<http://www.dicesp.com/paginas>> [05-05-2014]) contienen información, fundamentalmente de naturaleza sistemática, que no está contenida en ninguna otra obra lexicográfica del español.

generativo (basado en reglas) –frente al conocimiento formulaico, basado en la memoria, que es el único que se puede desarrollar en los casos de las restricciones rutinarias–, pues una misma relación léxico-sintáctica permite generar diferentes combinaciones⁸⁵. Tanto en las combinaciones asistemáticas como en las restricciones sistemáticas estamos ante la combinación de dos piezas léxicas que mantienen su significado estable; no obstante, el hecho de que estas combinaciones sean composicionales no quiere decir que –como acabamos de comprobar con el contraste entre, por ejemplo, *dar un empujón* o *un manotazo* y *hacer una llave*– su formación esté exenta de problemas.

Estas restricciones que, desde el punto de vista sistemático, los predicados imponen a sus argumentos pueden ser, a su vez, de dos tipos⁸⁶: si están fundamentadas únicamente en nociones de naturaleza física (o corpórea), se puede construir el paradigma extensional de acuerdo con el conocimiento que tenemos del mundo, de ahí que hayan sido denominadas restricciones sistemáticas extralingüísticas; estas guardan relación con las restricciones selectivas introducidas en la gramática generativa en los años sesenta, que consisten en rasgos semánticos del tipo “humano”, “animado”, “objeto material”, “objeto líquido” y otros similares. El verbo actualiza en estas combinaciones la acepción básica (preconceptual de acuerdo con el modelo cognitivo de Lakoff y Johnson); por ejemplo, el verbo *dar*, en su sentido básico de “entregar, hacer llegar, proporcionar” se combina con sustantivos que designan entidades materiales; esa restricción, por tanto, está fundamentada en una noción física –sustentada, a su vez, en la idea de desplazamiento de índole espacial característica del verbo–, por lo que las piezas con las que el predicado se combina se deducen fácilmente. Una vez que conocemos el significado del predicado, seamos hablantes nativos o no (en otras lenguas probablemente existe un término equivalente con igual extensión), podemos añadir piezas léxicas a esta clase aplicando nuestro sentido común.

Por el contrario, si la restricción no está basada únicamente en nociones físicas, se tiene que recurrir al conocimiento del idioma para obtener las piezas léxicas que conforman la extensión, de ahí que sean denominadas restricciones sistemáticas intralingüísticas. A diferencia del grupo anterior, la restricción viene dada en función de las propiedades del predicado, de su estructura interna. No

⁸⁵ Obsérvese que esta productividad se pierde si la direccionalidad de la selección de naturaleza semántica parte de los argumentos para llegar a los predicados, como sucede con los trabajos derivados de la Teoría Sentido-Texto. Sobre las ventajas de una direccionalidad frente a la otra, véanse los trabajos de Bosque (2001 y 2004).

⁸⁶ Para establecer esta distinción, partimos del trabajo de Bosque (2005 [2004]).

nos basta con nuestro conocimiento del mundo, y necesitamos un conocimiento profundo del verbo para obtener el paradigma extensional⁸⁷. Este no tiene por qué coincidir en otra lengua, de ahí que pueda ser una fuente de interferencias. En estos casos de restricción sistemática intralingüística, el predicado verbal selecciona sus argumentos restringiendo el conjunto de piezas léxicas que pueden denotar de acuerdo con sus rasgos semánticos, por lo que las unidades seleccionadas no son piezas aisladas –como en el caso de la combinatoria asistemática–, sino clases léxicas cuyos integrantes comparten ciertos rasgos semánticos (Bosque, 2005 [2004]). Así, el predicado *dar*, en el sentido de “entregar, hacer llegar, proporcionar”, se combina también con sustantivos que denotan “asistencia, sustento o protección”, como *ayuda, servicio, apoyo, protección, sustento, alojamiento, atención, auxilio, respaldo*, etcétera (Redes, 2005: 634 [2004]).

Pese a que este verbo se combina con sustantivos pertenecientes a diferentes clases léxicas que, a priori, partiendo del significado del verbo, no podemos señalar, el hecho de que el sentido figurado no sea más que una extensión natural del literal⁸⁸ –o, lo que es lo mismo, una proyección metafórica

⁸⁷ Alonso Ramos (1997) se refiere a estos conceptos que integran la interfaz léxico-sintaxis (dejando a un lado las unidades fraseológicas) con los términos *coocurrencia léxica libre* y *coocurrencia léxica restringida*, pero la segmentación del dominio que abarcan ambos conceptos no coincide del todo con la nuestra, a juzgar por los ejemplos aportados por la autora. La coocurrencia libre atañe a la semántica y se deduce de la definición de las unidades léxicas en cuestión; esta se corresponde con nuestras restricciones sistemáticas extralingüísticas, pues las combinaciones intralingüísticas –y por descontado la combinatoria rutinaria– no las podemos codificar a partir de la definición de las palabras involucradas, a menos que descompongamos la estructura semántica de la palabra y aportemos definiciones semánticas mucho más detalladas y completas; en tal caso, el concepto de Alonso Ramos se correspondería no solo con nuestra combinatoria sistemática extralingüística, sino también intralingüística. Solo teniendo esto en cuenta aceptaríamos que *dar datos* (ejemplo propuesto por la autora) es producto de la combinatoria libre del verbo, pues el verbo en cuestión se combina con sustantivos de información, como *explicación, testimonio, detalle, dato, pista, opinión...* En cambio, la coocurrencia léxica restringida atañe al léxico, no solo a la semántica, y se da cuando –en este caso– el verbo es seleccionado por otros lexemas. Independientemente de que no coincidamos en la direccionalidad de la selección, esta explicación conecta con nuestra combinatoria asistemática, pero en los ejemplos no se produce esta equivalencia. Para la autora, *dar un paseo, un beso* o *permiso* son muestras de coocurrencia restringida, mientras que para nosotros son sistemáticas intralingüísticas, es decir, podríamos codificar estas combinaciones a partir de la descomposición de la definición semántica del verbo; así, de acuerdo con la información obtenida en *Redes*, el verbo *dar*, en el sentido de “ejecutar”, selecciona sustantivos que denotan “desplazamiento, frecuentemente si es breve o repentino”, como *paseo*; selecciona también sustantivos que designan “ciertos movimientos impulsivos o resultantes de algún contacto”, como el *beso*; y, en el sentido de “proporcionar”, selecciona la clase léxica formada por sustantivos que denotan “autorización, aceptación y otras formas de asentimiento o aquiescencia”, como *permiso*.

⁸⁸ Esta relación explica que consideremos como combinaciones composicionales las combinaciones proyectadas por el verbo cuando este adopta sentidos figurados. Algunos investigadores incluyen parte de estas combinaciones en el grupo que denominan *colocaciones*, definidas como “un vínculo léxico-semántico condicionado por la sintaxis” (Koike, 2001: 176). No obstante, muchos de ellos consideran que son unidades estables que presentan una fijación mayor que la de las combinaciones libres, por lo que se establece que no siempre son composicionales

que toma como dominio de origen el Esquema de Imagen del CAMINO– nos permite descodificar el significado de la combinación cuando el verbo actualiza un significado figurado, pero no codificar la combinación. Es decir, estas diferencias en la naturaleza de la restricción del verbo –extralingüística o intralingüística– ponen de manifiesto que los usos figurados o metafóricos de las palabras pueden interpretarse como extensiones naturales de los usos literales, como el producto de una proyección a partir del significado básico que se da en el dominio físico para conformar un dominio abstracto a través del cual se amplía nuestro conocimiento del mundo desde lo conocido. No obstante, pese a que, como se acaba de argumentar, concebamos esta proyección como algo natural, propia de nuestra conceptualización cognitiva del mundo –lo que, sin duda, contribuirá a facilitar el proceso de descodificación de las estructuras–, la formación de estas últimas combinaciones (intralingüísticas) requiere un conocimiento lingüístico profundo –del sistema–, lo que puede suponer un incremento en el grado de dificultad en el momento de la codificación.

En suma, el marco teórico de nuestra propuesta taxonómica de los conceptos que integran la interfaz léxico-sintáctica, fundamentalmente en lo que se refiere a la subcompetencia léxico-combinatoria –centrada en la coocurrencia del verbo con su argumento tema (el nombre que funciona normalmente como su CD)–, está basada en el enfoque lexicográfico de *Redes*. Con ello pretendemos mostrar cómo este enfoque puede facilitar la compleja tarea en la descripción de un verbo como *dar*. La cuestión de la naturaleza de la restricción que los predicados imponen a sus argumentos y el concepto de selección de clases léxicas pueden ser una buena herramienta para describir rigurosamente el uso que se hace de este verbo en la interlengua.

3.2.5. El factor sociolingüístico

Además de los aspectos ya señalados, en el análisis de las propiedades que integran el conocimiento léxico del verbo no podemos pasar por alto que una unidad léxica es un elemento contextualizado en una situación comunicativa concreta que es producida en un acto pragmático específico (Cervero y Pichardo, 2000: 58). Al salto que se produce –con la llegada en los años setenta de los planteamientos teóricos funcionalistas– de la oración como unidad máxima del

desde el punto de vista semántico, aspecto con el que nosotros no estamos de acuerdo, pues estas combinaciones intralingüísticas las concebimos siempre como composicionales.

análisis lingüístico –proyectada por las características léxico-sintácticas del verbo– al discurso, se une el desarrollo de la competencia sociolingüística, que comprende el conocimiento y las destrezas necesarias para abordar la dimensión social del uso del verbo (véase *MCER*, 2002: 116); es decir, esta competencia capacita, por un lado, para distinguir los distintos significantes que la situación comunicativa requiere para un mismo significado, lo que implica manejar los diferentes niveles de uso (formal, neutro y coloquial) de la lengua. Aquí podemos rastrear la influencia de la teoría sistémico-funcional de Halliday (1977), y sus parámetros del contexto situacional (el campo, el tenor y el modo), que determinan el registro en cada situación: el contexto provoca la elección de un lexema y no otro –con el que comparte ciertos rasgos intensionales y, por tanto, el comportamiento sintagmático y la carga extensional–, lo que conlleva un conocimiento de las relaciones paradigmáticas del verbo –más concretamente nos referimos aquí a las estilísticas–, gracias al cual se logra ser más efectivo en la colocación con otras palabras a la hora de reflejar una idea; estas diferencias, como se ha dicho, están condicionadas por el contexto, por lo que ello no implica que se dé un significado diferencial. En suma, saber en qué registros y contextos debe usarse el verbo posibilita que la producción del HNN resulte natural a oídos nativos. Así, aunque *dar*, en principio, es un verbo neutro en español⁸⁹, cuando un HNN produce el enunciado “Aprendí que el museo fue establecido antes de su muerte y después Miró le *dio muchas obras*”⁹⁰, no se le puede asignar una marcación neutra al verbo, sino de cierta pobreza léxica, pues su grado de especialización semántica se halla por debajo de lo que un HN consideraría natural. Un verbo como *donar*, sin embargo, sí cumpliría esa función en la situación descrita y podría ser marcado de forma neutra.

Por otro lado, la competencia sociolingüística capacita para distinguir la expresión formal que la comunidad de hablantes de una variedad geográfica determinada ha sancionado⁹¹ para el significado en cuestión. Por ejemplo, al analizar los datos de corpus nativo, puede comprobarse⁹² que en el estándar americano resulta natural y frecuente la combinación *dar gusto* tanto en estructuras transitivas (*da gusto nadar en el mar*) como en impersonales (*así da gusto*)

⁸⁹ A un elemento se le aplica una marcación neutra cuando se usa en situaciones que de por sí no requieren un mayor grado de especialización semántica para que resulten naturales a oídos nativos (Montoro, 2010: 257).

⁹⁰ Ejemplo obtenido del corpus no nativo CEDEL2.

⁹¹ Como es usual en los estudios normativos, en este trabajo se emplea *sancionar* con el sentido de “Autorizar o aprobar cualquier acto, uso o costumbre” (*DRAE*, 2001).

⁹² Para conocer las cifras de los datos obtenidos del análisis de este ejemplo en los corpus CREA y en el de Davies, véase el capítulo 5. En él, además, profundizamos en el manejo de todos estos conceptos (*estándar, marcación neutra, etcétera*) y recuperamos los ejemplos citados.

y en ditransitivas (*me da gusto saber que vendrás*); en el estándar peninsular, en cambio, el valor en el grado de disponibilidad es el mismo en cuanto a las estructuras transitivas e impersonales, pero en las ditransitivas se reduce a contextos más restringidos, en ocasiones con cierta connotación sexual. Por consiguiente, un aprendiente que maneje la variante peninsular no deberá recurrir al uso ditransitivo de *dar gusto* en estructuras ditransitivas.

En esta competencia incluimos también el manejo de las coapariciones preferentes de palabras, entendidas como la opción combinatoria más frecuente, pero no exclusiva. Son combinaciones preferentes, por un lado, lo repetido o habitual, en el sentido de extendido sin tener por qué serlo; tal es el caso del sustantivo *regalo(s)* con el verbo *hacer* y no con *dar*. Si un corpus amplio como el CREA nos dice que *hacer* es más frecuente con *regalo* que *dar* (45 casos frente a 4 teniendo en cuenta que en todos ellos, desde la perspectiva situacional, los dos son intercambiables)⁹³, no nos está proporcionando ninguna información puramente lingüística; ambos predicados, a partir de una restricción fundamentada en nociones de naturaleza física, se combinan con sustantivos que designan entidades materiales, pero esta clase –formada por las entidades materiales que se pueden dar o hacer– no es lingüística (ni semántica ni pragmática). Simplemente la comunidad de hablantes parece preferir, al menos en la forma del infinitivo, el verbo *hacer* frente a *dar* cuando se trata de seleccionar este sustantivo; de hecho, *hacer* es la primera opción que aparece en *Práctico* en la entrada *regalo*, pese a que nada en el sistema impide que pudiera haber sido al contrario.

Por otro lado, hay quienes consideran que en el ámbito de lo sistemático intralingüístico las combinaciones son coapariciones preferentes de palabras (Higueras, 2006; Koike, 2001, entre otros); sin embargo, estas palabras coaparecen porque su comportamiento se ajusta a algún sistema, es decir, la combinación se da con frecuencia porque el verbo restringe como argumentos los elementos de una clase léxica en función de las propiedades semánticas del predicado, por lo que la frecuencia combinatoria radica en el hecho de que los elementos que forman la combinación están sujetos a ese sistema. Así, el esquema fraseológico semiproductivo *dar de* se combina frecuentemente con *comer* –expresión que contiene un elemento implícito: *dar algo de comer*–, y ello se debe a que ese predicado verbal selecciona los sustantivos que denotan “ingestión”; por consiguiente, dado que *comer* se incluye en esa clase, es esperable

⁹³ Consulta realizada de acuerdo con los parámetros fijados por Sinclair (1991): hacer dist/5 regalo y dar dist/5 regalo; esto es, la palabra *regalo* debe aparecer a una distancia medida en palabras no superior a 5 de la palabra *hacer* o *dar* (*de*).

que este infinitivo se combine frecuentemente con este verbo y no con otro. No obstante, hay que decir que, incluso en estos casos, el sistema a veces nos ofrece varias opciones y una de ellas es elegida por preferencias o inclinaciones de la comunidad de hablantes. Por continuar con el mismo ejemplo, este es el caso de *dar de* y *dar a*: ambos esquemas seleccionan verbos que denotan “ingestión” (Redes, 2005: 642 [2004]), pero las preferencias por *dar de comer, cenar, merendar* frente a *dar a comer, a cenar, a merendar* se pueden constatar⁹⁴.

En definitiva, conviene que el aprendiente conozca el grado de frecuencia del verbo y su integración en el sistema, para saber si está ante un elemento central en el sistema o marginal-periférico.

Así, hay factores que limitan dónde, cuándo y cuánto ciertas palabras se pueden usar. Desconocer estas restricciones puede implicar un uso inapropiado, por lo que deben ser consideradas en un análisis de interlengua. Un uso puede ser arcaizante, anticuado o neologismo; puede ser vulgar, coloquial, neutro o elevado; puede pertenecer al lenguaje de un grupo determinado; puede ser afectado, irónico, tabú o cómico; o desconocido en una determinada área geográfica, puede ser inusual. En otras palabras, la emisión de un enunciado está condicionada por factores extralingüísticos, como la relación entre los interlocutores (jerarquía o cercanía), el nivel de formación del hablante, su sexo y edad, la localización espacial en que se produce la emisión, el tema tratado (variedad diafásica, diastrática, diatópica) y la preferencia de uso en determinados tipos o géneros textuales.

3.2.6. Rasgos de la competencia léxica del verbo

Tras haber indagado hasta aquí en la implicación de toda esta serie de subcompetencias comunicativas en la adquisición del léxico, esquematizamos en la tabla que sigue⁹⁵ el conocimiento léxico del verbo –desde la perspectiva únicamente del aprendiente de español como emisor de un mensaje escrito–, atendiendo a todos los aspectos relevantes en cuanto a su forma, su significado y su uso. Así, en las dos columnas de la izquierda se incluyen las competencias involucradas en el conocimiento léxico del verbo; de esta manera, es fácil observar cómo este conocimiento se extiende por otros dominios de la competencia lingüística en una compleja interfaz. Nos hemos basado en las

⁹⁴ Sobre el índice de frecuencia de cada uno de estos ejemplos nos detenemos también en el capítulo 5. Sobre frecuencia y preferencia, véase Bosque (2005: LXXXI [2004] y 2001).

⁹⁵ Partiendo de Nation (2001: 27) y Baralo (2006: 2).

competencias recogidas en el *MCER* (primera columna de la izquierda), algunas de las cuales han sido eliminadas y otras reorganizadas, renombradas o añadidas (segunda columna de la izquierda), con la única intención de facilitar la labor de análisis del verbo en un estudio de interlengua. En la tercera columna se desglosa este conocimiento o competencia léxica en los diferentes tipos de capacidades requeridas para su desarrollo, de acuerdo con la competencia involucrada en cada caso; y en la columna de la derecha incluimos los niveles lingüísticos correspondientes a cada tipo de saber.

Competencias según el <i>MCER</i>	Competencias involucradas	Competencia léxica del verbo	Niveles lingüísticos
Competencia ortográfica	1. Competencia ortográfica	Conocer la correcta escritura del verbo	
Competencia gramatical	2. Competencia morfológica	a) Conocer las formas irregulares, tanto en los aspectos que afectan a la base como a los elementos flexivos b) Reconocer la raíz del verbo y conocer los mecanismos de formación de las palabras derivadas	Morfológico
Competencia semántica	3. Competencia semántica	a) Conocer los significados: básico y figurados. Conocer lo que está incluido en el concepto y reconocer las actividades que puede referir el concepto al que remite el verbo b) Conocer los límites entre la palabra y otras relacionadas	Semántico
Competencia gramatical	4. Competencia sintáctica	Conocer los vínculos que establece el verbo con las demás piezas léxicas por medio de sus marcas formales	Sintáctico
Competencia léxica ⁹⁶	5. Competencia léxico-sintáctica: 5.1. Subcompetencia léxico-sintáctica 5.2. Subcompetencia léxico-combinatoria: 5.2.1. Subcompetencia léxico-combinatoria rutinaria 5.2.2. Subcompetencia léxico-combinatoria sistemática	5.1. a) Reconocer la red argumental léxico-semántica del verbo (funciones semánticas) requerida para proyectar la estructura sintáctica (funciones sintácticas) b) Conocer las restricciones categoriales y los aspectos de gramaticalización en los que está involucrado 5.2. Conocer con qué palabras y tipos de palabras se emplea el verbo 5.2.1. Conocer las restricciones normativas, ajenas al sistema lingüístico: a) Conocer la combinatoria asistemática b) Conocer las unidades fraseológicas en las que participa el verbo 5.2.2. Conocer las restricciones que los predicados imponen a sus argumentos, producto del carácter restrictivo del sistema lingüístico: a) Conocer las restricciones extralingüísticas b) Conocer las restricciones intralingüísticas	Léxico-sintáctico
Competencia sociolingüística	6. Competencia sociolingüística	a. Conocer en qué registros se usa el verbo b. Saber si la expresión formal en la que participa el verbo es la sancionada para un significado determinado en la variedad geográfica manejada c. Conocer las coapariciones preferentes del verbo, frente a las demás posibles	Todos los niveles lingüísticos (morfológico, semántico, sintáctico, léxico-sintáctico)

Tabla 1. Conocimientos necesarios para la producción escrita del verbo

⁹⁶ Los contenidos de esta competencia no se corresponden en su totalidad con nuestra competencia léxico-sintáctica, esto es, el cambio no es meramente terminológico. En el *MCER* se tiende a simplificar los contenidos o presentarlos de modo laxo. Esta falta de correspondencia es aplicable al resto de las competencias.

Como se desprende de esta tabla, puede entenderse el conocimiento léxico como una competencia transversal, que contiene información codificada sobre la forma de las palabras, su significado literal o figurado, así como su flexión y derivación, su función sintáctica, su combinatoria y su valor comunicativo. Respecto de la combinatoria, gracias a la cantidad de etiquetas, tanto categoriales como conceptuales, con las que se almacena el verbo en el lexicon, puede reconocerse la red argumental léxico-semántica requerida para proyectar la estructura sintáctica de la oración, y pueden conocerse los vínculos que establece el verbo con las demás piezas léxicas por medio de sus marcas formales, al mismo tiempo que se conocen las restricciones combinatorias en las que entra el verbo; en cuanto al valor comunicativo, se refiere a las restricciones de uso que permiten adecuar el verbo al contexto situacional correspondiente de modo que suene natural a los oídos nativos. El conocimiento léxico es, por consiguiente, “una representación mental de gran complejidad, que integra diferentes aspectos y componentes cognitivos, algunos más automáticos e inconscientes otros más conscientes, reflexivos y experienciales” (Baralo, 2007: 41). Estos aspectos se van adquiriendo mediante un proceso acumulativo –los conocimientos nuevos se sustentan sobre los existentes– (Higueras, 2006; Baralo, 2006; Nation, 2001; Schmitt, 2000) que tiene lugar a lo largo de toda la vida, frente a la relativa estabilidad de la competencia gramatical:

La gramática de una lengua se adquiere y completa a una determinada edad, no así el lexicon mental. Este se amplía y reestructura constantemente a lo largo de la vida de los hablantes. Esto quiere decir que el lexicon mental crece paralelo al conocimiento del mundo que puede llegar a adquirir una persona (Luque Durán, 2001: 225).

Aunque, como señala Luque Durán, el lexicon mental nunca adquiere una estructuración terminada –lo que hace que existan diferencias en los lexicones mentales de diferentes personas–, se puede decir que los hablantes de una misma lengua comparten básicamente el mismo lexicon mental. Este razonamiento implica algo ya mencionado anteriormente: que conocer una palabra es una cuestión de grado –que va desde el reconocimiento hasta la producción en una situación comunicativa determinada–:

For any lexical aspect, learners can have knowledge ranging from zero to partial to precise. This would mean that all word knowledge ranges on a continuum, rather than being known versus unknown (Schmitt, 2000: 118).

Es una cuestión, por lo tanto, de grado de profundidad; y en este contexto, la profundidad –entendida como el conjunto de aspectos que se conocen del verbo y que indican el nivel de conocimiento alcanzado– se relaciona con la calidad del conocimiento del verbo por parte de los aprendientes, lo que supone el control de la producción del verbo.

Partiendo de estas consideraciones, en el análisis de interlengua centrado en un verbo concreto interesa averiguar en qué punto del proceso de la adquisición del verbo se encuentra el grupo de HNN al que correspondan las muestras analizadas –pese a las posibles diferencias individuales en el lexicon–, cuáles de estos diferentes tipos de conocimientos que comprenden el dominio del verbo –los que se acaban de especificar en este apartado– están adquiridos y cuáles se encuentran todavía pendientes de ser aprendidos.

Hemos visto a lo largo de este capítulo que las teorías más recientes de adquisición lingüística otorgan al léxico un papel primordial en el aprendizaje de una lengua, y que las nuevas propuestas didácticas coinciden en señalar que solo partiendo de un enfoque léxico es posible desarrollar una competencia comunicativa similar a la de un nativo, pues sin el vocabulario no se puede transmitir ni interpretar nada. Unido a ello, el léxico contiene información sobre la estructura y el significado oracional y, además, abarca y requiere el desarrollo de todas las competencias.

Acorde con estas nuevas tendencias, nuestro análisis de interlengua se integra en este marco de revalorización del léxico: mediante este estudio pretendemos descubrir cuál es el grado de conocimiento de un verbo frecuente por parte de un grupo de HNN concretos, con el fin de contribuir a la mejora de la descripción de la competencia léxica de este grupo en un estadio de interlengua específico. Es de suma importancia descubrir las características del lexicon –tanto su extensión como su profundidad– en los diferentes estadios de interlengua de los grupos particulares, pues

la competencia léxica es la base del desarrollo de las destrezas y habilidades comunicativas en una lengua extranjera [...]. La capacidad de un hablante no nativo para formular un enunciado lingüístico, oralmente o por escrito, y la capacidad de identificarlo, de reconocerlo y de interpretarlo, cuando lo escucha o lo lee, se sustenta en el conocimiento léxico y en su disponibilidad para usarlo en ambos procesos cognitivo-comunicativos (Baralo, 2007: 384).

De acuerdo con lo expuesto, estudios como el que presentamos aquí pueden desembocar en la actuación en el aula de LE y en la creación de materiales para ello (en el capítulo 9 se presentan algunas aplicaciones didácticas).

4. EL VERBO *DAR*: NATURALEZA LÉXICO-SEMÁNTICA Y PROYECCIÓN SINTÁCTICA

El verbo *dar* es uno de los más representativos del grupo de verbos frecuentes. De acuerdo con *A Frequency Dictionary of Spanish*, de Davies (2006: 13), este verbo ocupa la décima posición de los 1071 verbos registrados en este corpus de 20 millones de palabras; y en De Miguel (2011: 141), en Lozano Zahonero (2010: 88-89) y en la *Nueva Gramática de la Lengua Española. Manual* (2010: 670) es introducido como el verbo de apoyo más frecuente en español, el que tiene mayores posibilidades combinatorias; en esta misma línea, Herrero Ingelmo (2002: 6) presenta este verbo en el primer lugar del grupo de los verbos de apoyo de amplio espectro, es decir, aquellos que actualizan un mayor número de sustantivos predicativos.

Si en el capítulo 3 se ha presentado una propuesta formalizada de todos los aspectos que deben ser considerados en el análisis del conocimiento escrito del verbo en general, este capítulo está dedicado por completo a la caracterización del verbo *dar*. Con la misma intención manifestada en el capítulo anterior de desarrollar herramientas eficaces para la categorización de los datos analizados del corpus y para facilitar la explicación de los resultados obtenidos, el objetivo ahora es proponer una descripción de los diferentes significados que actualiza este verbo y de las diferentes estructuras en las que participa.

Para ello indagaremos en la estructura léxica del verbo *dar* a través de sus diferentes comportamientos léxico-sintácticos, de sus coocurrencias, ya que, como apunta De Miguel (2011: 40), donde verdaderamente se visualiza la estructura léxica es en la sintaxis. Así, esta exploración semántica y sintáctica del verbo *dar* se realiza, por un lado, aplicando a las coocurrencias del verbo *dar* contenidas en el corpus CEDEL2 los presupuestos teóricos de Bosque (2005, 2001b) sobre combinatoria léxica y los planteamientos de De Miguel (2008 y 2011) sobre el verbo de apoyo; y, por otro lado, haciendo uso de la información contenida en la entrada del verbo *dar* del diccionario combinatorio *Redes*, que nos ha resultado fundamental para observar y analizar las tendencias en el comportamiento de este verbo.

Por lo que respecta a la estructura del capítulo, en el apartado 4.1 abordamos la cuestión de la naturaleza léxico-semántica de *dar*, ya que –dejando a un lado las expresiones idiomáticas o fraseológicas– las combinaciones en las que participa no se forman, en su mayoría, aplicando libremente las reglas de la gramática, como sucede en *dar un caramelo*, sino que están determinadas por las características particulares de las piezas léxicas involucradas, como es el caso de

dar una explicación o dar un salto. En el apartado 4.2 se sistematizan en forma de tabla las conclusiones alcanzadas a partir de las cuestiones teóricas y metodológicas planteadas en el apartado 4.1. En estas tablas se categorizan los diferentes significados que el verbo actualiza en las construcciones V + N, y en las demás construcciones sintácticas en las que participa el verbo –registradas en el corpus CEDEL2–. Dichas tablas son herramientas imprescindibles para la clasificación de los datos, y nos sirven de apoyo en el capítulo 7, en el que se analiza el verbo.

4.1. CONSIDERACIONES SOBRE LA NATURALEZA LÉXICO-SEMÁNTICA DE *DAR*

Uno de los rasgos con los que se suele caracterizar los verbos frecuentes del tipo *dar*, *hacer* o *tomar* es el de la polisemia: distintas acepciones contenidas en el mismo lema. El verbo *dar* es un verbo muy polisémico, y este fenómeno plantea problemas léxicos y semánticos a la hora de establecer nuestros dos propósitos principales: 1) formular una definición que dé cuenta del comportamiento sintáctico de *dar*, lo cual permita 2) describir distintivamente los sentidos de este verbo. A estas dificultades nos referimos a continuación, en las secciones 4.1.1 y 4.1.2, respectivamente.

4.1.1. Hacia una definición de dar que se visualice en la sintaxis: problemas encontrados y soluciones para el análisis de los datos

Como paso previo a este primer propósito, conviene considerar los planteamientos de Coseriu en torno a la polisemia –a la que se refiere como la existencia de “varios significados unitarios y delimitables”– y a la necesidad de que sea correctamente identificada ante el riesgo de que pueda quedar oculta tras la etiqueta homofonía (1981: 205-206):

Antes de optar por la homofonía, hay que preguntarse si no se trata de variación: en todos los casos en que las homofonías no sean evidentes (...), hay que buscar primero aquello que las lenguas deben tener para ser tales, es decir, invariantes de significado o “significados unitarios”. Y solo cuando resulte absolutamente imposible “reducir” todas las acepciones de una forma a un valor unitario de lengua, será lícito admitir homofonía.

Estas consideraciones serán tenidas en cuenta en el análisis de los datos del corpus: siempre que en un uso de *dar* se pueda rastrear ese significado unitario, común, será interpretado como una acepción de una misma palabra, como un caso de polisemia. Ahora bien, describir esa polisemia, esto es, distinguir las diferentes acepciones de una palabra, es una tarea compleja, considerada, según Alonso Ramos (1997: 4), como uno de los problemas cruciales de la lexicografía (a este aspecto está dedicada la sección 4.1.2); la consecución de esta tarea requiere –como se acaba de señalar– una actividad previa no menos complicada: ofrecer una definición de una palabra polisémica que sea unitaria y que permita discriminar las acepciones de una misma palabra de otros lexemas homófonos⁹⁷. A continuación, proponemos una definición que conjuga el comportamiento del verbo tanto en combinatoria sistemática (que atiende, por lo tanto, a las restricciones del verbo extralingüísticas e intralingüísticas) como en combinatoria composicional asistemática (véase el capítulo 3 para la definición de estos conceptos); después, exponemos los problemas que esta definición ocasiona y las soluciones alcanzadas para el análisis de los datos del corpus.

4.1.1.1. Sobre los contenidos aspectuales y las idiosincrasias léxicas en el significado unitario de *dar*: la trayectoria y la meta en la transferencia y en la creación

En el capítulo 3 nos hemos referido al valor unitario de las palabras con la expresión *significado sistémico*, presente en el uso básico del verbo (literal o físico), del tipo *dar un libro*, y en los metafóricos (traslaticios o figurados) –considerados estos como proyecciones naturales de aquellos–, que se encuentran tanto en las CVP del tipo *dar cariño* como en las CVA⁹⁸ del tipo *dar un salto* (distinción que se avanzaba en el apartado 1.4). Por consiguiente, como ya se señaló allí, una

⁹⁷ La mayoría de los autores que centran sus trabajos en un análisis teórico de la combinatoria léxica del verbo *dar* (Alonso Ramos, 1997; Herrero Ingelmo, 2002; Mitatou, 2011) no reparan en la necesidad de proponer una definición unitaria a partir de la cual poder derivar y vincular las diferentes acepciones del verbo, y poder demostrar así que pertenecen a un mismo lexema; otros autores, como Cuartero Otal (2014), en cambio, rechazan este planteamiento al poner en duda la posibilidad de que este verbo sea tratado en una única entrada, en la medida de lo posible, monosémica, y se decantan por la existencia de muchos *dar* en español.

⁹⁸ También conocidos como *verbos ligeros* (del inglés *light verb constructions*) o *verbos soporte* (del francés *constructions à verbe support*). Bosque (2001b) señala que el término inglés *light* focaliza la menor carga semántica de estos verbos, mientras que los términos empleados en francés y español se centran más en su defectividad gramatical. A estas construcciones nos referimos en mayor profundidad en la sección 4.1.1.2.

palabra contextualizada contiene siempre dos fuentes de significado (Ruhl, 1989): el sistémico y el contextual –ya sea este básico o figurado, pleno o de apoyo–.

De Miguel (2008: 571) recurre al concepto de *infraespecificación*, de Pustejovsky (1995), para referirse al significado sistémico –si bien con ciertas variaciones con respecto a este– de los verbos polisémicos:

Infraespecificación (*underspecification*): “Falta de especificación de los signos lingüísticos que los capacita para intervenir en diferentes estructuras sintácticas y, en consecuencia, en distintas operaciones de composición semántica”.

De este modo, “las palabras no poseen significados cerrados y estables sino definiciones poco especificadas aunque potencialmente capacitadas para ampliarse, cuando aparecen en un determinado contexto” (De Miguel, 2011: 140). Por lo tanto, si el significado sistémico propuesto está presente en las diferentes acepciones de la palabra, ha de dar cuenta de todo su comportamiento sintáctico.

De Miguel (2008: 575) ofrece la siguiente definición infraespecificada de *dar*: “pasar algo de una fuente a una meta [sea *una explicación*, sea *un caramelo*]”. Bajo esta definición subyace el contenido de *transferencia*, que es retomado en su trabajo de 2011 (144), como se desprende de la afirmación “el contenido de transferencia o trayectoria está presente (...) en *dar*”. Téngase en cuenta que por *transferencia* se entiende acción y efecto de “pasar o llevar algo desde un lugar a otro” (DRAE, 2001). A nuestro parecer, estas reflexiones semánticas equiparan dos conceptos no idénticos –como son la *transferencia* y la *trayectoria* (la distinción de estos conceptos se analiza seguidamente)– y excluyen usos del verbo, como los de *dar un salto*, *dar una voltereta*, *dar un paseo*, *dar una clase* o *dar comienzo*, en los que el evento denotado por el sustantivo no preexiste, por lo que no se puede pasar o transferir.

La definición sistémica planteada en el capítulo 3 nos parece más abarcadora que la anterior, pues incluye los usos anteriores de *dar* (*salto*, *voltereta*, etcétera) en los que se comporta como verbo de apoyo, esto es, como verbo de creación, con significado próximo al del *hacer*⁹⁹; no obstante, su aplicación no está exenta de problemas, como se irá viendo a lo largo del capítulo.

⁹⁹ Como se anunciaba en el apartado 1.4, Bosque (2001b: 23) señala como característica semántica distintiva de los verbos de apoyo su significado abstracto, cercano al de *hacer*. Esta idea la desarrollamos en la sección 4.1.1.2.

A da B (a C) significa “*A* causa que *B* pase de no existir a existir (en *C*, de manera que los seres *C* disfruten o padezcan *B*)”.

En esta definición unificamos los 11 sentidos de *dar* planteados en Alonso Ramos (1997) –ya que la autora en este trabajo dedicado exclusivamente al comportamiento del verbo *dar* no propone una definición unificadora–, y nos inspiramos en la postura defendida por Gross (1975) –y por el grupo de léxico-gramática que dirigió en París– cuando planteamos la definición como un esquema sintáctico. Dicho esquema desemboca en una entrada léxica de *dar* que se corresponde con una oración simple, lo que nos permite conceder al verbo un comportamiento previsible en una oración. De esta manera, léxico y gramática quedan integrados en esta definición. La oración está compuesta por un predicado (*dar*) y unos argumentos que le son propios (*A*, *B* y *C*), por lo que la definición propuesta es aplicable de manera directa cuando se reproduce este esquema sintáctico.

Ya se ha hecho notar en el capítulo 3 que en *dar* es posible interpretar que tras ese cambio que se produce en *B* –de no existir a existir, o de no existir en *C* a existir en *C*– subyace un contenido de trayectoria implícita presente a lo largo de todo el Esquema de Imagen del CAMINO propuesto por Lakoff y Johnson (1999). Entendemos por *trayectoria* “línea descrita en el espacio por un cuerpo que se mueve, y, más comúnmente, la que sigue un proyectil” (DRAE, 2001). En relación con ello, Lakoff y Johnson proponen la siguiente estructura de Esquema de Imagen del CAMINO (1999: 33):

- Una entidad se desplaza
- Un lugar origen del desplazamiento
- Un destino planeado del desplazamiento de la entidad desplazada
- Una ruta que se extiende desde el origen al destino
- La trayectoria descrita por el movimiento de la entidad que se desplaza
- La posición de la entidad desplazada en un momento dado
- La dirección de la entidad desplazada en dicho momento
- La localización final de la entidad desplazada, que puede coincidir o no con la del destino planeado

De esta manera, se observa que el CAMINO surge de nuestra percepción de entidades que se mueven a lo largo de un trayecto, de una línea descrita en el

espacio, en el que hay un punto inicial (lugar de origen), un destino y un recorrido. Según este esquema, conceptualizamos directamente recurriendo a las categorías perceptivas, a una experiencia física que tiene que ver con dominios básicos —en este caso, el espacio—, como ocurre en *dar un caramelo a alguien*, o mediante proyecciones metafóricas que toman como dominio de origen el Esquema de Imagen, como en *dar una explicación a alguien*. Por consiguiente, de acuerdo con este planteamiento cognitivo, cuando esa línea —física o figurada— o trayecto alcanza su término, el evento ha tenido lugar, ya sea la dación de un caramelo (la transferencia) o la ejecución de un abrazo (la creación). En esa trayectoria o camino se presuponen —y así lo especifican Lakoff y Johnson en su esquema— un origen y la meta; el foco está en la última porque supone el punto final de ese recorrido y el estado resultativo —el caramelo está en manos de C; el abrazo se ha ejecutado— que surge como consecuencia natural del proceso anterior.

Así, estos dos elementos —la trayectoria y su meta¹⁰⁰— están presentes de manera explícita tanto en los eventos descritos por *dar un caramelo*, *dar una explicación* o *dar cariño a alguien*, como en los de *dar un salto*, *dar un ladrido* o *dar clase*.

Por consiguiente, se puede afirmar que la extensión de *dar* abarca, en primer lugar, los objetos (concretos o abstractos) que interpretamos que existen con anterioridad y que se desplazan (literal o figuradamente) hasta llegar al punto final que marca que los eventos descritos por *dar un caramelo*, *dar cariño* o *dar una explicación* han tenido lugar, y, por tanto, esos objetos —*el caramelo*, *la explicación* o *el cariño*— están ahora en C. Estos tres predicados están siendo interpretados como de cambio de estado y no como predicados de creación, lo que implica que interpretamos que no es necesario *dar una explicación* para que esta exista, para que se produzca, del mismo modo que señala Alonso Ramos (2004: 105) que no hay por qué dar o expresar una opinión para poder tenerla, lo cual justifica, de acuerdo con los criterios de la autora, la exclusión de la expresión *dar una opinión* del grupo de las CVA. Así, *el caramelo*, al igual que *las explicaciones*, *el cariño*, *la opinión* o *el alojamiento* se *tienen* antes de *darlos* (y así se registra en *Práctico*, 2006: 605 y 233). Estas combinaciones son asimismo interpretadas en el diccionario combinatorio *Redes* (2005: 634) como usos plenos de *dar*, dado que se asimilan al grupo proyectado a partir de *dar + sustantivos que designan entidades materiales*. De esta manera, tanto *dar un papel* como *dar alojamiento* y *dar cariño* se incluyen bajo el

¹⁰⁰ Nótese que no interpretamos *meta* como destinatario de la acción, sino como estado resultativo.

epígrafe de los sustantivos que se combinan con *dar* en el sentido de “entregar, ofrecer, hacer llegar o proporcionar”.

A estos eventos sí que se les puede aplicar la noción de *transferencia* señalada por De Miguel en su definición infraespecificada (2008: 575); en otras palabras, estos ejemplos sí que son contemplados en la definición de la autora; la *explicación*, *el cariño* o *un caramelo* se tienen y se pueden pasar o llevar de un lugar a otro, de *A* a *C*. La noción de *transferencia* motiva nuestra categorización de estos usos del verbo *dar* como plenos (frente a los usos como verbo de apoyo).

En segundo lugar, la extensión de *dar* abarca también los sustantivos de eventos en el que el evento descrito pasa de no existir a existir. Estos usos se interpretan en nuestro análisis como predicados de creación, como CVA; dado que *el beso*, *el salto*, *el ladrido* o *la clase* no preexisten, no se pueden transferir, *no pasan de un lugar a otro*, por lo que no se interpretan como un predicado de cambio de estado. En estos casos, *el beso*, *el ladrido*, *el salto* o *la clase* se crean o se desarrollan a medida que el sujeto los da, mientras van recorriendo la ruta (de acuerdo con los términos de Lakoff y Johnson), y una vez que se alcanza el estado final de la actividad –o, lo que es lo mismo, la meta en su trayectoria–, que surge como consecuencia natural del proceso anterior.

De este modo, se constata que la transferencia o, en otras palabras, el uso de *dar* como predicado de cambio de estado, presupone siempre una trayectoria, pero no a la inversa; una trayectoria no implica que se produzca una transferencia, sino que aquella puede subyacer a un predicado de creación. Por lo tanto, la trayectoria y el estado final alcanzado –más allá del cual no se puede seguir avanzando– son los contenidos que están siempre presentes en *dar*, pero al tratarse de información aspectual también se encuentran en todos los eventos considerados *efectuaciones* (Morimoto, 1998: 15), esto es, en las actividades que tienen establecido dentro de su propia estructura temporal un punto final, del tipo *hacer una tarta* o *construir una casa*.

La transferencia, sin embargo, queda *en principio* excluida de los predicados de creación; subrayamos *en principio* porque en aquellos eventos de creación en los que existe un destinatario puede interpretarse esta noción de *transferencia*, como en *dar un abrazo*, *dar un golpe* o *dar comienzo* a *C*¹⁰¹; de hecho, se llega a

¹⁰¹ Cuando De Miguel acompaña su definición infraespecificada de *dar* con los sustantivos *caramelo* y *explicación*, los presenta como muestras que ejemplifican la construcción con uso plenamente predicativo del verbo en el caso de *caramelo*, y la CVA en el caso de *explicación*. Pese a que *dar* una *explicación* pueda ser interpretado como una CVA, la definición continúa sin abarcar los eventos que no admiten la noción de transferencia, como *salto* o *voltereta*. Ya se ha argumentado que nuestra postura es otra; si atendemos a criterios semánticos (y no solo gramaticales), en *dar una explicación* no interpretamos que *dar* funcione como verbo de apoyo, pues actualiza su sentido

proponer (De Miguel, 2011: 144) la transferencia como el rasgo distintivo de *dar* en aquellos eventos en los que un sustantivo dado puede coaparecer con *dar* y con *hacer*, como en *dar/hacer un masaje*. De Miguel interpreta que con *hacer* la CVA denota un evento de creación, mientras que con *dar* es un evento de transferencia, y nótese que la autora, cuando añade *dar un masaje* que “es un evento de transferencia en el mismo sentido que *dar un abrazo*”, interpreta el evento de *dar un abrazo* como de transferencia y no de creación. Consideramos que aquellos casos en los que tienen cabida los dos verbos, como en el ejemplo del *masaje* o como en *dar/hacer un pase* (entrega del balón; recogidas ambas posibilidades combinatorias en *Práctico*, 2006: 949), ambos son eventos de creación –el pase solo existe una vez que se ejecuta–, pero en los eventos de *dar* el foco está en el destinatario –lo que no implica que tenga lugar una transferencia–, mientras que con *hacer* el foco recae en la acción desarrollada. Como se ha visto, toda definición de *dar* fundamentada únicamente en la noción de *transferencia* excluye los usos del tipo *la vaca da leche*, *dar un salto*, *un paseo*, *una voltereta*, que no aceptan el destinatario o pueden prescindir (como en *grito* o *ladrido*) de él.

Por todo lo expuesto hasta ahora, consideramos que la noción de *transferencia* constituye la idiosincrasia léxica de *dar* –pues es el sentido actualizado en todos los usos en los que el verbo *dar* funciona como verbo pleno–, pero que en las CVA la gramática prescinde de ella y selecciona *fundamentalmente*¹⁰² los rasgos aspectuales de su significado (Bosque, 2001b: 35) que sí que están presentes en todos los usos del verbo¹⁰³.

A continuación se presenta el esquema de la definición desglosada que abarca todos los usos de *dar*; *A da B* (a *C*) significa “*A* causa que *B* pase de no existir a existir (en *C*, de manera que los seres *C* disfruten o padezcan *B*)”:

- 1) *A* causa que *B* pase de no existir a existir: contiene los rasgos aspectuales de la definición presentes en todos los eventos de creación (las CVA) y de transferencia (las CVP): la trayectoria y el estado final. *Dar un abrazo* o *dar un salto*, y *dar un libro* o *dar cariño*.

pleno, el de transferir, y no el de ejecutar o hacer, por lo que necesariamente interpretamos que *explicación* es un objeto abstracto que preexiste a la acción denotada por el verbo.

¹⁰² Subrayamos *fundamentalmente* dado que, como se analiza en la sección 4.1.1.2, existen otros rasgos léxicos mínimos o subléticos en el significado del verbo que también participan en las CVA.

¹⁰³ Como se ha expuesto, hay autores que interpretan la noción de transferencia en aquellas CVA en las que existe o puede existir un destinatario (*dar un beso*, *dar una clase* o *dar un grito*). Aunque este proceso de inferencia lo consideramos natural, puede llevar a juicios erróneos a la hora de establecer criterios para distinguir las CVA de las CVP.

2) (en *C*, de manera que los seres *C* disfruten o padezcan *B*): sobre el destinatario –implícito o explícito– recae el rasgo léxico idiosincrásico de la transferencia, pues no hay transferencia sin destinatario (*dar un libro* o *dar cariño a alguien*), de ahí que eventos de creación con destinatario puedan llegar a ser interpretados, a nuestro juicio erróneamente, como efectuaciones de transferencia (*dar un abrazo* o *dar un golpe*). Este elemento no puede aparecer en muchas CVA (*dar un salto*).

Esto explica que aparentemente el verbo en las CVA se vacíe de significado; dado que el rasgo idiosincrásico de *dar* que se corresponde con la idea de transferencia no está presente, el contenido que se mantiene en ellas es principalmente¹⁰⁴ el aspectual. A este respecto, Alonso Ramos (2004: 96) afirma que “lo que hace que un verbo sea de apoyo es su carácter vacío en la colocación¹⁰⁵: el hecho de (...) no añadir su propio significado léxico al nombre con el que se combina”.

Pero, si el verbo se vacía, ¿qué hay en *dar* –que no tiene hacer– para que se combine en las CVA con los sustantivos eventivos salto, paseo o grito?; ¿y por qué es hacer y no dar el que se combina frecuentemente con declaración cuando su significado aspectual es el mismo que el de dar, y cuando es dar el que se combina con discurso, comunicación o rueda de prensa ¹⁰⁶ ?; ¿estas combinaciones son, entonces, azarosas y arbitrarias? Nos ocupamos de estas cuestiones a continuación.

4.1.1.2. Sobre la existencia de rasgos mínimos de *dar* que determinan la selección semántica de sus argumentos. La concordancia léxica y la redundancia

De acuerdo con la clasificación de las clases aspectuales de Morimoto (1998: 13-17), *dar* sería etiquetado como un tipo de evento, una situación

¹⁰⁴ Véase n. 6.

¹⁰⁵ Para Alonso Ramos (2004: 21) las CVA constituyen colocaciones, en donde el nombre es la base y el verbo, el colocativo (de acuerdo con la terminología de Hausmann, 1979). Son, por tanto, casos de coocurrencia restringida, por lo que las considera unidades semifraseológicas. En Sanchez Rufat (2010) defendemos que las propiedades características de las colocaciones no demuestran que estas combinaciones formen parte de la fraseología; la mayoría de ellas constituyen ejemplos de las restricciones que hemos denominado intralingüísticas.

¹⁰⁶ En una búsqueda del CREA, 378 son los resultados obtenidos de *hacer declaración(es)*, mientras que son 47 los resultados de *dar declaración(es)*; resulta llamativo que tan solo uno de esos 47 usos se corresponda con el habla de España (el resto se registra en diversos países hispanoamericanos). De acuerdo con estos datos, *dar declaración* es una combinación restringida a Hispanoamérica, donde también es más frecuente la combinación con *hacer*.

inherentemente dinámica –en oposición a estado–, que pertenece al grupo de las efectuaciones –y no al de actividades ni al de logros–, ya que *dar* se prolonga a lo largo del tiempo, como las actividades del tipo *correr*; pero a diferencia de estas, las efectuaciones tienen establecido dentro de su propia estructura temporal un punto final al que ya nos hemos referido, que marca la finalización del evento (recuérdense los conceptos clave *trayectoria* y *meta*). Los logros, en cambio, implican también un punto final pero solo marcan el momento en que se alcanza el término final de una situación, como *descubrir*.

Las efectuaciones descritas por *dar*, como se ha expuesto en la sección anterior, pueden referirse a un acto de transferencia –cuando el nombre es referencial o designa un objeto abstracto que se interpreta como preexistente (el caramelo o la explicación)– o a un acto de creación, cuando el nombre denota un evento y el verbo tiene un significado abstracto cercano al del verbo *hacer*; lo que, de acuerdo con Bosque (2001b: 23-24)¹⁰⁷, caracteriza a las CVA desde una perspectiva semántica:

Most current references to the semantic content of light verbs implicitly assume that it comes close to the meaning of an abstract verb of action, such as *do*. In fact, the meaning of *pay* in *pay attention* is not much different from that of *make* in *make a promise* or *give* in *give a cry*.

Esta proximidad significativa entre el grupo de verbos de apoyo y *hacer* no es casual. Desde una perspectiva de la significación aspectual, *hacer* y *dar* son

¹⁰⁷ Las propiedades, en su mayoría gramaticales, de las CVA han sido descritas para el español en De Miguel (2011 y 2008), en Alonso Ramos (2004), en Blanco Escoda (2000), en Bosque (2001b) y en Mendivil (1999), entre otros. En esencia, todos ellos coinciden al enumerar las pruebas que habitualmente se aportan para identificar las CVA; pero, como se desprende de los ejemplos ofrecidos en algunos de estos trabajos –y como De Miguel (2011) manifiesta abiertamente–, no todas las propiedades presentadas son exclusivas de las CVA, sino que pueden aplicarse también a verbos o usos predicativos, como, por ejemplo, el hecho de que un verbo de apoyo puede suprimirse y la nominalización resultante conserva la carga semántica de la CVA original (piénsese en *El cuadro de Dalí y pintar* o *La opinión de Pedro y dar*), o el hecho de que exista una pieza léxica con la que parafrasear la CVA y también exista en la CVP (como en *encarar por subir los precios* u *opinar por dar una opinión*). Por otro lado, hay casos de CVA a los que no se les puede aplicar todas las propiedades; así, no siempre existe una pieza léxica con la que se pueda parafrasear toda la CVA (como en *dar una voltereta*), o en una CVA un sintagma preposicional no siempre puede recibir dos análisis, como complemento del verbo o del sustantivo predicativo (como en *El empujón que Pedro dio a Juan*, pero **El empujón a Juan que Pedro dio*). No es nuestro objetivo tratar en profundidad cada una de estas pruebas y propiedades, pues en el análisis de los datos –como se ha referido– el criterio para categorizar *dar* como verbo de apoyo es el semántico: nos referimos a la pérdida de la noción de la transferencia y el mantenimiento de un valor próximo al de *hacer*. No obstante, a la hora de categorizar nuestros datos sí que tomamos en consideración que, además de este criterio semántico, en las CVA los eventos denotados por el verbo y el nombre han de compartir el sujeto.

efectuaciones; constan, por lo tanto, de dos estadios: “un proceso de desarrollo, dirigido hacia una determinada dirección, y un estado final, que surge como consecuencia natural del proceso anterior” (Morimoto, 1998: 15). Así, *hacer* es “crear algo” (De Miguel, 2008: 575), o, aplicando el esquema sintáctico de Gross, *A* hace *B* = “*A* causa que *B* pase de no existir a existir” (ya sea *una tarta*, *un análisis* o *una declaración*). Contrástese con la definición sistémica de *dar* propuesta: *A* da *B* (a *C*) = “*A* causa que *B* pase de no existir a existir (en *C*, de manera que los seres *C* disfruten o padezcan *B*)”; de este modo, si prescindimos de la parte contenida entre paréntesis –que, como se ha expuesto, se corresponde con el rasgo léxico más idiosincrásico de *dar*, el destinatario requerido en toda transferencia (ausente en algunas combinaciones, como en *dar un salto* o *un paseo*)–, *hacer* y *dar* comparten una misma definición sistémica, que es la que se actualiza en las CVA y que está caracterizada por los rasgos aspectuales propios de las efectuaciones.

Si conectamos este hecho con nuestras hipótesis de partida –formuladas sucintamente en el apartado 1.4–, planteábamos que a la hora de utilizar un sustantivo eventivo, la selección del verbo de apoyo puede ser una fuente de error para los aprendientes de una L2, ya sea por motivos intralingüísticos (la coincidencia aspectual existente en *dar* y *hacer* –y demás verbos de apoyo–) o interlingüísticos (interferencia de la L1). Probablemente, en la mayoría de las lenguas haya un equivalente de *dar* que, en el sentido de *proporcionar* o *transferir*, se utilice con sustantivos que designan entidades materiales; ciertamente este es el caso de *to give*, en inglés. Es posible que esas lenguas utilicen ese mismo verbo con *amor*, *cariño*, *alojamiento* u *opinión*, pues estos objetos abstractos también se proporcionan; y en el caso de que no sea así en la L1, una vez que el sentido básico (físico o literal) de *dar* es adquirido en la L2, estos usos plenos aunque figurados se pueden interpretar fácilmente como extensiones naturales del uso pleno literal. Sin embargo, es probable que cada lengua sancione un verbo de creación o de apoyo diferente para un sustantivo dado, pues, como se ha señalado, verbos como *dar* o *hacer* comparten el mismo contenido aspectual próximo a *ejecutar*, *hacer efectivo* o *producir*, y –como observa Alonso Ramos (1997: 22-23)– un sustantivo que denota un evento tan solo requiere este significado aspectual para realizarse, para existir (ya sea *un abrazo* y *un salto* o *un análisis* y *una declaración*), de manera que el significado más idiosincrásico del verbo no se necesita; por ello, en principio, cualquiera de los dos verbos –*dar* o *hacer*– podría ser empleado para acompañar a un sustantivo eventivo. Por lo tanto, debemos considerar *dar* como la forma verbal específica que nos proporciona el lexicón para rellenar un contenido aspectual próximo a *realizar* o simplemente *hacer*

cuando hablamos, por ejemplo, de *paseo*, *grito* o *suspiro*. En cambio, en inglés los paseos se toman y en francés se hacen (*take* y *faire*, respectivamente); en francés *los gritos* y *los suspiros* se empujan (*pousser*), no se dan.

Este planteamiento pone de manifiesto, por un lado, que el sustantivo de evento y el verbo de apoyo constituyen un predicado complejo en el que el verbo no predica en exclusiva ni de manera fundamental (De Miguel, 2011: 139)¹⁰⁸ –de ahí la defectividad semántica atribuida a los verbos de apoyo (Alonso Ramos, 2004; Bosque, 2001b; Blanco Escoda, 2000), considerados como mero soporte sintáctico del nombre que expresa un predicado semántico–:

El verbo de apoyo *dar* [en *dar una paliza*] no es más que una herramienta léxica, empleada con fines morfológicos y sintácticos para permitir la construcción de la oración y la particularidad principal de las CVA reside en el hecho de que los actantes semánticos del nombre y el propio nombre se realizan como actantes sintácticos del verbo (Alonso Ramos, 2004: 19).

El verbo *dar* [en *dar un beso*] no se analiza, pues, como un predicado, sino como un soporte de las marcas de tiempo y de persona del predicado nominal (Blanco Escoda, 2000: 99)

Por otro lado, del hecho de que el nombre solo requiera información aspectual para poder existir se podría deducir que es una cuestión de hábito, de norma –y no del sistema– que cada comunidad de hablantes seleccione un verbo de apoyo u otro; ciertamente, esto no es así. La elección del verbo para cada sustantivo eventivo no es arbitraria en cada lengua, como así lo considera Alonso Ramos (1997: 22 y 2004: 25):

entiendo por verbo de apoyo todo verbo combinado con un nombre predicativo en función de primer complemento (objeto directo o complemento preposicional), que no es escogido por el hablante sobre una base semántica, sino más bien de una manera arbitraria en función del nombre, y cuyo papel es: a) expresar las marcas de modo, tiempo y persona, y b) proporcionar posiciones sintácticas para que los actantes del nombre puedan aparecer en un contexto oracional.

¹⁰⁸ Gross (1989) precisamente distingue los verbos de apoyo de los verbos con uso ordinario a partir de la semántica del nombre. Los verbos de apoyo únicamente se combinan con nombres predicativos y los verbos predicativos, con nombres concretos.

Ni los verbos de apoyo –tras perder su contenido léxico idiosincrásico al acompañar a un sustantivo eventivo– se vacían de significado léxico, como así lo considera Blanco Escoda (2000: 100) al afirmar que *dar*, en *dar un beso*, es “un verbo semánticamente vacío que ‘conjuga’ el nombre al que acompaña”; y lo interpreta de igual modo Alonso Ramos (1997: 22 y 2004: 87):

A pesar de que un verbo de apoyo no es necesariamente vacío, es decir, que puede tener un significado léxico, en cambio, en el contexto de la colocación ha de ser necesariamente vacío. Entiendo que un verbo es vacío en el contexto de una colocación si no está seleccionado por su propio significado léxico y solo contribuye semánticamente inscribiendo en el tiempo el predicado expresado por el nombre¹⁰⁹.

De igual manera que el contenido semántico de una construcción se altera si se sustituye el verbo *dar* en su sentido pleno por otro verbo, como en *dar un caramelo* y *quitar un caramelo*, tampoco se mantiene si se cambia el verbo de apoyo por otro verbo: *dar un abrazo* frente a *recibir un abrazo*, o *dar un salto* frente a *iniciar un salto*. Esto demuestra que el verbo de apoyo también es responsable del significado del evento denotado por el verbo más el nombre, por lo que no está vacío.

Stein (1991) y De Miguel (2008, 2011) también rechazan la idea de que estos verbos estén vacíos de significado y funcionen como simples conectores. Butt y Geuder (2001) defienden que los diferentes significados de *to give* forman un *continuum* a partir del significado básico o literal hasta el significado más abstracto del verbo de apoyo, que no es completamente vacío.

De Miguel (2011: 141) argumenta que el escaso significado predicativo de estos verbos (recuérdese que en las CVA el verbo se deshace de su significado léxico más idiosincrásico; en el caso del verbo *dar*, el de *transferencia*, que requiere la existencia de un destinatario) se va rellenando en función del sustantivo con el que aparezca, por lo que en las CVA se pierde el sentido de *transferencia*, pero se ganan otros sentidos; de ahí que la autora no hable de vaciado sino de rellenado. Este hecho se relaciona con la mencionada capacidad potencial de ampliación de significado que caracteriza la definición sistémica, poco especificada, de estos verbos cuando aparecen en un contexto determinado. Así, *dar* no solo significa *ejecutar* (esto es, “causar que *B* pase de no existir a existir”, lo que equivale a su definición como verbo de apoyo, que a su vez se corresponde con la parte fija e imprescindible de la definición sistémica), sino *ejecutar sonidos* cuando coaparece

¹⁰⁹ Para mayor claridad, hemos eliminado de la cita las llamadas de notas al pie.

con sustantivos como *grito, pitido, aullido, ladrido...* (Redes: 641). Asimismo, *dar* significa *ejecutar desplazamientos breves o repentinos* cuando acompaña a *un paseo, una vuelta, una voltereta, un salto o un paso* (Redes: 640); y precisamente por el contenido ampliado que surge de esta última acción de rellenado, se explica por qué *dar* no se combina con *viaje, recorrido o excursión*, sustantivos que se interpretan como desplazamientos largos (ver *infra*); el verbo, por lo tanto, no es elegido por el hablante de manera arbitraria en función del nombre.

De esta manera, *dar* no se combina libremente con todos los sustantivos de eventos que requieren un verbo para ser ejecutados; de ser así, bastaría con tener en cada lengua un solo verbo de apoyo. Como se infiere a partir de los anteriores ejemplos, el sistema es más restrictivo de lo que aparenta ser; el verbo *dar* ha de contar en su definición con unos rasgos subléxicos (término y concepto tomados de De Miguel, 2008) que determinan su posibilidad de combinación con ciertos sustantivos y que solo se hacen visibles en la sintaxis. De acuerdo con esto, y como se ha visto en los anteriores ejemplos *dar un grito* y *dar un paseo*, el verbo realiza una aportación léxica, no solo aspectual, a la combinación y –como demuestra De Miguel (2011: 144)– participa en la selección semántica de los argumentos.

Así, la selección léxica de *dar* está fundamentada en esos rasgos subléxicos contenidos en su estructura interna que son compatibles con la de los nombres con los que se combina, por lo que las CVA son interpretadas por De Miguel (2011: 140) como “un proceso gramatical y semántico determinado por procesos generales de concordancia de los rasgos léxicos contenidos en la hipotética estructura interna de los nombres y los verbos que entran en combinación”. Según esta concordancia de rasgos léxicos, una palabra con el rasgo subléxico [+desplazamiento breve] admite el predicado *dar*, como *vuelta, paseo, salto o paso*; en cambio, una palabra con el rasgo [-desplazamiento breve] no admite *dar*, como *viaje, excursión o marcha*, que se interpretan como desplazamientos largos en oposición a los anteriores. En estas combinaciones, la selección léxica es llevada a cabo por los dos elementos, es mutua: *paseo* selecciona *dar* como verbo de apoyo necesario para que el evento pueda existir; y *dar* selecciona *paseo*, pues es el verbo de apoyo que al ir acompañado del sustantivo *paseo* actualiza el significado de “ejecutar movimientos breves” presente en su estructura interna.

La concordancia léxica que De Miguel aplica a las CVA (2008 y 2011) es un concepto presente en Bosque (2005: CXLII [2004]). Este lo considera como un rasgo distintivo de la combinatoria léxica: “las relaciones de concordancia constituyen la textura que permite hacer explícitas las conexiones que la sintaxis

exige”. De Miguel aprovecha, asimismo, el concepto de *redundancia* (propuesto por el mismo autor), al que se refiere como

exponente de un fenómeno general de concordancia de rasgos léxicos, que se manifiesta en español en otros muchos procesos, y que consiste en la homogeneidad o la homogeneización de ciertas partículas del significado interno de las palabras que acaban proporcionando información redundante (De Miguel, 2008: 567).

La *redundancia*, entonces, se interpreta como una exigencia del sistema lingüístico; es una forma de concordancia de rasgos léxicos (la concordancia puede llevarse a cabo, además, con marcas morfológicas, de función, posicionales) que está presente en la combinatoria léxica, aunque no siempre se produce (Bosque, 2005: CXLII [2004]; no ocurre, y sería lo esperable, en **dar una declaración* (combinación no sancionada por los hablantes peninsulares [véase n. 10]), o en **dar una llave* (movimiento brusco), caso al que aludimos en el capítulo anterior).

En esta concordancia de rasgos se fundamenta, precisamente, *Redes*, donde no se proponen palabras aisladas como argumentos seleccionados por un predicado, sino una serie de grupos que forman las clases léxicas. Cada clase aparece encabezada por un epígrafe que se corresponde con el rasgo semántico seleccionado por el verbo y presente, por consiguiente, tanto en las palabras que integran la clase léxica como en la hipotética estructura interna del verbo, y de ahí la concordancia y la consiguiente redundancia de rasgos. De este modo, en *Redes* se especifican los complementos de *dar*, que no se deducen directamente de su significado, pero que sí se contemplan en la definición sistémica ampliable por el contexto dado; dicho de otro modo, a partir de la información contenida en el diccionario se infieren los rasgos mínimos de su definición que determinan la posibilidad de combinación del verbo con ciertos sustantivos.

De este modo, nos basamos en Bosque (2005 [2004]) y en De Miguel (2008 y 2011) cuando afirmamos que la selección léxica de clases semánticas por parte del verbo *dar* en sus usos figurados –ya sea como verbo pleno o como verbo de apoyo– es un exponente del fenómeno de concordancia de rasgos que implica redundancia. Por lo tanto, tras comprobar que *dar* no coaparece con *paseo*, con *salto* o con *voltereta* por azar, sino porque se combina con el paradigma de los sustantivos que denotan un movimiento breve o repentino, se infiere la existencia del rasgo subléxico [+desplazamiento breve] en la estructura interna del verbo, por lo que se presupone la redundancia de este rasgo. De esta manera

se comprueba que la definición sistémica que hemos propuesto está potencialmente capacitada para ampliarse cuando aparece en un contexto concreto.

A diferencia de estas combinaciones, las que se construyen con las antiguas restricciones selectivas introducidas en la gramática generativa en los años sesenta (persona, objeto...), como en *dar un caramelo, un libro, una camisa...*, son posiblemente, como sugiere Bosque (2005: CXLII [2004]), las únicas que no muestran concordancia de rasgos subléxicos –ni *caramelo*, ni *libro*, ni *camisa* designan objetos inherentemente *dables* o susceptibles de ser dados–, de ahí que a estas relaciones no las consideremos intralingüísticas sino extralingüísticas. El conocimiento del mundo y el sentido común nos permiten construir este paradigma extensional de *dar*, basado en restricciones de naturaleza física.

En definitiva, la selección léxica pone de manifiesto una forma particular de concordancia, y las CVA se pueden considerar un caso más de esta concordancia que se da en la combinatoria léxica general, aunque, como señala Bosque (2005: CXLII), no se conocen todavía los rasgos específicos que la hacen posible; no obstante, se intuyen en la información combinatoria que proporciona *Redes*. También intenta delimitar estos rasgos –partiendo precisamente de algunos de los presupuestos de Bosque– De Miguel (2008 y 2011), que, en la hipótesis que maneja sobre las CVA, se inspira en los mecanismos generativos de significado propuestos por Pustejovsky (1995), a partir de los cuales las palabras reciben un número ilimitado de interpretaciones según el contexto. De acuerdo con esta hipótesis, la definición infraespecificada de las palabras contiene distintas informaciones codificadas en diversas estructuras¹¹⁰. En relación con todo ello, para analizar las coapariciones de *dar* en el corpus *CEDEL2* nos hemos servido del modelo propuesto por Bosque en *Redes* (2005 [2004]) porque

¹¹⁰ Otra hipótesis de explicación de la combinatoria léxica que se produce en las CVA, pero que no está basada en la concordancia de rasgos, es la presentada por el enfoque lexicográfico del *Dictionnaire explicatif et combinatoire (DEC)*, dirigido por Mel'čuk, según el cual las funciones léxicas son las que dan cuenta de todos los casos de coocurrencia restringida. Alonso Ramos (1997), apoyándose en este enfoque, explica la combinatoria de *dar* por medio de la función léxica *Oper₁*, que se aplica a un nombre predicativo para proporcionar un verbo de apoyo (o el significado *hacer*); para la autora es el nombre predicativo el que selecciona el verbo que le atribuya las marcas de tiempo, de número y de persona (Alonso Ramos, 1997: 22). En este marco lexicográfico, la direccionalidad de la selección léxica defendida es otra diferente a la planteada en *Redes* (sobre cuestiones de direccionalidad, véase Bosque [2004b]). Cabe añadir que, pese a la distinta direccionalidad de *Redes* y el *DEC* (y con este mismo enfoque el *DICE*), ambos constituyen dos modos diferentes de estudiar un mismo fenómeno, por lo que se complementan. En el *DEC* se muestra cómo una serie de funciones léxicas establece un gran número de conexiones semánticas a partir del vocabulario seleccionado, convierte las bases de las colocaciones en lemas; en *Redes* se interpretan los colocativos, que son convertidos en lemas, como predicados que seleccionan argumentos que pertenecen a clases léxicas restringidas.

se halla más en consonancia con nuestros planteamientos: reconocemos al verbo como predicado y, por tanto, elemento léxico con capacidad para seleccionar argumentos (clases léxicas restringidas) y para atribuirles propiedades o designar relaciones entre ellos (como muestra De Miguel, 2011: 144); en aquellos contextos en los que el verbo funciona como verbo de apoyo, esta capacidad es compartida con el nombre eventivo con el que constituye un predicado complejo, por lo que la selección léxica es mutua, y cada uno aporta su significado a la construcción.

Como conclusión, a partir de la definición sistémica de *dar* propuesta en la sección 4.1.1.1, y de acuerdo con las cuestiones tratadas en esta sección 4.1.1.2, estamos en disposición de distinguir en los datos los usos del verbo cuando se comporta como verbo pleno de los usos en los que funciona como verbo de apoyo.

Todavía es posible dar un paso más a la hora de categorizar los usos de *dar*. En función de los nombres –agrupados en clases léxicas– con los que coaparezca, ya actualice un sentido pleno o de apoyo, el verbo se rellena de significado, por lo que, si se establecen las clases léxicas, es posible especificar la definición de los diferentes sentidos de *dar*. Esto permite mayor precisión a la hora de clasificar cada uso que se hace del verbo en el corpus.

4.1.2. Descripción de los sentidos de *dar*: problemas encontrados y soluciones para el análisis de los datos

Pese a que es posible percibir a partir de la definición sistémica de *dar* el vínculo semántico común entre los diferentes usos del verbo en las construcciones composicionales del tipo V + N, los límites entre los sentidos figurados de esta palabra –tanto entre los plenos como entre los de apoyo– no se demarcan con facilidad; por el contrario, pueden llegar a ser escurridizos; ello se debe a que la separación que se establece entre el uso físico y los figurados no siempre es mínima (como en los ejemplos en [1]), sino que a menudo es máxima (como en los ejemplos en [2]):

- [1] Uso físico: dar (proporcionar) un libro \longrightarrow Usos figurados: dar (proporcionar) alojamiento, seguridad, espacio, oportunidad, cariño
- [2] Uso físico: dar (proporcionar) un libro $-\ - - \longrightarrow$ Usos figurados: dar (hacer efectivo) un salto, una clase, un grito, comienzo, las gracias; dar (suscitar) hambre, envidia, pena, ganas

Este hecho ocasiona que el tratamiento lexicográfico de *dar*, tomando varios diccionarios de uso cualesquiera, como pueden ser el *DRAE*, el *Gran Diccionario de la Lengua Española* (Larousse) y el *Diccionario Salamanca de la lengua española*, sea muy distinto. En el primero se recogen hasta 52 acepciones del lema *dar* –sin sumar las unidades fraseológicas en las que participa–, en el *Gran Diccionario* se ofrecen 71 y en el *Diccionario Salamanca* se proponen 34 acepciones. No está claro cómo se distribuyen todos los usos de *dar* en esas 52 acepciones del *DRAE* o en las 71 y en las 34 del *Gran Diccionario* y del *Salamanca*; no parece posible, por tanto, comparar sistemáticamente las formas en que se distribuyen las acepciones en cada uno de los diccionarios.

Así, el *DRAE* entiende que la combinación *dar importancia* es un ejemplo de la acepción de *conceder, otorgar* (la número 6), mientras que en el *Gran Diccionario* ejemplifica la acepción 18, *atribuir una cierta relevancia o valor a una acción o suceso*, que es distinguida de la de *conceder una cosa a una persona* (la número 11). En el *Gran Diccionario*, *dar por inocente* y *dar por vencido* son ejemplos que pertenecen a dos acepciones distintas; el primero, a la de *considerar o declarar a una persona o una cosa en una condición o estado* (la número 23) y el segundo, a la de *reconocer una persona que erraba en una cosa* (la número 54). En el *DRAE*, ambos ejemplos pertenecerían a una misma acepción, la de *suponer, declarar, considerar* (la número 8). Se podrían añadir muchos ejemplos análogos.

Pese a estas diferencias en el tratamiento lexicográfico a la hora de segmentar y clasificar los diferentes usos figurados, y a pesar de que los usos figurados en [2] puedan percibirse bastante alejados del sentido físico, la conciencia del hablante no pierde toda relación entre el uso físico y el figurado; y, aunque la cuestión de cómo aprenden los hablantes los sentidos figurados de las palabras es sumamente compleja (Bosque, 2005: CXII [2004]), en el capítulo 3 y en la sección 4.1.1.1 de este capítulo se ha planteado una hipótesis sobre los mecanismos por medio de los cuales interpretamos los usos figurados como extensiones naturales del sentido físico a partir de los presupuestos de la psicología cognitiva. El hecho de que nos haya sido posible proponer una definición sistémica que abarque ambos tipos de sentido es una prueba más de la conexión que existe entre las diferentes acepciones del verbo *dar*. Así, esta hipótesis de trabajo –basada en los planteamientos semánticos de Lakoff y Johnson (1999) y los de Ruhl (1989), que plantea, por lo tanto, que tras el Esquema de Imagen del CAMINO se puede hallar esa definición sistémica del verbo– permitiría aportar una explicación plausible a por qué las proyecciones metafóricas de *dar* a partir de su significado básico son las que son.

Aunque las interpretaciones figuradas no introducen nuevos significados de *dar*, recurrimos a paráfrasis del verbo con el objetivo de facilitar la organización y clasificación de los datos. Así, no se pretende crear nuevas definiciones del lema *dar* cuando nos referimos a este verbo como *proporcionar* al ir seguido de sustantivos que denotan asistencia, sustento o protección (*Redes*: 634); o a *dar* como *ejecutar* cuando coaparece con sustantivos que denotan desplazamiento breve (*Redes*: 640); no obstante, estas paráfrasis –referidas a un uso del verbo pleno y a un uso como verbo de apoyo, respectivamente– pueden ocultar a simple vista que la definición sistémica subyace a ambas combinaciones, por lo que se corre el riesgo de interpretar una acepción como un significado nuevo, esto es, como un lexema distinto. Veamos esto con un ejemplo concreto obtenido del corpus: la combinación “*A da alojamiento a C*” es producto de la anterior definición parafraseada de *dar*, esto es, *dar* = *proporcionar* cuando va seguido de sustantivos que denotan asistencia; esta es la definición de una acepción de *dar* con sentido pleno figurado, y no es más que una forma abreviada de la siguiente definición de *dar*, sistémica, rellenada: “*A causa que B* –siendo *B* un sustantivo que denote asistencia, sustento o protección– *pase de no existir a existir en C* (de manera que *C* disfrute de dicha asistencia, sustento o protección denotada por el sustantivo dado)”. Como se puede observar, este sentido está contenido en la definición sistémica y es activado solo en los contextos particulares aquí descritos, esto es, por ciertos nombres que denotan asistencia, sustento o protección, como pueden ser *consejo*, *ayuda*, *servicio* o *atención*. Por consiguiente, la estructura interna de *dar* debe de contener el rasgo subléxico [+sustento] que, como se ha señalado en la sección anterior, concuerda con el de la clase semántica con la que coaparece.

Así, es evidente que *dar alojamiento* o *cariño* no es producirlo o ejecutarlo; y que *dar un salto* no es proporcionarlo; no obstante, estos usos –*producir* o *ejecutar* y *proporcionar*– se originan a partir de la definición sistémica, por lo que este hecho, unido al de que estas dos definiciones de *dar* (o paráfrasis) se apliquen a contornos diferentes, confirmaría que se trata de dos sentidos (o acepciones) distintos de un mismo lema. Una manera de comprobar si dos usos de *dar* pertenecen al mismo sentido es recurrir al criterio de Green-Apresjan (citado en Alonso Ramos, 1997: 27 y 2004: 77). Si coordinamos dos nombres que pertenecen a diferente clase semántica y comprobamos que no es posible elidir el verbo, ello sugiere que se trata de dos sentidos distintos. Por ejemplo, *dar cariño* y *dar consejo*: *Durante mi infancia, me dio cariño y consejo*. *Cariño* pertenece a la clase semántica de sustantivos que designan “diversas inclinaciones y sensaciones afectivas que se caracterizan por transmitirse a los demás”, mientras que *consejo*

pertenece a los sustantivos que denotan “asistencia o sustento”. Pertenecen pues, a diferente clase léxica, pero, dado que el verbo se puede elidir, existe una indicación de que se trata de dos muestras diferentes de la selección léxica de una misma acepción del verbo: *dar* con el sentido de entregar o proporcionar. Lo mismo es aplicable a *dar ánimo* (sustantivo que denota estímulo) y *dar explicación* (sustantivo de información): *Esa noche me dio las explicaciones necesarias y mucho ánimo para superarlo*. Sin embargo, esta operación no se puede realizar fácilmente con *dar consejo* y *dar un paseo*: **Esa tarde nos dimos varios consejos antes de partir y un paseo*¹¹¹. El primer *dar* tiene el sentido de proporcionar, mientras que el segundo, el de ejecutar.

Si el objetivo del análisis de los datos es descubrir la profundidad del conocimiento de este verbo por parte del grupo concreto de HNN, presentar todos estos sentidos como definiciones de *dar* no vinculadas conlleva interpretaciones erróneas. La extensión de este verbo a las nociones inmateriales (*cariño, paseo*) es parte de nuestro conocimiento de esa palabra –operativo y tácito– como HN, pero no podemos deducir que el aprendiente de un nivel avanzado ya ha llegado a la conclusión, en este estadio de interlengua, de que en realidad *dar* es una sola cosa. Aunque no es necesario conocer –ni explícita ni implícitamente– este sentido unitario para usar el verbo en todas sus combinaciones, sí que permite al HNN conectar más fácilmente una combinación nueva a los usos ya conocidos; es decir, permite aprender mediante la configuración de redes semánticas en lugar de adquirir el verbo por medio de paráfrasis desvinculadas del significado sistémico. No obstante, se ha de tener en cuenta que aunque se le proporcionara al HNN la definición sistémica de *dar* – que nunca está presente en los diccionarios; estos presentan el listado de acepciones–, tampoco sería capaz por sí solo de inferir automáticamente toda la lista de palabras con las que *dar* se puede combinar (o, lo que es lo mismo, rellenar esa definición sistémica, poco especificada), o prever que cualquier añadido que le haga a esa lista resultará natural a los oídos de los hablantes nativos. Como ya se ha señalado, no todo lo que pase de no existir a existir (en C) puede ser dado. Recuérdese que el sistema lingüístico fija muchas más opciones de las que solemos reconocer explícitamente (Bosque, 2005: LXXIX [2004]), y el sistema establece, por poner un ejemplo, que *dar* y no *hacer* se combina con los sustantivos que designan eventos que se interpretan como

¹¹¹ No obstante, es posible que el problema esté en las especificaciones, pues, a nuestro juicio, resulta admisible decir *me dio un paseo* (agradable, por la parte antigua, en coche...) y un consejo que me salvó la vida; o (¿)Nos dimos varios consejos antes de partir y un paseo para disimular la pena.

unidades informativas, muy frecuentemente relativas a la presentación de contenidos (*Redes*, 2005: 640 [2004]), por lo que *damos conferencias, clases o una comunicación*; pero, en cambio, *hacemos una declaración*, y es que, como observa Alonso Ramos (2004: 94), no es posible controlar semánticamente todos los nombres que aparecen con el verbo. Esta última combinación constituye un ejemplo de lo que hemos denominado en el capítulo 3 combinatoria asistemática. Con esto quedan respondidas las preguntas planteadas al final de la sección 4.1.1.1.

De acuerdo con lo aquí expuesto, es preciso especificar a partir de una definición común –se sobreentiende la sistémica– las extensiones figuradas de *dar*, para disponer de un patrón de comportamiento léxico-sintáctico con el que poder contrastar los datos, que sirva para poder discriminar los usos adquiridos de un mismo lexema de los no adquiridos, y así determinar cuáles de las redes léxico-sintácticas que nacen de *dar* están fijadas en el lexicón de este grupo de aprendientes y cuales están todavía por tejer. Las implicaciones didácticas de estos resultados son evidentes.

En suma, para ser coherentes en este estudio de interlengua con el propio proceso de adquisición del léxico, no perdemos de vista la definición sistémica, aunque usemos paráfrasis para agilizar el análisis de los datos y facilitar su presentación. Si el HNN, a medida que aprende, va vinculando en el lexicón por medio de redes semánticas los diferentes usos de *dar* (recuérdese la hipótesis del funcionamiento del lexicón de Baralo, 2001), resulta contraproducente aprender cada paráfrasis como usos aislados, desconectados entre sí semánticamente pues, la carga y el esfuerzo memorísticos resultan mayores si las acepciones no se conectan.

Por otro lado, en un análisis como el nuestro centrado en el uso que los participantes hacen de todas las estructuras en las que participa *dar*, y no solo de la construcción más frecuente V + N, los límites entre los significados pueden llegar a ser todavía más borrosos, como sucede, por ejemplo, en las construcciones V + sintagma preposicional (SP) del tipo *le dio por reír* o *se dio contra la pared*, dado que se puede llegar a transformar de tal manera el potencial actancial o valencial del verbo –esto es, la estructura argumental: la cantidad y cualidad obligatoria u opcional de los actantes– que ello acabe dificultando la tarea de rastrear el valor sistémico que subyace a una estructura dada.

Recuérdese que planteamos la definición sistémica de *dar* como un esquema sintáctico que se corresponde con el de una oración simple. De esta manera, léxico y gramática quedan integrados en esta definición; si este esquema se altera, el comportamiento de *dar* ya no es previsible en una oración aplicando

la definición sistémica. Si la oración está compuesta por un predicado (*dar*) que lleva un complemento regido, los argumentos que le son propios son otros, ya no necesariamente *A*, *B* y *C* de la definición sistémica, por lo que esta es difícilmente aplicable de manera directa cuando se reproduce este otro esquema sintáctico: V + SP.

Mitatou (2011: 18) se refiere a estos otros esquemas como construcciones menos transparentes desde una perspectiva semántica, y en nuestro análisis abarca las construcciones preposicionales, de participio absoluto o pronominales. Por lo general, en estas construcciones es más complicado recuperar de la definición sistémica el argumento interno del verbo, el elemento *B*, que desempeña la función semántica *tema*, esto es, el participante afectado por el suceso, el que se altera, sufre algún proceso o se desplaza por una ruta que se extiende desde el origen al destino describiendo así una trayectoria (en términos de Lakoff y Johnson). No obstante, aun así, es posible interpretar a partir del contexto en el que aparecen estos usos la noción de *trayectoria* y la de *estado final resultante* –referidas al proceso por el que algo pasa de no existir a existir en el contexto situacional–, esto es, el contenido aspectual de *dar*, por ello, integramos estos usos dentro del mismo lexema. Aunque, en ocasiones, incluso este contenido aspectual es difícil de localizar, nos refuerza en nuestra decisión el hecho de que estas construcciones en todos los diccionarios de uso consultados también aparecen bajo la misma entrada o lema. A continuación, a modo de ejemplo, se recogen algunas construcciones con *dar(se)* que no se corresponden con la estructura V + N (tema), trabajada hasta ahora, y las ordenamos de acuerdo con una escala de transparencia que va –a partir de la definición sistémica– de más a menos transparente. Este orden no es estricto, pues depende de la percepción o interpretación de cada hablante:

[3] Darse + SN: Dicho de una cosa: suceder o existir *Lo único que hacen es legalizar una situación que ya se da* (una situación que ha pasado de no existir a existir).

[4] Dar + con + SN: encontrar a alguien o algo. *Fue complicado pero dimos con él* (Él pasa de no existir a existir a ojos del hablante).

[5] Dar + de + infinitivo (verbos que designan ingestión): ejecutar acciones que denotan ingestión. *Dio de comer a los niños* (Tras la acción verbal, los niños pasan de existir de una manera “no estar comidos” a existir de otra “estar comidos”).

[6] Dar(se) + por + participio: considerar. *Le di por muerto* (Él pasa de existir de un modo –“vivo”– a existir de otro –“muerto”– a ojos del hablante).

[7] Dar + que + infinitivo: provocar una acción. *Esa noticia da que pensar* (Esa noticia causa que se produzca un cambio de estado que va de la no reflexión a la reflexión).

[8] Darse + a + SN (sustantivos que designan actividades si se practican de forma continua o entusiasta): entregarse. *En esa época se dio a la violencia* (Él pasa de existir de un modo en una época a existir de otro modo en esta otra época en el que la violencia cobra protagonismo).

[9] Dar + a + SN (sustantivos que denotan ubicación): dicho de una cosa: tener orientación hacia, mirar. *La ventana daba justo al Guggenheim* (El Guggenheim pasa a poder ser observado debido a la ubicación de la ventana, por lo que esta permite hace que el Guggenheim pase a existir ante los ojos del hablante)

4.2. TABLAS

A lo largo del apartado 4.1 se han analizado los problemas de naturaleza léxico-semántica del verbo *dar* que surgen a la hora de realizar un estudio de interlengua por tratarse de un verbo muy polisémico cuya definición sistémica está muy poco especificada. Hemos visto cómo las propiedades léxicas del verbo pueden no estar explícitas en su definición, pero la sintaxis las descubre. A continuación, se presentan tres tablas fundamentadas en los aspectos teóricos anteriormente tratados en las que se recoge de forma esquemática toda la información relevante para el análisis de los datos, acerca de los usos y estructuras del verbo *dar*¹¹².

En la primera tabla (tabla 2) se recogen las diferentes clases semánticas con las que coaparece *dar* y se especifica si la clase es seleccionada por un predicado *dar* con sentido pleno o de apoyo. En esta tabla solo se atiende a aquellas construcciones que tienen el mismo potencial actancial que el contenido en la definición sistémica: V + N (argumento interno y tema), con las funciones semánticas *agente* (desempeñada normalmente por el sujeto) y *experimentador* o *destinatario* (desempeñadas por el objeto indirecto) opcionales. Este esquema sintáctico abarca, además de estructuras transitivas y ditransitivas, las estructuras pasivas, impersonales, pronominales y las intransitivas –en las que sigue existiendo el tema, aunque este papel semántico sea desempeñado por la función sintáctica sujeto y no por el OD, como en *Me da la sensación de que esta opinión no es compartida por todos*–. Para esta tabla nos hemos servido de las clases léxicas

¹¹² Habría sido muy útil contar con clasificaciones estables por estructuras sintácticas para afrontar análisis como el que realizamos en esta investigación. Hemos hecho una clasificación de acuerdo con nuestros propósitos, por lo que al no ser el objetivo principal del análisis –lo que requeriría una investigación más amplia, más profunda y con otra orientación– se corren ciertos riesgos al tomar decisiones durante la tarea de sistematización de los resultados. No obstante, la clasificación que aquí ofrecemos la consideramos valiosa –aun con márgenes de indefinición– en tanto que abarca todos los usos de los participantes de los corpus analizados.

propuestas en *Redes* en la entrada del lema *dar*, que han sido en su mayoría transferidas a nuestro trabajo: en numerosos casos las listas de sustantivos que conforman la clase (en *Redes*) han sido ampliadas por los datos del *CEDEL2*, y en algún caso aislado, modificadas. También han sido creadas algunas clases semánticas nuevas (los sustantivos son recogidos en el capítulo 7; aquí solo nos referimos a las clases a las que pertenecen).

La mayoría de las combinaciones composicionales del corpus responden a restricciones del verbo sistemáticas, esto es, los sustantivos pertenecen a una clase léxica caracterizada por un rasgo semántico que, de acuerdo con los presupuestos teóricos defendidos por Bosque y De Miguel, deben ser compatibles con algún rasgo subléxico presente en la estructura interna de *dar*. Los rasgos subléxicos de *dar* se pueden deducir de esta tabla. Por otro lado, hemos registrado asimismo algún uso de *dar* + N en combinatoria composicional (por lo que se mantiene el significado sistémico) asistemática, esto es, ajena, a nuestro juicio, a la descripción del sistema lingüístico.

La tabla presenta en la columna de la izquierda la definición sistémica, para que se aprecie que todos los usos pertenecen a un mismo lexema; de esta manera, la tabla permite reflejar la baja carga intensional contenida en la definición sistémica de *dar* y explicar así su extensión máxima. En las dos columnas del centro se recogen los diferentes tipos de combinaciones a las que pertenecen los datos: en la segunda, comenzando por la izquierda, se indica si la construcción está basada en un uso pleno del verbo o en un uso de apoyo –los distinguimos mediante las paráfrasis *proporcionar (transferir)* o *producir[se] (hacer)*–; en la tercera columna, comenzando por la izquierda, se presentan las diferentes clases semánticas a las que pertenecen los sustantivos del corpus, por lo que son las encargadas de sacar a la luz el rasgo subléxico que debe de estar en la estructura interna del verbo *dar* y que es activado en ese contexto, por medio de la concordancia de rasgos. En la primera columna de la derecha, se incluye un ejemplo de cada combinatoria (extraído del *CEDEL2*), para facilitar la comprensión. Y, para terminar, en la última fila nos referimos a las pocas combinaciones composicionales asistemáticas encontradas.

Definición sistémica	<i>Dar</i> en CVP (proporcionar) o en CVA (producir[se])	Clases semánticas seleccionadas	Ejemplos
<i>(A)</i> da <i>B</i> (a <i>C</i>) = <i>A</i> hace que <i>B</i> pase de no existir a existir (en <i>C</i> , de manera que los seres <i>C</i> disfruten o padezcan <i>B</i>)	1. Proporcionar	Nombres que denotan razón de ser, contenido, denominación, destino y otras propiedades de las cosas que se tienen por fundamentales.	También son famosas Maui, Kauai y la Big Island, que da nombre al archipiélago.
	2. Proporcionar	Nombres que denotan asistencia, sustento o protección.	Estas parejas les darían mucho apoyo a los niños.
	3. Proporcionar	Nombres que denotan garantía o confianza. También algunos que designan nociones que se asocian con la estabilidad, o la ausencia de perturbación.	Quedarme en casa me da seguridad.
	4. Proporcionar	Algunos nombres que designan períodos o intervalos.	El último día no nos dio mucho tiempo a disfrutar.
	5. Proporcionar	Nombres que denotan ocasión, oportunidad o situación favorable para hacer algo.	Yo siempre le doy una oportunidad al cine nacional.
	6. Proporcionar	Nombres que denotan capacidad, generalmente de actuar. También otros que expresan difusión, expansión o promoción.	Un estado democrático es el que da libertades a los ciudadanos.
	7. Proporcionar	Nombres que denotan autorización, aceptación y otras formas de asentimiento o aquiescencia.	Seguía teniendo fiebre todos los días, por lo que no me podían dar el alta.
	8. Proporcionar	Nombres que designan lo que ilustra algo o da razón de su existencia.	Es el propio partido el que da el ejemplo de discriminación.
	9. Proporcionar	Nombres que designan la forma en que algo se manifiesta o el aspecto que adquiere.	Tenemos algunos problemas que dan una imagen poco agraciada de la ciudad.
	10. Proporcionar	Nombres que designan diversas inclinaciones y sensaciones afectivas que se caracterizan por transmitirse a los demás.	La adopción nace del deseo de dar amor a un niño.
	11. Proporcionar	Nombres que designan valores y cualidades de las personas o las cosas; frecuentemente si aluden a alguna característica esencial de su presencia o comportamiento.	Había una sola cámara, lo cual le daba más realismo.
	12. Proporcionar	Nombres que denotan relevancia, prominencia.	Al día siguiente nadie dio importancia a lo ocurrido.
	13. Proporcionar	Nombres que designan el resultado de algo, especialmente si es beneficioso o exitoso.	Nos dieron las calificaciones del examen de Selectividad.
	14. Proporcionar	Nombres que denotan estímulo, ánimo o fuerza.	Estos fondos dan un poco de aliento a la economía local.
	15. Proporcionar	Nombres de información.	Se me helaron los huesos cuando el médico me dio el diagnóstico.
	16. Proporcionar	Cuantificadores.	Es un gran profesional que lo dio todo por su equipo.
	17. Proporcionar	Nombres que se refieren a tiempo, actividad u objetos que pueden usufructuarse susceptibles de ser asignados (cita, hora, mesa).	Me dieron hora para las 21:15.
	18. Proporcionar	Adjetivos sustantivados que denotan propiedades positivas o negativas, frecuentemente se manifiesta en	Hay que exprimir la vida al máximo aprovechando todo lo bueno que esta nos da.

	construcciones de relativo: lo bueno, lo malo.	
19. Proporcionar	Nombres que aluden a un reconocimiento por una labor u otra circunstancia (medalla, premio, reconocimiento, mención, regalo, presente).	Le van a dar ese título a Tiger Woods.
20. Producir (se)	Nombres que denotan miedo, inquietud, zozobra.	Los cohetes y los fuegos me dan un poco de miedo.
21. Producir (se)	Nombres que designan otros estados de aflicción, relacionados con la infelicidad, enojo o rechazo.	Si no puedo encenderme mi cigarrillo, me da un poco de rabia.
22. Producir (se)	Nombres que denotan desplazamiento, breve o repentino.	Los españoles hemos dado un paso hacia el futuro.
23. Producir (se)	Nombres que designan formas del trato social.	En la carta le daba las gracias.
24. Producir (se)	Nombres que designan sensaciones, emociones percibidos como placenteros.	Es capaz de darles la felicidad que tanto añoran.
25. Producir (se)	Nombres <i>sensación e impresión</i> , así como otros que designan alteraciones y reacciones físicas adversas, generalmente repentinas.	Me dio la impresión de que el público no diferencia la realidad de la ficción.
26. Producir (se)	Nombres que designan eventos interpretados como unidades informativas relativas a la presentación de contenidos y otros que expresan eventos que se ofrecen a la contemplación pública.	Me gustaría ir a Francia o Irlanda a dar clases de Español.
27. Producir (se)	Nombres que denotan comienzo y final.	Los reclusos dieron origen a esta zona.
28. Producir (se)	Nombres que denotan deseo o necesidad.	Hay tantas cosas que ver que dan ganas de liarse la mata a la cabeza y mudarse.
29. Producir (se)	Nombres que designan ciertos movimientos bruscos, impulsivos o resultantes de algún contacto.	Me puse a dar golpecillos contra el suelo para salir del letargo.
30. Producir (se)	Nombres que denotan molestia en grado elevado.	La pequeña le daba la lata a mi madre.
31. Producir (se)	Nombres derivados de verbos o asociados a ellos, que denotan término, desenlace y otras nociones relativas al fin de las acciones o procesos.	Era incapaz de dar solo una respuesta.
32. Producir (se)	Nombres que designan sonidos.	Dio un grito y se alejó.
33. Producir (se)	Nombres, a menudo en plural, que denotan dificultad (dificultad, problema, complicación).	Trabajar en un lugar cerrado, donde muchos fuman, nos puede dar problemas de salud.
34. Producir (se)	Nombres que designan fenómenos atmosféricos o relacionados con el clima (temperatura, lluvia, nieve, aire).	Orense sigue dando temperaturas muy altas.
35. Producir (se)	Nombres que designan horas.	El encantamiento no se rompe porque dan las 12:00.
37. Producir (se)	Nombres que denotan mejora, limpieza o aseo (reforma, lavado, barrido, planchazo, limpieza, baño, ducha).	Va a darle una limpiadita a la camioneta.
38. Producir (se)	Nombres que se refieren a cosas que se originan en el agente de acuerdo con la capacidad o naturaleza del agente (vaca y leche, árbol y sombra, sol y calor).	Los árboles daban suficiente sombra.
39. Producir (se)	con otros nombres derivados de verbos o asociados con ellos. Admiten paráfrasis con el verbo del que se derivan o con el que se asocian (dar contratos, seguimiento, relevo).	

Tabla 2. Combinatoria composicional de *dar* en la construcción V + N

En la siguiente tabla (tabla 3) se recogen otros usos del verbo con esquema sintáctico-semántico diferente en los que, como se ha señalado, resulta más complicado rastrear ese significado sistémico contemplado en la estructura *A da B a C*, aunque sí que se puede interpretar cierto vínculo con este. Para una mejor exposición de la información, no incluimos en esta tabla la definición sistémica:

Tipo de construcción	Clase	Significado	Ejemplos
1. Construcción de participio absoluto	1. Participio de <i>dar</i> + SN	Teniendo en cuenta lo que dicho nombre expresa.	Además, está permitido bañarse, lo que no da ninguna pereza dada la temperatura del agua.
2. Pronominal	1. <i>Darse</i> + SN	Sucedér, existir.	Lo único que hacen es legalizar una situación que ya se da de todos modos.
	2. <i>Darse</i> + SN	Entregarse con sustantivos que denotan exceso en una actividad cualquiera, de manera puntual: un <i>banquete</i> , un <i>homenaje</i> , <i>festín</i> , <i>atracción</i> , <i>panzada</i> .	Me di un banquete con mis amigos.
3. Construcción preposicional	1. <i>Dar a</i> + SN	Estar una cosa abierta u orientada hacia una dirección.	El sexto piso tenía una tremenda ventana que daba justo al Guggenheim.
	2. <i>Dar contra</i> , <i>en</i> o <i>a</i> + SN	Chocar, impactar.	Luego sintió que caía y con un golpe seco su cuerpo fue a dar contra el suelo polvoriento
	3. <i>Dar con</i> + SN	Encontrar a alguien o algo.	Conseguimos dar con la fecha en la que a los tres nos venía bien.
	4. <i>Dar por</i> + infinitivo ¹¹³	Iniciarse una acción inesperada.	Estábamos aterrizando y me dio por reír, no podía creerme que estaba en la India.
	5. <i>Dar por</i> + adjetivo o participio	Considerar cierta cosa decididamente como ese adjetivo expresa; construcción que denota el momento a partir del cual se considera que comienza la situación que resulta del evento denotado por el participio.	Nunca me he dado por vencida.
4. Construcción con complemento introducido por <i>como</i>	1. <i>Dar algo como</i>	Clasificar, asociar, equiparar algo.	El siglo XX es un siglo de guerras y de crisis de todo tipo que dio como resultado la creación de grandes brechas o abismos insuperables

¹¹³ Aunque no haya acuerdo entre los gramáticos, esta estructura no la consideramos perífrasis verbal aunque tan solo sea porque el infinitivo es conmutable por *eso* (Gómez Torrego, 1999: 3342).

5. Adjetivo o participio	1. Dado/a 2. Dado/a 3. Dado/a (precedido de <i>venir</i>)	1. Inclinado, que tiene tendencia a algo. 2. Concreto, determinado, establecido. 3. Causa o vínculo necesario.	1. La película me ha sorprendido gratamente aun cuando nunca he sido muy dada a este tipo de películas. 2. Acudí a la hora dada. 3. Esto viene dado por la situación actual.
--------------------------	--	--	--

Tabla 3. Combinatoria de *dar* en otras construcciones sintácticas diferentes a la de V + N.

Por último, en la tabla final (tabla 4) se incluyen las construcciones fraseológicas en las que participa el verbo, con diversas escalas de lexicalización. En estas estructuras el significado sistémico puede llegar a desaparecer –si se desconoce o no se logra rastrear el sentido original, literal, del que procede la combinación idiomática–, como sucede en numerosas locuciones. Algunas de estas construcciones son locuciones por sí mismas (locuciones prototípicas o locuciones más estrictas¹¹⁴). Otras no son más que usos lexicalizados de perífrasis verbales o de construcciones factitivas que, aplicados al caso de *dar*, son los que se conocen como *esquemas fraseológicos (semi)productivos*; mientras que en las construcciones preposicionales los complementos preposicionales tienen una función sintáctica, en los esquemas fraseológicos semiproductivos todo el segmento constituye una unidad léxica compleja; se trata de combinaciones en las que el verbo en forma no personal está restringido léxicamente, en el sentido de limitado a unos pocos verbos de determinada clase semántica (*Nueva Gramática de la Lengua Española. Manual*, 2010: 534-535).

¹¹⁴ La *locución*, entendida en sentido estricto, cumple los siguientes requisitos recogidos por Piera y Varela (1999: 4408): a) sus componentes léxicos son invariables; b) sus componentes léxicos no son conmutables por otros; c) no admiten alteraciones de orden; d) no es posible extraer sus componentes, esto es, desplazarlos fuera de la locución; e) la subcategorización no es pertinente, esto es, un componente puede no tener, o no admitir, complementos de la categoría que normalmente requiere; f) las funciones semánticas de los componentes no son las previsible, aunque las sintácticas sí lo son; g) no se puede insertar nada entre dos complementos léxicamente determinados.

Tipo de construcción	Clase	Significado	Ejemplos
1. Locución	1. Dar (se) vuelta	Girar(se)	El ladrón se dio vuelta
	2. Dar igual	Ser indiferente	Me da igual que sea legal
	3. Darse cuenta [de]	Advertir	No me di cuenta de ello
	4. Dar a luz [a]	Dicho de una mujer: parir	Dio a luz a una niña
	5. Dar el brazo a torcer	Ceder	Decidieron no dar el brazo a torcer
	6. Dar lugar [a]	Ocasionar	Eso da lugar a malentendidos
	7. (No) dar abasto	(No) bastar	No da abasto con tanto trabajo
	8. Dar vida [a]	Representar	Da vida a un hombre sencillo
	9. Para dar y regalar	En abundancia	Había discotecas para dar y regalar
	10. Dar una lección [a]	Disciplinar	El terrorista dio una lección al gobierno
	11. No dar para más, (no) dar de sí	(No) rendir; ser capaz de producir más o mejor resultado o de ensancharse	Hice Filología Hispánica; la lengua y la literatura dan mucho de sí
	12. Dar la gana	Querer hacer algo con razón o sin ella	Te puedes ir de viaje cuando te dé la gana
	13. Dar de alta [a]	Autorizar	Pasadas tres semanas me dieron el alta
4. Esquema fraseológico semiproductivo	1. Dar de (a) + infinitivo	Ejecutar acciones que denotan ingestión.	Soy capaz de darles de comer, y de darles la felicidad que tanto añoran
	2. Dar a + infinitivo	Ejecutar acciones que denotan conocimiento o difusión de información, percepción e intelección, elección o ingestión	La productora no ha tenido suficiente dinero para darla a conocer
	3. Dar que + infinitivo	Provocar o dar motivo para que se haga una acción (lo que el infinitivo expresa)	La trayectoria vital de una de las mujeres que más ha dado que hablar

Tabla 4. Combinatoria de *dar* en construcciones fraseológicas

5. LA CUESTIÓN DE LA NORMA EN ELE

La cuestión de la norma lingüística, además de vertebrar numerosos estudios en torno a los usos producidos por los hablantes nativos de una lengua, resulta clave para enfrentar procedimientos básicos en el ámbito de la ASL. En ambos casos, es imprescindible alcanzar una explicación lo suficientemente sólida del concepto que permita un análisis fiable de los usos correspondientes. Habitualmente, el concepto de norma se vincula al de naturalidad, por un lado, y se halla indisolublemente unido al de corrección, por otro. Mientras que la naturalidad mide la representatividad de los datos en función del conocimiento que el hablante tiene de su lengua (Bosque, 2005: XLVIII [2004]), la corrección se refiere, desde una perspectiva muy general, a lo que se produce libre de errores o conforme a unas reglas establecidas.

De acuerdo con ello, se suele distinguir entre dos orientaciones que la norma puede tomar: por un lado, se habla de una norma prescriptiva, reguladora, que centra las políticas lingüísticas con el conjunto de pautas que definen los usos de una lengua considerados prestigiosos¹¹⁵; y, por otro, existe una norma descriptiva, referida a los usos sancionados¹¹⁶ como comunes o normales por una comunidad de hablantes (Coseriu, 1967). La primera de estas orientaciones se relaciona normalmente con la idea de corrección, y la segunda, con la de naturalidad. Pese a este aparente enfrentamiento, como reconoce también Fernández (1997: 29), la norma reguladora debe estar atenta a lo que es “normal”, natural en la comunidad y adecuar a ello sus prescripciones. Así pues, cualquier aproximación al concepto procuraría una integración de ambas nociones –naturalidad y corrección– que, por descontado, suele resultar enormemente compleja. Se trata, en suma, de la búsqueda de un equilibrio necesario en la constitución de un referente lingüístico.

¹¹⁵ El *DRAE* (2001) lo constata: “4. f. Ling. Conjunto de criterios lingüísticos que regulan el uso considerado correcto. 5. f. Ling. Variante lingüística que se considera preferible por ser más culta”. La referencia es pertinente por el papel fundamental que desempeña la Real Academia Española en la prescripción de usos lingüísticos.

¹¹⁶ Recuérdese que como es usual en los estudios normativos, en este trabajo se emplea sancionar con el sentido de “Autorizar o aprobar cualquier acto, uso o costumbre” (*DRAE*, 2001).

En dicha búsqueda, el concepto de *error* adquiere especial relevancia, pues afecta a toda desviación involuntaria de la norma establecida en una comunidad de hablantes; en relación con las nociones anteriores, puede referirse tanto a lo que no es correcto como a lo que no es natural. El error, además, es particularmente importante en el terreno de la enseñanza-aprendizaje de LE. Corder (1967) fue de los primeros en indicar la importancia de los errores por la información que proporcionaban acerca del proceso de adquisición de la lengua, y se refirió a la necesidad de analizar e identificar todas las estructuras producidas por el aprendiente, pues el error podía estar encubierto en una estructura superficialmente bien formada pero inadecuada en el contexto dado. Por primera vez, como señalamos en el capítulo 1, la lengua del aprendiente era el principal foco de atención de los investigadores, y no ya la L1 y la L2 como se venía haciendo en el marco del AC –con el objeto de predecir y evitar las transferencias negativas de la L1 en el proceso de adquisición de las estructuras de la L2. Es en ese momento, con la llegada del AE, cuando se hace necesario fijar en los estudios de ASL una norma o modelo lingüístico que permita describir la interlengua y categorizar los datos de la producción no nativa, es decir, determinar con criterio lo que es error y lo que no lo es; una norma que, asimismo, proporcione la posibilidad de contemplar la enorme variabilidad de realizaciones de la lengua –diatópicas, diafásicas y diastráticas.

En el camino hacia ese modelo lingüístico de referencia adecuado a las pretensiones de la ASL, en el que se busca describir el proceso de adquisición de la lengua y los factores que afectan a los aprendientes en ese proceso, el enfoque del AE se terminó desvaneciendo prácticamente del todo en los años ochenta¹¹⁷ (Ellis, 1994). Atendiendo al objetivo mencionado, y considerando la interlengua como una entidad con valor propio, los corpus de aprendientes constituyen en la actualidad, según se explicó en el capítulo 2, una adición muy valiosa a las fuentes de datos de ASL (Guo, 2006: 11) pues, al operar con una gran cantidad de datos, permite la obtención de unos resultados generalizables y representativos. Asimismo, el corpus de aprendientes incorpora la ventaja de que parte de datos

¹¹⁷ Para una reflexión sobre las debilidades de esta metodología, véase Schachter y Celce-Murcia (1977: 441).

procedentes de un uso natural de la lengua, y no obtenidos de análisis experimentales o de la introspección, frecuentes en la investigación en ASL (Granger, 2002).

De esta manera, los estudios basados en un análisis de corpus de aprendientes –a diferencia de los fundamentados en el AE– ofrecen la posibilidad de examinar desviaciones a gran escala y permiten descubrir, además de las dificultades de los aprendientes, sus aciertos mediante la introspección del investigador pero también gracias a la disponibilidad de los contextos situacionales y de uso, que nunca se pierden de vista (Ellis, 1994). No obstante, para que esta labor resulte completa, conviene contar –tal como ocurre con el CEDEL2– con un corpus de control nativo de iguales características (género textual, temas tratados, características sociales de los participantes, etcétera), que aporta objetividad –más allá de la intuición nativa– a los usos orales o escritos de los hablantes nativos (Nesselhauf, 2004). A partir de este corpus de control, puede establecerse de manera relativamente fiable qué usos lingüísticos no nativos resultan naturales y cuáles no; el procedimiento, que opta más por la vertiente descriptiva de la norma que por la prescriptiva, se basa en que la frecuencia con que se documenta un uso en el corpus de HNN se corresponda con la que presenta el corpus de HN. El criterio de la frecuencia, no obstante, no asegura completamente la naturalidad de un uso, según argumentamos en el apartado 5.2.

He ahí, en efecto, una de las principales controversias en la investigación basada en el análisis de corpus de aprendientes: la determinación de la norma lingüística –como también señala Kaszubski (2000)–, esto es, los usos lingüísticos con los que los datos de la interlengua son contrastados y juzgados. Según señalamos a continuación (apartado 5.1), Leech (1998) se muestra de acuerdo con la tendencia más reciente, que consiste en operar con el concepto de lengua normativa más descriptivo que prescriptivo, por lo que a partir de un análisis contrastivo de la interlengua sería posible conocer no solo los errores de los HNN, sino también los aciertos y los usos que tienden a sobreutilizar e infrautilizar. Por su parte, Quirk (1995) y Widdowson (1998) ponen de manifiesto los riesgos de prescindir de otros estándares en el ámbito de la

enseñanza de lenguas, condición también aplicable a las investigaciones en el análisis de la interlengua.

El presente capítulo, por lo tanto, tiene como objetivo proponer la aplicación de un concepto de norma que integre ambos enfoques: uno descriptivo y otro más regulador (prescriptivo). Dicho concepto, además, pretende superar el obstáculo que resulta de la controversia anteriormente descrita, y que dificulta investigaciones como la presente, sustentada en el análisis del CEDEL2. Se trata, en suma, de la constatación de que disponer de un corpus de control nativo es necesario para analizar la naturalidad en la producción no nativa; y, por otro lado, de poner de manifiesto que el enfoque descriptivo es insuficiente para emitir juicios acerca de la adecuación de los usos lingüísticos.

5.1. LA NORMA DESCRIPTIVA

En la actualidad, lo más frecuente en los estudios sobre la lengua del aprendiente basados en un corpus es disponer de un corpus nativo de control que se concibe como modelo de lengua, pues representa el estado lingüístico ideal (véase, para el inglés, los estudios de Altenberg y Granger, 2001; Nesselhauf, 2005; Guo, 2006; y Prieto *et al.*, 2009 para el español) que debe ser alcanzado por los HNN. En este principio se basa una nueva corriente de estudios en la que se maneja un concepto de lengua más descriptivo que prescriptivo, lo que responde a la conciencia cada vez más extendida en el aula de LE de que la meta en el aprendizaje reside en alcanzar una competencia lo más próxima posible a la del nativo¹¹⁸, esto es, lograr la capacidad de llevar a cabo una actuación comunicativa natural.

De este enfoque de ACI basado en un corpus se desprenden, mediante la comparación entre un corpus de HN y uno de HNN, las diferencias y las semejanzas lingüísticas existentes entre ambos (véase la explicación de Leech sobre este enfoque [1998: xvii] recogida en el apartado 2.2). Por primera vez en

¹¹⁸ Se trata de una aspiración no exenta de obstáculos, según se verá a lo largo del capítulo.

los estudios de interlengua, al amparo de la comparación de ambos corpus, se obtienen los errores que cometen los HNN, al tiempo que se recogen también los usos correctos; en relación con estos últimos, debe precisarse que la corrección desde un punto de vista gramatical no implica la adecuación de un elemento: las semejanzas en el uso (frecuencia, intencionalidad) apuntan a un dominio por parte de los HNN, mientras que las diferencias indican carencias. Según esta perspectiva, si el dominio lingüístico del aprendiente se equipara con el logro de una competencia casi nativa, aquellos rasgos empleados por los HN y ausentes en los HNN habrán de ser aprendidos o utilizados –en aquellos casos en los que conozcan las formas y sus usos contextuales pero no la frecuencia–; y viceversa: aquellos usos correctos formalmente que son frecuentes entre los no nativos y, en cambio, no aparecen en el corpus nativo deberán ser, si no erradicados, usados con menor asiduidad.

Que un corpus nativo de control se conciba como norma descriptiva o modelo de lengua para los HNN ha supuesto sin duda alguna un importante avance en los estudios de interlengua y en la posterior aportación de sus hallazgos al aula: el foco ahora está puesto en el uso real que los HN hacen de la lengua, por lo que las intuiciones del investigador acerca de la naturalidad o corrección de los usos lingüísticos de los HNN pueden ser contrastadas y, por tanto, confirmadas o rechazadas. Asimismo, los rasgos y procesos –presentes en la producción no nativa– permitidos por el sistema lingüístico, pero que no han sido sancionados o no forman parte de la variante lingüística considerada preferente¹¹⁹ por la mayoría de los miembros que integran la comunidad de hablantes de una lengua, se consideran desviaciones de la norma descriptiva concebida como referente lingüístico.

Obsérvense las siguientes muestras obtenidas del CEDEL2 referidas a la combinatoria léxica de los predicados *dar de* y *dar a*:

[1] Ella va a dar de cenar a los niños

[2] Luis daba a beber a los caballos

¹¹⁹ Por *preferente* entendemos la opción más frecuente, pero no exclusiva. Para una reflexión sobre la interpretación de la preferencia, véase Bosque (2001a: 29-31).

En el diccionario combinatorio *Redes* (2005: 642 [2004]) aparece información al respecto que puede sintetizarse de la siguiente manera: 1) *dar de* + verbos que denotan ingestión: *comer/ beber/ cenar/ merendar/ desayunar/ mamar*; 2) *dar a* + verbos que denotan ingestión: *comer/ beber/ ?cenar/ ?mamar/ ?merendar/ ?desayunar*¹²⁰. Estas construcciones aportadas por *Redes* ponen de manifiesto que el sistema tiene dos predicados con igual significado –*dar de* y *dar a*–, cuyo potencial combinatorio es, en principio, el mismo, a saber: seleccionan sustantivos que denotan ingestión. Sin embargo, si realizamos una búsqueda en un corpus nativo de referencia más extenso (CREA), obtenemos los siguientes resultados: *dar de mamar*: 28 resultados / *dar a mamar*: 1; *dar de comer*: 158 / *dar a comer*: 2; *dar de beber*: 21 / *dar a beber*: 7 (4 de los resultados pertenecen a un mismo hablante); *dar de cenar*: 4 / *dar a cenar*: 0.

Así, observamos que *dar a comer, a cenar, a beber y a mamar* son todas ellas opciones del sistema del español que o bien no son sancionadas por los HN, pese a que el sistema lingüístico lo permite, o son concebidas como opciones no preferentes –nótese que *dar a cenar* ni siquiera aparece en el corpus, y *dar a mamar* una sola vez– frente a *dar de comer, de cenar, de beber y de mamar*, que son las coapariciones preferentes y, por tanto, naturales para los oídos nativos¹²¹. Que los docentes y autores de materiales conozcan estos resultados permitirá que tanto el input como la producción lingüística de los aprendientes se filtre con esa norma descriptiva, dado que ahora sí se cuenta con usos lingüísticos reales, y no con la intuición nativa, que no siempre se corresponde con el uso real que los nativos hacen de su lengua; de este modo, volviendo al ejemplo anterior, en un intento de aproximar su producción a la del nativo, el aprendiente seleccionará el

¹²⁰ Los infinitivos precedidos por el signo de interrogación no aparecen en *Redes*, pero ello no implica que la combinación no exista, pues en este diccionario no se pretende proporcionar paradigmas completos, sino definir y ejemplificar paradigmas representativos de las combinaciones léxicas que se agrupan y se analizan en las entradas del diccionario (Bosque, 2005: LXV [2004]).

¹²¹ Una posible explicación de esta dualidad del sistema (*dar de* y *dar a*) –aunque no se señala en *Redes*– es que con *dar de* se explicita únicamente la ingestión (por eso puede aparecer sin sustantivo- complemento), mientras que con *dar a* debe seleccionarse un alimento concreto (lo que explicaría que *dar a mamar* resulte raro); nótese la diferencia entre *le dio de comer* frente a *le dio a comer coles*.

predicado *dar de*, aunque pueda tener almacenado como conocimiento receptivo el predicado *dar a* con algunos sustantivos del paradigma, como *comer* y *beber*. Pese a estas ventajas mencionadas, este principio es cuestionable: su aplicación a rajatabla puede llevar a errores.

5.2. DEBILIDADES DE LA NORMA DESCRIPTIVA

Que una determinada construcción o uso lingüístico abunde entre los HN y que, al mismo tiempo, sea muy escasa entre los HNN –y viceversa– impide hablar de una competencia similar entre ambos grupos. Ahora bien, las semejanzas no implican necesariamente dominio lingüístico, ni la presencia –puntual o reiterada– de un determinado fenómeno lingüístico en el corpus de HN implica necesariamente que su uso sea modélico. Los datos indican únicamente eso, frecuencia, y este es un concepto estadístico, es decir, está en función del número de apariciones de un determinado rasgo en los textos. Frecuencia y naturalidad suelen darse simultáneamente, pero no coinciden necesariamente siempre: un determinado uso puede estar documentado pocas veces pero resultar muy natural para oídos nativos (Bosque, 2005: XLVIII [2004]), y algo frecuente puede ser interpretado como natural o no por HNN, y hasta puede incluir errores, como se recoge a continuación.

La lengua nativa no está libre ni de marcas¹²² –véanse los ejemplos [9] y [10] más abajo– ni de errores o, dicho de otro modo, desviaciones involuntarias de la norma establecida en una comunidad de hablantes. Por ello, la producción nativa no puede erigirse, en un estudio de interlengua, como único referente a ciegas, sin matizaciones¹²³, lo que implica que el estudioso, a nuestro parecer, debe ser capaz de reconocer en su estudio dichas marcas y errores.

¹²² Como es sabido, el concepto de *marca* ha evolucionado mucho desde su origen en el seno de la Escuela de Praga. Aquí nos interesa la perspectiva según la cual los elementos menos marcados se consideran más naturales, más frecuentes que los más marcados (en el sentido en que lo emplean, precisamente en el ámbito de la ASL, Larsen-Freeman y Long, 1991).

¹²³ En Sánchez Rufat (en prensa) se realiza un análisis de corpus contrastivo sobre el uso de la coma en oraciones subordinadas antepuestas. En este estudio se parte precisamente de esta

Por lo que respecta, en primer lugar, al error, repárese en los siguientes ejemplos extraídos del corpus de control del CEDEL2¹²⁴:

[3]. Se *dá* [sic] por muerta

[4]. Hay veces que las parejas gay *les* dan más amor a un niño

Muestras lingüísticas con este tipo de deficiencia ortográfica o morfosintáctica (de concordancia, en este caso) ponen de manifiesto que no toda la producción nativa que conforma el corpus de control está exenta de problemas. Aunque no sean frecuentes en el corpus, pues el porcentaje de aparición es bajo –por lo que podrían considerarse errores más bien aislados, individuales (o tipográficos) o no representativos de la producción nativa en su totalidad–, estos datos integrarían la descripción de la lengua nativa del grupo de control analizado, por lo que la producción no nativa estaría siendo juzgada con respecto a ellos.

Los HN conocedores de la variante lingüística que se considera preferible por ser más culta¹²⁵ saben que estos usos no forman parte de ese referente lingüístico. Pero si se establece como criterio para categorizar un dato de los aprendientes como aceptable-natural el de su aparición en al menos 5 ocasiones en un corpus nativo de control (como fija Neeselhauf [2005] en su estudio sobre las colocaciones), usos similares a los de [3] y [4] podrían ser etiquetados de esta manera, como aceptable-natural.

asunción propia de los recientes trabajos en ASL (Guo, 2006: 5) que el corpus nativo ha de funcionar como referente o modelo. Contrario a las expectativas, el análisis muestra resultados sorprendentes: los nativos adolecen de un mayor desconocimiento del uso adecuado de la coma en este contexto sintáctico, y esa diferencia es significativa –tal y como se constata en el análisis–. Esto demuestra la necesidad sentida durante dicha investigación de recurrir a una norma reguladora de los contextos sintácticos concretos en los que el uso de la coma resulta adecuado, y, al mismo tiempo, prueba que no bastó establecer como única norma un corpus de control escrito de similares características.

¹²⁴ Las cursivas de los ejemplos son nuestras.

¹²⁵ Véase la n. 1.

Con esta misma línea argumentativa puede abordarse, por poner un ejemplo, el queísmo¹²⁶, fenómeno presente y frecuente en ambos corpus. De los usos no nativos de *darse cuenta* en el corpus CEDEL2, se presenta este rasgo en un 30 % de los casos, frente al 18 % en los HN. Obsérvense los siguientes ejemplos de HN ([5] y [6]) y de HNN ([7] y [8]):

[5] Me he dado cuenta que lo construyo con cada acción que hago

[6] De repente, nos dimos cuenta que no era posible viajar

[7] Y me di cuenta que los italianos son tan aficionados

[8] También me dí [sic] cuenta que quiero trabajar con latinos

Aunque la frecuencia pueda ser un indicador de dominio en la producción no nativa, especialmente cuando la frecuencia es amplia (Guo, 2006: 179), no constituye una garantía de dominio. De hecho, existen razones suficientes para que las muestras correspondientes a ese 18 % de queísmo no formen parte del referente lingüístico al que el alumno se ha de aproximar. Un estudioso de la interlengua, aunque opere con un modelo de lengua más descriptivo, puede optar por referirse a estos datos como usos no prescriptivos. Esto es así, no solo por el aliento de focos con vocación prescriptiva como la Real Academia Española, sino también por la “conciencia normativa” o el apego más o menos riguroso a lo lingüísticamente correcto –que suele identificarse, como ya se ha dicho, con la variante más culta– que la mayoría de los hablantes posee. En el caso del queísmo, aunque sea un fenómeno que pueda estar extendiéndose, aparece en el *DRAE* como “indebido” (véase la nota 11) pues, en el intento de equilibrar lo natural con lo correcto, no se considera lo suficientemente generalizado y aceptado en la población hispanohablante. Por su parte, muchos de los nativos que incurren en este uso no tendrían inconveniente en aceptarlo como algo inadecuado tras una explicación satisfactoria (pruebas de sustitución pronominal,

¹²⁶ El *DRAE* (2001) lo define de la manera siguiente: “m. Gram. Empleo indebido de la conjunción *que* en lugar de la secuencia *de que*”. La última edición del *DRAE* (2014) lo define como “Uso, normativamente censurado, de la conjunción *que*, en lugar de la secuencia *de que*, como expresión introductora de ciertos complementos oracionales; p. ej., Me da la sensación que no han venido en lugar de Me da la sensación de que no han venido.

por ejemplo), en razón de la mencionada “conciencia normativa” que los acerca a valores como la corrección o el prestigio lingüístico; ello es aún más esperable en los sujetos cuyas muestras integran el corpus de control manejado, en tanto que se encuentran cursando estudios superiores. He ahí el peligro de considerar sin cuestionamientos estos usos nativos como adecuados: aunque el objetivo del aprendiente sea lograr una competencia lo más próxima posible a la de un HN, debe tenerse en cuenta que ese HN dispone de su concepto de norma, en el que, de algún modo u otro, la prescripción está presente.

En relación con las marcas, si el dominio del aprendiente es juzgado en función de la proximidad de su producción con respecto al referente lingüístico tomado, hay que comprobar si el corpus de control representa fielmente la variedad de lengua seleccionada como modelo lingüístico, o si, por el contrario, incluye marcas o rasgos marcados que lo desvinculan de ese modelo. Porque, en efecto, otro de los problemas que presenta el objetivo de alcanzar una competencia lo más próxima posible a la del HN lo constituye el hecho de que la lengua que se aprende contenga numerosos estándares. En tal caso, la norma descriptiva tendría que vincularse a uno de esos estándares o, en el ámbito del español, a un estándar panhispánico no practicado, en realidad, en ninguna área nativa concreta. Obsérvense los siguientes ejemplos:

[9] Le doy un *puntaje*

[10] Trato de alejarme pero *me da cosita*

Al tomar como norma descriptiva una variedad concreta, surge la posibilidad de contemplar en el corpus las variaciones diatópicas, diafásicas y diastrácticas que se alejen de ella¹²⁷. Así, ante el sustantivo *puntaje*¹²⁸ que aparece

¹²⁷ Al igual que existen diferentes variedades lingüísticas insertas en el corpus CEDEL2, en el aula de LE el alumno está expuesto a un *input* integrado por diferentes variedades: por un lado, el profesor maneja una variedad diatópica más o menos acusada, y, al mismo tiempo, se acoge a la variación diafásica adecuada a cada situación comunicativa; y, por otro lado, los audios y vídeos presentes en los materiales utilizados en el aula suelen presentar, cada vez con más frecuencia, las mismas variaciones, pues se busca desarrollar y ampliar la competencia receptiva del estudiante al tiempo que se pretende la coherencia –la adscripción a una sola variedad– en la competencia productiva.

en el ejemplo [9], el investigador, si fija como referente lingüístico el habla formal de los hablantes cultos del Cono Sur de Latinoamérica, ha de etiquetarlo como natural, dado que su uso está extendido en Argentina, Chile o Uruguay –en estas zonas, por tanto, a este sustantivo le corresponde un uso no marcado–. Sin embargo, en España o México esta voz no se reconoce como perteneciente a la variedad extendida entre los hablantes cultos¹²⁹, por lo que, en el caso de que la variedad de referencia seleccionada fuera el español peninsular o mexicano, estaríamos ante un claro caso de sustantivo marcado desde una perspectiva diatópica. De acuerdo con ello, los hallazgos de este tipo de estudios de lingüística de corpus son aplicables a las aulas de LE. El aprendiente, si tiene esa posibilidad, selecciona una variedad lingüística de acuerdo con sus intereses y, en función de la variedad que elija, unos elementos y no otros serán marcados. Así, y siguiendo con el mismo ejemplo, si el HNN está aprendiendo la variante peninsular, el anterior sustantivo no formará parte del referente lingüístico, pues no es una voz que deba adquirirse para asemejarse a la producción nativa de esta zona geográfica; y, si de alguna forma, durante su formación se topa con él, necesitará saber que es un elemento marcado geográficamente.

De la misma manera, en [10] tanto el investigador como el aprendiente etiquetarán –en el lexicon, en el caso de este, y en la interlengua, en el caso de aquel– esta secuencia con la marca diafásica correspondiente, ya que *darle (a uno) cosa (algo)* es una expresión marcada en la situación comunicativa a la que pertenecen las muestras, que son redacciones de clase. Se trata, en efecto, de un uso coloquial presente en una situación que exige, si no una variedad formal, al menos neutra. Así, sustantivos como *vergüenza* o *apuro* resultan más precisos y apropiados en este tipo de texto; y, por otro lado, el diminutivo aporta una intensidad afectiva igualmente inesperada, pues estaríamos ante un rasgo de estilo muy poco frecuente.

¹²⁸ DRAE (2001) s. v. *puntaje*: “América m. Arg., Chile, El Salv., Hond. y Ur. Conjunto de puntos obtenidos en algún tipo de prueba”.

¹²⁹ Téngase en cuenta la diferencia entre *entender* y *reconocer* un uso. Aquí aplicamos *reconocer* cuando el hablante identifica un rasgo como perteneciente a la norma que hayamos seleccionado, independientemente de que dicho hablante también lo use o de que simplemente sea capaz de interpretarlo por el contexto aportado.

Como acabamos de ver, un corpus de control nativo no está libre de contener todas estas desviaciones (fallo ortográfico, queísmo, dialectalismo¹³⁰, coloquialismo) del propio modelo que, de acuerdo con el concepto de norma descriptiva, representa. Dichas desviaciones ponen de manifiesto que una lengua debe analizarse y enseñarse sobre la base de algún tipo de norma prescriptiva. No cabe duda de que, cuando se lleva a cabo un análisis de la interlengua, disponer de un corpus de control nativo de características similares al de HNN (muestras que pertenezcan al mismo género discursivo, coincidencia de temas sobre los que versan las muestras, participantes con iguales características sociales, como la edad o la formación) es muy beneficioso. Permite saber qué es lo que los nativos producen típicamente en una situación comunicativa determinada, y el análisis contrastivo de la producción no nativa revela si hay determinados rasgos o procesos lingüísticos que los aprendientes tienden a sobreutilizar o infrautilizar en ese mismo contexto. Por ejemplo, un análisis del verbo *dar* en el corpus CEDEL2¹³¹ indica que los aprendientes de nivel avanzado sobreutilizan las estructuras ditransitivas e infrautilizan las estructurales pronominales del verbo (como *darse*, con el sentido de *suced*er una cosa), así como las formadas con un complemento de régimen (como *dar por* + *infinitivo*; *dar con* + *algo o alguien*; *dar por* + *participio*). En cuanto a los sentidos del verbo, sobreutilizan los de *proporcionar* (*dar cariño*) y *producir(se)* (*dar sed*), y, en cambio, infrautilizan el de *ejecutar* (*dar un salto*). Son resultados de un análisis contrastivo que, además de contribuir al conocimiento de la interlengua de un grupo particular de aprendientes, ofrecen aplicaciones al proceso de aprendizaje de español como lengua extranjera.

A pesar de ello, para poder describir la interlengua de un grupo de aprendientes, a los resultados de esta comparación, que son interpretados únicamente en clave de frecuencia –y ya se ha señalado que una amplia frecuencia no refleja necesariamente dominio–, hay que añadir los que se obtengan de un análisis que determine la adecuación de las muestras lingüísticas

¹³⁰ El español se puede concebir como un complejo dialectal (Borrego, 1999), por lo que empleamos aquí el término *dialecto* para referirnos a cada una de las variedades diatópicas del español.

¹³¹ Más adelante abordamos el asunto con la profundidad que requiere en el capítulo correspondiente al análisis del verbo (7).

al sistema lingüístico del español y al contexto situacional en el que aparecen, es decir, a la variedad del español seleccionada como modelo; y, en relación con esto, un corpus nativo también puede presentar problemas (errores aislados o no tan aislados, como es el caso del queísmo) y ciertas particularidades (marcas) de distinto orden. El enfoque aportado por Leech (1998: xvii), sintetizado en el fragmento que reproducíamos en el apartado 2.2 y que exponíamos en el apartado 5.1, requería, por tanto, ser matizado y completado.

5.3. NORMA Y ESTÁNDAR

De lo expuesto anteriormente, nos interesa retomar, por un lado, la idea de que, en un estudio sobre la lengua de aprendientes basado en un corpus, disponer de un corpus de referencia nativo es, por sí solo, insuficiente para emitir juicios lingüísticos; y, por otro lado, se desprende de los ejemplos tratados que tanto la descripción de una lengua como el desarrollo de sus posibilidades de aplicación –como constituye la enseñanza de español como lengua extranjera– se impregnan del concepto de *norma* (Demonte, 2003: 1). En el establecimiento de esa norma, de esa variedad de referencia, debe acudir al enfoque prescriptivo para solucionar problemas como los descritos en relación con el manejo de un corpus de control. Y dicha prescripción, en razón de las consideraciones de los especialistas y de la “conciencia normativa” de los hablantes a la que antes aludíamos, suele vincularse, como ya se ha mencionado, a variantes que se consideran preferibles por ser más cultas.

Así las cosas, en el momento de seleccionar esa variante lingüística preferible entran en conflicto las variedades que existen en la lengua. En un estudio como el que llevamos a cabo con el corpus CEDEL2, se está operando con una lengua, el español, que es oficial en veintidós países, por lo que plantearse qué variedad de lengua ha de erigirse como norma para contrastar y analizar los datos es una cuestión compleja de por sí. Más aún –como veremos más adelante– si se tiene en cuenta la enorme heterogeneidad respecto de la procedencia de los participantes del corpus. Son, así, veintidós las academias de

la lengua que existen, cada una de ellas con una norma caracterizada por un rasgo propio de prestigio que representa lo que es correcto. Cada una de estas normas se establece con carácter sociolingüístico frente a la norma ideal de referencia (Borrego, 2001), esto es, lo que Demonte (2003: 4) denomina “el estándar en estado puro”. Las primeras reflejan el origen de los usuarios, constituyen la variable de prestigio en un ámbito determinado; la segunda, a la que de ahora en adelante nos referiremos como *estándar*, es una koiné (Benincá, 1993: 248), una variedad común a un conjunto de dialectos en la que se descartan las peculiaridades y se seleccionan piezas léxicas y morfológicas transparentes¹³² y de consenso; esta definición refleja por sí misma la dificultad que se plantea a la hora de tener que seleccionar por consenso la opción que resulta más transparente o más prestigiosa.

Para la fijación del español estándar, ya hace algunos años que se está experimentado un giro hacia una idea más vaga e incluyente (Borrego, 1991); en otras palabras, se ha trazado un panorama más panhispánico. El prestigio ya no se identifica con un dialecto determinado, esto es, “no aspira ya a ir asociado a rasgos típicos del castellano peninsular como la pronunciación de la *c* y *ç* como interdentales, de la *s* como apicoalveolar o el uso del leísmo de persona” (Hernández Alonso, 1996: 197, citado en Demonte, 2003: 7). Si a este cambio de actitud se suman las características de un corpus como el CEDEL2, en el que los informantes no nativos aprenden el español tanto en EE. UU. como en el Reino Unido y los nativos proceden de muchos dialectos del español (desde el español de México al de Argentina pasando por el español peninsular), resulta lógico y esperable que las muestras de lengua no sean juzgadas de acuerdo con una norma sociolingüística concreta, y que, por lo tanto, se tomen los dos estándares (americano y peninsular)¹³³ como el referente con el que las muestras, tanto de

¹³² La noción *transparente* que aparece en los trabajos de Benincá (1993) o Crystal (1995) resulta un tanto más superficial que la que nosotros manejamos. Aportamos una definición más precisa a continuación, a partir de la diferenciación entre los conceptos *entendimiento* y *reconocimiento*.

¹³³ La cuestión acerca de si deben aceptarse uno o dos estándares es recurrente en el mundo hispanohablante (Demonte, 2003: 6). No es nuestro objetivo intentar aclarar esta cuestión. No obstante, debe tenerse en cuenta que el estándar americano, si es que existe, representa a veintiún países, veintiuna normas de las que se extraen los rasgos y procesos que configuran el estándar; mientras que el estándar peninsular se corresponde con un solo país, por lo que el

los HNN como de los HN, serán contrastadas. Este referente es el que debiera ser seleccionado tanto en un estudio de interlengua de las características descritas anteriormente como en el aula de español (ya sea en un contexto de LE o de L2). Si bien es cierto que la mayoría de los docentes adopta como variedad para enseñar la propia (resulta hasta cierto punto inevitable no hacerlo), se intenta – como señala Demonte (2003: 7) en consonancia con la anterior explicación de Benincá (1993)– que dicha variedad esté formada por un léxico y una sintaxis no específica, y que las peculiaridades fónicas no se manifiesten de forma llamativa, aunque persistan rasgos tanto fonéticos como prosódicos que identifican la zona de la que se procede; de hecho, es frecuente que estas matizaciones se apliquen también en la interacción con hablantes de otras variedades del español. Según eso, estas características comprenden el material lingüístico más próximo a un estándar puro. Si a la enseñanza de este estándar se une el beneficio de exponer a los alumnos a la mayor cantidad de variantes posible (o normas sociolingüísticas), se logra que estos adquieran, como también observa Vázquez (2008), una competencia receptiva amplia y una producción unívoca no impuesta, resultado de la elección del aprendiente según factores de motivación o necesidades.

Así, en el desarrollo de un estudio de interlengua puede tomarse como referente la unión de los estándares peninsular y americano que, aunque no son del todo homogéneos, se pueden unificar en el análisis y constituir un referente lingüístico de carácter panhispánico. Ahora bien, a efectos prácticos, tomar como modelo un estándar conlleva una serie de complicaciones y dificultades. Las lenguas estándar son abstractas, y no están descritas¹³⁴; los rasgos y procesos de una variedad estándar no configuran un sistema, un todo exhaustivo y homogéneo, sino que surgen (Crystal, 1995) por contraste con los rasgos y procesos considerados regionales, rurales, marginales, inapropiados o

esfuerzo homogeneizador para configurarlo es mucho menor. En este trabajo asociamos la variedad estándar con la mencionada configuración de carácter panhispánico.

¹³⁴ Se definen más por lo que no son que por lo que son, lo cual se debe a que las lenguas estándar son objetos que, por definición, están siempre incompletos (Demonte, 2003: 5), en proceso continuo de configuración y pactando consigo mismos: “son susceptibles de cambios que dependen más de la voluntad de los usuarios que de propiedades objetivas, y constituyen entidades heterogéneas (sociales, convencionales, políticas, lingüísticas) tanto en su origen como en sus límites y contenido”.

incorrectos. Para concretar, por tanto, ese estándar panhispánico se ha de establecer una serie de criterios para que el análisis pueda realizarse de manera sistemática.

En primer lugar, si –como mantiene Crystal (1995)– poder ser entendido por el mayor número posible de hablantes ha de ser un criterio de estandarización para el referente lingüístico escogido, un ejercicio de introspección por parte del investigador basta para juzgar los datos sin marcas de especificidad, transparentes, como pertenecientes al estándar panhispánico –y, por tanto, categorizarlos como adecuados. La noción *transparente* se relaciona con la diferencia entre *entender* y *reconocer* que apuntábamos en el apartado 5.2 (véase la n. 15). No señalamos como transparente el uso o proceso que pueda ser interpretado fácilmente en un contexto dado, sino aquel que los hablantes reconocen como perteneciente a una zona suficientemente representativa del español, independientemente de que sea utilizado en su propia variedad. Así, consideramos insuficiente el criterio de Crystal (1995) que se refiere a la comprensión mutua para establecer la estandarización: volviendo al ejemplo (9), cualquier hablante del español puede entender fácilmente el significado del sustantivo *puntaje*, pero no reconocerlo como un uso suficientemente representativo de uno de los dos estándares principales. El término, por tanto, sería no transparente, en razón de su pertenencia a una norma sociolingüística concreta y, por tanto, restringida a ciertas zonas geográficas. Si, no obstante, se duda de la transparencia de un uso a causa de las restricciones, limitaciones y a la parcialidad atribuible al desconocimiento propio, debe recurrirse a otras fuentes para poder categorizar los datos de forma sistemática.

Es cierto que, por las razones ya esgrimidas, no existen gramáticas ni diccionarios dedicados únicamente a ese estándar general que incluyan todos los rasgos y procesos panhispánicos (estos se describen en parte en algunas gramáticas normativas). Sí es cierto que, en lo que respecta al sistema gramatical, la *NGLE* (2009) puede considerarse como obra de referencia para juzgar los datos de un corpus desde una visión panhispánica. En su introducción, se describe como una obra panhispánica, la primera gramática académica que refleja todas las variedades del español, pues constituye el producto de una acción

conjunta de las veintidós academias; implica, por tanto, la articulación de un consenso que fija la norma común para todos los hispanohablantes, armonizando la unidad del idioma con la fecunda diversidad en que se realiza. La obra, en definitiva, admite recoger una muestra del español de todas las áreas lingüísticas con sus variantes geográficas y sociales, con lo que reflejaría así la unidad y la diversidad de la lengua española. Aunque resulta discutible que tal pretensión, acaso demasiado ambiciosa, haya podido lograrse por completo, la *NGLE*, por la amplitud de sus contenidos y por su vocación panhispánica, se convierte en una herramienta muy útil para la adscripción de los datos al estándar, pues permite adscribir al estándar panhispánico todos aquellos rasgos o procesos que no aparezcan marcados por su pertenencia a una o más variedades.

Tómese, por ejemplo, el componente sintáctico, el más impermeable a los cambios¹³⁵. En él, el estándar (sintáctico) se corresponde en su mayoría con la sintaxis normativa de todos los dialectos del español; en otras palabras, las estructuras suelen estar presentes en todas las variedades, y la única diferencia reside en su valor en la escala de disponibilidad. El siguiente ejemplo ilustra la manera en que pueden procesarse las estructuras sintácticas en un análisis de interlengua. Obsérvese la siguiente muestra extraída del corpus de HNN del CEDEL2:

[11] Me da gusto saber que

A partir de un primer ejercicio de introspección, consideramos que, mientras que en el estándar americano nos parece natural y frecuente la combinación *dar gusto* tanto en estructuras transitivas (*da gusto nadar en el mar*) como en impersonales (*así da gusto*) y en ditransitivas (ejemplo [11]), en el estándar peninsular, en cambio, el valor en el grado de disponibilidad es el mismo en cuanto a las estructuras transitivas e impersonales, pero en las ditransitivas se reduce a contextos más restringidos, en ocasiones con cierta

¹³⁵ Fue fijado a finales de la Edad Media (Demonte, 2003: 11); las pocas variaciones que persisten tienen por lo general una larga historia y representan estrategias alternativas posibles que se sienten todas ellas como normales.

connotación sexual. Por lo tanto, en el español peninsular la construcción ditransitiva sería –y hablamos siempre desde la perspectiva de la introspección– la estructura preferente únicamente en este contexto, aunque su uso, como comprobaremos a continuación, no llega a ser del todo exclusivo de dicho contexto. En cualquier caso, si nuestras intuiciones se confirmasen, los tres usos de la estructura deberían adscribirse al estándar panhispánico, puesto que se documentan en al menos uno de los dos estándares principales.

Para verificar si nuestras intuiciones acerca del uso de esta expresión son correctas, hemos de comprobar el uso que verdaderamente hacen los HN de esta estructura. El corpus de control del CEDEL2 no es lo suficientemente amplio como para albergar la estructura, pero con una búsqueda de la expresión (*me da gusto*) en un corpus mucho más extenso como el CREA podemos confirmar nuestras intuiciones: se obtienen 17 resultados, 15 de ellos de diferentes dialectos americanos, y 2 de ellos se corresponden, en principio, con el dialecto peninsular. De estos, uno se refiere al contexto antes mencionado: *A mí me da gusto, mi madre me da gusto, es una cerda*; y el otro, en cambio, no se refiere a este ámbito: *Creo que me da gusto también escribirte (...) Porque lo hago por decisión propia, porque quiero*. Ahora bien, hay que tener en cuenta que este ejemplo está extraído de una obra del escritor español Víctor Alba, quien estuvo exiliado en México durante más de veinte años. De igual modo, el ejemplo recogido en *Redes* para esta combinación procede también de México: *qué gusto me daría que los señores tomaran como ejemplo a este amante esposo*. Dado que la información contenida en los diccionarios combinatorios¹³⁶ y en los corpus de referencia no es exhaustiva, pues pueden no contener usos muy utilizados por los usuarios de una lengua o incluir otros en un número tan bajo de apariciones que no se puedan hacer interpretaciones sobre su aceptabilidad (Nesselhauf, 2005: 49), y dados los riesgos, ya expuestos, que entraña prescindir de una norma reguladora, ha de recurrirse a una herramienta como la *NGLÉ* para determinar finalmente la estandarización de los datos. En ella se presentan ejemplos de la expresión *dar gusto* tanto en estructuras transitivas como intransitivas indistintamente, sin ningún tipo de especificación geográfica,

¹³⁶ Véase la n. 6.

por lo que sus usos transitivo, ditransitivo o impersonal deben, en definitiva, interpretarse en el análisis como pertenecientes al estándar panhispánico, en tanto que los tres se extienden por, al menos, una de las orillas del Atlántico. En definitiva, las comprobaciones a través de un corpus extenso como el CREA y de un diccionario combinatorio como *Redes*, más la aplicación de la herramienta que supone la *NGLÉ*, confirman las intuiciones iniciales.

En lo que se refiere al vocabulario estándar, diccionarios como el *DRAE*, el *Diccionario panhispánico de dudas (DPD)* o el *Diccionario de coloquialismos y términos dialectales del español* resultan muy útiles para fijarlo. Este vocabulario se presenta en formas léxicas sentidas como transparentes para la mayoría de los usuarios del español, y las obras mencionadas contienen información acerca de la estandarización de algunas voces y construcciones, al tiempo que recogen también estructuras y piezas léxicas marcadas geográfica y diastráticamente, que llegan a estar incluso indicadas como marginales, inapropiadas o incorrectas. El siguiente ejemplo ilustra este proceso de etiquetación de datos léxicos como estándar o no estándar. A juzgar por la información obtenida del *DRAE*, tanto *coche* como *carro* forman parte del estándar, pues cada pieza pertenece al estándar del español hablado en cada orilla del Atlántico: *coche* en el estándar peninsular y *carro*¹³⁷ en el estándar americano; la mayoría de hispanohablantes deberían reconocer ambos lemas como pertenecientes cada uno de ellos a una zona lo suficientemente representativa del español, independientemente de cuál sea el lema utilizado en su propia variedad o dialecto. Diferente es el caso de voces como *puntaje*, que forma parte de la norma culta argentina, chilena, salvadoreña, hondureña y uruguaya. El *DRAE* y el *Diccionario de coloquialismos y términos dialectales del español* ayudan en este caso a categorizar estos usos de manera sistemática. Así, mientras que en el anterior ejemplo *carro* es etiquetado como perteneciente al estándar panhispánico, *puntaje* se etiquetaría como restringido a determinadas zonas.

Dejando a un lado el criterio estrictamente diatópico, un terreno como la combinatoria léxica se muestra muy susceptible a la variación diafásica y a la

¹³⁷ De una búsqueda del lema *carro* en el *DRAE* (2001) se obtiene: “9. m. Am. coche (l vehículo automóvil)”.

variación diastrática. A la primera ya nos hemos referido *supra* a través de la expresión *me da cosita*. A la segunda podemos aproximarnos a través del grado de selección léxica: una gran especialización semántica suele asociarse a un nivel sociocultural alto (Montoro, 2010). Obsérvese el siguiente caso:

[12]. Aprendí que el museo fue establecido antes de su muerte y después Miró le *dio muchas obras*

El uso del verbo *dar* en este contexto no puede etiquetarse como neutro¹³⁸, sino que representa cierta pobreza léxica, pues su grado de especialización semántica se halla por debajo de lo que HN consideraría natural. Un verbo como *donar*, sin embargo, sí cumpliría esa función en la situación descrita y podría ser etiquetado como forma neutra. El estándar manejado, por tanto, debe atender también a las diferencias diafásicas y diastráticas contenidas en cada uno de los estándares que lo componen.

Asimismo, hay variaciones en el uso individual de ciertas combinaciones por parte de los nativos. Tomemos, en función de los intereses de nuestro estudio, las combinaciones con verbos frecuentes. Cuando se pide a HN – incluso de la misma variedad diatópica y con semejante destreza lingüística– que juzguen una determinada combinación, a menudo no la encuentran ni completamente aceptable ni completamente inaceptable (Mel'čuk, 1996: 171, Stengers, 2009). El concepto de error, tan unido al de norma, es, en este ámbito, más difícil de captar (Nesselhauf, 2005: 39) y, por lo tanto, es más complejo establecer una tipología con los dominios de cada categoría claramente delimitados. Debido a esta variabilidad, tanto en el tipo de variación (diatópica, diafásica, diastrática) a la que puede estar sujeta una combinación léxica, como en

¹³⁸ A un elemento se le aplica una etiquetación neutra cuando se usa en situaciones que por sí no requieren de un mayor grado de especialización semántica para que resulten naturales a oídos nativos (Montoro, 2010: 257). Blanco Escoda (2000: 101) señala a este respecto que un verbo soporte, como *dar*, puede vehicular informaciones de tipo diastrático de nivel estándar (o neutro), como en *Juan le da un golpe a María*, respecto a posibilidades de tipo familiar o posibilidades de nivel de lengua más elevado como *Juan le mete un golpe a María* o *Juan le asesta un golpe a María* respectivamente. Para Alonso Ramos (1997: 32) el alto potencial combinatorio del verbo *dar* radica precisamente en que es un verbo neutro estilísticamente.

el grado de aceptación de ciertos usos, se puede establecer una triple categorización de los datos para describir este aspecto de la interlengua. En primer lugar, pueden distinguirse los usos (formas y significados) del verbo que se sobreutilizan o se infrautilizan de los que se usan con igual frecuencia con la que lo hacen los nativos; para ello, como ya se ha indicado, es suficiente disponer de un corpus de control nativo, el cual nos permite calcular frecuencias. En segundo lugar, hay que clasificar esos datos de acuerdo con su aceptabilidad. Para ello, no es suficiente contar con un corpus de control ni con uno más extenso como el CREA, sino que es necesario aplicar el estándar que hemos seleccionado para considerar los usos aceptables o inaceptables. Según el grado de aceptabilidad, se puede establecer una escala formada por dos categorías: aceptable-natural y aceptable-restringido. Los usos clasificados como naturales no difieren en nada del estándar, esto es, no presentan ninguna desviación ortográfica, semántica, gramática, combinatoria, diatópica, diastrática ni diafásica; en cambio, los usos restringidos son muestras de usos correctos desde una perspectiva formal, pero considerados no preferentes por los hablantes nativos¹³⁹ (como los casos *de dar a* vistos en el apartado 5.2). En tercer lugar, el término *inaceptabilidad* y otros términos relacionados (*error, fallo, incorrección*) toman como único sentido en un estudio de este tipo aquellas formas o usos que difieren del estándar en cuanto a los aspectos ortográfico, semántico, gramatical, combinatorio, diatópico, diastrático y diafásico. Dado que el análisis tiene como objeto de estudio la interlengua, no se deben inferir afirmaciones del tipo “los errores son malos” o “es un error, por lo que necesariamente debe ser corregido o erradicado en este nivel de dominio”.

5.4. CONCLUSIONES

Teniendo en cuenta todas las consideraciones anteriores, para la descripción de la interlengua de un grupo de aprendientes resulta de enorme

¹³⁹ Véase la n. 5.

utilidad disponer de un corpus de control nativo de iguales o muy similares características, que permite, al contrastar las frecuencias de los usos de uno y otro corpus, determinar qué elementos sobreutilizan o infrautilizan los aprendientes. No obstante, puede ocurrir que un determinado uso esté poco documentado en el corpus de control pero resulte muy natural para el hablante nativo, y también es posible que un uso frecuente sea interpretado por un alto porcentaje de nativos como no natural o incluso, mediante el filtro de la “conciencia normativa” que todo hablante posee, como erróneo. Para superar estos obstáculos, es necesaria la aplicación de una norma reguladora que precise el grado de adecuación de los usos no nativos, en tanto que el contraste con el corpus de control, como se ha dicho, no siempre resulta suficiente. Piénsese, por ejemplo, en el uso del queísmo, aludido: aunque su presencia en un corpus de control sea relativamente amplia, tanto los focos especializados como un buen número de hablantes lo señalarían como no adecuado.

Para un corpus como el CEDEL2, en el que tanto los HN como los HNN manejan referentes diatópicos distintos, la unión de los estándares peninsular y americano en un estándar panhispánico constituye un referente lingüístico operativo para categorizar los datos de la interlengua. Para ello, sugerimos, una vez establecida una primera categorización basada en los índices de frecuencia, una doble categorización: el primer paso consiste en clasificar los datos de acuerdo con su aceptabilidad, con lo que se distingue entre secuencias aceptables –que, a su vez, se subdividen en estructuras naturales y restringidas– e inaceptables; y el segundo paso se centra en los datos inaceptables y consiste en localizar la fuente de la inaceptabilidad. El estándar propuesto, por lo tanto, contendrá los usos considerados aceptables en al menos una de las dos esferas que lo componen. Además, la determinación de un estándar permite filtrar los usos desde las perspectivas diatópica, diafásica y diastrática. Piénsese en los ejemplos aportados aquí en relación con las características del CEDEL2: *puntaje*, desde el punto de vista diatópico, no pertenece al estándar porque se restringe a ciertas zonas de América –es decir, el término puede ser entendido pero no es reconocido por un hablante como propio de la totalidad del estándar americano–; *dar (a alguien) cosa (algo)*, desde el punto de vista diafásico, no pertenece al

estándar propuesto porque este toma exclusivamente una variedad formal o, al menos, de marcación neutra; y, en fin, *dar obras*, desde el punto de vista diastrático, no pertenece al estándar porque el contexto lingüístico exige un mayor grado de especialización semántica en el verbo seleccionado. En los casos en los que el investigador albergue dudas sobre la estandarización de algún elemento y sienta la necesidad de contrastar sus intuiciones, puede recurrir a herramientas como un corpus más amplio (CREA), diccionarios apropiados (*Redes, DPD...*) o la *NGLÉ*, mediante un proceso similar al que hemos expuesto a propósito de las estructuras ditransitivas con la expresión *dar gusto*. En el capítulo correspondiente a la metodología (6) se aborda con profundidad cada una de las fuentes seleccionadas para contrastar los datos, y los criterios de uso establecidos para el análisis.

6. METODOLOGÍA

Es momento de describir con más profundidad los corpus comparados y analizados en este trabajo –el corpus de aprendientes y el corpus nativo de control–. Asimismo, nos referimos a las herramientas empleadas en la investigación y a los procedimientos seguidos en el análisis.

6.1. LOS DATOS Y HERRAMIENTAS

6.1.1. *El CEDEL2*¹⁴⁰

El Corpus Escrito del Español como L2 (CEDEL2) (Lozano, 2009; Lozano y Mendikoetxea, 2013) es un corpus escrito informatizado de aprendientes de L1 inglés-L2 español que contiene muestras de lenguas de todos los niveles de dominio (inicial, intermedio y avanzado); los aprendientes son clasificados de acuerdo con el test de diagnóstico de la Universidad de Wisconsin (1998), que es independiente y estandarizado, y se puede realizar en línea. El CEDEL2 dispone de un corpus nativo diseñado de igual manera con el objetivo de que sirva para establecer comparaciones entre la lengua de los HNN y la de los HN.

Como se avanzaba en el apartado 2.2.2, este corpus nace en 2006 en el grupo de investigación WOSLAC (Word Order in Second Language Acquisition Corpora), en la Universidad Autónoma de Madrid y la Universidad de Granada, con el objetivo de determinar el papel que desempeñan las interfaces en el desarrollo de la interlengua. El proyecto tiene entre sus objetivos la compilación de dos corpus de aprendientes: el CEDEL2 y el WriCLE (Written Corpus of Learner English), este último constituido por muestras de lengua de hablantes de L2 inglés con español como L1. Un análisis contrastivo del español de los anglófonos y del inglés de los hispanohablantes permitirá determinar si las carencias o déficits de un grupo son resultado de la interferencia de la L1 o consecuencia de patrones de desarrollo universales (Lozano, 2013: 6)¹⁴¹.

¹⁴⁰ La información incluida en este apartado sobre el CEDEL2 proviene de las diversas presentaciones sobre el corpus que su director (Lozano) y la directora del proyecto en el que este se origina (Mendikoetxea) han llevado a cabo (en Lozano, 2009; y en Lozano y Mendikoetxea, 2013).

¹⁴¹ Esta investigación se centra en el estudio de una sola interlengua, por lo que queda pendiente para un futuro trabajo una comparación con la interlengua del corpus WriCLE que permita contrastar y completar los resultados de este análisis.

Hemos recurrido al CEDEL2 –y no al otro gran corpus informatizado de aprendientes del español, el corpus oral SPLLOC (Spanish Learner Language Oral Corpora) (Mitchel *et al.*, 2008), de la Universidad de Southampton– porque constituye una fuente de datos más reales o naturales al haber sido creado de acuerdo con los diez principios del diseño de corpus propuestos por Sinclair (véase al respecto Sinclair, 2005), garantía, por lo tanto, de la calidad metodológica en la conformación del corpus. Estos criterios, que posibilitan comprobar los requisitos de calidad en el diseño y la elaboración de los corpus, son los siguientes (Lozano y Mendikoetxea, 2013: 7): 1) selección del contenido de acuerdo con criterios externos, como es la función comunicativa de los textos, y no la lengua utilizada; 2) representatividad del corpus del estado de lengua seleccionado; 3) contraste con un corpus de control diseñado a tal efecto; 4) criterio estructural basado en la sencillez; 5) almacenamiento por separado de las etiquetas y del texto en bruto; 6) los textos deben comprender actos de habla completos, independientemente de su tamaño; 7) el diseño y la composición del corpus deben estar documentados; 8) equilibrio entre las muestras de lengua oral y escrita si el corpus pretende incluir ambas variedades; 9) el control del tema en un corpus solo debe imponerse por el uso de criterios externos; y, finalmente, 10) los textos deben ser homogéneos, por lo que se deben descartar los que no son representativos de la variedad seleccionada.

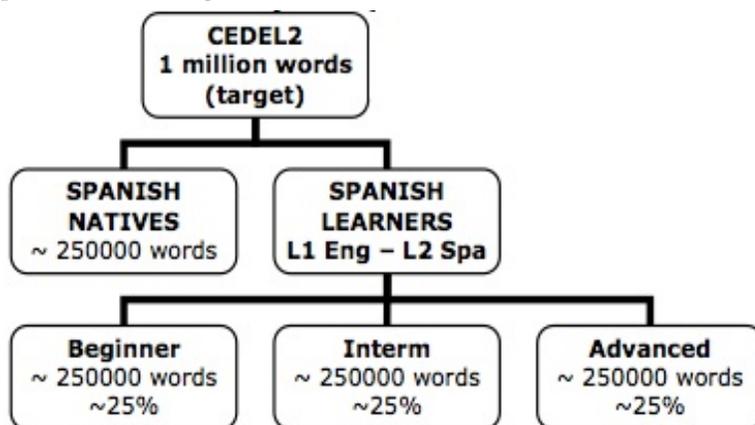
Así, el CEDEL2, se trata de un corpus que, en resumidas cuentas, pretende ser representativo de la lengua escrita únicamente (a la representatividad nos referimos más adelante, en este mismo apartado), por lo que los resultados solo se extrapolarán a la interlengua escrita –no obstante, está demostrado que el conocimiento lingüístico en la interlengua es sistemático, independientemente de que sea representado por el canal oral o escrito (Lozano y Mendikoetxea, 2013: 13)–; por otro lado, el contenido del corpus es seleccionado por criterios externos, es decir, está diseñado para poder contestar a cualquier pregunta de investigación en L2 referida a cualquier estructura lingüística, dado que no se observa ningún desequilibrio de unas estructuras lingüísticas sobre otras.

En el último recuento del que tenemos constancia, realizado en marzo de 2011, el CEDEL2 contaba con 750000 palabras en formato electrónico, ya que los datos se están obteniendo por medio de una aplicación en línea. Se siguen recopilando datos –el objetivo es llegar al millón de palabras–, por lo que todavía no se encuentra disponible para toda la comunidad investigadora ni para el resto de usuarios interesados; aun así, ya se han utilizado los datos existentes en algunos trabajos (presentados en el capítulo 2), a los que esta investigación se suma. Es un corpus considerado suficientemente amplio como para aportar un

panorama fiable de cualquier uso lingüístico: “the learner subcorpus will eventually contain c. 750000 words, a reasonable sample size for a learner corpus” (Lozano y Mendikoetxea, 2013: 16), por lo que produce datos más fiables y naturales que los obtenidos tradicionalmente de corpus más pequeños y diseñados siguiendo una metodología *ad hoc*¹⁴².

En la siguiente figura (tomada de Lozano y Mendikoetxea, 2013: 10) se muestran los criterios estructurales del corpus y el objetivo perseguido en el tamaño de la muestra en lo referente al número de palabras y al porcentaje.

Figura 2. Diseño del corpus CEDEL2



La estructura presentada en el esquema refleja la sencillez del diseño del corpus –principio fundamental de todo corpus diseñado adecuadamente (Sinclair, 2005)–, pues está basado en criterios muy reducidos en número y claramente separables, como son la subdivisión en tres subcorpus de aprendientes en función del nivel de dominio lingüístico y la existencia de un corpus de control nativo.

Los corpus de control permiten un contraste fiable entre los datos de la interlengua y la lengua nativa, que se erige como norma descriptiva en muchos de los estudios de corpus (véase al respecto de la norma el capítulo 5); precisamente, los dos subcorpus son diseñados en condiciones comparables para que el corpus

¹⁴²No obstante, es cierto que el ICLE (Granger *et al.*, 2009), uno de los corpus de aprendientes más grandes que existe en la actualidad, cuenta con casi 3,3 millones de palabras; lógicamente, “the art of sampling has a bearing on the extrapolability of the results: the larger the sample, the more reliable the findings” (Lozano y Mendikoetxea, 2013: 16), si bien conviene que matizar que estas cifras del ICLE se corresponden con los datos de aprendientes de inglés de 16 L1 diferentes, por lo que la diferencia entre el tamaño de los subcorpus de aprendientes de inglés del ICLE y del subcorpus del español del CEDEL2 no es tal (como se especifica a continuación).

nativo sea tratado como norma o modelo lingüístico; a este objetivo se refiere Leech (2001: 333) cuando afirma que “Optimally we also need targeted corpora –corpora targeted to represent as closely as possible the learner’s future communicative needs, (...)”. En nuestra investigación hacemos uso del corpus de control únicamente para detectar rasgos de la interlengua relacionados específicamente con la L1 o con patrones de sobreutilización o infrautilización, como defiende Tono (2003: 803); para analizar estas cuestiones en el corpus de aprendientes, sin lugar a dudas el corpus de control es una herramienta de gran valor. No obstante, una lengua debe analizarse también sobre la base de algún tipo de norma prescriptiva que permita determinar la adecuación de las muestras lingüísticas al sistema lingüístico del español y al contexto situacional en el que aparecen, es decir, a la variedad del español seleccionada como modelo (recuérdense nuestras consideraciones en torno a la norma expuestas en el capítulo 5).

En otoño de 2010, en el momento en el que iniciamos este estudio, los textos sumaban unas 730000 palabras (de acuerdo con la gráfica presentada en Lozano y Mendikoetxea [2013: 16] sobre la evolución del CEDEL2 con respecto al número de palabras) no etiquetadas –por lo que trabajamos con el material puro, sin procesar– a las que hubo que restar las palabras pertenecientes a participantes de nivel inicial e intermedio, dado que este es un estudio de un nivel avanzado de interlengua. En el análisis hemos contado con prácticamente 370000 palabras, de las cuales unas 172000 palabras pertenecen al corpus de aprendientes de nivel avanzado –de acuerdo con el test de diagnóstico de la Universidad de Wisconsin– y aproximadamente 198000, al corpus nativo. Esta cantidad coincide, como señalábamos en la n. 3, con el tamaño de cada uno de los 16 subcorpus del ICLE, esto es, 200000 palabras de inglés escrito producido por hablantes de 16 L1 diferentes, que es considerada como una cantidad suficiente para realizar análisis válidos (Nesselhauf, 2005: 45-46). Como se observa en el gráfico 1, hay un equilibrio en el porcentaje de cada corpus referido al número de palabras: aproximadamente $\frac{1}{4}$ del número total de palabras pertenece al subcorpus nativo, y prácticamente otro cuarto pertenece al subcorpus de aprendientes de nivel avanzado.

Gráfico 1. Proporción de palabras en el análisis del CEDEL2



El número de participantes es de 601 nativos –lo que equivale a un 35 % del total de los participantes– y de 361 aprendientes de nivel avanzado –lo que se corresponde con un 21 %–. Estos datos indican que los textos de los participantes nativos son de menor extensión. No obstante, aunque es relevante tener una estimación del equilibrio del corpus en cuanto al tamaño del texto y la cantidad de palabras con las que contribuye cada participante, lo verdaderamente importante, como señala Sinclair (2005), es que cada texto esté completo desde una perspectiva informativa, esto es, de acuerdo con su función comunicativa, independientemente de la extensión de cada texto.

Samples of language for a corpus should wherever possible consist of entire documents or transcriptions of complete speech events, or should get to this target as possible. This means that samples will differ substantially in size (Sinclair, 2005: 7).

y

There is no virtue from a linguistic point of view in selecting samples all of the same size. True, this was the convention in some of the early corpora, and it has been perpetuated in later corpora with a view to simplifying aspects of contrastive research. Apart from this very specialised consideration, it is difficult to justify the continuation of the practice. The integrity and representativeness of complete artifacts is far more important than the difficulty of reconciling texts of different dimensions (Sinclair, 2005: 6).

Prácticamente todos los nativos han recibido educación universitaria; la mayoría procede de España, pero hay también muchos hablantes de variedades de Sudamérica y América Central. Por su parte, la mayoría de los aprendientes viene de universidades e institutos de EE.UU. (77 %); otros son alumnos universitarios del Reino Unido (6 %) y alumnos en España que se encuentran realizando su tercer año universitario en el extranjero (5 %); hay también un pequeño porcentaje de aprendientes de español de otros países, como Australia, Canadá o Nueva Zelanda (2 %). El 10 % restante lo conforman aprendientes que provienen de otros países o que no han sido especificados en la aplicación en línea.

Así, los textos corresponden a aprendientes que han adquirido la LE en un contexto de aprendizaje formal, para cuya producción no ha habido ninguna restricción en el contenido lingüístico ni temático; por eso, el corpus se considera un corpus representativo, como ya se ha señalado al comienzo del apartado 6.1.1. El corpus representa el estado de lengua elegido para este análisis: los aprendientes participaron voluntariamente y pudieron elegir libremente un tema de redacción de entre 12 propuestos, unos descriptivos y otros más argumentativos –por lo que las muestras de lengua son variadas y no reflejan las restricciones léxicas de un solo tipo de texto (Guo [2006: 55] se refiere a que “different genres or text types will involve a different lexis”; véase Lozano y Mendikoetxea [2013: 9] para algunos clichés léxicos propios del ensayo argumentativo)–. Los títulos de los temas propuestos, que han sido seleccionados de los manuales de enseñanza de español, son los siguientes: ¿cómo es la región donde vives?; habla de una persona famosa; resume una película que has visto recientemente; ¿qué hiciste el año pasado durante las vacaciones?; ¿cuáles son tus planes para el futuro?; describe un viaje que has hecho recientemente; cuenta una experiencia que hayas vivido; habla del problema del terrorismo en el mundo; ¿qué opinas de la nueva ley anti-tabaco?; ¿crees que las parejas gay tienen el derecho de casarse y adoptar niños?; ¿crees que la marihuana se debería legalizar?; analiza los principales aspectos de la inmigración.

Estos títulos sugieren temas muy abarcadores; para abordarlos, no es necesario, por tanto, emplear usos lingüísticos restringidos. Asimismo, dado el carácter neutro de los tipos de texto utilizados, es fácil que los problemas encontrados se reproduzcan en otros tipos de texto, tal y como argumenta Nesselhauf (2005: 46) a propósito de su análisis de colocaciones basado en corpus: “as the text type investigated is highly neutral, i.e. not specialised and on

a medium level of formality, the difficulties identified are likely to occur in most other text types as well”¹⁴³.

El CEDEL2 contiene información precisa y detallada de cada estudiante y de las condiciones en las que es elaborada la redacción, lo que es esencial en una investigación sobre la lengua del aprendiente. Así, cada participante completa, además del test de diagnóstico¹⁴⁴, otros dos impresos en línea: el primero, relacionado con las circunstancias de aprendizaje, abarca datos personales (edad, sexo, información sobre la institución en la que se encuentra aprendiendo español y tipo de curso que realiza), lingüísticos (L1, L1 del padre y de la madre, lengua vehicular en casa, edad y duración de la primera inmersión en un país de habla hispana, así como la duración de otras posibles estancias en un país hispanohablante) y, por último, una autoevaluación de su nivel de dominio estimado en cada una de las cuatro destrezas en español y en otras lenguas. El segundo impreso que ha de rellenar el participante incluye la propia redacción en español e información adicional relacionada con las condiciones de producción de dicha redacción (investigación previa sobre el tema, lugar de redacción –en clase o en casa–, herramientas lingüísticas empleadas, como el diccionario, el sistema de autocorrección, ayuda de algún experto o nativo...).

Toda esta información permite filtrar los datos y descartar los textos que no cumplen los criterios del corpus, así como indagar sobre las posibles causas de los resultados “extraños” o “llamativos”. Partiendo de esta información, y teniendo en cuenta que los datos a nuestra disposición en el momento del análisis eran una versión preliminar del subcorpus del español CEDEL2 que no se correspondía en todos los aspectos con los criterios de diseño del corpus, fue necesario fijar los siguientes criterios para determinar la inclusión de redacciones en nuestro análisis: a) el inglés es la única lengua nativa o la lengua nativa dominante del aprendiente (es la lengua vehicular en casa y es la lengua

¹⁴³ Además de los aspectos citados relacionados con el tipo de texto y temas tratados, la representatividad de un corpus también está vinculada con la existencia de muestras lingüísticas de diferentes estadios de desarrollo de la interlengua; en el diseño del CEDEL2 se optó por un diseño transversal estándar, dada la dificultad logística de tomar un diseño longitudinal, que implica recoger muestras de un mismo grupo de aprendientes durante varios años, a medida que su nivel de dominio se incrementa a lo largo del tiempo (Lozano y Mendikoetxea, 2013: 9). Es nuestro objetivo explorar este aspecto transversal de la estructura del corpus en un futuro trabajo, en el que analizaremos el desarrollo del conocimiento de *dar* desde un nivel inicial hasta un nivel avanzado.

¹⁴⁴ Los informantes, en efecto, han sido clasificados en el nivel avanzado, aunque en este tipo de pruebas siempre existe cierto margen para que puedan darse pequeñas diferencias de nivel. No obstante, como afirma Nesselhauf al respecto de su análisis de las colocaciones léxicas basado en un corpus, “As suggestions for language teaching have to be made for similarly heterogeneous groups, however, the difference in individual proficiency-levels does not seem particularly problematic for the purposes of the present study”.

considerada como nativa para el aprendiente, presumiblemente porque es la lengua de instrucción escolar); b) el aprendiente cursa bachillerato o la universidad, independientemente de su edad; c) los datos personales están disponibles y no falta ninguna información relevante; y d) no ha recibido ninguna ayuda de un nativo o experto en español para elaborar la redacción.

6.1.2. *Antconc*

Antconc es el programa desarrollado por Laurence Anthony (2005) que utilizamos en esta investigación para obtener cálculos lexicométricos¹⁴⁵. Hay dos funciones principales en Antconc (versión 3.2.3.m): concordancias (o concordance), que es la herramienta más importante disponible para el lingüista de corpus (McEnery y Hardie, 2012: 35), encargada de presentar las concordancias o la forma buscada en su contexto (ya sea palabra, morfema, sintagma, expresión multipalabra, etcétera), y el listado de palabras o Wordlist, que produce listados de palabras de diferente tipo.

Antes de profundizar en las funciones del programa empleadas en el análisis, conviene recordar que esta investigación sigue una metodología mixta, característica de los estudios generativos de adquisición de la L1 y de los estudios de adquisición de la L2 desde los años 80 del siglo pasado (Cruz Piñol, 2012: 33):

En la actualidad, se reconoce que los corpus (...) no se pueden analizar válidamente sin la intuición y la facultad interpretativa del analista, que usa conocimientos de la lengua (como hablante nativo o no nativo competente) y conocimientos acerca del lenguaje (como lingüista) (Villayndre, 2009).

Ciertamente, funciones muy concretas pueden ser facilitadas por el ordenador –como los programas de concordancias o los indexadores automáticos–, pero, como señala Bosque (2001a: 11), “los listados de datos que proporcionan tales programas informáticos distan mucho de constituir en sí mismos una aportación directa a la ciencia del lenguaje”. En esta misma línea, Nesselhauf (2005: 41) añade:

Whereas fully automatic analyses are often not advisable for the investigation of learner language (due to the high number of deviations usually present), automatic

¹⁴⁵ Véase la página *web* de Anthony para el uso del *software* en publicaciones de investigación <<http://www.antlab.sci.waseda.ac.jp/software.html>> (19-05-14).

analysis can nevertheless assist the manual work involved in learner language analysis.

Estamos de acuerdo con Nesselhauf en que una de las limitaciones de gran cantidad de estudios de corpus tiene que ver precisamente con el excesivo apoyo en el análisis informatizado, lo que ha llevado a que la mayoría de los rasgos que pueden ser extraídos con facilidad de manera automática ya han sido examinados en los corpus de aprendientes de inglés (no es este el caso en la interlengua de aprendientes de español, dado que los corpus no cuentan todavía con las herramientas de búsqueda y las anotaciones necesarias). Esta limitación desaparece si se parte de una pregunta de investigación específica en lugar de explotar las posibilidades del corpus directamente. Así, una manera de hacer uso de la extracción automática puede ser restringir el análisis a ciertas combinaciones de palabras como las coapariciones con verbos frecuentes (Nesselhauf, 2005: 43), a lo que, precisamente, está dedicado el presente estudio de interlengua de español.

En nuestro trabajo, el análisis automatizado se limita al uso de las concordancias, con las cuales únicamente pretendemos cribar un corpus extenso para obtener un conjunto de datos; a partir de ellos, posteriormente, realizaremos una selección que nos permita alcanzar los objetivos de nuestra investigación. Las concordancias, como explica Sinclair (1991: 32), “is a collection of the occurrences of a word-form, each in its own textual environment”. En el caso de que las formas lingüísticas de un corpus estén anotadas, es posible buscar no solo una forma lingüística, sino también una etiqueta como, por ejemplo, una búsqueda por las palabras marcadas con la etiqueta de la categoría gramatical “verbo”. Scott (1999: 13), al respecto de las concordancias, defiende que

it is through changing the shape of data, reducing it and then re-casting it in different format, that the human capacity for noticing patterns comes to the fore (...) Human beings are good at noticing, and particularly good at noticing visual patterns.

En definitiva, observar la lengua a través de las concordancias es un proceso distinto al de estudiarla a través de textos (Sinclair, 1991; Hunston, 2002).

Cuando el corpus no está lematizado, como es el caso del CEDEL2, para poder localizar en él todas las formas del paradigma verbal es necesario introducir en la búsqueda todas las posibilidades de flexión verbal, y así se hizo

en esta investigación dado que la técnica del asterisco, utilizada para maximizar los resultados en una sola búsqueda, tampoco funciona en el caso del verbo *dar* por su raíz irregular–; al mismo tiempo, tuvimos que atender a los posibles problemas ortográficos de los aprendientes al escribir el verbo, para introducirlos también en la búsqueda. No está garantizado que algunas formas ortográficas erróneas de *dar* no hayan sido capturadas¹⁴⁶.

Cuando la forma *dar* y todas las formas flexionadas de este lema (incluidas aquellas formadas con pronombres enclíticos) fueron objeto de búsqueda en Concordance, se convirtieron en *the node words*, esto es, las palabras seleccionadas que aparecen en el centro de la pantalla, con todas las líneas dispuestas verticalmente en torno al núcleo. Este es el formato de concordancia más común, denominado *Key Word in Context* (KWIC). En la siguiente imagen (figura 3) puede verse la manera en que aparecen las líneas de concordancia (KWIC) en el *Concordance* de Antcon:

¹⁴⁶ Hasta que a los corpus de aprendientes no se les apliquen métodos de detección de variantes ortográficas que permitan etiquetarlos con esta información, las búsquedas seguirán teniendo estas limitaciones. En Rayson y Baron (2011) –introducido en el apartado 2.2.2 (véase n. 44) – se presenta una nueva aplicación de un enfoque híbrido para detectar los errores ortográficos en los datos de los aprendientes de inglés de ICLE; usan una versión modificada del programa *The variant detector* que nace como resultado de aplicar técnicas del ámbito del procesamiento natural del lenguaje al análisis automático de esos datos, lo que genera una cantidad mayor de datos –que la obtenida por medio del proceso de edición manual– para el análisis, y resultados más precisos que los obtenidos al introducir posibles errores ortográficos en la búsqueda de datos.

Figura 3. Concordancias de *dar* (del CEDEL2) en Antconc

The screenshot shows the AntConc 3.2.3m application window. The title bar reads "AntConc 3.2.3m (Macintosh OS X) 2011". The menu bar includes "File", "Global Settings", "Tool Preferences", and "About".

On the left, the "Corpus Files" pane lists two files: "Cedel_2_Advanced1.txt" and "Cedel_2_Advanced2.txt".

The main window displays a concordance table with the following columns: "Hit", "KWIC", and "File". The "Hit" column contains line numbers from 1 to 33. The "KWIC" column contains the word "dar" in context, with the word highlighted in yellow. The "File" column lists the source file for each hit, which is "Cedel_2_Advanced1.txt" for all entries.

At the bottom, the search control panel shows:

- Search Term: (Advanced)
- Concordance Hits: 240
- Search Window Size: 50
- Buttons: Start, Stop, Sort
- Kwic Sort: Level 1 (1R), Level 2 (2R), Level 3 (3R)

Para la presente investigación, dado que aún no se encontraba el corpus disponible, recibimos los textos por medio de una copia en formato Excel (véase figura 4), que tuvimos que convertir a TXT para poder analizar los datos con un *concordancer*. Una vez que se obtuvieron las líneas de concordancias, estas fueron exportadas a una tabla de Excel para proceder al análisis de los datos (véase figura 5).

Figura 4. Tabla de Excel facilitada por Cristóbal Lozano con las redacciones, información adicional sobre sus condiciones de producción, y datos personales y lingüísticos de los participantes

INITIALS:	BEO	dk	DAR	LM	JAH	JEL
UNIVERSITY:	Occidental Col.	Saint Louis Un	Saint Louis un	Saint Louis Un	Open University	Boston Universi
SEX:	female	male	male	female	female	female
DEPARTMENT:	English Litera		Spanish			
AGE:	20	28	23	24	33	20
DEGREE:	English Lit.	Masters	M.A.	Masters	BA Language stu	BA in history
EMAIL:	bory@oxy.edu	douggkollasch@	dranalli@loyol	lemila@hotmail	jennyahurst@yahoo	com
YEAR_OF_COURSE	3rd	3rd	2nd	2nd	2nd	3rd
NATIVE_LANGUAGE	English	English	English and It	English	English	english
STAY_SPANISH_C	Yes	Yes	Yes	Yes	Yes	Yes
STAY_SPANISH_W	Spain and Mexi	Spain	Spain	Spain	Spain	Madrid, Spain
STAY_SPANISH_WI	S- Spring 06 M	1999, 2003, 20	2002, 2005, 20	Fall 2001, Spr	2 years ago	this year
STAY_SPANISH_HI	S- 4months M-1	6 weeks-6 mont	4 months, 6 we	7 months total	2 weeks	since september
FATHERS_NATIVE	English	English	Italian	English	English	english
MOTHERS_NATIVE	English	English	English	English	English	english
LANGUAGE_SPOKE	English	English	English and It	English	English	english
AGE_STARTED_SP	12	14	15	12	30	13
YEARS_STUDYING	8	14	8	12	3	8
ABILITY_SPEAKI	Intermediate	Advanced	Very advanced	Advanced	Intermediate	Advanced
ABILITY_UNDESR	Intermediate	Very advanced	Very advanced	Advanced	Advanced	Very advanced
ABILITY_READIN	Advanced	Very advanced	Very advanced	Advanced	Advanced	Very advanced
ABILITY_WRITIN	Advanced	Advanced	Very advanced	Advanced	Intermediate	Advanced
OTHER_LANGUAGE	No	No	Yes	No	No	No
ABILITY_LANGUA			Italian			
ABILITY_SPEAKI			Very advanced			
ABILITY_UNDERSTANDING_LANGUAGE1:			Very advanced			
ABILITY_READING_LANGUAGE1:			Very advanced			
ABILITY_WRITING_LANGUAGE1:			Very advanced			
ABILITY_LANGUAGE2:			French			
ABILITY_SPEAKING_LANGUAGE2:			Advanced			
ABILITY_UNDERSTANDING_LANGUAGE2:			Advanced			
ABILITY_READING_LANGUAGE2:			Very advanced			
ABILITY_WRITING_LANGUAGE2:			Advanced			
PLACEMENT_RAW	43	43	43	43	43	42
PLACEMENT_PERC	100	100	100	100	100	98
TITLE_COMPOSIT	4. ¿Qué hicist	4. ¿Qué hicist	4. ¿Qué hicist	10. ¿Crees que	11. ¿Crees que	3. Resume una p
RESEARCH_FOR_E	No	No	No	No	No	No
HOW_LONG_WRITE	0,5	.3		0.5	1	0,75
WHERE_WRITE_ES	Outside class	Outside class	Outside class	Outside class	Outside class	Outside class
LANGUAGE_TOOLS	NO	NO	NO	NO	NO	NO
SPELLCHECKER:						
NATIVE_HELP:						
BILINGUAL_DICTIONARY:						
THESAURUS:						
GRAMMAR:						
OTHER_RESOURCES:						
ESSAY:	El año pasado durante las vacaciones del verano, trabajaba	El año pasado pasé las vacaciones de verano en España,	El año pasado, pasé mis vacaciones de verano en la	Creo que las parejas gay deben tener el derecho de casarse y	La marihuana es una droga blanda, utilizado por mucha gente,	este fin de semana estreno la nueva película de

En la tabla de Excel presentada en la figura que aparece a continuación se observan, en la columna Q, las líneas de concordancias obtenidas en el *Concordance* de Antconc; cada una de las demás columnas presentan los diferentes criterios que hemos fijado aquí para analizar los datos:

Figura 5. Concordancias importadas en una tabla de Excel una vez iniciado el análisis de los datos

The screenshot shows an Excel spreadsheet titled 'Análisis 08-20.xls'. The interface includes the ribbon (Gráficos, SmartArt, Fórmulas, Revisar) and the 'Inicio' tab. The spreadsheet contains a table with columns labeled A through Z. Column Q contains text extracted from a concordance search. A red circle highlights a specific row in column Q, which contains the text: 'Concordancias'.

A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R
																Concordancias	
																Idea de mi padre en el centro-este de Italia. La aldea es pequeñísima dado que tiene sólo Cedel_2_Advance	
																le a mi aldea adoptada. En cuanto llegué, mis primos, tíos y amigos me dieron la bienvenida Cedel_2_Advance	
																el comedor contando todo lo que yo había hecho durante el año pasado, de un paseo por Cedel_2_Advance	
																que reflejaba el color espléndido del cielo. Alquilamos un barquito y dimos una vuelta Cedel_2_Advance	
																y. Sería más fácil controlar, como en Amsterdam, donde el gobierno ha dado permisos Cedel_2_Advance	
																go que los que adoptan a niños piensan muy antes de adoptar si pueden dar a los adopt Cedel_2_Advance	
																pasadas para ocuparme que me ponen muy apasionado, pero tampoco que no me dan una Cedel_2_Advance	
																viaje porque para mí, es solamente el principio. Viajar a España me dio ganas de ver el Cedel_2_Advance	
																is que están queridos por nadie. Si los gays pudieran adoptar, podrían dar buenas hogar Cedel_2_Advance	
																concordancia, pue on Ananda, su nuevo amigo. Cuando llegué me saludó en castellano y me dio cuenta de Cedel_2_Advance	
																RSI X B sujeto tá o de la conversaciOn diaria. Tiene mucho más valor que las clases que damos en los in Cedel_2_Advance	
																nieve, las cosas se empeoran. Los administrativos de la universidad se daban cuenta de Cedel_2_Advance	
																ón, Pizarro y muchos más se llevaban bien con estas nuevas crías. Les dieron la oportur Cedel_2_Advance	
																mal. Los judíos de España, llamados sefardíes, y los moros no nos daban el respeto que Cedel_2_Advance	
																speto que nos debieron. Nosotros hicimos mucho para ambos grupos. Les dimos la opr Cedel_2_Advance	
																A sujeto tá o oportunidad de salvarles la alma. Como no creían en el demonio, les dimos la opción Cedel_2_Advance	
																A sujeto tá o se debe compartir, les quitamos su carga pesada y por supuesto les dimos otra opor Cedel_2_Advance	
																tra un inmigrante del oeste de Europa. Además, he leído artículos que dan detalles de Cedel_2_Advance	
																gue y quedan dos meses de este trabajo. Mientras terminar el trabajo, dado que hay m Cedel_2_Advance	
																de lugares de interés en ese pueblo. El nombre de ese pueblo es Praia da Forte, lo cual Cedel_2_Advance	
																blar con ellos fue interesante porque tenían nuestra misma edad y nos dio una perspec Cedel_2_Advance	
																por la cola solamente por ser polaco. Estaba bien alegre. Por fin, me dio el billete, nos Cedel_2_Advance	
																regresaremos a Los Estados Unidos. "El año pasado yo viaje a Salvador da Bahia en Bra Cedel_2_Advance	
																Aceptam a brasileña. También quise aprender un nuevo idioma y el portugués me dio esa oport Cedel_2_Advance	
																de lugares de interés en ese pueblo. El nombre de ese pueblo es Praia da Forte, lo cual Cedel_2_Advance	
																blar con ellos fue interesante porque tenían nuestra misma edad y nos dio una perspec Cedel_2_Advance	
																A Hemos al canto y estoy tan agradecida por la oportunidad que la universidad me dio para viajar y Cedel_2_Advance	
																el visto carteles o cuando oigo canciones en inglés aquí en España. Me da el sentido de Cedel_2_Advance	
																el capitalismo. Pero, a la vez, ¿cómo podemos decir que no debemos dar a los de un p Cedel_2_Advance	
																es muy antigua y grande así que hace siempre frío durante el invierno dado las hojas te Cedel_2_Advance	
																blar con ellos fue interesante porque tenían nuestra misma edad y nos dio una perspec Cedel_2_Advance	
																buscábamos un restaurante marroquí. Por suerte, aquí en España no había ninguno Cedel_2_Advance	

6.2. LOS PROCEDIMIENTOS

Nos centramos ahora en los aspectos metodológicos considerados en el análisis de los datos, y lo hacemos a través de varias secciones: la primera se refiere a algunas cuestiones preliminares que han de ser tenidas en cuenta al llevar a cabo un análisis de un corpus escrito de aprendientes; la segunda se centra en los tipos de construcciones analizados; en la tercera se discuten las fuentes seleccionadas para contrastar los datos; y las dos últimas secciones están dedicadas a los criterios para determinar el grado de aceptabilidad e inaceptabilidad de los datos.

6.2.1. Cuestiones preliminares sobre la metodología del análisis de interlengua basado en corpus

Debido a que los datos analizados en esta investigación se obtienen a partir de redacciones de clase —recuérdese que son elaboradas bajo condiciones de producción normales¹⁴⁷—, los resultados del análisis no informan del conocimiento receptivo de los aprendientes. Ni siquiera es posible afirmar que se pueda determinar con precisión el conocimiento productivo escrito de *dar* por parte del grupo de aprendientes en un estadio de la interlengua concreto como el avanzado, pues, por un lado, como también observa Nesselhauf (2005: 42) en su trabajo sobre las colocaciones léxicas, a partir de un análisis de corpus solo se pueden extraer conclusiones sobre los datos que conforman el corpus, por lo que desconocemos si en aquellas construcciones en las que puede participar el verbo *dar* pero que no han sido empleadas por los aprendientes habrían tenido dificultades; en otras palabras, la ausencia de una determinada estructura en el corpus impide averiguar si el alumno la conoce productivamente. Asimismo, como quiera que la extracción de los datos se ha hecho de manera mecanizada para explotar las posibilidades de análisis proporcionadas por un corpus extenso, no se han buscado en el corpus aquellas construcciones en las que *dar* debiera aparecer y otro verbo ha sido seleccionado erróneamente en su lugar, lo cual requeriría una extracción manual de los textos. Ya se ha apuntado (en el apartado 1.1) que en esta investigación se pretende demostrar que un enfoque de corpus informatizado aplicado a los estudios de interlengua del español puede ayudar a encontrar las semejanzas y diferencias entre el español escrito (en lo que se refiere al verbo *dar*) de un grupo de HN y el de un grupo de HNN; y precisamente una de las ventajas de disponer de un corpus informatizado es que permite trabajar con una gran cantidad de textos, por lo que los resultados de los análisis son más significativos que los de un análisis íntegramente manual que, por lo general, se aplica a corpus menos extensos. No obstante, un análisis manual de los textos que detectara usos como los incluidos a continuación (obtenidos del CEDEL2) completaría, sin lugar a dudas, nuestros resultados:

[1] Nos dimos cuenta de la importancia que ponen en relajarse y mantener amistades

[2] Al final se hicieron las 18:30

¹⁴⁷Entendemos por *condiciones normales* las que se desarrollan cuando el ejercicio por el que se obtienen los datos está muy poco pautado; no hay restricciones significativas ni formales.

En español, no *se pone importancia en algo*, sino que *se da importancia a algo*. De la misma manera que no *nos hacen las [X] horas*, sino que *nos dan las [X] horas*; pese a que sí se nos hace tarde.

Tal vez, en algunos de estos casos cabría incluso la posibilidad de reemplazar el verbo frecuente por otro con un mayor grado de especialización semántica –del tipo a *Nos dimos cuenta de la importancia que le atribuyen / conceden a relajarse y a mantener amistades*–; pero, dadas las características tipológicas de los textos con los que se está trabajando, se considera que el verbo menos marcado en este contexto y, por tanto, más natural y frecuente, es el verbo objeto de nuestro estudio que, en este tipo de texto, aporta neutralidad y naturalidad al discurso, pero no así pobreza¹⁴⁸.

Por otro lado, también se ha de considerar que “correct production by a student is not evidence that the student has mastered the language in question, only evidence that the student has produced the correct language on that particular occasion” (Lewis, 1993: 10). En resumidas cuentas, como afirma Granger (2002: 6), “it is difficult to control the variables that affect learner production in a non-experimental context”. En relación con ello, la metodología basada en corpus, al igual que los sistemas de obtención de datos por medio de técnicas de elicitación, como la narración de eventos a partir de dibujos, reconstrucción de textos, actividades diseñadas para explorar aspectos específicos de la lengua del aprendiente –como los pronombres clíticos (véase SPLLOC 1) o cuestiones de tiempo y aspecto (véase SPLLOC 2)–, son fuentes de datos que constituyen muestras de realización o actuación, no de competencia –como son los juicios de gramaticalidad, por ejemplo–, por lo que solo permiten deducir cómo es la competencia de forma indirecta. Estas cuestiones hacen que, en lo que se refiere al examen de la competencia de los aprendientes en el uso escrito del verbo, los resultados de un análisis de corpus no puedan ser interpretados de manera exhaustiva.

Pese a todas estas observaciones (limitaciones) referidas a la metodología basada en un corpus, y a pesar de que existen otros métodos de obtención de datos, como se acaba de señalar, se ha seleccionado este enfoque (el de corpus) porque se considera el más apropiado para abordar la cuestión del uso del verbo frecuente, un área de estudio apenas trabajada en ELE; ha permitido, de hecho, aportar un panorama más general de tal conocimiento en este estadio de interlengua, lo cual parece más importante que determinar con exhaustividad –

¹⁴⁸ De igual manera, Blanco Escoda (2000: 101) se refiere a las informaciones de tipo diastrático que el verbo *dar* puede vehicular como de nivel estándar con respecto a otras posibilidades de tipo familiar o más elevado.

por medio de alguna prueba de elicitación– el conocimiento de un número reducido de construcciones preestablecidas en las que el verbo *dar* participa.

Con todo, no ignoramos las reivindicaciones de trabajos como los de Lozano y Mendikoetxea (2012) y Gilquin y Gries (2009) acerca de la importancia de los datos obtenidos por medio de juicios de HN, incluso en aquellos casos en los que también hay corpus disponibles¹⁴⁹, y acerca de la necesidad de desarrollar una metodología que combine datos basados en corpus y datos basados en la elicitación o datos experimentales para encontrar las pruebas convergentes; en otras palabras,

Using converging evidence to triangulate results is paramount in current second language research: corpus-based results can, and should, be validated against corpus-external findings, and combining naturalistic and experimental data is crucial to gain insight into the relation between the two types of data (Lozano y Mendikoetxea, 2012).

De esta manera, Lozano y Mendikoetxea (2012) concluyen con un importante planteamiento: hay “a real need to use converging evidence on the same linguistic phenomenon in SLA”¹⁵⁰.

6.2.2. Estructuras consideradas en el análisis

En este trabajo se han examinado todas las estructuras proyectadas por el predicado verbal *dar*, esto es, aquellas en las que la red léxico-sintáctica del verbo es reconocible, así como las construcciones gramaticalizadas –en las que el verbo se convierte en un auxiliar– y las construcciones lexicalizadas –como son las locuciones verbales y los esquemas fraseológicos semiproductivos– en las que *dar* participa; recuérdese que en el capítulo 3, las unidades fraseológicas de las que forma parte el verbo se consideran un conocimiento integrado en la competencia léxica del verbo; concretamente, son las unidades que caracterizan la subcompetencia léxica-rutinaria.

¹⁴⁹ A la cuestión de los juicios de los HN nos referimos en la sección 6.2.3.

¹⁵⁰ Estas observaciones ponen de manifiesto la necesidad de abrir una línea de trabajo para futuros estudios que completen los resultados de este análisis de corpus, por medio de la aplicación de diferentes pruebas de elicitación que, por un lado, permitan contrastar los resultados obtenidos en esta investigación y, por otro lado, permitan determinar si los HN conocen las construcciones que estos verbos pueden proyectar pero que no han empleado en el corpus; y que, al mismo tiempo, ofrezcan también la posibilidad de investigar con qué grado exacto los aprendientes conocen las estructuras que ellos mismos producen en el corpus.

Ninguna de las estructuras en las que puede aparecer este verbo ha sido descartada a priori. No obstante, hay construcciones en las que participa *dar* que no han sido encontradas en ninguno de los dos corpus (nativo y no nativo). Como este hecho no proporciona información alguna para un estudio comparativo, han sido excluidas de toda consideración. Aun así, los datos analizados se corresponden con estructuras sintácticas muy variadas, como se desprende de la información recogida en las tablas 2, 3 y 4 del capítulo 4: construcciones transitivas, ditransitivas, intransitivas, impersonales, pronominales, pasivas, construcciones de participio absoluto, construcciones preposicionales, locuciones y esquemas fraseológicos semiproductivos.

De los resultados obtenidos de la primera búsqueda con el programa Antconc, se aislaron los usos relevantes del verbo –que pertenecen a alguna de las construcciones mencionadas– de los datos no relevantes –que no se corresponden con las anteriores construcciones–. Como ejemplo, la locución conjuntiva *dado que* es uno de los resultados obtenidos de la búsqueda del lema *dar* que ha sido excluida del estudio¹⁵¹; esta construcción pertenece –junto con los resultados de búsqueda que forman parte de los títulos de obras artísticas (como en el caso de la película *Te doy mis ojos*) y de nombres propios (*Costa da Morte* o *Leonardo di Caprio*)– a la categoría de los resultados no relevantes.

Por otro lado, las estructuras transitivas en las que aparecen como argumentos del verbo complementos directos coordinados pertenecientes a diferentes clases semánticas se han contabilizado como usos diferentes.

Véase el siguiente ejemplo:

[3] En cuanto llegué, mis primos, tíos y amigos me dieron la bienvenida [sic] y una comida absolutamente fantástica (3 HNN).

En esta construcción, dos son las combinaciones examinadas: *dieron la bienvenida* –siendo esta combinación una muestra de la combinatoria de *dar* con la clase de sustantivos que designan “formas de trato social: saludos, recibimientos y despedidas”– y *dieron una comida* –interpretando *comida* como un sustantivo que se asimila a los que designan “entidades materiales”, como un plátano, un plato de pasta o una mochila con un bocadillo–. Esta oración, concretamente, presenta un empleo erróneo del recurso de la coordinación si se aplica el criterio de Apresjan (véase en el apartado 4.1 2 las matizaciones realizadas a propósito de

¹⁵¹ Aunque *dado que* procede del participio *datus* –al igual que el adjetivo *dado*, que sí se ha analizado en esta investigación–, hemos descartado este uso porque lo consideramos demasiado alejado de la categoría verbal.

este criterio), dado que no se puede elidir uno de los dos verbos cuando se coordinan dos usos de *dar* que actualizan dos sentidos diferentes; en el primer caso, *dar* se utiliza en el sentido de “ejecutar o hacer efectivo”, y en el segundo, en el sentido de “proporcionar”.

En suma, el análisis realizado de todas las construcciones en las que participa *dar* cubre una importante porción de la interlengua española que no ha sido analizada en los estudios existentes sobre la lengua de los aprendientes de español.

6.2.3. Fuentes seleccionadas para contrastar los datos

En esta investigación se han acometido dos tipos de análisis: uno cuantitativo, que pretendía obtener el índice de frecuencia de uso del verbo *dar*, en general, y en cada una de las diferentes construcciones que proyecta, para comprobar si el uso que hacen los HNN es el natural –semejante al uso que los HN hacen del verbo– o, por el contrario, estos tienden a sobreutilizarlo o infrautilizarlo, y, en caso afirmativo, determinar qué construcciones son las desviadas del uso nativo. Para este análisis, se requería únicamente el subcorpus de control contenido en el CEDEL2 de iguales características al subcorpus de aprendientes. Por otro lado, se ha llevado a cabo un análisis cualitativo de cada uso para determinar si es el apropiado; de este modo, se ha distinguido, en primer lugar, entre usos del verbo aceptables e inaceptables. A su vez, los aceptables se han subdividido en estructuras naturales¹⁵² y estructuras restringidas. Por motivos ya discutidos en el capítulo 5, ha sido necesario aplicar una norma reguladora que precise el grado de adecuación de los usos no nativos, en tanto que el contraste con el corpus de control no siempre resulta suficiente. Dicha norma la representan las fuentes mencionadas en el citado capítulo, que abordamos ahora en profundidad.

Habría sido ideal que cada combinación analizada en el corpus hubiera sido juzgada por un gran número de HN o contrastada en un corpus muy extenso. Con respecto a la primera posibilidad, debido a la gran cantidad de combinaciones que han sido objeto de análisis en este estudio (unas 600), no

¹⁵² La naturalidad suele ir de la mano de la frecuencia, pero no son conceptos equivalentes. Hay combinaciones poco documentadas en un corpus nativo que pueden resultar muy naturales a oídos nativos; por ello, se ha seleccionado el concepto de *naturalidad* y no el de *frecuencia* cuando estas dos cuestiones no coinciden. Para una reflexión sobre estos dos conceptos, véase Bosque (2001 y 2004).

resultaba viable que una gran cantidad de informantes nativos juzgara cada una de ellas; además, en un estudio de estas características los datos obtenidos de la introspección deben ser completados con los obtenidos por otras fuentes, como se argumenta en el apartado 6.2.3.1. En lo que se refiere a la segunda cuestión, son bien conocidas las limitaciones de los corpus, que pueden o no contener combinaciones muy utilizadas por los usuarios de una lengua, o incluir otras en un número tan bajo de apariciones que resulte imposible hacer interpretaciones sobre su aceptabilidad (Nesselhauf, 2005: 49). Además, “ningún corpus imaginable contendrá todos los datos que puedan considerarse representativos en un estudio lingüístico cualquiera suficientemente amplio” (Bosque, 2005: CLIX [2004]).

Así las cosas, se ha adoptado un enfoque más práctico con el objetivo de completar el desarrollado por Nesselhauf (2005) en su análisis de las colocaciones del inglés. Se han seleccionado cuatro tipos de fuentes diferentes para contrastar los datos lingüísticos y determinar el grado de aceptabilidad de la combinación: introspección propia, diccionarios combinatorios y otras obras de consulta de carácter panhispánico, y extensos corpus nativos de referencia; por último, la introspección propia se contrasta con la de otros HN (a estas últimas – introspección propia y de otros HN– nos referimos a continuación y en el apartado 6.2.4).

6.2.3.1. Introspección

La competencia lingüística es un conocimiento operativo e inconsciente en los hablantes nativos de una lengua, es decir, está interiorizada, forma parte de su propia naturaleza; por ello, es lógico preguntar a los HN por sus reacciones ante determinadas combinaciones de palabras. Como señala Bosque (2011: 47),

Suele decirse que los datos obtenidos de un corpus, es decir, de un conjunto de textos, son siempre más objetivos que los obtenidos de la introspección. Se olvida a veces, sin embargo, que cuando el lector es hablante nativo filtra sistemáticamente por ella los textos que aparecen ante su vista. Reacciona, pues, ante ellos, en lugar de limitarse a dar fe de su existencia, a diferencia de lo que haría, por ejemplo, un ordenador.

Es probable que en los textos del corpus, no solo en el de los HNN, sino también en el de los HN, aparezcan combinaciones que no se ajustan a lo que es

esperable de acuerdo con lo que establece el sistema lingüístico, ya sea por hipercorrección, por la inmediatez propia del discurso coloquial o por la falta de control en la redacción (Montoro, 2010: 257-258). Podemos distinguir estos usos restringidos, incluso inadecuados, de los usos naturales del español, dado que, como usuarios nativos del español y como lingüistas, llevamos a cabo un proceso implícito de selección, interpretación y análisis cuando hacemos frente a los datos que proporcionan los corpus textuales. En relación con esto, Bosque (2011: 47) se refiere a las particularidades que percibimos los HN en los textos (sentidos infrecuentes, juegos de palabras, usos estilísticos diversos, cruce de combinaciones...) gracias al contraste implícito que realizamos con las previsiones de nuestro sistema lingüístico interiorizado. De hecho, como reclama Bosque (2011: 47), los ordenadores no son capaces de interpretar las combinaciones encontradas una sola vez en el corpus; no pueden distinguir una pauta estable en el sistema gramatical de un efecto estilístico buscado por el autor o un simple error. Ha de decidirlo el hablante o el lingüista en función del conocimiento interiorizado que tienen del idioma. Ya se ha señalado (en la conclusión del apartado 6.2.1) la importancia para Lozano y Mendikoetxea (2012) de los datos obtenidos de la introspección, incluso en aquellos casos en los que se cuenta con la existencia de los datos de un corpus.

Estos argumentos que subrayan la importancia de los datos obtenidos por los juicios nativos contrastan con los de aquellos autores (Sinclair, 1991; Mel'čuk 1996; Hunston, 2002; o Stengers, 2009, entre otros) que ponen de manifiesto el desacierto en el que incurren en ocasiones los HN a la hora de juzgar las probabilidades de coaparición de una palabra con otra.

Speakers differ in what they find acceptable, individual speakers are often uncertain about whether they find a combination acceptable or not, and, additionally, there can be differences between the acceptability judgments of speakers and their own use (Nesselhauf, 2003: 228).

Y

It emerged in the course of the analysis that native speaker intuition on the acceptability of word combinations is usually stricter than native speaker production (Nesselhauf, 2003: 241)

En este orden de cosas, Bosque (2011: 48), pese a defender la capacidad del hablante a la hora de decidir sobre la aceptabilidad o naturalidad de los usos

en función del conocimiento interiorizado que tiene del idioma, reconoce que para ciertas tareas, en particular para obtener paradigmas léxicos exhaustivos, la introspección resulta ser un recurso insuficiente. Por ello, aconseja completar los datos obtenidos por la introspección nativa con los datos obtenidos de los textos.

6.2.3.2. Otros corpus textuales

Los corpus textuales ponen de manifiesto la existencia de variantes en las que puede que, como investigadores, no hayamos pensado; o pueden mostrarnos la presencia de estructuras o redes léxicas cuya existencia desconocíamos, o con las que no estábamos muy familiarizados (Bosque, 2011: 48). Normalmente, las construcciones más comunes pueden ser encontradas en el corpus de control del CEDEL2; si una estructura usual se encuentra en el corpus de aprendientes, muy probablemente también será encontrada en el corpus de control. Ahora bien, el hecho de que algo exista en el corpus de aprendientes pero no pueda ser localizado en el de control puede deberse a que el uso sea incorrecto o a que el corpus de control no sea lo suficientemente extenso. Para subsanar este último problema, se ha recurrido en este análisis a dos de los corpus más extensos del español, el CREA y el Corpus del español, que han servido de corpus de refuerzo¹⁵³.

El CREA escrito es un banco de datos compilado por la Real Academia Española, que está compuesto por 154 millones de palabras procedentes en su mayoría de libros y textos periodísticos de todos los países de habla hispana desde 1975 hasta 2004; es un corpus representativo del español común, pues el tipo de lengua que aparece –el de la prensa– no pertenece a una variedad idiomática distinta de la que corresponde al resto de los hablantes, ni pone de manifiesto, como señala Bosque (2005: CLXI), la existencia de un sistema gramatical o léxico distinto. Por su parte, el Corpus del español, compilado por Davies, es más heterogéneo, consta de 100 millones de palabras procedentes de textos de diferente tipo –lengua hablada, ficción, prensa, literatura– y de diferentes épocas históricas –desde el siglo XIII hasta el XX–, de los cuales hemos seleccionado únicamente los de carácter escrito del siglo XX.

¹⁵³ No nos detenemos en aportar información detallada de estos corpus. Véanse sus localizaciones en línea: <<http://www.rae.es/recursos/banco-de-datos/crea>>, <<http://corpus.rae.es/creanet.html>> y <<http://www.corpusdelespanol.org/>>.

El primero es más extenso, y el segundo permite realizar búsquedas más avanzadas porque está lematizado, lo que resulta beneficioso para nuestra investigación. En cualquier caso, tanto el CREA como el Corpus del español son herramientas muy útiles para la descripción lingüística, y tienen un gran potencial como fuentes de datos de contraste.

6.2.3.3. Diccionarios combinatorios

Asimismo, a la hora de determinar la aceptabilidad o naturalidad de los usos del verbo *dar* en los HNN, el español cuenta con la inestimable ayuda de dos diccionarios combinatorios de reciente creación¹⁵⁴ que ya han sido mencionados en varias ocasiones en los anteriores capítulos: *Redes* y el *Diccionario combinatorio práctico del español contemporáneo* o *Práctico*. Ambos han sido redactados a partir de juicios de informantes nativos y de un corpus de prensa española y americana de la última década del siglo XX y principios del siglo XXI compuesto por más de 250 millones de palabras. En los dos se separan los usos físicos del lema de sus usos figurados, que se presentan como extensiones naturales de los primeros. Mientras que *Redes* contiene alrededor de 8000 entradas y 200000 combinaciones, *Práctico* incluye cerca de 14000 entradas y 400000 combinaciones. Este último es un diccionario concebido para el uso, por lo que no posee la carga conceptual de *Redes*; en ambos se “traza una frontera entre esas expresiones manidas y otras agrupaciones de palabras, igualmente acuñadas, que se sustentan en bases semánticas restringidas y están integradas en el sistema lingüístico” (Bosque, 2006: XV). Esto es, los dos diccionarios distinguen entre lo que hemos denominado *combinatoria sistemática* de la *asistemática*. La vía que se sigue en *Redes* consiste en presentar verbos, adverbios, sustantivos y adjetivos y establecer relaciones del tipo predicado-argumento. Se describen los predicados y los criterios semánticos mediante los cuales estos restringen sus argumentos; estos criterios semánticos se presentan como índices de esas combinaciones y se completan con otras combinaciones no obtenidas automáticamente. En *Práctico*

¹⁵⁴ El *DiCE* es otro diccionario combinatorio en proceso de creación que proporciona información semántica y combinatoria de las unidades léxicas; pero, debido al criterio de direccionalidad seguido en esta obra que convierte las bases en lemas –frente al criterio adoptado en *Redes*, que convierte en lemas los colocativos–, resulta menos adecuado para nuestro análisis, aunque en el fondo el *DiCE* y *Redes* sean dos formas de expresar el mismo hecho. Aquí nos hemos centrado mayoritariamente en el análisis de la relación entre el predicado *dar* y sus argumentos, por lo que un diccionario que seleccione los predicados verbales en las entradas y permita la obtención de sus argumentos se erige en una fuente de datos muy rica.

el camino es otro: los grupos de combinaciones –esto es, las clases léxicas seleccionadas por el predicado– se ilustran con un número representativo de combinaciones, pero no se define el grupo o la clase a la que pertenecen; lo que se pretende es que sea un diccionario de fácil manejo, por lo que se presta menos atención a la delimitación y caracterización de los grupos y ello permite incluir muchas más entradas. Pese a estas diferencias, ambos diccionarios comparten la idea de que en la mayor parte de los casos las combinaciones no se eligen individualmente, que se pueden ampliar fácilmente y que la elección, por tanto, obedece a criterios semánticos restrictivos (Bosque, 2006: XIV).

Aparte de estos diccionarios, no ha habido prácticamente intentos de describir la combinatoria léxica del español en profundidad. Pese a su aportación –que a diferencia de los diccionarios combinatorios del inglés, por ejemplo, permiten obtener paradigmas léxicos–, en español, al igual que sucede en otras lenguas, como anota Nesselhauf (2005: 281-282), los diccionarios (normales y combinatorios) no pretenden ser exhaustivos, sino que incluyen solo una parte de la combinatoria léxica producida por los HN, por lo que no pueden ser la única fuente para determinar la aceptabilidad (o el grado de aceptabilidad) de las colocaciones. De ahí la necesidad de completar los datos obtenidos de los diccionarios combinatorios con otras fuentes de datos.

6.2.3.4. Obras de consulta generales de carácter panhispánico

En definitiva, con el objetivo de realizar un análisis cuantitativo y cualitativo del verbo lo más minucioso posible, se han combinado las fuentes de filtración de los datos obtenidos del corpus CEDEL2. Se ha recurrido, así, al subcorpus de control del CEDEL2, a corpus textuales extensos de referencia, como son el CREA y el Corpus del español, a los diccionarios combinatorios del español *Redes* y *Práctico* y, dado que la lingüística con corpus ha de ser una lingüística en la que el corpus esté al servicio del investigador (Bosque, 2005: CLVIII [2004]), los datos han sido filtrados también por el investigador, así como por otros HN.

Con todo, hemos encontrado usos erróneos o conflictivos que, lógicamente, no encontramos en el corpus de control, en los extensos corpus de referencia ni en los diccionarios combinatorios, y sobre los cuales la introspección nativa no era capaz de establecer un resultado claro acerca de cuál es la percepción mayoritaria. Ante esta situación, a estas fuentes de datos se han unido las obras de consulta generales de carácter panhispánico como el *DRAE*,

el *DPD* o la *NGLE*, mencionadas en el capítulo 5, donde destacamos –en la tarea de establecer el carácter estandarizado o marcado de los usos– la utilidad de recurrir a las autoridades en el uso correcto de la lengua, como la RAE, ya que esta recoge la conciencia lingüística amparada en cierta mayoría (culta) de hablantes; y, dada su vocación panhispánica, estas obras suelen indicar las variaciones a este estándar panhispánico.

Con todas estas fuentes se incrementan las garantías de objetividad de los resultados.

6.2.4. Criterios para determinar el grado de aceptabilidad de los datos

Un análisis de corpus de aprendientes requiere una tipología del error predefinida (Granger, 2007). Con el objetivo de establecerla, en este apartado se describen los pasos seguidos para clasificar los usos verbales de acuerdo con su grado de aceptabilidad; y en el siguiente apartado se describe el procedimiento de anotación de los usos inaceptables.

En la propuesta tipológica del error colocacional presentada en Alonso Ramos *et al.* (2010b: 3214), se reconoce la necesidad de determinar una escala de los diferentes grados de aceptabilidad del error para realizar un análisis en profundidad: “it is to be investigated whether, for instance, a three grade scale ‘bad’ –‘improvable’– ‘good’ is enough or whether a more fine-grained scale is required”. En el trabajo sobre colocaciones en aprendientes de español de Prieto *et al.* (2009), se establece una clasificación, considerada preliminar, entre colocaciones correctas, incorrectas y mejorables, estableciendo como únicos criterios para la categorización la intuición de los investigadores y los datos obtenidos del corpus de control del CEDEL2. Más precisa resulta la tipología de Nesselhauf (2003, 2005), en la que se fijan 5 categorías en una escala de aceptabilidad de las colocaciones –claramente aceptable, esencialmente aceptable, confusa, esencialmente inaceptable y claramente inaceptable– basada en su aparición en diccionarios combinatorios del inglés y en el Corpus Británico Nacional; en aquellos casos en los que el investigador no es capaz de formular un juicio claro de acuerdo con los criterios fijados, Nesselhauf recurre al juicio de dos hablantes nativos que clasifican la colocación en una de las cinco categorías. En el análisis de corpus de Guo, en cambio, no se predefine una tipología del error pues, en palabras del autor (2006: 223),

Learner's errors in my research are not a central interest in themselves because it is my belief that error are unavoidable in the process of language acquisition (...). What is more important to me is to identify which part of the English language, in terms of verbs, is produced more like the NSs, which part is less like the NSs and which part is grossly deviant from the Ns use.

A nuestro juicio –independientemente de que no se busque analizar el grado de adecuación o inadecuación de cada uso y a pesar de que en ese trabajo (Guo, 2006) no se haya fijado un concepto de norma algo más prescriptiva que la que constituye el estado de lengua del corpus de control y de los grandes corpus nativos de referencia–, la lengua usada del mismo modo en el que lo hacen los HN equivale a usos naturales, la usada de manera no tan exacta a como lo hacen los HN es lengua restringida y la que supone una desviación completa del uso nativo es inaceptable.

Consideramos que, en efecto, si se quiere realizar un estudio riguroso y profundo de la interlengua, se ha de diseñar una escala que abarque los diferentes grados de aceptación en cuyos extremos encuentran los usos etiquetados como inaceptables y los usos naturales. Ya se ha mencionado (apartados 6.2.3.1 en adelante) las ventajas y limitaciones de cada una de las fuentes de datos empleadas, y se ha señalado la necesidad de que los datos obtenidos de cada una de ellas se interpreten como complementarios. Ahora bien, hay que fijar los criterios que van a ser aplicados en el procedimiento de anotación de los datos de forma clara para que a cada etiqueta le corresponda un dominio bien acotado desde una perspectiva lingüística.

De los usos del verbo *dar* aceptables –esto es, aquellos que no difieren del estándar en cuanto a la ortografía, semántica, gramática (morfología y sintaxis), combinatoria y marcas diatópicas, diastráticas y diafásicas (como se ha planteado en el capítulo 5)–, se han considerado aceptable-naturales los que satisfacen al menos uno de los criterios que se explicitan seguidamente.

1) El primer criterio al que se recurre en este estudio es el de la introspección propia; no obstante, en aquellos casos que no presentan ninguna desviación del estándar pero en los que vacilamos acerca de si constituyen una expresión natural o preferente –la más frecuente sin ser exclusiva– o un uso restringido –correcto desde una perspectiva formal pero considerado no preferente por hablantes nativos–, el proceso de categorización de datos continúa y se pasa al siguiente criterio (recuérdese los casos de *dar de mamar* y *dar a mamar* tratados en el capítulo anterior).

2) Son consideradas secuencias aceptable-naturales aquellas que ocurren en la misma forma¹⁵⁵ y con el mismo significado intencional en uno de los dos diccionarios combinatorios de los que disponemos. Si aparece en *Práctico*, automáticamente será clasificado el uso como natural, pues cada uso combinatorio del verbo que incluye se ejemplifica únicamente con las combinaciones más representativas (Bosque, 2006: XVII). En *Redes* la secuencia ha de aparecer con la marca de frecuencia (++) , que indica combinación sumamente frecuente, o con la marca (+), que señala que la combinación es bastante frecuente. Se recurre al siguiente criterio de obtención de datos si el uso del verbo que está siendo analizado no aparece (recuérdese que el hecho de que no aparezca en los diccionarios combinatorios, como ya se ha señalado, no implica que la combinación no exista) o aparece en este diccionario sin marca alguna –lo cual indica que se trata de una combinación atestiguada que suele resultar aceptable a los oídos de un hablante nativo– o con la marca (INDOC), que se refiere a aquellos usos indocumentados añadidos por el lexicógrafo que el corpus utilizado para la elaboración del diccionario no proporciona, pero que resultan naturales a oídos de los hablantes encuestados durante el proceso de elaboración de dicho diccionario. También se recurre al siguiente criterio si un uso no aparece en *Redes* pero puede añadirse sin dificultad a una de las clases léxicas que lo constituyen.

3) De igual manera, se clasifican como combinaciones de tipo aceptable-natural aquellas que ocurren en los corpus de referencia en la misma forma y con el mismo significado intencional al menos en 5 textos diferentes. Lógicamente, es necesario filtrar esos datos, asegurarse de que no proceden de un mismo autor o de una misma área geográfica y de que no contienen ninguna otra desviación del estándar o error (recuérdese lo visto en el capítulo 5 a propósito del queísmo en el corpus de hablantes nativos). Para esto último, contamos con obras de

¹⁵⁵ Por *misma forma* se entiende que la combinación en la redacción del aprendiente ha de ser idéntica a la encontrada en el diccionario o en el corpus de referencia, tanto en los lexemas que incluye como en las categorías y marcas funcionales seleccionadas (preposición del complemento regido, número del sustantivo, presencia o ausencia de determinativo). No serán tenidas en cuenta las cuestiones relativas al tiempo y al aspecto (recuérdense las consideraciones a este respecto expuestas en el capítulo 3. También se alejan de nuestro objeto de análisis las cuestiones relacionadas con la estructura informativa, como el hecho de que el español sea una lengua ProDrop y el inglés no, lo que conduce a apariciones explícitas o ausencias injustificadas desde la perspectiva informativa que nosotros no consideramos en este trabajo. Por otro lado, dado que resulta complicado establecer criterios para distinguir eficazmente las erratas de los errores ortográficos, se ha decidido no atender a este aspecto en el análisis, ya que, por ejemplo, hay participantes que manifiestan no poder añadir tildes a sus textos por cuestiones relacionadas con el teclado utilizado, y otros que informan de que no van a poner ninguna porque les resulta incómodo.

referencia desde el punto de vista normativo como las de la RAE. Somos conscientes de que la barrera numérica (la de los 5 textos) que delimita las combinaciones naturales de las restringidas no deja de ser arbitraria. No obstante, seguimos a Nesselhauf (2005: 50) cuando se refiere a este número de la siguiente manera: “It is rather high, as a pilot study revealed that lower occurrences were often not judged acceptable by native speakers”. Es de esperar que al tratarse de un verbo tan frecuente, las combinaciones naturales superarán sin dificultad las cinco apariciones en el corpus de control. En cambio, en otro tipo de combinaciones con predicados no tan comunes, o en aquellos casos en los que el elemento infrecuente sea el argumento, es posible que encontrar cinco ejemplos sea complicado. Recuérdense los ejemplos aportados en el capítulo 5, en el que se mostraba cómo de la búsqueda de *dar de mamar* se obtenían en el CREA veintiocho apariciones, y de *dar a mamar*, solo una. Estos datos se corresponden con nuestras primeras intuiciones, pues la primera combinación nos parecía desde un primer momento natural, y en la segunda vacilábamos¹⁵⁶.

Como se ha mencionado, en un estudio de estas características se ha de tener presente las limitaciones que las fuentes presentan. Por un lado, una combinación de palabras documentada una sola vez puede resultar absolutamente natural para el hablante nativo –recuérdese que *frecuencia* y *naturalidad* son dos conceptos diferentes (Bosque, 2005: CLIX [2004])–, por lo que el hecho de que un uso no aparezca en el CREA o en el Corpus del español 5 veces o más no implica que sea inaceptable, que no exista o que esté restringido. Por otro lado, los diccionarios combinatorios no recogen todas las combinaciones posibles en una lengua, por lo que si un uso no aparece en el diccionario o incluso si el sustantivo argumento del verbo no se puede añadir con facilidad a ninguna de las clases léxicas ya establecidas, puede todavía tratarse de una combinación común.

¹⁵⁶ Se descarta que esta única aparición en el CREA pueda corresponderse con un uso inaceptable del verbo, dado que la pieza léxica seleccionada puede añadirse a una de las clases léxicas ya establecidas en *Redes*: *dar a* se combina con “algunos verbos que denotan ingestión”, por lo que *mamar* se sumaría a *comer*, *beber*, *tomar* o *probar*. Por otro lado, la necesidad de establecer varios criterios para filtrar los datos del corpus se evidencia en el siguiente ejemplo: de la búsqueda en el corpus CREA de *dar a beber*, que la intuición, en un primer momento, nos llevaba a considerar como poco natural, se obtuvieron siete resultados, por lo que automáticamente este uso se habría de clasificar como aceptable-natural. Al examinar los resultados más detalladamente, se comprobó que, de esos siete, cuatro provenían del mismo autor, por lo que de acuerdo con el criterio de las 5 apariciones en el corpus de referencia no se podría considerar este uso como natural. No obstante, en *Redes* sí que aparece marcado como uso bastante frecuente (+), por lo que en este estudio la coaparición *dar a beber* sería clasificada finalmente como natural. Esto demuestra la necesidad de aplicar diversas fuentes de datos en un estudio sobre el grado de aceptabilidad de los usos.

Así las cosas, todas las combinaciones que no pueden ser juzgadas como aceptable-naturales de acuerdo con los criterios señalados (la introspección, los diccionarios combinatorios y los corpus de referencia), incluyendo aquellas en las que no se puede determinar con seguridad si el sentido intencional de la combinación del aprendiente coincide con el sentido en la(s) fuente(s) de datos, se presentan a HN para que realicen un juicio de aceptabilidad (introspección de los informantes): “la introspección propia ha de estar contrastada con la ajena, y cuanto más riguroso sea este contraste, mayores serán las garantías de objetividad” (Bosque, 2011: 47).

Los informantes son hablantes nativos del español peninsular y americano y han pasado toda su vida en un país de habla hispana. El perfil de estos informantes coincide con el de los autores del corpus: son universitarios o graduados. En este punto se sigue el procedimiento de Nesselhauf (2005: 49-50): las combinaciones se presentan a los nativos en contexto, para que el sentido buscado por el aprendiente se refleje lo más claramente posible. El contexto consiste en la oración en la que la combinación ocurre; a menudo se entrega la oración anterior y posterior, de ahí que en las instrucciones para los informantes nos refiramos al texto. Las oraciones que forman el contexto son en ocasiones modificadas y corregidas en lo que se refiere a errores ortográficos, errores semánticos y gramaticales fuera de la combinación, para evitar que los errores de los informantes puedan trasladarse de la posible desviación léxica a una puramente gramatical u ortográfica. A veces incluso la oración en la que se encuentra la combinación ha sido reformulada en parte para hacer el sentido buscado más claro, sin modificar en modo alguno la propia combinación. Las combinaciones fueron subrayadas en la oración, para tener la certeza de que son el objeto juzgado, y no lo son otras partes de la oración o la totalidad del texto.

A los informantes se les pidió que juzgaran la combinación en una escala de 3 puntos: aceptable natural, aceptable poco natural, inaceptable. Si la combinación es etiquetada con algún valor diferente al de aceptable natural, los informantes han de aportar ellos mismos la opción que consideran más natural para expresar el sentido buscado. Cuatro nativos, 2 procedentes de diferentes puntos peninsulares (Cádiz y Salamanca) y 2 de países americanos distantes geográficamente (Argentina y Méjico) llevaron a cabo los juicios de las combinaciones. Si se da coincidencia en los juicios, el resultado constituye el etiquetado final. Si dos hablantes del mismo estándar¹⁵⁷ (americano o peninsular)

¹⁵⁷ Ya se ha referido en el capítulo 5 que no es el objetivo de este trabajo intentar aclarar si se deben aceptar uno o dos estándares en el mundo hispanohablante. Debido a la enorme diversidad en el origen de los HNN y HN del corpus, la elección de estándar considerada como

consideran una combinación como aceptable-natural, tenemos, de acuerdo con la consideración ya expuesta de un estándar panhispánico, motivos suficientes para etiquetarla como muestra lingüística natural en una zona suficientemente representativa del español. Si solo uno de los hablantes en cada una de las zonas separadas por el Atlántico califica una combinación como natural mientras que los demás la consideran inaceptable o poco natural, estamos ante un caso de combinación restringida a una zona geográfica, por lo que, al desviarse de nuestro estándar panhispánico, es considerada inaceptable. Si no hay diferencia clara entre los dos estándares, y no existe acuerdo entre los nativos en el grado de aceptabilidad, se hace una media. Para ver las combinaciones de los juicios que resultaron en los juicios finales, véase la siguiente tabla¹⁵⁸. Las anotaciones para el grado de aceptabilidad se realizan de esta manera:

- Aceptable natural: +
- Aceptable no natural: (+)
- Inaceptable: *

Juicio 1 (Méj.)	Juicio 2 (Arg.)	Juicio 3 (Sal.)	Juicio 4 (Cád.)	Media
+	+	+	+	+
+	+	+	(+)	+
+	+	+	*	+
+	+	(+)	(+)	+
+	+	*	*	+
+	+	(+)	*	+
*	*	*	*	*
+	*	*	*	*
+	(+)	(+)	(+)	(+)
(+)	(+)	(+)	(+)	(+)
(+)	(+)	(+)	*	(+)
(+)	(+)	*	*	(+)
(+)	*	*	*	*
+	(+)	+	(+)	(+)
(+)	*	(+)	*	*
+	*	+	(+)	(+)

Tabla 5. Juicios de aceptabilidad¹⁵⁹

más apropiada es la panhispánica, pese a las dificultades a la hora de operar con este concepto. Se recuperan de nuevo los conceptos de *reconocimiento* y *entendimiento* ya tratados en el capítulo anterior. Para determinar si un uso es natural, poco natural, o inaceptable, los informantes han de plantearse si identifican dicho uso como perteneciente a una zona suficientemente representativa del español; se está hablando, pues, de reconocimiento y no de entendimiento ni de realizar una categorización en función de que también usen dicha combinación en su variedad lingüística.

¹⁵⁸ Fueron once los usos que se sometieron a su consideración; en el capítulo 7 (apartado 7.2.1) nos referimos al grado de acuerdo entre los cuatro informantes sobre esos usos.

¹⁵⁹ Los juicios 1 y 2 se corresponderían con un mismo estándar, hispanoamericano o peninsular, y el 3 y el 4, con el otro.

Para establecer las medias, se han concedido valores numéricos a los diferentes grados de la escala: aceptable natural se corresponde en una escala del 1 al 3 con el valor máximo, el 3; aceptable poco natural, con el 2; inaceptable, con el 1. En el caso de que la diferencia no esté basada en el estándar de una zona u otra, y la media de los cuatro informantes esté en uno de los límites entre un grado y otro, como sucede por ejemplo en $+(+)(+)$ o $(+)**(+)$, la media se establecerá en el grado más bajo.

Como existe la posibilidad de que los estándares de aceptabilidad a la hora de juzgar las producciones no nativas por parte de los informantes varíen (Mel'čuk, 1995; Nesselhauf, 2005; Guo, 2006; Stengers, 2009, Alonso Ramos *et al.*, 2010b; entre otros), esta metodología excluye al menos la posibilidad de que una combinación usada con frecuencia entre HN sea juzgada como inaceptable o restringida, ya que, si es frecuente, probablemente aparezca en *Redes* o *Práctico*, u ocurra al menos cinco veces en el corpus de control, por lo que, antes de llegar a pasar a esta última fuente de datos, será etiquetada como aceptable-natural¹⁶⁰. Aun así, debido a que el número de informantes es bajo, los juicios de aceptabilidad han de considerarse como una aproximación, juzgada como suficiente para los objetivos marcados en este estudio. No obstante, hay trabajos que muestran que la correlación entre los datos del corpus y los juicios de los nativos en coapariciones léxicas es buena (Lapata *et al.*, 1999; Shei, 1999; Hoffmann y Lehmann, 2000). De hecho, antes de enviar los datos a los cuatro informantes nativos, se ha realizado una prueba piloto con varias combinaciones que demuestra acuerdo entre los nativos en aquellas combinaciones inexistentes, y por tanto, inaceptables, y en aquellas frecuentes en el mundo panhispánico, y por tanto, aceptables-naturales. Por último, las obras de referencia básicas de nuestro estándar normativo han sido usadas de manera transversal, más que como un paso establecido previo o posterior a otro. Esto nos ha ayudado a determinar si usos frecuentes en los corpus (cinco resultados o más) pudieran ser no normativos (recuérdese el ejemplo de *darse cuenta que*); esto es, estas obras de consulta nos han informado, en definitiva, de variaciones o desviaciones de la norma, ya sean diatópicas, diafásicas o diastráticas.

Por otro lado, al comienzo de este apartado se ha presentado la primera distinción a la hora de categorizar los datos normales (aceptables) desde un

¹⁶⁰ Ya se ha visto como en *dar a beber* los corpus indicaban uso restringido (tres coapariciones en el CREA), pero *Redes* señalaba que era una combinación frecuente y natural. De esta manera, es improbable –aunque no imposible– que fallen las dos fuentes, esto es, que una combinación natural no aparezca en los corpus de referencia ni en los diccionarios combinatorios.

punto de vista gramatical, semántico, combinatorio y sociolingüístico: las secuencias aceptables se subdividen en estructuras aceptable-naturales y aceptable-restringidas. Al describir los criterios para clasificar una combinación en el grupo aceptable-natural, se han puesto de manifiesto los empleados para etiquetar secuencias como aceptable-restringidas. De esta manera, las estructuras aceptable-restringidas se definen por lo que las demás –los usos naturales y los inaceptables– no son¹⁶¹. De esta manera, las estructuras aceptable-restringidas no pertenecen a la categoría de las inaceptables; se sigue tratando, pues, de estructuras normales –respetan la norma estándar del español panhispánico– desde el punto de vista diatópico, diastrático, diafásico, morfosintáctico, semántico, combinatorio.

Se ha señalado la necesidad de establecer diversas fuentes de datos y criterios de aplicación para que sean empleados en aquellas ocasiones en las que a través de la introspección no es posible obtener un juicio firme y claro. Por consiguiente, si una construcción es aceptable pero no cumple ninguno de los criterios anteriores, constituye un uso aceptable-restringido. A continuación se recogen de manera resumida los pasos dados en el proceso de categorización de una muestra lingüística como restringida y como natural.

Un uso es considerado aceptable-restringido si no ocurre en la misma forma y con el mismo significado intencional en uno de los dos diccionarios combinatorios mencionados. En *Redes* no ha de aparecer con la marca de frecuencia ++ (combinación sumamente frecuente) o + (combinación bastante frecuente). Si en *Redes* aparece con marca (-), sin marca o con la marca de indocumentado (INDOC), puede ser tanto natural como restringido, dependiendo de los resultados de la búsqueda en un corpus de referencia y de los resultados de los informantes. Y lo mismo se aplica a aquellos casos en que la combinación no aparece en los diccionarios pero el sustantivo puede añadirse sin dificultad en alguna de las clases léxicas que configuran los paradigmas extensionales del verbo. Si en el corpus de referencia aparece más de cinco veces, automáticamente la combinación es etiquetada como natural; pero, si aparece menos de cinco veces o no aparece, eso no indica necesariamente que estemos ante una combinación restringida por ser poco común; hay que dar un paso más: los informantes la han de juzgar como aceptable-restringida para etiquetarla de este modo.

¹⁶¹ Al realizar un análisis de los datos sistemático, cada etiqueta debe corresponderse con un concepto que tiene un valor determinado, discreto y diferencial en el sistema y que se caracteriza, por lo tanto, por lo que los demás conceptos no son.

En caso de vacilación, tanto las obras de la RAE (*NGLE*, *DPD*, *DRAE*) como la *Gramática descriptiva de la lengua española* o *El diccionario de coloquialismos y términos dialectales del español* nos ayudan a decantarnos por el uso restringido, natural o inaceptable, dado que en estas obras se suele señalar cuándo un uso es poco común o está restringido a ciertas zonas o situaciones. Recuérdese que estos materiales de consulta ayudan a establecer el estatus estandarizado y, por consiguiente, marcado de ciertos usos.

6.2.5. Criterios para determinar la inaceptabilidad de los datos

Si en la sección anterior, referida a la aceptabilidad de los datos, se destaca la posibilidad de que existan en los usos del corpus dos grados distintos de aceptabilidad (aceptabilidad natural y aceptabilidad restringida), en esta sección se discute brevemente sobre la necesidad de aplicar esta variabilidad a los datos inaceptables y, a continuación, se presenta una tipología del error.

En Alonso Ramos *et al.* (2010b: 3213) se plantea desde una perspectiva teórica –ya que los planteamientos no se aplican a ningún estudio empírico– la necesidad de marcar diferentes grados en la inaceptabilidad de un error, pues “not all errors are equally unacceptable. While some of them are even hard to understand for a native speaker (...) others can be ‘tolerated’”. Es cierto que desde una perspectiva comunicativa unos errores intervienen más que otros en el fracaso de la interacción, lo que resulta relevante en el aula de lenguas extranjeras. Sin embargo, en un estudio de interlengua, esta distinción en el grado de inaceptabilidad no es relevante; el error, como se ha visto en el capítulo 5, es concebido como una desviación del estándar –ya sea esta de tipo semántico, combinatorio o diastrático, por nombrar algunos–, y como tal es considerado; por consiguiente, un error en el uso del verbo en cualquiera de los niveles lingüísticos coloca la construcción automáticamente en el grupo de los usos inaceptables.

Con el objetivo de ofrecer un análisis del error lo más detallado posible, que contribuya a ofrecer una descripción precisa de la lengua del aprendiente y que, al mismo tiempo, pueda ser utilizado en la enseñanza de ELE, se presenta a continuación una propuesta tipológica del error –diseñada expresamente para esta investigación– originada a partir del marco del AE propuesto por Corder (1971), que, a su vez, se complementa con algunos de los planteamientos de la hipótesis de funcionamiento del lexicón mental no nativo, de Baralo (2001), que pretende dar cuenta de los diferentes elementos que se activan en el

procesamiento productivo de la L2. Esta propuesta taxonómica del error se configura una vez que los errores son recopilados, y no antes, para no influir en el análisis con los tipos de errores que previsiblemente nos podríamos encontrar.

Así, en la tipología se distinguen tres dimensiones fundamentales (basadas en Corder, 1971) que se corresponden con las tres fases del análisis llevado a cabo en el capítulo 7: localización del error (*identificación* en términos de Corder), descripción (que, además de la *descripción*, incluye la *clasificación* de Corder) y explicación desde una perspectiva psicolingüística (coincide con la *explicación* de Corder). En las líneas que siguen, nos detendremos en cada una de ellas.

6.2.5.1. Localización del error

El primer paso en el análisis de los datos consiste, como es lógico, en detectar el error en la muestra. Una vez que el error es descubierto, es necesario indicar su localización exacta dentro de la combinación. Este aspecto es importante, en primer lugar, para determinar en qué parte de las combinaciones sintagmáticas proyectadas por estos verbos los aprendientes tienen más problemas; y, en segundo lugar, facilita la descripción del error.

Para establecer la clasificación que a continuación se presenta, se toma como punto de partida la organización interna del sintagma verbal y la relación entre el predicado y sus argumentos, en la que también se va a fundamentar la descripción del error, como se detalla en el apartado 6.2.5.2. Una manera de formalizar la organización interna del sintagma verbal y dar cuenta de su formación consiste precisamente en atender a la jerarquía estructural del sintagma verbal: el núcleo del sintagma se une, en primer lugar, con su complemento directo, y, en una segunda fase de la formación del sintagma verbal, a esos elementos (núcleo + CD) se les aplica el sujeto para construir el sintagma (Haegeman, 1992: 78-88; Rodríguez Ramalle, 2005: 43-52)¹⁶². Sujeto y CD parecen relacionarse de manera distinta con el núcleo verbal: los verbos establecen relaciones más estrechas con sus complementos que con su sujeto,

¹⁶² Esta manera de formalizar la organización interna del sintagma verbal y dar cuenta de su formación nos interesa por su *capacidad descriptiva* para ilustrar las diferentes relaciones que se dan dentro del sintagma, lo cual nos permite segmentar de manera nítida, desde una perspectiva lingüística, los diferentes puntos del sintagma en los que se puede localizar el error. En cualquier caso, la adopción de otra perspectiva sobre la organización del sintagma no habría afectado a los resultados de nuestro análisis (véase nota 25). Así pues, no pretendemos con ello sumarnos a los argumentos explicativos de ninguna teoría acerca de la manera en que se produce la formación del sintagma, lo cual excede a los objetivos de este trabajo.

por lo que el primero está sujeto a más restricciones semánticas y sintácticas¹⁶³, lo que posiblemente convierte al CD (normalmente el tema) en un elemento de más difícil manejo para un aprendiente de L2 (esta hipótesis aquí formulada – presentada en diferentes términos en el capítulo 1 a propósito de la enorme capacidad extensional de este verbo– es contrastada en el capítulo 7). Esto explica que el sujeto sea considerado el argumento externo y el CD el argumento interno¹⁶⁴.

Como se ha visto en el capítulo 3 a propósito de la hipótesis del funcionamiento del lexicón mental no nativo de Baralo (2001, nos centramos de nuevo en ella en el apartado 6.2.5.2.1), se considera que –independientemente del diferente grado de relación entre el verbo y el complemento, y el verbo y el sujeto–, ambas posiciones (la de sujeto y CD) están pautadas por los requisitos léxico-sintácticos que debe satisfacer el núcleo verbal, por lo que son obligatorias. El complemento indirecto también es exigido por el verbo, ya sea de forma explícita o implícita, cuando este actualiza el sentido de “proporcionar o transferir” (*dar un libro / alojamiento / cariño a alguien*) y, en algunas ocasiones, cuando actualiza el sentido de “ejecutar”, esto es, cuando funciona como verbo de apoyo (*dar un beso / un puñetazo a alguien*). En cambio, los adjuntos no son

¹⁶³ Escandell (2004) aporta diversas pruebas que muestran esta diferente relación; así, la proforma *lo hizo* afecta únicamente a la secuencia léxica formada por el verbo y su complemento, pero no sustituye nunca al conjunto verbo y sujeto, dejando fuera al complemento. En la misma línea, ciertos procesos de anteposición de constituyentes que pretenden poner de relieve un constituyente oracional, afectan de manera distinta a la unión formada por verbo y complemento y por verbo y sujeto. Esta última unión no se puede focalizar excluyendo al CD. Por último, la construcción de ciertos compuestos evidencia una relación entre el verbo y el CD, pero nunca con el sujeto.

¹⁶⁴ Esta perspectiva no resulta, a nuestro juicio, incompatible con la aproximación cognitivista al análisis del verbo reflejada en otros capítulos; son diferentes aportaciones sobre un mismo hecho lingüístico que responden a los distintos enfoques de los lingüistas (al fin y al cabo, el modelo cognitivista nace del generativista). En el capítulo 4 veíamos cómo la definición sistémica de *dar* reflejaba una relación estable y necesaria entre V y N (argumento interno y tema), y, en cambio, las funciones semánticas *agente* (desempeñada normalmente por el sujeto) y *experimentador* o *destinatario* (desempeñadas por el CI) resultaban opcionales en la definición (por ello figuran entre paréntesis), pues pueden estar implícitas (o no estar en absoluto en el caso del CI); estas diferentes relaciones se reflejaban también en el Esquema de Imagen del CAMINO, pues el nombre, que se refiere a la entidad que se desplaza a lo largo de un trayecto hasta alcanzar su término o meta (estado resultativo), es el único elemento contenido en la estructura interna de *dar* (de acuerdo con este Esquema), y, sin embargo, ni el agente ni el destinatario se contemplan en dicha estructura (véase el apartado 4.1.1.1). Esto refleja, por lo tanto, otras maneras (diferentes a la presentada por Haegeman [1992] al comienzo de este apartado 6.2.5.1) de describir la diferente relación que existe entre el verbo y los otros tres elementos potencialmente contenidos en el sintagma verbal. Tomar solo una postura de aproximación al verbo sería eliminar otras posturas que, a nuestro juicio, son también muy enriquecedoras, como señalábamos en el capítulo 1; o que pueden resultar útiles desde una perspectiva descriptiva de acuerdo con los objetivos planteados en cada fase de la investigación, como veremos a propósito de la hipótesis de funcionamiento del lexicón no nativo, de Baralo (2001) (véase también n. 23).

exigidos por el núcleo del sintagma, y dado que precisamente en este estudio se pretende determinar el conocimiento del núcleo *dar* por parte de los aprendientes, estos complementos se dejan fuera del análisis.

Por lo tanto, de acuerdo con una organización jerárquica –producto de la aplicación de los principios generales que, según esta perspectiva de análisis lingüístico, regulan la unión de dos o más piezas léxicas en torno a un elemento nuclear, verbal en este caso–, el verbo solo rige al sintagma nominal en la posición de complemento y al sujeto, pero nunca a un adjunto ni, por poner un ejemplo, a un sintagma preposicional modificador del núcleo del complemento directo. Así, de acuerdo con lo argumentado, en una oración del tipo *Mañana por la mañana Luis dará el golpe del siglo*, solo se examina en este estudio el verbo *dar* como núcleo del sintagma verbal, *Luis* como sujeto y *el golpe* como núcleo del sintagma nominal con función de CD. Ni *mañana por la mañana*, que ocupa la posición de adjunto, ni *del siglo*, que ocupa la posición de sintagma preposicional que modifica al núcleo del complemento *golpe*, forman parte de las muestras de lengua analizadas. A este último tipo de complemento se hace referencia con mayor profundidad en la sección “6.2.5.1.1”.

Por consiguiente, en esta primera dimensión de la tipología –la localización– se marca si el error concierne a todo el sintagma verbal o, por el contrario, solo afecta al núcleo verbal o a alguno de los elementos que este rige, ya sea complemento (directo o indirecto) o sujeto. A continuación se presenta un ejemplo de error de cada una de estas localizaciones, y una tabla con la dimensión de la localización en la tipología del error.

[4] Núcleo: los moros no nos daban el respeto que nos debieron (15 HNN).

[5] Complementos: Si los gays pudieran adoptar, podrían dar buenas hogares a estos huérfanos (10 HNN).

[6] Sujeto: Por lo menos, como profesora me ha dado la oportunidad de pasar los veranos en España (76 HNN).

[7] Sintagma: Manchester daba la luz a muchos grupos musicales (153 HNN).

Localización	Núcleo del sintagma verbal
	Complementos
	Sujeto
	Totalidad del sintagma verbal

Tabla 6. Dimensión de la localización en la tipología del error

6.2.5.1.1. Error relevante frente a error no relevante

En relación con el vínculo que establece el núcleo con el resto de los elementos proyectados por él en el sintagma, distinguimos entre errores relevantes y no relevantes. Los relevantes se localizan en el núcleo del sintagma verbal o en alguno de los argumentos exigidos por el núcleo. Si el problema no radica directamente en el núcleo verbal, sino en los argumentos, solo consideramos errores relevantes aquellos en los que el problema está directamente relacionado con el núcleo del argumento (como es, en el caso de las construcciones V + N, el núcleo del sujeto, el núcleo del CD o el del CI si lo hubiera), esto es, cuando el núcleo nominal presenta algún problema de forma o significado, o selecciona un determinante actualizador o prescinda de él erróneamente, o cuando un complemento exigido por el verbo presenta una preposición o prescinda de ella erróneamente.

En el caso de los argumentos que adoptan la forma de un sintagma verbal (en lugar de un sintagma nominal) –con lo que las relaciones semánticas con el verbo continúan siendo iguales y tan solo cambian las propiedades formales de la oración–, no se tienen en cuenta los errores que se encuentran en el interior de este sintagma verbal (no relevantes), dado que los elementos que lo integran ya no son exigidos por el verbo *dar*, sino por otro núcleo verbal subordinado. Así, en el ejemplo

[8] me dio la oportunidad de realizar que la vida es más sencilla (201 HNN)

hay un error en la selección del verbo *realizar*, en lugar del correcto *darse cuenta de*. Este error se encuentra dentro del sintagma de los argumentos seleccionados por el verbo *dar*, pues se localiza en el CD, pero no se considera como error relevante al no tratarse de un fallo por desconocimiento del núcleo del complemento directo, puesto que el aprendiente conoce que el nombre *oportunidad* en este contexto sintáctico no puede prescindir de un determinante y de un complemento modificador introducido por la preposición determinada *de* seguida de un infinitivo. Este infinitivo es incorrecto en sí mismo –es decir, por lo que el aprendiente cree que significa– y no por desconocimiento de los requisitos semánticos del núcleo del complemento *oportunidad*; en resumidas cuentas, el alumno tiene un problema con el verbo *realizar* y no con la proyección sintagmática del núcleo nominal *oportunidad*. Por este motivo, no es tratado como error relevante. Lo mismo sucede con el error en la siguiente secuencia:

[9] Comimos un plato muy español: paella. A mi familia le gustaba mucho. Mis padres se dan cuenta de que es muy importante para aprender otro lenguaje. Comprender que mis estudios en Madrid son los más importantes de mi vida en la universidad (81 HNN).

El error se localiza dentro del suplemento¹⁶⁵ del núcleo verbal locutivo *darse cuenta de*, pero dado que el propio suplemento es un sintagma verbal, no es clasificado como error relevante. Nótese que el error está relacionado con uno de los argumentos del verbo subordinado, el sujeto del verbo *ser*, que no está expresado en la oración y no se infiere del contexto; la expresión *una estancia en el extranjero* podría encajar en esta posición.

Diferente es el caso del siguiente ejemplo:

[10] los homosexuales pueden dar el amor tan bien como los demás (39 HNN).

El núcleo del sintagma nominal que realiza la función de CD, *amor*, no requiere un determinativo actualizador en esta construcción; los nombres continuos pueden denotar la extensión de los predicados correspondientes; en otras palabras, pueden constituir por sí solos expresiones referenciales (Bosque y Demonte, 1999: 895-896). Por lo tanto, en este caso sí se considera que hay un desconocimiento del uso del nombre *amor* en este contexto.

En suma, los errores que se localizan en el interior de los argumentos del verbo —a excepción del núcleo— son etiquetados como errores no relevantes. Como se ha referido en la sección anterior, no se atiende a los errores en los adjuntos, pues no son elementos exigidos por el núcleo del sintagma. Este es el caso, por ejemplo, de *dar una vuelta por bicicleta*, donde el error se localiza en la preposición, núcleo del adjunto.

Hasta aquí, se ha hecho referencia a los errores (relevantes) que se dan en el sintagma proyectado por el verbo. Más tarde, atendiendo al objetivo específico de la investigación, el interés reside en poder distinguir entre los errores

¹⁶⁵ La preposición que introduce el suplemento, aunque tiene como propiedad léxica la obligatoriedad de ir acompañada de otro sintagma, no exige un valor semántico al complemento exigido, a diferencia de la preposición en *Viene desde Madrid* o *Viene con la llave*; en el primer ejemplo, al haber una preposición de dirección, ha de aparecer un complemento que denote lugar; en el segundo, la preposición, al ser instrumental, hace que el complemento se interprete como un instrumento. Así, la preposición del suplemento, a nuestro juicio, carece de significado, por lo que no determina el papel temático del sintagma nominal que aparece como su término y, por tanto, no desempeña la función de núcleo. La preposición es, en este contexto, una marca formal fija cuya selección depende del régimen del verbo y no de motivos semánticos (Rodríguez Ramalle, 2005: 43, 262).

relevantes que están generados por el desconocimiento del verbo *dar* y los que están generados por desconocimiento de los argumentos. Así, en

[11] aunque me gustó esa película, y doy una marca de 6.5 de 10 por la [...] (38 HNN)

estamos ante un error de desconocimiento del sustantivo –marca por nota o puntuación– que funciona como núcleo del argumento CD. En cambio, en la secuencia

[12] Ese recuerdo siempre lo tendré y siempre me dará una sonrisa (125 HNN)

hay un desconocimiento de la restricción léxico-semántica del verbo *dar*, pues la sonrisa no se da, sino que se arranca o se provoca.

6.2.5.2. Descripción del error

En apartados anteriores se ha expuesto la manera en que la competencia léxica escrita del verbo está conformada por la interacción de varios tipos de conocimiento que se corresponden, por un lado, con los asociados a diferentes niveles de análisis lingüístico (el semántico, el gramatical [morfología y sintaxis] y el que hemos denominado combinatorio por abarcar la interfaz léxico-sintaxis, junto a la fraseología); por otro lado, esta competencia está integrada por los conocimientos relacionados con el factor sociolingüístico –en el que se contempla la interrelación entre lo lingüístico y lo extralingüístico–; y, por último, al tratarse de la producción escrita, la competencia se vincula a un conocimiento ortográfico que permite al usuario reconocer la forma escrita de la palabra y usarla por escrito (véase la n. 14 para las razones por las que este tipo de conocimiento ha sido finalmente desestimado en el análisis).

Así las cosas, desde una perspectiva descriptiva, un uso del verbo *dar* puede ser erróneo si presenta una desviación de la norma en cualquiera de los aspectos señalados: semántico, gramatical, combinatorio o sociolingüístico. Esta clasificación del error se corresponde con el primer tipo de taxonomía descriptiva usada por Corder (1971): el lingüístico. Contar con una hipótesis explicativa sobre el modelo de organización y funcionamiento del lexicón mental no nativo nos permite establecer más subdivisiones en la taxonomía del error justificadas desde una perspectiva psicolingüística.

6.2.5.2.1. El modelo de funcionamiento del lexicón utilizado en la descripción del error: el lexicón y el sistema computacional

Una de las tipologías del error más reciente es la de Alonso Ramos *et al.* (2010b), que, aplicada a las colocaciones, se origina también a partir de los estadios en el tratamiento del error de Corder (1971) –la localización, descripción y explicación–. En el nivel descriptivo se distingue entre error léxico, gramatical y de registro (descripción del error según el nivel lingüístico en el que se produce); los dos primeros son a su vez desglosados de acuerdo con la anotación multinivel diseñada en el Corpus Falko, un corpus de aprendientes de alemán (Lüdeling *et al.*, 2005). Esta subdivisión está basada en los fenómenos que describen la desviación; por ejemplo, en los errores léxicos se señala si el problema se debe a un caso de sustitución, de creación, de síntesis, de análisis o de confusión en el sentido. Esta clasificación se corresponde con el segundo tipo de taxonomía descriptiva usada por Corder (1971): estrategia superficial.

Sin perder del todo de vista estas descripciones del error –lingüística y según la estrategia superficial–, para nuestra categorización del error en el nivel descriptivo, ya anunciamos en el capítulo 3 (y en apartados anteriores de este capítulo) que nos servimos de la propuesta del funcionamiento del lexicón realizada por Baralo (1997, 2001); como ya se ha señalado (véase n. 23), pretendemos aprovechar la gran capacidad descriptiva de esta teoría acerca del procesamiento del lenguaje en L2, gracias a la cual nos ha sido posible establecer la distinción entre los errores de tipo léxico y de tipo computacional (o sintáctico).

Como se expone en el capítulo 3, Baralo (2001) recurre al modelo generativo de estructura del componente léxico (Chomsky, 1995) –que supone la existencia de un módulo cognitivo que distingue entre el lexicón y el sistema computacional del lenguaje, o entre la información léxica y el componente sintáctico que procesa dicha información– para analizar las relaciones entre la información del lexicón mental y los principios formales del sistema lingüístico, y demuestra la existencia de un vínculo muy directo también en la interlengua entre saber vocabulario y saber construir oraciones en las que aparecen las palabras que lo integran. El lexicón es un conjunto de elementos léxicos organizados que constituyen las unidades atómicas de la sintaxis. Estos elementos contienen dos fases: una formal y una semántica, indispensables para el funcionamiento de las palabras dentro de la comunicación. La organización,

según la autora, permite proceder del significado a la forma en el proceso de codificación (Baralo, 2001: 7-8). Ahora bien, cada ítem almacenado en el lexicon contiene un sistema articulado de rasgos que especifica no solo las propiedades fonéticas y semánticas (forma y significado), sino también morfológicas y sintácticas que son idiosincrásicas para el ítem (uso) (Baralo, 1997: 60). Esto quiere decir que, si el hablante tiene disponible *dar* para la codificación de un mensaje, es porque parte del significado que le permite acceder a su forma, la cual está almacenada con la etiqueta correspondiente a su categoría gramatical, verbo, a la que podrá aplicar las reglas de flexión propias de los verbos. Al categorizar el verbo como una acción, se adquiere la estructura léxico conceptual correspondiente, que tiene toda la información necesaria para proyectar la estructura sintáctica de la oración: el hablante será capaz de reconocer el verbo como una acción transitiva que contiene tres argumentos en su estructura léxico conceptual: el agente, el tema y el destinatario. Saber usar este verbo implica saber proyectarlo en el eje sintagmático y construir oraciones en las que el agente se proyecte como el sujeto de la oración, el tema como el objeto directo y el destinatario como el objeto indirecto (Baralo, 2001: 11).

Además de las piezas que poseen significado léxico –por lo que pueden seleccionar a sus complementos–, el lexicon también está conformado por las categorías funcionales, que carecen de contenido léxico y cuyo contenido es más bien gramatical, relacional (Rodríguez Ramalle, 2005); en nuestro estudio están representadas por marcas como la flexión verbal y nominal que poseen rasgos funcionales como el género, número o la persona, por medio de los cuales establecen vínculos con las categorías léxicas que también poseen tales marcas formales, mediante relaciones de concordancias de rasgos. Tanto las categorías léxicas como las funcionales se registran en el lexicon con sus rasgos de selección.

Baralo (2001: 21) manifiesta el desconocimiento que todavía existe en cuanto al modo en el que el hablante adquiere el conocimiento de las entradas léxicas, sean palabras simples, raíces o afijos, con todos sus rasgos semánticos, categoriales, fonológicos. No obstante, como se acaba de señalar, sí sabemos que la adquisición de una entrada léxica supone, por un lado, la adquisición de un concepto relacionado con una forma, e implica, por otro, saber usar la pieza. Esto último supone, como ya se ha defendido en el capítulo 3, tener presente, entre otros aspectos, que las palabras tienen una carga cultural, por lo que no se puede prescindir de la dimensión sociolingüística; en otras palabras, almacenamos la pieza, no solo con sus particularidades formales, fonológicas, sintácticas y semánticas, sino también como depositaria de ciertas propiedades

idiosincrásicas de la lengua, presentes en la propuesta de Aitchison (1994) de descripción de adquisición de una palabra.

Esta distinción entre el conocimiento formal y semántico (de la pieza aislada) y el conocimiento combinatorio y sociolingüístico (de la pieza en un contexto de uso) está en la base de los dos niveles que hemos distinguido en la producción escrita del verbo: el del lexicón y el computacional. Esta segmentación resulta útil a efectos de la categorización de los datos, aunque el límite entre ambos niveles está necesariamente difuminado, dado que en el lexicón almacenamos la pieza con toda su información de uso. Así, del lexicón se recuperan las piezas (primer nivel) con las que se construyen los sintagmas en el sistema computacional (segundo nivel); es decir, este último se corresponde con la creación de estructuras mediante la fusión de dos o más piezas simples; es ahí donde se hace visible la interacción de los componentes semántico, gramatical y combinatorio. Esta fusión, como se ha señalado, está impuesta desde el principio del proceso de producción por las exigencias selectivas de las categorías léxicas y de las categorías funcionales contenidas en la información registrada en el lexicón, de ahí que, como ya se ha referido, no se pueda establecer ningún corte nítido entre los dos estadios fijados.

6.2.5.2.2. Descripción del error: error léxico y error computacional

A partir de los dos estadios descritos en Baralo (2001) en los que se desarrolla la producción del verbo –su recuperación del lexicón y su combinación con otras piezas en el sistema sintáctico, de acuerdo con una situación comunicativa concreta– distinguimos dos tipos de errores en este nivel descriptivo del error: léxico y computacional. A menudo, no es posible detectar la fuente del error con precisión, pues no tenemos acceso al contenido del lexicón de los aprendientes; el único estadio observable es el que se corresponde con la construcción del sintagma en el sistema computacional: el error puede deberse a una falta de información en el lexicón (es decir, a un conocimiento incompleto¹⁶⁶ acerca de la categoría léxica o funcional) o a un almacenamiento de información errónea, y aun así manifestarse solo en el nivel sintáctico de creación de estructuras.

¹⁶⁶ Ya se ha visto en el capítulo 3 que el conocimiento léxico es resultado de un proceso de adquisición muy lento y gradual.

De este modo, con vistas a poder realizar un análisis sistemático de los datos, se ha optado por establecer como *error léxico* solo el directamente relacionado con la recuperación del lexicón de la pieza léxica aislada o la categoría funcional (es decir, sin atender a la información –contenida también en el lexicón– relativa a la construcción de los sintagmas) y como *error computacional* aquel que está relacionado con los rasgos de selección de la pieza léxica o de la categoría funcional (problemas en la concordancia de rasgos). Ambos son, a su vez, desglosamos a continuación (esta clasificación se corresponde con los dos tipos de taxonomía descriptiva de Corder [1971]: lingüística y según la estrategia superficial)¹⁶⁷.

Entendemos por *error léxico* cualquier problema en la selección paradigmática. Se corresponde con dificultades en la adquisición de alguno de los aspectos de la categoría extraída: el semántico, el morfológico o confusión con el valor sociolingüístico de la pieza extraída. También se incluyen aquí las deficiencias en las locuciones y demás unidades fraseológicas en aquellos casos en los que hay un problema con su forma, significado o contexto de uso. Existen cuatro subclases de errores léxicos:

1) Por ausencia. Este error solo puede darse cuando la categoría léxica que contiene la información semántica relevante para construir el sintagma (el verbo en el sintagma verbal, el sustantivo en el sintagma nominal, la preposición cuando es la marca del sintagma que garantiza la identificación de una función oracional o sintagmática determinada –como en el complemento circunstancial, en el complemento indirecto o en el complemento del nombre–) no está presente (ni implícita ni explícitamente). Es decir, el núcleo del sintagma, que es el elemento que manifiesta y controla las propiedades de todo el sintagma, no está. El siguiente ejemplo ilustra la ausencia de la preposición *a* que introduce el sintagma nominal con la función de CI, y que es, a su vez, la encargada de asignarle al sintagma el valor temático de *experimentante*:

[13] va siempre en aumento el número de personas que les da asco el cigarro (97 HNN).

En estos casos, la preposición actúa como verdadera preposición, es decir, no solo otorga caso a sus complementos, sino que además les asigna papel temático (diferente es la preposición que introduce en ocasiones el sintagma que

¹⁶⁷ Recuérdese que tanto unos como otros pueden localizarse en el núcleo verbal –verdadero foco de estudio de esta investigación–, en el complemento o en el sujeto.

funciona como CD o suplemento; en ambos casos son consideradas únicamente marcas de caso, según el criterio de numerosos autores; véase n. 23)¹⁶⁸.

2) Por la forma. Consiste en un error en la forma de la pieza debido a una confusión en las marcas formales (problema en la flexión verbal o nominal extraídas)¹⁶⁹. A continuación se ejemplifican dos casos de error en la forma de la unidad léxica extraída; en el primero el error se localiza en el complemento, y en el segundo en la locución *darse cuenta de*:

[14] Me da curage ver que a los estudiantes de Penn State nada más les interesa [...] (94 HNN).

[15] pero la gente va a dar cuenta que es demasiado caro pelear con una planta (146 HNN)

3) Por el significado. En este caso, se produce una confusión del significado de la pieza léxica seleccionada, esto es, la pieza léxica extraída existe pero tiene otro significado. Se trata de un caso de impropiedad semántica. En el siguiente ejemplo el aprendiente confunde el significado de la pieza subrayada por el de *sensación*:

[16] cuando oigo canciones en inglés aquí en España me da el sentido de que haya una pérdida de autenticidad de la cultura (29 HNN)

4) Por el contexto comunicativo. El error se produce por una falta de adecuación diatópica, diafásica o diastrática. Recuérdese a ese respecto el ejemplo [8] de este capítulo: *dar una marca* frente a *dar una nota*.

Por lo que respecta al error computacional, se refiere a cualquier problema relacionado con la selección sintagmática del núcleo. En el error computacional se incluyen las dificultades que atañen a la interfaz léxico-sintáctica, esto es, por un lado, la red argumental léxico-semántica del verbo; y, por otro lado, también abarca todos los errores relacionados con el conocimiento combinatorio, tanto sistemático como rutinario (a excepción de la fraseología), y afecta al conocimiento relacionado con los vínculos que establece el verbo con las demás piezas léxicas por medio de sus marcas formales (conocimiento sintáctico). Así, se distinguen tres tipos de error computacional:

¹⁶⁸ No obstante, dejamos constancia de la falta de acuerdo en torno a la función de las preposiciones en la bibliografía consultada.

¹⁶⁹ Véase n. 14 para una explicación sobre la falta de consideración de las cuestiones ortográficas en este análisis.

1) Error en las exigencias selectivas sintácticas. En el lexicon “no solo se encuentran almacenadas las categorías léxicas y funcionales, sino que además estas aparecen caracterizadas por medio de su entorno sintáctico” (Rodríguez Ramalle, 2005: 65), pero, como se ha señalado, las características sobre el entorno sintáctico de la palabra solo son visibles en el sistema computacional. Dependiendo del contexto sintáctico, un sustantivo en un sintagma nominal determinado proyectado por él puede seleccionar o prescindir de complementos y determinante¹⁷⁰; por ejemplo, el sustantivo, si tiene un comportamiento gramatical de naturaleza continua, puede prescindir de un determinante actualizador, ya que por sí mismo tiene la capacidad de denotar la extensión de los predicados correspondientes, esto es, puede constituir por sí solo una expresión referencial (Bosque y Demonte, 1999: 895-896). En el siguiente ejemplo, se observa cómo desde una perspectiva sintáctica el sustantivo *amor*, por los motivos argumentados anteriormente, no requiere un determinante; se trata de un problema de adición de determinante (frente a los casos de omisión):

[16] El amor puede hacer muchísimas cosas, y los homosexuales pueden dar el amor tan bien como los demás (39 HNN)

La relación con el determinante es menos estrecha que con los complementos (al igual que ocurre entre el verbo con el sujeto y los complementos), pues no restringe el tipo de determinante que acompaña al sustantivo; en cambio, ciertos nombres sí que tienen la capacidad de seleccionar la categoría e interpretación de los complementos que lo modifican¹⁷¹.

Asimismo, se incluyen como errores de las exigencias sintácticas del verbo los problemas relacionados con el número del sustantivo con el que se combina el verbo –como el hecho de que *dar* se combine que sustantivos no contables que denotan sustento o protección (como *ayuda* o *alojamiento*)– y con la preposición que rige –o no rige– el verbo en determinadas construcciones, como sucede en relación con la frecuente omisión de la preposición *de* del suplemento de *darse cuenta*. Son también problemas de las exigencias sintácticas, esta vez del sustantivo, las incorrecciones relacionadas con la aparición obligada de la preposición *a* delante del CD, ya sea porque introduzca un rasgo de animación

¹⁷⁰ En este estudio, los determinantes no son interpretados como núcleo del sintagma nominal: la presencia o no de determinante es una propiedad de la construcción; así, en rasgos generales, para que un sustantivo sea sujeto en español, necesita que el determinante le aporte la referencialidad.

¹⁷¹ Para una relación de los diferentes argumentos y ejemplos que confirman estos vínculos diferentes, véase Rodríguez Ramalle (2005: 46).

requerido en una construcción concreta que, como señala Rodríguez Ramalle (2005: 257), no está necesariamente asociado a referentes animados (como en *Le dio a Jenni*, imaginando que alguien llama así a su ordenador); o ya sea porque, de acuerdo con la estructura informativa en un contexto dado, la preposición tiene la propiedad de presentar la información conocida por los hablantes (compárese *Al final tuvo que dar uno de sus perros al centro de acogida de animales, ¿llegaste a conocer alguno?* frente a *Al final tuvo que dar a uno de sus perros al centro de acogida de animales; precisamente a Hugo, el que más te gustaba a tí*). Dado que el CD se manifiesta en español mediante la marca sintáctica de la posición y mediante la preposición *a* ante los complementos de persona y otros que se le asimilan (NGLE. *Manual*: 16), un error en esta preposición lo consideramos vinculado con un desconocimiento de las exigencias sintácticas del sustantivo en un contexto determinado.

Desde una perspectiva léxico-sintáctica, del comportamiento gramatical de un verbo depende el número de sintagmas nominales que coaparecen con él dentro de su proyección sintagmática. Ya se ha señalado que el verbo *dar* se construye potencialmente con tres sintagmas nominales: el sujeto que representa a quien da, el complemento directo que representa lo dado y el complemento indirecto que se corresponde con a quien se da. En el siguiente ejemplo, el verbo *dar* no aparece acompañado de ningún sintagma que desempeñe la función de sujeto, cuando la construcción lo demanda:

[17] Por lo menos, como profesora me ha dado la oportunidad de pasar los veranos en España (76 HNN)

2) Error en las exigencias selectivas semánticas. Ya se ha expuesto en varias ocasiones que las piezas léxicas no solo seleccionan sintácticamente sus argumentos, asignándoles al mismo tiempo un papel temático, sino que además exigen que estos posean propiedades semánticas específicas. Un problema de este tipo, aunque se deba a un desconocimiento de la información de la pieza léxica contenida en el lexicón, solo puede observarse en el nivel sintáctico de creación de estructuras sintagmáticas, de ahí que sea clasificado como error computacional. El siguiente ejemplo ilustra un problema de esta naturaleza; el sustantivo *deseo* y otros que designan lo que se desea o se espera obtener, como *capricho* o *antojo*, se combinan con el verbo *conceder*, usado en el sentido de “dar u otorgar”¹⁷²; es cierto, sin embargo, que el verbo *dar* en el sentido de “producirse

¹⁷² Información obtenida de *Redes*.

o suscitarse en alguien” se combina con *deseo*, pero en estructuras intransitivas, como en *Me dieron deseos de hablar con ella*:

[18] y por eso tengo planes de cuidarla bien, y darle todos los deseos de su corazón (234 HNN)

3) Error en las exigencias selectivas funcionales. Las categorías funcionales, como ya se ha mencionado, carecen de contenido léxico –su significado es más bien de carácter relacional– pero están representadas mediante marcas formales como la flexión verbal, la flexión nominal o mediante los pronombres clíticos, los cuales establecen lazos con las categorías léxicas que también poseen esas marcas formales. Estos lazos se producen a través de concordancia de rasgos. Véase el siguiente error de concordancia en la construcción proyectada por el participio absoluto, que selecciona un sujeto con el cual debe concordar en género y número:

[19] No esperaba mucho de la ciudad de Bilbao dado su fama de ser un centro industrial (138 HNN)

En suma, como se ha señalado en la sección 6.2.5.1 de este capítulo (localización), los errores pueden encontrarse en el núcleo del sintagma verbal, en el complemento, en el sujeto y en el sintagma en su totalidad. De esta manera, en los ejemplos presentados hasta ahora, se ha visto que los errores léxicos y computacionales descritos pueden darse en el núcleo verbal, en el complemento y en el sujeto. Ahora bien, el error también puede localizarse en toda la construcción y no en un solo elemento en concreto; es decir, en este tipo de error identificar los elementos individuales desviados resulta imposible. Así, en lo que se refiere al error en el sintagma, se ha categorizado como léxico cuando la construcción tiene un significado diferente al buscado por el usuario. Por otro lado, el error es computacional si los elementos del sintagma se encuentran presentes pero desordenados, si el sintagma es incomprensible debido al modo en el que las piezas léxicas han sido unidas en el eje sintagmático (y no se puede corregir intercambiando elementos aislados) o si los elementos no se fusionan o coordinan de la manera adecuada, es decir, si las relaciones de concordancia de rasgos no son las adecuadas a lo largo del sintagma. A continuación, se muestra un ejemplo de cada tipo, léxico [20] y computacional [21]:

[20] Si la sociedad les da [a los homosexuales¹⁷³] el permiso de existir, como en España o en otros países (143 HNN)

[21] es fascinante mirar a los niños que tienen tanta imaginación y energía que te pueden dar [por “les puedes dar”] una cosa muy sencilla como una hoja de papel y juegan con ella para tres días (54 HNN)

En definitiva, la taxonomía de los errores cometidos en el uso de dar por un grupo de aprendientes de español, por un lado, atiende a los conocimientos – sobre los diferentes niveles lingüísticos y extralingüísticos que se activan en su producción– y, por otro lado, establece una diferenciación entre los errores léxicos y los computacionales que permite, a su vez, referirse a los fenómenos que describen la desviación (ausencia del núcleo, confusión en el significado, confusión en las restricciones semánticas, etcétera).

La estructuración del lenguaje verbal, la organización de las lenguas en niveles no es “un tema fácil ni el acuerdo entre los investigadores, al respecto, es unánime, más bien ocurre lo contrario. Así, es fácil imaginar que no existe una concepción teórica incontestable que pueda ser aplicada al ámbito que nos ocupa ahora” (Penadés, 2003: 21), y así lo reflejan las enormes diferencias en las clasificaciones del error existentes, como las de Vázquez (1991), Santos (1993) y Fernández (1997). La nuestra es una propuesta taxonómica más, que nace de un intento por clasificar los errores en el uso de dar de manera sistemática y que pretende cubrir en futuros análisis de interlengua todos los errores en el conocimiento productivo escrito de un verbo por aprendientes de una L2; y, para ello, esta clasificación del error se sirve, como se ha visto, de un marco teórico variado, que responde al complejo panorama descrito por Penadés.

Además, para completar esta descripción atendiendo no solo a la naturaleza formal de la lengua, sino también al uso que se hace de ella, hemos adoptado un marco teórico psicolingüístico. El hablante, para construir enunciados lingüísticos, parte del lexicón mental, de la información contenida en cada unidad léxica; pese a que todos los errores son solo visibles en la combinación, en la sintagmática, distinguimos el hecho de que estas unidades son signos formados por significante y significado –primer nivel de análisis, caracterizado por los errores léxicos– que en el acto comunicativo no aparecen aislados, sino combinados –segundo nivel de análisis, caracterizado por los errores computacionales–, de acuerdo con la información léxico-sintáctica de cada unidad léxica contenida también en el lexicón mental.

¹⁷³ Hemos añadido el contenido de los corchetes para aportar más cantidad de contexto, y así facilitar, en la medida de lo posible, la interpretación.

Desde esta perspectiva combinada de la lingüística teórica interna y la lingüística aplicada –por medio de someter a los niveles de estructuración de la lengua un modelo de organización y funcionamiento del lexicón–, se pueden clasificar los errores de manera sistemática tal y como se ha propuesto en esta sección. A continuación, se incluye una tabla con la dimensión descriptiva de la tipología del error:

ERROR LÉXICO	ERROR COMPUTACIONAL
Omisión	Exigencia selectiva sintáctica
Forma	Exigencia selectiva semántica
Significado	Exigencia selectiva funcional
Contexto	

Tabla 7. Dimensión descriptiva de la tipología del error

6.2.5.3. Explicación del error

Tras describir las diferentes maneras en las que un uso no nativo del verbo *dar* puede desviarse de un uso nativo, la cuestión ahora es determinar de dónde viene el material lingüístico que estos aprendientes usan para crear estas desviaciones. Y es que, como señalan Dulay *et al.* (1982: 145), “the accurate description of errors is a separate activity from the task of inferring the sources of those errors”. Este análisis explicativo ayuda a determinar qué elementos son el principal foco de problemas en la producción escrita del verbo y a predecir los posibles errores, más allá de los ocurridos en este corpus; también puede proporcionar información sobre el proceso mental desarrollado en la producción del verbo, aunque las inferencias sobre este proceso fundamentadas únicamente en datos producidos de manera natural deben ser interpretadas como posibilidades, y no certezas.

En esta última dimensión de la tipología se clasifican los errores de acuerdo con un criterio etiológico-lingüístico. Tanto esta clasificación como las referidas a la descripción y localización del error han sido diseñadas para analizar un corpus concreto y un contenido lingüístico determinado, por lo que no están guiadas por una pretensión de generalidad, esto es, no se pretende que esta taxonomía sirva para todos los propósitos, tal y como advierte Tono (2003: 801) tras señalar que

As shown in the history of error analysis, categorizing learner errors is a laborious and often fruitless job, for there are various ways of classifying errors, depending

on research interest and theories involved and it is often the case that the classification is only as valid as the theory it is based on. Also, most people have different perspectives on error types, thus leading to very low inter-rater (or classifier) reliability.

Así pues, no nos hemos podido servir directamente de clasificaciones ya establecidas dentro del análisis de errores debido, entre otras cosas, a la confusión en los aspectos descriptivos y explicativos (véanse las clasificaciones de Vázquez, 1991; de Santos, 1993; y de Fernández, 1997) y a la falta de precisión y de especificación en la definición de las categorías de errores, especialmente en lo que se refiere al error léxico. Tono (2003) mantiene que los objetivos de la investigación deben ser la primera consideración a la hora de valorar una taxonomía del error, lo cual implica lo inadecuado de tomar prestada o diseñar una clasificación del error sin tener fijados los objetivos del análisis. En esta fase del tratamiento del error se establece una distinción entre dos tipos de errores: interlinguales e intralinguales¹⁷⁴.

Los errores interlinguales se deben a una transferencia de la L1 o una L3 a la L2. Esta distinción concierne tanto a los errores léxicos como a los errores computacionales. Así, los errores interlinguales se producen por importación: el aprendiente crea un significado, una forma o una construcción en la L2 a partir de una construcción, significado o forma de la L1 o de otra lengua. En el ejemplo

[22] Me da el sentido de que haya una pérdida de autenticidad de la cultura española (29 HNN)

el aprendiente confunde *sentido* por *sensación* debido a que el español distingue dos formas para lo que en inglés son dos acepciones de la misma forma, *sense*. En este caso, estamos ante un ejemplo de importación de significado.

Para obtener información cuantitativa y cualitativa sobre la influencia de L1 en los errores del uso de *dar* en este grupo de aprendientes, todos los errores

¹⁷⁴ En propuestas tipológicas del error más recientes, como la de Granger (2007: 467), la clasificación se centra en la dimensión descriptiva; en Nesselhauf (2003), tras describir el error colocacional en el corpus, se introduce un apartado sobre el papel de la L1 en la producción de las colocaciones; en su trabajo de 2005, Nesselhauf, aunque no establece de forma explícita ninguna tipología del error en este nivel explicativo, describe a partir de cada error colocacional detectado las diferentes causas intralingüísticas e interlingüísticas que lo provocan; así, distingue entre elementos formales y semánticos de la L2 y de la L1 que intervienen en la construcción de colocaciones en la interlengua.

son examinados para detectar si ha habido influencia. Dado que el análisis está basado en datos de producción, la L1 se puede asumir pero no determinar con certeza, por lo que todos los enunciados sobre la influencia de la L1 deben ser interpretados como que dicha influencia es probable.

Por otro lado, los errores intralinguales no pueden atribuirse a la lengua materna ni a otra lengua que el alumno haya especificado en su perfil en la encuesta (quedaría pendiente en un futuro trabajo comprobar si aprendientes de este nivel de diferentes L1 cometen estos errores, para poder determinar si estos son consecuencia de patrones universales). Los motivos de este tipo de error pueden ser muy diversos. En el caso de los errores léxicos, estos pueden deberse a una derivación errónea, esto es, el aprendiente produce una forma inexistente en L2 como resultado de una derivación errónea por analogía con otra forma de la L2; a una sobregeneralización, es decir, a una selección de una unidad léxica más genérica o vaga de la que se requiere (casos de pobreza léxica); a un desconocimiento de las marcas diatópicas, diafásicas o diastráticas que alejan del estándar la producción del aprendiente; o el problema puede estar ocasionado por una selección léxica errónea (de una palabra, de una locución o de un morfema), sin ninguna razón aparente. En los errores intralingüísticos computacionales no se realiza una distinción explicativa más detallada, son errores que se producen sin una razón clara y sin intervención de la L1, por desconocimiento del sistema del español. En el siguiente ejemplo

[23] No esperaba mucho de la ciudad de Bilbao dado su fama de ser un centro industrial (138 HNN)

el participio debería concordar en género y número con el sujeto seleccionado, pero posiblemente por desconocimiento de este tipo de estructura –de las restricciones funcionales que se producen en esta construcción–, y dado que el género no marcado en español es el masculino, el aprendiente posiblemente considere que el primer elemento de la construcción es fijo, invariable.

A continuación, se presenta una tabla con la distinción entre los diferentes tipos de error léxico y computacional realizada desde una perspectiva etiológica-lingüística.

	ERROR LÉXICO	ERROR COMPUTACIONAL
Interlingüístico	<ol style="list-style-type: none"> 1. Importación de la forma 2. Importación del significado 	<ol style="list-style-type: none"> 1. Importación de la construcción sintáctica 2. Importación de la restricción semántica
Intralingüístico	<ol style="list-style-type: none"> 1. Sobregeneralización 2. Selección léxica errónea 3. Desconocimiento de aspectos diatópicos o diafásicos. 	Intralingüístico

Tabla 8. Dimensión explicativa de la tipología del error

En suma, la metodología descrita a lo largo de este capítulo se ha diseñado específicamente para ser empleada en el análisis de los datos del corpus, que se desarrolla en el siguiente capítulo.

En los estudios lingüísticos basados en corpus no existe un único enfoque; hay análisis basados fundamentalmente en datos numéricos y estadísticos, y otros apenas se apoyan en ellos a la hora de analizar y describir la lengua. Como la mayoría de estudios de corpus, esta investigación se encuentra entre estos dos extremos en el espectro de los enfoques de lingüística de corpus: el enfoque cuantitativo y el enfoque cualitativo. Los estudios cuantitativos son útiles para describir la lengua, pero de ellos se obtiene poca interpretación; en los estudios cualitativos se recurre más a la comprobación de hipótesis y a la interpretación personal, los datos numéricos desencadenan la verdadera investigación; no obstante, se analizan los datos con tanto detalle que se corre el riesgo de perder de vista el panorama general (Guo, 2006: 222).

Para explotar las ventajas de los dos enfoques y evitar los problemas particulares de cada uno de ellos, esta investigación los combina. Se aplica un enfoque cuantitativo en un primer estadio del análisis, cuando se busca cuantificar la frecuencia del lema verbal, en general, y la frecuencia de uso de cada estructura y significado del verbo *dar*. Esta frecuencia sirve como punto de partida para un análisis cualitativo basado en una categorización de los datos –presentada en este capítulo– que es resultado de un razonamiento lógico y de una interpretación personal –basada en estudios relevantes del campo de ASL y de la psicolingüística–.

Como se ha señalado, el foco de nuestro interés en este análisis de la interlengua es identificar qué parte de la interlengua del español, en relación con el verbo *dar*, se produce más como la lengua de los nativos y qué parte se aleja más de la lengua de los nativos. Por consiguiente, la comparación entre los dos corpus es la clave de toda la investigación. Dado que los usos lingüísticos en el corpus de aprendientes oscilan entre aceptables-naturales e inaceptables, el análisis de los datos no nativos conlleva mucho más tiempo que el análisis del

corpus nativo, y requiere el diseño de todo un aparato metodológico que permita no solo describir el uso de *dar* en función del uso que hacen los nativos de este verbo, sino también identificar, describir y, por fin, explicar los errores más representativos que comete este grupo concreto de aprendientes.

Esta propuesta metodológica facilita la descripción y la explicación de la interlengua, lo cual nos puede proporcionar información sobre el proceso mental desarrollado en la producción del verbo, y nos permite predecir lo que los aprendientes necesitan para completar su adquisición de este verbo.

7. EL USO DEL VERBO *DAR*: ANÁLISIS DE CORPUS CONTRASTIVO

En este capítulo se examina el uso escrito de *dar*, uno de los verbos más representativos del grupo de los verbos frecuentes, los cuales constituyen una de las áreas más productivas de la lengua, pues son elementos clave en la comunicación oral y escrita.

Ya se han argumentado en los capítulos 2 y 6 las ventajas de llevar a cabo una investigación que combina un enfoque cuantitativo y cualitativo de los datos. La estructura de este capítulo está determinada por este doble enfoque. El apartado 7.1 está dedicado al análisis de la frecuencia de los datos, que ofrece, en primer lugar, los datos cuantitativos generales del uso del lema *dar* (7.1.1). A continuación, se realiza un estudio cuantitativo y descriptivo de las estructuras sintácticas en las que participa el verbo en cada corpus y de los diferentes sentidos del verbo seleccionados en función del contexto de uso. Con el objetivo de llevar a cabo una investigación exhaustiva separamos en dos apartados distintos los contenidos del nivel léxico-sintáctico involucrados en el conocimiento del verbo *dar*: por un lado, aquellos que tienen que ver con las diferentes construcciones sintácticas fruto de las restricciones categoriales del verbo y de los procesos de gramaticalización y lexicalización en los que participa (7.1.2); y, por otro lado, aquellos contenidos relacionados con la semántica del verbo y las palabras y tipos de palabras con las que se combina (7.1.3). Por último, partiendo de los criterios establecidos para determinar el grado de aceptabilidad de los datos o la inaceptabilidad de los datos, se cuantifican los errores y los aciertos localizados en los usos de *dar* (7.2).

Todos los apartados avanzan de un análisis cuantitativo a uno cualitativo por el que se pretende abandonar el enfoque fundamentalmente descriptivo que caracteriza la presentación de las frecuencias, con lo cual, a su vez, se avanza de lo descriptivo a lo explicativo. Para ello tenemos en cuenta los datos cuantitativos, pero a partir de su explotación estadística determinamos conclusiones fundamentadas en nuestra interpretación, que se postula desde nuestra perspectiva basada en estudios de otras áreas de ASL, como la psicolingüística, y de nuestra experiencia docente en ELE.

En definitiva, este análisis pretende descubrir las semejanzas y diferencias en el uso de *dar* entre el español de un grupo concreto de HNN y el español nativo que representa el estado de lengua al que se ha de aproximar el HNN. La metodología empleada está basada en corpus (véanse capítulos 1 y 2 a este respecto); este es una fuente de datos que constituye muestras de realización o actuación, no de competencia –como son los juicios de gramaticalidad, por

ejemplo–, por lo que solo permiten deducir cómo es la competencia de forma indirecta. Estas cuestiones hacen que, en lo que se refiere al examen de la competencia de los aprendientes en el uso escrito del verbo *dar*, los resultados de un análisis de corpus no puedan ser interpretados de manera exhaustiva; y que la información que pueden proporcionar sobre el proceso mental desarrollado en la producción del verbo, al estar fundamentados únicamente en datos producidos de manera natural, sean interpretados como posibilidades, y no como certezas; la información obtenida abre, por lo tanto, futuras vías de análisis más profundos (a estas nos referimos en el capítulo 8).

7.1. RELACIONES DEL ÍNDICE DE FRECUENCIA ENTRE LOS DOS CORPUS. ANÁLISIS CUANTITATIVO Y CUALITATIVO

La inmensa mayoría de los análisis de interlengua contrastivos se refieren a la divergencia entre las frecuencias en la lengua del aprendiente y la lengua nativa con los términos *sobreutilización* e *infrautilización* (del inglés *overuse* y *underuse*), terminología muy extendida en el ámbito de los estudios contrastivos de interlengua desde la publicación pionera de Granger (1998), *Learner English on Computer*. Ahora bien, como observan Guo (2006: 175) y Hasselgard y Johansson (2011: 55), la interpretación de estos conceptos puede inducir a juicios erróneos dado que se utilizan normalmente para indicar que cierto elemento es producido por los HNN de manera diferente –y se infiere que, por tanto, equivocada– a como lo hacen los HN:

when people talk about ‘overuse’ or ‘underuse’ for a particular item, they imply that learners are using such an item wrongly. When people say a particular word is ‘overused’ by learners, they are implying that on some of the occasions it should no be used. By the same token, when people talk about the ‘underuse’ of the word, what they are implying is that learners do not use the word when they should use it (Guo, 2006: 175).

Que los HNN utilicen un elemento con más frecuencia que los HN no indica necesariamente que ese elemento esté mal usado o que deba usarse menos. Conviene, por lo tanto, matizar estos dos conceptos sobregeneralizadores – *infrautilización* y *sobreutilización*–.

La siguiente tabla, adaptada de Guo (2006: 176), nos permite, por un lado, establecer qué usos del verbo *dar* son frecuentes y cuáles son infrecuentes en los

HN y los HNN; y, por otro lado, presenta una especificación a partir de los dos conceptos anteriores que contribuye a describir la lengua del aprendiente con mayor precisión.

HNN	HN	Interpretación Posible
1. Alta frecuencia de X	Alta frecuencia de X	X representa la parte de la lengua de los HNN más ampliamente compartida ¹⁷⁵
2. Alta frecuencia de X	Baja frecuencia de X	A menos que la diferencia se deba a una variación en el tema del ensayo, existe una sobreutilización de X por desconocimiento del uso preferente en los HN
3. Baja frecuencia de X	Alta frecuencia de X	A menos que la diferencia se deba a una variación en el tema del ensayo, indica una infrautilización de X por desconocimiento del uso preferente en los HN y que, por tanto, otros elementos se están usando inadecuadamente en su lugar ¹⁷⁶
4. Baja frecuencia de X	Baja frecuencia de X	X no se usa con frecuencia en español en general o en este tipo de textos; posiblemente los HNN apenas conocen este uso en español ¹⁷⁷
5. No casos de X	Baja frecuencia de X	La diferencia se debe al tema o es un caso de infrautilización; posiblemente X no sea suficientemente conocido por los HNN, quienes estarán usando en su lugar otras expresiones posiblemente inadecuadas (opciones no preferentes)
6. Baja frecuencia de X	No casos de X	La diferencia se debe al tema del texto o hay una inadecuación por sobreutilización al usar algo que los HN no usan; desconocen el uso preferente. Posiblemente no conocen bien la estructura que utilizan, ya que los HN no la usan, por lo que pueden contener errores
7. No casos de X	Alta frecuencia de X	Si esta divergencia no se debe al tema, sugiere una falta de dominio por parte de los HNN
8. Alta frecuencia de X	No casos de X	Si no se debe al tema, y el uso es correcto, los HNN deben considerar otros modos de expresarse

Tabla 9. Relación entre los índices de frecuencia nativos y no nativos

Volveremos a estos esquemas de relación –que están sujetos a otras interpretaciones posibles– al examinar las diferencias y semejanzas en las

¹⁷⁵ Como se analiza en los siguientes apartados, una frecuencia similar no garantiza semejanzas en el uso detallado.

¹⁷⁶ Probablemente las ocurrencias de X en este esquema se correspondan con ensayos de un nivel de dominio de español alto (sobre la función evaluativa del ACI se reflexiona en el capítulo 9).

¹⁷⁷ Ídem.

frecuencias del corpus. Por el momento, debe señalarse que, para poder interpretar los ocho tipos de relaciones de frecuencia entre los dos corpus que figuran en la Tabla 9, es necesario fijar primero los valores de las nociones *frecuencia alta* y *frecuencia baja*. Un porcentaje del 5 % o más se considerará frecuencia alta, y uno menor que el 5% se considerará frecuencia baja. Aunque este límite es arbitrario, está basado en nuestra experiencia y coincide con el marcado por Guo (2006: 182), quien trabaja con un corpus de tamaño muy similar al nuestro.

Los tipos de relaciones de frecuencias 2, 3, 5, 6, 7 y 8 revelan las áreas en las que se producen las desviaciones de frecuencia entre la lengua nativa y no nativa. Aparentemente, los tipos de relaciones de frecuencia 1 y 4 son de semejanza, pero la semejanza no tiene por qué indicar un uso detallado similar, pues la 4 puede deberse a desconocimiento de los aprendientes, y tanto el esquema 1 como el 4 pueden esconder diferencias estadísticamente significativas en el índice de frecuencia –pese a que estemos refiriéndonos a una alta frecuencia de un ítem en ambos grupos o a una baja–. Por ejemplo, como se ve en la tabla 9 los HN utilizan la estructura ditransitiva con una frecuencia del 30,6 % y los HNN la emplean con una frecuencia del 48,4 %; estas frecuencias son altas (superan el 5 %), pero se diferencian en un 18 %, lo que es estadísticamente significativo ($p < 0,05$).

En principio, desde una perspectiva comparativa entre los dos corpus, cuanto mayor sea la frecuencia de un ítem en el corpus nativo, más certeza habrá en cuanto a la significatividad de la ausencia o presencia del ítem en el corpus de aprendientes. De igual modo, cuanto menor sea la frecuencia del ítem en el corpus nativo, menos seguridad tendremos para referirnos a la significatividad de su presencia o ausencia entre los aprendientes. En relación con ello, los esquemas 1 y 3 diagnostican mejor los problemas del aprendiente, junto a los esquemas 2 y 8, ya que los rasgos de alta frecuencia presentes en los HNN representan lo popular y homogéneo del español colectivo de los HNN (Guo, 2006: 178-179).

Pese a que esta perspectiva de interpretación de la relación entre las frecuencias alta y baja permite describir el español de los HNN con mayor detalle frente a lo que supondría limitarse a la sobreutilización o infrautilización de un elemento, no nos permite conocer con exactitud si las diferencias son estadísticamente significativas; así lo señalan McEnery y Hardie (2012: 51):

To better understand and characterise the frequency data arising from a corpus, corpus linguistics appeal to statistical measures which allow them to shift from

simply describing what they see to testing the significance of any differences observed. Most things that we want to measure are subject to a certain amount of random fluctuation. We can use significance tests to assess how likely it is that a particular result is a coincidence, due simply to chance.

De este modo, aplicaremos siempre que sea posible el test de significatividad Chi-cuadrado a las diferencias entre las frecuencias, con un 5 % como nivel crítico de significatividad estadística ($p < 0,05$). Como la mayoría de los aspectos que medimos están sujetos a cierto grado de fluctuación azarosa, por medio de este test es posible averiguar si los resultados obtenidos son, por tanto, estadísticamente significativos. Ello sucede cuando hay un 95 % de posibilidades de que nuestro resultado no se produzca por simple coincidencia. El asterisco junto a las cifras a lo largo del análisis marca diferencias entre los dos grupos estadísticamente significativas.

Esto quiere decir que los datos serán doblemente filtrados: en primer lugar, analizaremos los datos para determinar las diferencias y semejanzas en las frecuencias (alta o baja) de los usos en ambos corpus, clasificándolos en uno de los ocho esquemas de relación expuestos en la tabla 9; a continuación, comprobaremos si esas diferencias o semejanzas son o no son significativas, aplicando el test de Chi-cuadrado.

7.1.1. Frecuencia total del lema dar: ¿tienden los HNN a utilizar este lema más o menos que los HN?

7.1.1.1. Análisis cuantitativo

El primer paso del análisis consiste en averiguar si este grupo de HNN de ELE utiliza el verbo *dar* con la misma frecuencia que los HN. Para determinar la frecuencia de uso del lema en ambos corpus recurrimos, en primer lugar, al programa Antconc (versión 3.2.3.m) para obtener todas las ocurrencias a través de las concordancias (o *concordancer*). Como ya se ha señalado, el CEDEL2 no es un corpus lematizado en el momento en que realizamos esta investigación, por lo que para poder localizar en él todas las formas del paradigma verbal se han introducido en la búsqueda todas las posibilidades de flexión verbal. Al mismo tiempo, se ha atendido a los posibles errores ortográficos en los que hayan podido incurrir los aprendientes al escribir el verbo, para introducirlos también en la búsqueda. No obstante, no es posible garantizar que algunas formas

ortográficas erróneas de *dar* no hayan sido capturadas¹⁷⁸. A continuación, se excluyeron de las concordancias obtenidas todos los datos irrelevantes para este estudio (como los nombres propios, el sustantivo *dado* o la conjunción *dado que*). En números absolutos, los resultados obtenidos son 360 usos del lema en los HN y 218 en los HNN. Tras obtener estos datos, contamos las frecuencias normalizadas (ocurrencias por 100000 palabras) del lema *dar* en los dos corpus¹⁷⁹. Los resultados se recogen en la tabla 10:

Lema	HNN	HN
Dar	127,6*	183,1

Tabla 10. Frecuencia de *dar* en HN y HNN

Por cada 100000 palabras, los HN emplean *dar* en 183,1 ocasiones, frente a las 127,6 de los HNN. Así, por cada ocurrencia de *dar* en el corpus no nativo, el lema ocurre 1,4 veces más en el nativo.

La tabla muestra, por tanto, una diferencia estadísticamente significativa ($p < 0,05$) entre la frecuencia de uso de *dar* de los HNN y los HN; los HNN usan menos el lema *dar* que HN.

7.1.1.2. Análisis cualitativo

Esta primera fase del análisis permite obtener unos datos cuantitativos interesantes; como señalaba Sinclair (1991: 79) y como se desprende de algunos estudios, como el de Altenberg y Granger (2001) referido al uso del verbo *to make* por aprendientes de inglés franceses y suecos, los HNN tienden a usar menos estos verbos que los HN. Puede haber varias razones para ello. En primer lugar, es posible que los HNN tiendan a evitar estos verbos por la complejidad de su uso, que radica en cuestiones de la intralengua (extensa red de construcciones sintácticas en la que participan estos verbos y los diferentes

¹⁷⁸ Véase en el anexo 1 el listado de formas verbales introducidas en la búsqueda.

¹⁷⁹ A partir de la suma de las apariciones del lema *dar* en el corpus (frecuencia absoluta) se calcula su *porcentaje* de uso. Dado que los dos corpus no son exactamente iguales en tamaño –el corpus nativo cuenta con 196623 palabras y el no nativo con 170849 palabras, por lo que el primero excede en unas 25000 palabras al no nativo–, necesitan normalizarse a una base común. En esta ocasión el porcentaje (%) no transmite la frecuencia de uso de la palabra de manera significativa (0,00183 %), por lo que se calcula una *frecuencia normalizada* a partir de una base de normalización comparable con la medida del corpus, como es 100000.

sentidos que actualizan en cada una de ellas) y de la interlengua (diferencias con el verbo equivalente de la L1). Por tanto, a causa de esta dificultad, los HNN pueden, conscientemente, buscar otras fórmulas que, al no ser las preferentes de los HN, resulten menos naturales o inadecuadas, como señala Sinclair (1991); así, en *Íbamos caminando, vio el charco y realizó un salto para esquivarlo*, el verbo frecuente es sustituido por otro con el que *dar* resulta intercambiable solo en algunos contextos, como los deportivos. En segundo lugar, es posible que, por desconocimiento, y de manera inconsciente, los HNN utilicen ítems léxicos inadecuadamente en lugares donde se debería emplear el verbo *dar*. Ya hicimos referencia (apartado 6.2.1) a algunos ejemplos de este tipo hallados en el corpus, como *Nos dimos cuenta de la importancia que ponen en relajarse y mantener amistades* o *Al final, se hicieron las 18:30*, donde el aprendiente elige por error otros verbos frecuentes del español como *poner* y *hacer* en lugar de *dar* (a este intercambio en la interlengua inglesa se refieren también Chi Man-lai *et al.* [1994]).

Que se den estas posibilidades no implica que no puedan producirse casos en el corpus de los HNN en los que se utilice otro verbo u otra estructura adecuadamente y que pudieran ser reemplazados por el verbo *dar*, como cuando decimos *Ese día me aconsejó algo que no olvidaré* en lugar de *Ese día me dio un consejo que no olvidaré*, o *Se estaba preparando para pronunciar un discurso*, en lugar de *dar un discurso*, expresiones –la de *dar* y la de *pronunciar*– perfectamente intercambiables en un registro neutro, como el que caracteriza una redacción de clase. Hay que tener en cuenta que esta intercambiabilidad puede darse por igual en el corpus no nativo y en el nativo, por lo que no puede descartarse que una parte de esa diferencia de frecuencia que existe en el uso del lema se deba al empleo de otros usos correctos y naturales por parte de los HNN.

A continuación profundizamos más en los usos de *dar*, considerando que este verbo expresa una amplia variedad de significados y participa en una gran cantidad de estructuras. Las diferencias y semejanzas entre las muestras de los HNN y las de los HN en el uso de estos significados y estructuras nos permite describir el uso que hace del verbo el grupo de aprendientes.

7.1.2. Construcciones sintácticas del verbo y su frecuencia

Hemos establecido 12 estructuras sintácticas principales de uso del verbo *dar* en las que hemos agrupado los datos, según la descripción ofrecida en el capítulo 4. La tabla 11 recoge estas estructuras e ilustra cada una de ellas con un ejemplo del corpus. Dado que unos pocos ejemplos pueden clasificarse en más

de una categoría (como darse cuenta de, que es locución pronominal; o dar a + SN, como en la ventana da al norte, que es construcción preposicional e intransitiva), parecía conveniente agrupar los datos atendiendo a una jerarquización de las categorías; los criterios que hemos seguido son los siguientes: si un ejemplo es pasivo, locución, construcción absoluta, construcción preposicional, construcción introducida por como o esquema fraseológico semiproductivo, será automáticamente clasificado en esta categoría, independientemente de que pueda ser a su vez pronominal, transitivo, intransitivo, ditransitivo o impersonal. Entre estos últimos, si un uso es pronominal o impersonal, también anula al resto: transitivo, ditransitivo o intransitivo. Así, el siguiente ejemplo, que es impersonal y transitivo, es categorizado como impersonal: *¿Por qué no dar la oportunidad a un niño sin parientes de vivir en un buen hogar?*, y *Se dio un acontecimiento extraño* es marcado como pronominal, cuando es también intransitivo. De esta manera, si un uso es etiquetado como transitivo, intransitivo o ditransitivo es porque no pertenece a ninguno de las demás construcciones.

Estructuras	Ejemplos ¹⁸⁰
1. Transitiva	Algunas pistas dan mucho miedo
2. Intransitiva	Me da la sensación de que esta opinión no es compartida por todos
3. Ditransitiva	Los fuegos me dan un poco de reparo
4. Impersonal	Una vez legalizado el matrimonio entre homosexuales, toca dar un paso más
5. Pasiva	El hecho de que en España se haya dado un paso adelante es significativo
6. Construcción absoluta	Dada la cantidad de españoles que pueblan la sociedad española
7. Pronominal	Lo único que hacen es legalizar una situación que ya se da de todos modos
8. Construcción preposicional	Al principio fue un poco complicado pero dimos con él
9. Construcción con complemento introducido por <i>como</i>	Es un siglo de guerras y de crisis de todo tipo que dio como resultado la creación de grandes brechas o abismos insuperables
10. Adjetivo	En un momento dado, yo solicité una persona nativa
11. Locución	El padre se da cuenta de que su socio lo ha engañado
12. Esquema fraseológico semiproductivo	La trayectoria vital de una de las mujeres que más ha dado que hablar en la España que va de los 60 a los 80

Tabla 11. Tipos de construcciones sintácticas en las que participa *dar*

¹⁸⁰ En el CD adjunto se incluyen todas las concordancias obtenidas a partir de la búsqueda del *dar*.

Para presentar el español de los HNN en relación con el español nativo, se clasifica cada uso de *dar* del corpus en una de estas doce construcciones sintácticas.

7.1.2.1. Análisis cuantitativo

En la tabla 12 se recogen las frecuencias de *dar* por cada construcción sintáctica. *RF* representa el recuento de frecuencia o frecuencia absoluta y *FN* representa la frecuencia normalizada. El porcentaje (%) se refiere a la frecuencia absoluta dividida por la frecuencia total de *dar* en el corpus correspondiente.

Estructuras	HNN RF	HNN FN	HNN %	HN RF	HN FN	HN %
1. Transitiva	39	22,9	17,8	62	31,5	17,2
2. Intransitiva	5	2,9	2,3	14	7,1	3,9
3. Ditransitiva	106*	62*	48,4*	110	55,9	30,6
4. Impersonal	12	7	5,9	30	15,3	8,3
5. Pasiva	3	1,7	1,4	8	4	2,2
6. Construcción absoluta	3	1,7	1,4	5	2,5	1,4
7. Pronominal	1*	0,6*	0,5*	12	6,1	3,3
8. Construcción preposicional	2	1,2	0,9	12	6,1	3,3
9. Construcción con <i>como</i>	0	0	0	3	1,5	0,8
10. Adjetivo	0	0	0	3	1,5	0,8
11. Locución	45	26,3	20,5	93	47,4	25,8
12. Esquema fraseológico semiproductivo	2	1,2	0,9	8	4	2,2
Totales	218	127,5	100	360	183	100

Tabla 12. Frecuencia de *dar* en las diferentes estructuras

Los siguientes gráficos presentan el porcentaje de cada estructura usada por los HNN y por los HN:

Gráfico 2. Frecuencias de *dar* de cada estructura en el corpus no nativo

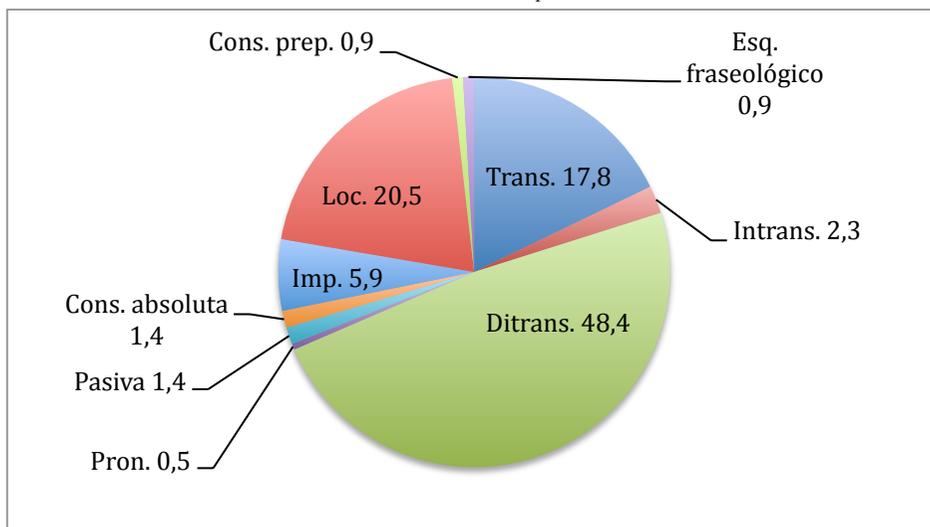
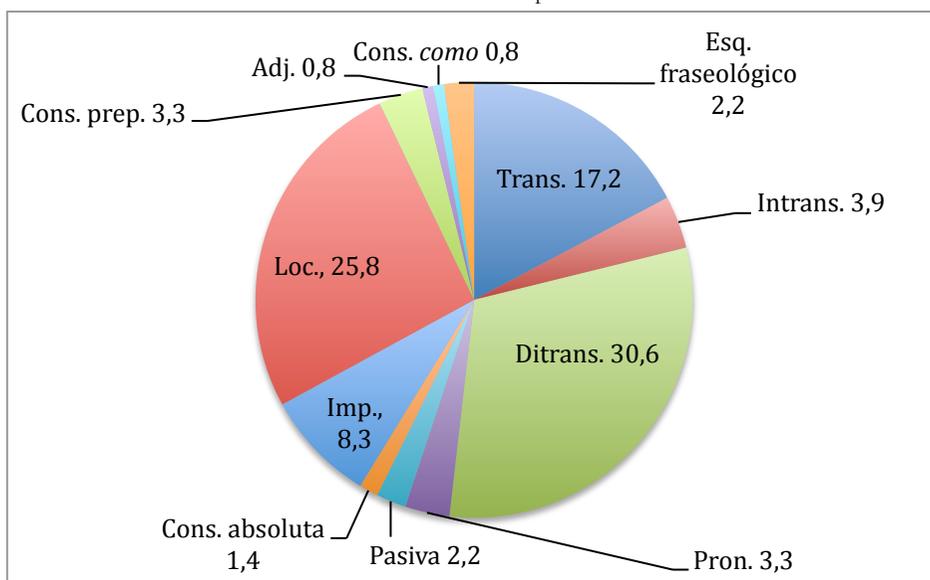


Gráfico 3. Frecuencias de *dar* de cada estructura en el corpus nativo



En primer lugar, atendemos a aquellas construcciones con el verbo *dar* de mayor potencial comunicativo, y, por ello, más frecuentes; nos referimos a las construcciones V + N, que suman más del 50 % de las estructuras empleadas en ambos corpus –de hecho, *dar* ocupa una de las primeras posiciones de los verbos más frecuentes del español por la asiduidad con la que utilizamos las estructuras

en las que el verbo actualiza los sentidos de “proporcionar” y “producir” y va acompañado de un nombre–.

Esta construcción V + N se encuentra en las estructuras sintácticas que se proyectan a partir de la definición sistémica del verbo manteniendo el mismo potencial actancial: A da B (a C) significa “A causa que B pase de no existir a existir (en C, de manera que los seres C disfruten o padezcan B)”. El verbo y el nombre B (argumento interno y tema) son los elementos de la oración que mantienen una relación más estrecha, por lo que su coaparición es necesaria y, por tanto, estable; en cambio, las funciones semánticas *agente* (desempeñada normalmente por el sujeto A) y *experimentador* o *destinatario* (desempeñadas por el CI C) pueden no estar presentes, de hecho, el CI es prescindible. Así, si nos centramos en las construcciones en las que se mantiene esta relación V + N (tema), se han de abarcar no solo las estructuras transitivas y ditransitivas, sino también las pasivas, impersonales e intransitivas. Tomemos como ejemplo las oraciones recogidas en la Tabla 12: 1) Transitiva: *Algunas pistas dan mucho miedo*. En esta oración, el tema es el CD. 2) Intransitiva: *Me da la sensación de que esta opinión no es compartida por todos*, donde el tema es el sujeto. 3) Ditransitiva: Los fuegos artificiales me dan *un poco de reparo*. El tema es el CD. 4) Impersonal: *Una vez legalizado el matrimonio entre homosexuales, toca dar un paso más*. El tema es el CD. 4) Pasiva: *El hecho de que en España se haya dado un paso adelante es significativo*. El tema es el sujeto. Vemos, por tanto, que en todas estas diferentes estructuras sintácticas se mantiene el tema, independientemente de la función sintáctica que desempeñe este papel semántico, mientras que los papeles *agente* y *destinatario* pueden ser prescindibles.

En el corpus no nativo estas construcciones suman el 75,8 % del total, mientras que en el corpus nativo suman el 62,2 %*. Esta diferencia estadísticamente significativa ($p < 0,05$) demuestra que las observaciones anotadas al comienzo del capítulo no son verdaderamente paradójicas, como también advierten en su estudio Altenberg y Granger (2001: 178): los HNN pueden al mismo tiempo infrautilizar un verbo frecuente y sobreutilizar ciertas estructuras.

7.1.2.2. Análisis cualitativo

El anterior resultado sugiere varias consideraciones. Por un lado, el hecho de que los HNN usen más estas estructuras que los HN puede deberse a que se sientan más cómodos con el esquema V + N. Es posible que el principio de los “lexical teddy bears” de Hasselgren (1994) se cumpla no con respecto al verbo

dar en su totalidad, pero sí en lo referente a estas estructuras que presentan una forma similar en la L1 y que se proyectan también en la L1 a partir de un significado básico del verbo equivalente, pues también en inglés (*someone gives something (material) (to somebody)*), lo que provoca que los HNN las sientan como estructuras seguras y las sobreutilicen. En relación con este resultado, también explicarían esta sobreutilización estudios como el de Cobb (2003), que demuestran que los HNN, cuando tienen que realizar una tarea escrita sobre un tema para un lector anónimo, confían en el léxico restringido de la lengua oral, determinado por el contexto, en lugar de hacer uso del vocabulario extenso que caracteriza la escritura nativa. El verbo *dar* seguido de un nombre, sin duda, formaría parte de este léxico restringido de la oralidad. De acuerdo con Cobb (2003), a medida que el nivel de dominio avanza, estos ítems frecuentes están menos presentes en los textos escritos, pero este proceso de cambio parece ser muy lento. Laufer (1998) atribuye este hecho a las convenciones del sistema académico que enfatiza la corrección y no premia la riqueza léxica, por lo que, de acuerdo con esto, los aprendientes sentirían que la construcción *dar + N* es un valor seguro ante la alternativa de buscar continuamente un verbo con un grado de especialización semántica más alto, más propio del texto escrito, como sería el caso del verbo *conceder* en lugar de *dar* en *Sería más fácil de controlar, como en Ámsterdam, donde el gobierno ha dado permisos especiales a ciertos cafés* (6 HNN).

Ya hemos matizado que el mayor uso de una estructura por parte de los HNN no supone que en algunas de las ocasiones en que aparece *dar + N* no deba emplearse dicha construcción. Así, esta mayor frecuencia no tiene por qué implicar un mal uso necesariamente, pero sí que se debe plantear la posibilidad de que responda en ocasiones a casos en los que los HN tengan otros usos como preferentes, lo que convierte los usos no nativos en poco naturales o inadecuados –o, como acabamos de ver, poco específicos semánticamente–. Esto sucede, por poner un ejemplo, cuando los HN prefieren el uso de un verbo en lugar de emplear el sustantivo derivado del verbo en coaparición con el verbo *dar*, como en [*en Año Nuevo*] *Emery y yo nos dimos besos a media noche porque es la tradición* (225 HNN), en lugar de *nos besamos*; estas dos estructuras no son intercambiables en todos los contextos (véase otros ejemplos al respecto en el apartado 7.1.2.1.1 a propósito de las ditransitivas y transitivas).

No obstante, ya utilicen los HNN estas estructuras adecuada o inadecuadamente (se irá viendo a lo largo del análisis), lo que sí es evidente es que esta mayor utilización de la construcción *V + N* implica una menor utilización en algunas de las demás construcciones que tienen otra red actancial, como se refleja en la tabla y gráficos anteriores en el caso de los esquemas

fraseológicos semiproductivos o en el de las construcciones preposicionales, por poner algunos ejemplos. Esto no resulta sorprendente, dado que estas estructuras formales son diferentes en inglés, y en ellas además es mucho más complicado rastrear ese significado básico de *dar*, por lo que resulta más compleja la codificación, como sucede en *Al final, dimos con él*, que no se puede transferir directamente del inglés: *Finally, we found him*. Asimismo, estas estructuras son sintácticamente más complejas que las construcciones V + N porque requieren más elementos para su elaboración.

No es de extrañar, por tanto, que el resto de las estructuras sintácticas (las que no son V + N), que son menos frecuentes en el español general –como se desprende de los datos analizados–, sean utilizadas menos por los HNN de lo que las utilizan los HN: pueden ser sentidas como complejas –y evitarse por las razones intralingüísticas e interlingüísticas mencionadas (también presentadas el apartado 1.4)– y desconocidas –porque aparecen menos en el *input* al que están expuestos los HNN y porque, en consecuencia, tienen menos oportunidades de contrastar su uso en el *output*, menos necesidad de utilizarlas, en comparación con las estructuras frecuentes–.

De hecho, del 38 % restante de las estructuras nativas (porcentaje que se corresponde con las estructuras que no son V + N), el 25,8 % se corresponde con el uso de locuciones, por lo que todavía un 12,2 % está repartido entre las demás estructuras (construcciones absolutas, preposicionales, introducidas por *como*, pronominales, estructuras fraseológicas semiproductivas y adjetivo). En el caso de los HNN, estas cifras son significativamente más bajas ($p < 0,05$); del 24,2 % restante, el 20,5 % son locuciones, por lo que las cinco estructuras restantes se localizan en tan solo en el 3,7 %*.

Esta diferencia estadísticamente significativa en el uso de las estructuras que no se asimilan a la construcción V + N (12,2 % y 3,7 %*) ilustra uno de los esquemas de relación introducidos en la tabla 9, el de “3. Baja frecuencia en los HNN vs. alta frecuencia en los HN”, con la cifra del 5 % establecida como límite entre la alta y la baja frecuencia. Estas diferentes estructuras sintácticas, en su conjunto, sí que aparecen con frecuencia en el corpus nativo, por lo que, de acuerdo con nuestras asunciones basadas en la intuición, la infrautilización de estas estructuras se debe, por un lado, a que otros elementos se están usando en su lugar inadecuadamente y, por otro lado, a que se desconocen –posiblemente la mayoría no son conscientes de su uso en español– por lo que, probablemente, las ocurrencias de estas estructuras en el corpus no nativo se correspondan con redacciones de un nivel de dominio de español más alto.

Estos datos coinciden con los resultados de Cobb (2003), en los que se demuestra que los aprendientes de inglés usan menos estructuras V + preposición *out*, lo que sugiere que estas construcciones multipalabra en las que participa el verbo son descubiertas por los aprendientes de manera muy lenta.

A continuación, dedicamos un análisis más pormenorizado a cada una de las estructuras mencionadas en este apartado.

7.1.2.3. Interpretación de las relaciones de frecuencia en cada una de las estructuras sintácticas entre el corpus nativo y el no nativo

A partir de la tabla de frecuencias sobre *dar* en las diferentes estructuras (tabla 12) y de los dos gráficos anteriores (gráficos 2 y 3), se observa también que el *ranking* de usos de las estructuras es el mismo en los dos corpus: la estructura ditransitiva es la más utilizada, seguida de las locuciones y de la transitiva. No obstante, si en lugar de centrarnos en el *ranking* atendemos a la frecuencia de las estructuras, y al uso que se hace de cada una de ellas, se observan importantes diferencias entre un grupo y otro. En este apartado, interpretamos las relaciones de frecuencias entre los dos corpus con respecto a las estructuras empleadas basándonos en las asunciones presentadas en la introducción del capítulo (los ocho esquemas diferentes de la tabla 9).

7.1.2.3.1. Alta frecuencia en HNN *vs.* alta frecuencia en HN

Hay cuatro estructuras que se incluyen en este esquema de “alta frecuencia *vs.* alta frecuencia”: a) ditransitiva, b) transitiva, c) impersonal y d) locución. Cada una de ellas suma más del 5 % de las estructuras empleadas. En principio, las estructuras incluidas en este esquema representan los aspectos más comunes de la lengua de los HNN.

a) Ditransitivas y b) transitivas¹⁸¹

	HNN		HN	
	FN	%	FN	%
Ditransitiva	106*	48,4*	110	30,6
Transitiva	39	17,8	62	17,2

Tabla 13. Frecuencia de *dar* en las estructuras ditransitivas y transitivas

Resulta lógico el hecho de que las ditransitivas sean más abundantes que las transitivas en los dos grupos, dado que las construcciones en las que participa *dar* se refieren o a una transferencia (rasgo idiosincrásico de *dar*), y no hay transferencia sin destinatario –nos referimos a los usos de *dar* en los que el significado que se actualiza es el de “proporcionar”, como en *dar cariño o alojamiento a alguien*–, o se refieren a un evento de creación (CVA) que, en la mayoría de los casos, puede tener un destinatario (aunque no tiene lugar ninguna transferencia, pues el objeto denotado por el sustantivo no preexiste), como en *dar una bofetada, un abrazo, una clase o los buenos días a alguien*. Según la tabla 2, en el capítulo 4, solo dos de las catorce clases léxicas que se combinan con *dar* cuando funciona como verbo de apoyo no permitirían (o muy raramente) la aparición de este complemento; nos referimos a los grupos 22 y 34, esto es, los sustantivos que designan un “desplazamiento breve” y los sustantivos que designan “fenómenos atmosféricos” como en *Alguien da un salto* y *El tiempo ha dado nieve para esta semana*. A modo de ejemplo, en el corpus no nativo, de las 66 CVA utilizadas, solo 14 no admitirían un CI, es decir, el 21,2 % de las CVA empleadas; en el caso de los HN, de las 82 CVA usadas, 23 no admiten un CI, lo que supone el 28 % de todas las CVA. Son, por tanto, una minoría (en torno a una cuarta parte) las CVA que no admiten un CI. Esta diferencia en los porcentajes (21,2 % y 28 %) también nos muestra que los HN utilizan en sus redacciones un 7 % más que los HNN las dos anteriores clases léxicas combinadas con el verbo de apoyo *dar* (a las clases léxicas se dedica todo el apartado 7.1.3)

Así, teniendo en cuenta que la mayoría de los eventos de creación tienen en su red argumental una casilla para este CI, y que los eventos de transferencia implican siempre un destinatario, es esperable que las oraciones ditransitivas superen en número a las transitivas; y pese a que el CI pueda estar, en ocasiones, implícito, es más frecuente que aparezca –frente a que no exista– el destinatario

¹⁸¹ Ambas estructuras comparten la transitividad, por lo que podemos agruparlas. Lo único que las diferencia es la presencia explícita del CI en las ditransitivas.

sobre el que recae el rasgo idiosincrásico de *dar* de la transferencia o el destinatario de un evento de creación. En el caso de los HNN, por ejemplo, de los 38 usos transitivos, 13 son eventos de creación que no aceptan ningún destinatario, como *dar un paseo, un salto, un suspiro, una voltereta o una vuelta*, por lo que solo pueden formar parte del grupo de los transitivos. De los 25 restantes, 15 son usos del verbo con el sentido de “proporcionar” y 10 con el de “producir”; estos ejemplos sí podrían admitir un destinatario, como en *Estaba en mi oficina en la facultad, preparándome para dar una clase* (HNN 165) o *Muchas personas hablan de Dios cuando dan consejos* (HNN 219). No obstante, los usos ditransitivos (esto es, con CI) de *dar* con el sentido de “proporcionar” son 99 (frente a los 15 anteriores que prescinden del CI), y son 32 los casos en los que aparece el CI con *dar* como verbo de creación (frente a los 20 que no llevan el CI). Estos datos confirman la preferencia de *dar* –cuando actualiza los significados de “proporcionar” y “producir”– con destinatario: de los usos de *dar* que pueden llevar CI, el 81 % lo llevan, frente al 19 % que no lo llevan. En el caso de los HN la preferencia por los usos con CI explícito también es clara: 110 son las estructuras que se utilizan con CI frente a las 39 que podrían llevarlo explícito pero no lo llevan; dicho de otro modo, un 65 % presenta CI frente al 35 % que no lo hace aunque podría. Así, en ambos grupos el porcentaje de usos con CI explícito es mayor que el de usos con CI implícito, aunque los HN prescinden del CI más que los HNN. Esta diferencia radica, en parte, en el hecho de que los HNN sobreutilizan las combinaciones de *dar* con las clases de sustantivos que denotan “sensaciones físicas, psíquicas o anímicas” (*ataque de nervios, alegría, felicidad, miedo*), como se describe y explica en el apartado 7.1.3. En dichas estructuras siempre existe un experimentador de dichas emociones, que es el CI.

En relación con estos resultados, si seguimos analizando los porcentajes, observamos que estos dos tipos de estructuras, las ditransitivas y las transitivas, suman el 47,8 % del total de las estructuras en los HN, y el 66,2 %* en los HNN. Aunque estas estructuras son frecuentes en ambos corpus (superan el 5 % de aparición fijado), hay una diferencia en la frecuencia que es estadísticamente significativa ($p < 0,05$). Las estructuras transitivas se usan en ambos grupos por igual, en un 17 % de los casos, por lo que la diferencia radica en las ditransitivas.

Hemos visto al comienzo de este apartado (7.1.2.1.1) que es esperable que las ditransitivas sean, en el español general (no ya solo en nuestros corpus), más frecuentes que las transitivas. En relación con lo mencionado anteriormente acerca de las estructuras V + N en las que se engloban las transitivas y las ditransitivas, consideramos que estas son estructuras con la que los HNN se sienten cómodos, por lo que es posible que las ditransitivas, que de por sí se

utilizan mucho (y mucho más que las transitivas), sean utilizadas incluso en contextos en los que los HN tendrían otra opción preferente, como es el caso de los ejemplos en los que encontramos paráfrasis con *dar* en lugar del verbo del que se deriva el sustantivo o en lugar de un verbo asociado. En usos como estos pueden esconderse las causas de su sobreutilización:

[1] Me da gusto saber que puedo ayudarles a mejorar sus vidas (116 HNN). En lugar de *Me gusta saber*.

[2] el gobierno dice que [el cannabis] es "una droga" que actúa como una entrada a las drogas mas duras. (...) Es una mentira del gobierno para dar miedo a los jóvenes (145HNN). En lugar de *asustar*.

[3] [en Año Nuevo] Emery yo nos dimos besos a media noche porque es la tradición (225 HNN). En lugar de *nos besamos*.

Como se puede observar, la diferencia entre los porcentajes del uso de la construcción V + N se origina, en parte, en la diferencia de uso de las ditransitivas. En el apartado 7.1.3 profundizamos más en estas estructuras cuando analicemos las semejanzas y diferencias en las palabras y tipos de palabras con la que se combina *dar*.

c) Impersonales

	HNN		HN	
	FN	%	FN	%
Impersonales	12	5,9	30	8,3

Tabla 14. Frecuencia de *dar* en las estructuras impersonales

Se entiende por estructuras impersonales aquellas que cuentan con un predicado impersonal léxico, es decir, aquellas en las que se da una ausencia del argumento correspondiente al sujeto cuando está determinada por la naturaleza semántica de los predicados, como en *Está nevando*. No obstante, la tradición establece –y así también se considera en este estudio– como estructuras impersonales aquellas en las que la acción se predica de alguna entidad genérica, como en *Compré el vuelo por Internet; toda una maravilla, te dan un billete electrónico y ya está* (10 HN). Así pues, en este sentido amplio de impersonal, todas las oraciones del corpus en las que *dar* presente un sujeto tácito de interpretación inespecífica han sido etiquetadas como impersonales.

Así, se forman en español oraciones impersonales con verbos conjugados en la tercera persona del plural, como es el caso del ejemplo anterior. En estas oraciones se interpreta un sujeto tácito que alude siempre a personas; este recibe interpretación inespecífica, que suele hacer referencia a conjuntos más amplios de individuos, con un significado próximo al genérico. La interpretación genérica puede darse también con la primera persona del plural, como en *Esta experiencia me ha hecho creer en el valor extraordinario de convivir con gente paciente que le ayuden a uno aprender por medio de la conversación diaria. Tiene mucho más valor que las clases que damos en los institutos* (12 HNN); y con la segunda del singular, compatible con el sujeto expreso: *Tendemos a vivir como si no fuera a ocurrir nunca y cuando ocurre de esta manera, por la calle, el dolor al que no has dado importancia te sobreviene y cada vez es más intenso* (94 HN).

También se han clasificado como oraciones impersonales las oraciones con *se* impersonal, como *Me entristece [el telediario] en el sentido de que consigo darme cuenta de la politización que se le da a la información* (376 HN), así como las oraciones en las que el sujeto del infinitivo carece de un controlador, como en *dijeron que era un sitio para comer o algo así, que no para dar regalos* (111 HNN); lo característico de esta interpretación es que el sujeto del infinitivo recibe una lectura no específica, no referencial, similar a la que posee el sujeto en una oración impersonal del tipo de *Se come bien aquí (la gente, cualquiera)* (Rodríguez Ramalle, 2005: 305). Esta construcción, por lo tanto, establece vínculos interpretativos con las oraciones impersonales (a diferencia del controlador que debe recibir sus rasgos de un argumento de la oración principal, como en *Confío en ir*).

Las estructuras impersonales son usadas con frecuencia en ambos corpus, pero los HNN la usan menos que los HN. A continuación, comprobamos cómo estos cuatro tipos de impersonalidad se distribuyen en cada uno de los corpus.

	3. ^a RF	%	1. ^a RF	%	2. ^a RF	%	se RF	%	inf. RF	%	Total RF	%
HN	16	53,3	1	3,3	6	20	2	6,7	5	16,7	30	100
HNN	2	16,6	4	33,3	0	0	0	0	6	50	12	100

Tabla 15. Frecuencia de cada tipo de impersonalidad

Los HNN prefieren usar el infinitivo con un sujeto no especificado: lo usan en un 50 % de los casos, mientras que los HN lo hacen en un 16,7 %. Es comprensible que los HNN, con L1 inglés, utilicen más estas estructuras con respecto a las otras: los infinitivos en inglés funcionan como en español, es decir,

existe una equivalencia entre los infinitivos con *to* y sin *to* en inglés y los infinitivos en español; pero, además, el gerundio en inglés también se comporta como los infinitivos en español: puede actuar como predicado o como complemento, por lo que es probable que el aprendiente trate de trasladar de su L1 al español el infinitivo y el gerundio con sujetos inespecíficos. El infinitivo es posiblemente la forma más neutra de expresar la impersonalidad, junto con el *se* impersonal, pero este es mucho más complejo de utilizar –además de que no existe un equivalente formal en la L1 y de que la exposición de los HNN a esta estructura no debe de haber sido muy abundante, como demuestra el hecho de que su presencia en un corpus nativo no es especialmente alta–, lo que explica que los HNN no lo utilicen ni una sola vez.

La 1.^a persona del plural la utilizan más los HNN, lo que indica que, en géneros textuales equivalentes, los HNN muestran más implicación personal que la de los HN (Cobb, 2003: 416); en efecto, la 1.^a persona indica mayor implicación personal que la 2.^a o la 3.^a.

En cuanto a las personas 3.^a y 2.^a del plural, hay una cuestión relacionada con el registro que conviene matizar. Tanto la 3.^a del plural como la 2.^a del singular, que suman el 73,3 % de los usos impersonales nativos, son más propias del registro coloquial o familiar, principalmente la 2.^a. Dado que ese registro es el que los hablantes más utilizamos, es lógico pensar que a diario los HN usen estas formas. En el momento en que tengan un desliz en el cambio de registro (del familiar al neutro o formal), esas formas pueden trasladarse. Así pues, si consideramos las redacciones de los HN desde el punto de vista del registro, convendría subrayar que esas formas no son las más adecuadas para un registro neutro¹⁸².

Según lo argumentado, la tendencia de los HNN (orden decreciente: 1.^o infinitivo [50] > 2.^o 1.^a del plural [33] > 3.^a del plural [16] > 2.^a del singular [0] = con *se* [0]) se puede explicar por tres razones: reflejo de la L1, grado de implicación personal y grado de dificultad. En los tres primeros casos, las estructuras son fácilmente reconocibles en la L1; la 2.^a del singular probablemente tenga también un aire de registro muy familiar en inglés que los alumnos quieren evitar; en el caso del *se*, forma una estructura muy compleja, y un alumno no la usa si no la tiene totalmente asimilada. De hecho, en el análisis

¹⁸² Como se desprende de estos datos, el corpus nativo no puede concebirse como modelo de producción en todos los aspectos. Ädel (2006: 206) hace una interesante reflexión al respecto: “if we take it for granted that learners aim to achieve as professional a style of writing as possible, we should not make recommendations to learners based on native-speaker student usage, but rather would use professional native-speaker writing as the target”. De ahí que las cuestiones normativas prescriptivas deban ser tenidas en cuenta en los análisis de corpus contrastivo.

longitudinal de Laufer (1998) se demuestra que los aprendientes no aplican en las actividades libres de producción escrita el conocimiento léxico que demuestran haber desarrollado desde una perspectiva receptiva, por lo que existe la posibilidad de que en este nivel nuestros HNN ya conozcan la estructura con *se* desde una perspectiva receptiva, pero no la usen en la producción, o al menos en la producción abierta.

d) Locuciones

	HNN FN	%	HN FN	%
Locuciones	45	20,5	93	25,8

Tabla 16. Frecuencia de *dar* en las locuciones

Las locuciones son la segunda estructura más utilizada por ambos grupos. A continuación se presentan todas las locuciones utilizadas en los corpus y la frecuencia de uso de cada una de ellas:

Locuciones	No Nativo RF	No Nativo %	Nativo RF	Nativo %
1. Dar(se) vuelta	0	0	3	3,2
2. Dar igual	1	2,2	5	5,4
3. Darse cuenta de	36	80	73	78,4
4. Dar a luz	5	11,1	0	0
5. Dar el brazo a torcer	0	0	2	2,1
6. Dar lugar a	0	0	3	3,2
7. (No) dar abasto	1	2,2	0	0
8. Dar vida a	0	0	1	1,1
9. Para dar y regalar	0	0	1	1,1
10. Dar una lección	0	0	1	1,1
11. No dar para más, (no) dar de sí	0	0	1	1,1
12. Dar la gana	2	4,4	2	2,1
13. Dar de alta	0	0	1	1,1
Total	45	100	93	100

Tabla 17. Frecuencia de cada tipo de locución

A simple vista observamos que la locución más usada –con bastante diferencia– en los dos corpus es *darse cuenta*, y que su frecuencia de uso es muy similar: 78,4 % entre los nativos y 80 % entre los no nativos¹⁸³. En el 20 % restante se concentran las demás locuciones. Los HN usan 10 tipos de locuciones más, en un porcentaje bajo. En cambio, los HNN usan tan solo 4 tipos más, la mayoría también en un porcentaje bajo. Estas cifras muestran, por un lado, que son pocas las locuciones producidas en el corpus no nativo (5 en total), por lo que el español productivo de este grupo en lo que se refiere a las locuciones con *dar* no es tan extensivo. Por otro lado, el hecho de que los HN hagan uso de una gama más amplia de locuciones, pero cada una de ellas en un porcentaje bajo, hace que, posiblemente, los HNN no tengan suficientes oportunidades para escribirlas o necesidad para utilizarlas. En relación con ello, se puede afirmar que la mayoría de los HNN no conocen estos usos del español (*dar lugar a* o *dar igual*). De hecho, las redacciones de aquellos HNN que usan locuciones menos frecuentes (*dar abasto* o *dar igual*) son aquellas que parecen reflejar un mejor dominio lingüístico (dentro del nivel establecido).

Llama la atención que haya ítems en el español no nativo que no se den en el nativo –lo normal es lo contrario–. En este caso tenemos dos tipos de locuciones usadas solo por los HNN: *dar a luz* y *dar abasto*, esta última utilizada solamente en una ocasión. Estas no son, en absoluto, locuciones inusuales o no naturales; en un corpus nativo más extenso, estas combinaciones sí se encuentran, como en el Corpus del Español, donde se obtienen 37 resultados de *dar abasto*, o en el CREA, donde aparecen 24; en este corpus aparece además *dar a luz* en 317 ocasiones. No obstante, pese a que estas combinaciones sí se utilizan, son mucho menos frecuentes en español que *darse cuenta*. Por ello, teniendo en cuenta el tamaño del CEDEL2 y los esquemas de relación de frecuencias que hemos establecido, estas locuciones no han tenido muchas oportunidades de aparecer, por lo que es razonable que los elementos correspondientes al esquema de uso “bajo uso en HNN vs. no uso en HN” no estén bien producidos por los HNN, dado que estos, por las razones aportadas, no han tenido muchas oportunidades de exponerse a ellos ni, por tanto, de adquirirlos (de hecho, hay errores en prácticamente todos ellos, como se describe y explica en el apartado 7.2). Hay que tener también en cuenta que la locución *dar a luz* se aplica en una esfera más específica, lo que explicaría su no aparición en el corpus nativo.

¹⁸³ Ya se ha señalado que una alta frecuencia de uso de un elemento o una frecuencia similar en los dos grupos –como es el caso– no implican que el uso que se hace de ese elemento sea adecuado. A la adecuación de los usos le dedicamos el apartado 7.2.

Es llamativo también el uso de *dar la gana* en ambos corpus; por una cuestión de registro, no se espera que esta locución coloquial (cf. *DRAE*, 2001) aparezca en una redacción de clase, caracterizada por un registro neutro.

El caso de *dar(se) vuelta* es especial, dado que es una locución propia del español de América; *vuelta* forma parte del verbo y toda la locución funciona unitariamente como un verbo transitivo; el complemento que expresa lo que se gira o vuelve es directo y, por tanto, si es de cosa, no va precedido de preposición: “Dio vuelta la cabeza para buscar a su colega” (véase al respecto el *DPD*). En la Península se utiliza más *dar(se) la vuelta*, que es la opción también más elegida por los HNN, y que no hemos categorizado como locución, ya que *vuelta* es, en este caso, CD, *dar* funciona como verbo de apoyo y lo que se gira o vuelve constituye el CI.

7.1.2.3.2. Baja frecuencia en los HNN vs. baja frecuencia en los HN

Las estructuras clasificadas en este esquema de frecuencia son las intransitivas, pasivas, construcciones absolutas, pronominales, preposicionales y esquemas fraseológicos semiproductivos. En líneas generales, si atendemos a nuestra interpretación sobre las relaciones de frecuencia, baja frecuencia en un corpus nativo indica ítems que o no son muy frecuentes en español en general o no lo son en el tipo de texto considerado (aunque conviene contrastar los datos con corpus nativos más extensos), pero que no por ello dejan de ser naturales y característicos de la producción nativa. Que los aspectos que caen en esta categoría no son muy empleados por los HN quiere decir que el HNN tampoco tendrá muchas ocasiones de producirlo, o mucha necesidad para hacerlo. Por otro lado, como observa Guo (2006: 177), es probable que, si no son tan frecuentes, los HNN no sean muy conscientes de su uso en español. Analizamos a continuación cada una de estas estructuras por separado.

a) Intransitivas

	HNN		HN	
	FN	%	FN	%
Intransitivas	5	2,3	14	3,9

Tabla 18. Frecuencia de *dar* en las estructuras intransitivas

En estas estructuras, el papel temático *tema* que participa en la red argumental de *dar*, en lugar de ser desempeñado por la función sintáctica de CD, es el sujeto de la oración.

En todas las intransitivas que usan los HNN el verbo se utiliza en el sentido de “producir(se) o suscitar(se) en alguien”, y el tema lo constituyen sustantivos que pertenecen a tres clases semánticas: la que está formada por los sustantivos *sensación* e *impresión*, así como otros que designan “alteraciones y reacciones físicas adversas, generalmente repentinas” (como en los ejemplos [1] y [2]); la que está formada por los sustantivos que denotan “deseo o necesidad” (véanse [3] y [4]); y, por último, la de los sustantivos que denotan “miedo o inquietud” (como en [5]):

[1] cuando oigo canciones en inglés aquí en España, me da el sentido [por sensación] de que hay una pérdida de autenticidad de la cultura española (29 HNN)

[2] espero que nunca me pase otra vez en mi vida, porque me dio un ataque de nervios (88 HNN)

[3] cuando sientes el olor a cigarro, te dan ganas de fumar (98 HNN)

[4] También hay MacDonald's por si le da hambre, y un montón de restaurantes chinos (117 HNN)

[5] hay tumbas de cristianos perseguidos. Era muy interesante pero me daba miedo un poco porque eran subterráneos (204 HNN)

En el caso de los HN, el verbo en las intransitivas actualiza también el sentido de “producir(se) o suscitar (se) en alguien” (en el 57,1 % de las intransitivas), y el tema-sujeto pertenece a las mismas clases semánticas que en el caso de los HN ([1], [2], [5], [7] y [8]), aunque debe añadirse dos: la de los sustantivos que designan “otros estados de aflicción y diversas sensaciones anímicas relacionadas con la infelicidad o el enojo” [9] y la de los sustantivos que designan “horas” [11]. Se recogen, a continuación, los usos de *dar* nativos intransitivos:

[1] me da la sensación de que esta opinión no es compartida por muchos fumadores (42 HN)

[2] que tiene tantas cosas que ver y hacer que dan ganas de liarse la manta a la cabeza y mudarse sin pensarlo (51 HN)

[3] Los días transcurrían deprisa y nos dio tiempo a hacer amistad con varios de los pasajeros del crucero (56 HN)

[4] Me habría gustado viajar al exterior pero no me dieron los tiempos, espero

este año poder hacerlo (91 HN)

[5] Me dio la impresión de que el público no puede distinguir la realidad de la ficción (102 HN)

[6] también me dio tiempo a encontrarme con una amiga y pasar un día en grata compañía (110 HN)

[7] pero te daba la sensación de que las vacaciones no se acababan nunca (215 HN)

[8] Trato de alejarme pero me da cosita ya que el problema no es nuestra relación (229 HN)

[9] el primer día no la cogieron pero el segundo día les dio mucha pena y volvieron a ver si estaba allí (231 HN)

[10] El último día no nos dio mucho tiempo a disfrutar de nada porque nos tuvimos que levantar pronto (262 HN)

[11] el encantamiento no se rompe porque dan las 12 de la noche (292 HN)

[12] el período vacacional es más largo que para el resto así que me dio tiempo a hacer muchas cosas (292 HN)

[13] alquilamos un coche y recorrimos lo que nos dio tiempo de la isla, no sin algunos contratiempos y pérdidas (306 HN)

[14] A veces, me da ganas de decirle a todo el mundo que no sean tan agonías (332 HN)

Por otro lado, los HN también producen oraciones intransitivas en las que el verbo actualiza el sentido de “proporcionar”, como cuando se combina con algunos sustantivos que designan “periodos o intervalos”, en concreto con el sustantivo *tiempo* ([3], [4], [6], [10], [12] y [13]); estas estructuras suman el 42,9 % de las intransitivas.

Resulta, por tanto, que el uso en los HN y en los HNN de las estructuras intransitivas en las que el verbo actualiza el sentido de “producir(se) o suscitar(se) en alguien” es similar: los HNN lo combinan con los sustantivos *sensación*, *ganas*, *ataque de nervios*, *hambre* y *miedo*, y los nativos con *sensación*, *ganas*, *impresión*, *cosita* (por apuro) y *pena*. Excepto el último sustantivo, los demás pertenecen a las tres clases semánticas señaladas, y son todos ellos combinaciones naturales. En los dos casos en los que los HNN combinan *dar* con el sustantivo *pena* lo producen con una estructura transitiva, lo cual revela que tal uso no está muy extendido entre los HNN.

Un uso interesante que hacen los HN es el que afecta a la expresión de las horas [11]; en español, las expresiones referidas a las horas pueden construirse con el verbo *dar*, que en español actual concuerda con el complemento de tiempo, y con un dativo; a veces admite un SN antepuesto que funciona como

sujeto (*reloj* o *campana*) y que bloquea la concordancia con el sintagma de significación temporal y la aparición del dativo (Fernández Soriano y Táboas, 1999: 1764); pero no es lo más frecuente, ni es este el caso en el corpus. Dado que no se utiliza con mucha frecuencia (un solo caso en el corpus nativo de control), y puesto que es un uso bastante particular y complejo, no es descartable que este uso no sea conocido por los HNN, ya que no existen muchas posibilidades de producirlo.

Pese a que el uso de las intransitivas cuando el verbo actualiza el sentido de “producir(se)” es similar en ambos grupos, la diferencia es que, mientras que en los HNN este uso supone el 100 % de los casos, en los nativos constituye el 57 %; el 43 % restante engloba usos en los que el verbo actualiza el sentido de “proporcionar”. Esto indica que el español productivo de los HNN en lo que se refiere a las intransitivas con *dar* en el sentido de “proporcionar” no es precisamente amplio; y, más concretamente, que la ausencia de uso de la estructura específica que corresponde a *dar* más el sustantivo *tiempo* puede estar provocando la producción de construcciones inadecuadas o no preferentes en lugares en los que dicha estructura tendría que aparecer.

b) Pasivas

	HNN FN	%	HN FN	%
Pasivas	3	1,4	8	2,2

Tabla 19. Frecuencia de *dar* en las estructuras pasivas

Las pasivas con el verbo *dar* no son muy utilizadas en ninguno de los dos corpus, comparadas con otras estructuras; no obstante, pese a la semejanza en la frecuencia también hallamos diferencias en su uso. Los HN emplean únicamente la pasiva con *se*; los casos se corresponden con la pasiva refleja, en la que el verbo actualiza prácticamente por igual el sentido de proporcionar ([2], [3], [4] y [6]) y el de producir ([1], [5] y [7]). A continuación se incluyen las muestras nativas¹⁸⁴:

[1] El hecho de que en España se haya dado un paso adelante es significativo (HN)

[2] las familias formadas por homosexuales deberían estar permitidas o se podrían

¹⁸⁴ Hemos eliminado una de las muestras porque el sintagma verbal resulta incomprensible.

dar las mismas facilidades que a las familias formadas por padres heteros (142 HN)

[3] Antes de nada, es imprescindible la educación que se da a los niños en las escuelas (145 HN)

[4] Por otra parte, hay gente que la desprecia solo por la imagen que se da de ella (146 HN)

[5] se abandonan las armas y por lo tanto se da un paso muy importante para alcanzar la paz en dicha zona (183 HN)

[6] Día a día se daban más declaraciones que trataban de justificar que no era ""justo"" (275 HN)

[7] Al denominarlo ""problema"" parece estar dándose la sensación de que los inmigrantes son un inconveniente (346 HN)

En cambio, los HNN emplean el verbo *ser* en 2 de las 3 ocasiones en las que usan la pasiva:

[1] Después de mucho tiempo y grandes luchas, los derechos necesarios fueron dados (77 HNN)

[2] Otros derechos que acompañan el matrimonio deben ser dados, como el de compartir el seguro sanitario y de vida (80 HNN)

[3] los turistas vienen a nuestra ciudad desde otros lugares porque aquí se les da la oportunidad de participar en varias actividades (105 HNN)

Es probable que este uso de la pasiva con *ser* se deba a la influencia de la L1; la combinación *rights* con *give* en pasiva se utiliza en inglés con más frecuencia que en voz activa (en el British National Corpus se encuentran 8 ejemplos de *given rights*, y, en cambio, solo 3 con la forma activa *give rights*); no obstante, los ejemplos [1] y [2] pertenecen al mismo aprendiente. La pasiva con *se*, con el verbo con el sentido de “proporcionar”, aparece en una ocasión [3], que corresponde a una redacción elaborada con un nivel de dominio lingüístico muy alto; esta pasiva, al igual que se dijo a propósito de las impersonales con *se*, resulta una estructura compleja para los HNN –de ahí su escasa producción–.

c) Construcciones absolutas

	HNN		HN	
	FN	%	FN	%
Construcción absoluta	3	1,4	5	1,4

Tabla 20. Frecuencia de *dar* en las construcción absolutas

Esta estructura se utiliza por igual, por lo que respecta a frecuencia, en el corpus nativo y en el no nativo. La baja frecuencia en la producción nativa indica que probablemente los HNN no tienen muchas necesidades de utilizarla, y ello contribuye al desconocimiento de la estructura.

Vale la pena describir el uso de esta construcción de manera cuantitativa con más detalle atendiendo a los errores de su producción, aunque nos estemos adelantando al apartado de la inaceptabilidad de los datos (7.2). La construcción absoluta es de las menos utilizada por los HNN –aunque no por ello es poco natural para oídos nativos– y se caracteriza por ciertas particularidades formales o estructurales que hay que conocer. El participio *dado*, *-da* se emplea con frecuencia en construcción absoluta cuando va seguido de un sustantivo, con el sentido de “teniendo en cuenta lo que dicho sustantivo expresa” (véase el *Diccionario Panhispánico de dudas*). En esta construcción, el participio debe concordar en género y número con el sustantivo. Los siguientes ejemplos muestran los problemas que los HNN e incluso los HN tienen para construir estas oraciones:

- [1] dada la peculiar organización impositiva (HN 75)
- [2] dada la cantidad de fumadores que pueblan la sociedad española (HN 224)
- [3] Todo el día intento que sea lo más eficaz posible dado al motivo de la escasez de tiempo libre (HN 226)
- [4] dada la temperatura del agua (HN 235)
- [5] una variedad de registros a los que no nos tiene acostumbrados, dado sus papeles algo estereotipados de guaperas perdonavidas e irresistible (HN 285)
- [6] dado las bajas temperaturas del exterior (HNN 31)
- [7] dado su fama de ser un centro industrial (HNN 138)
- [8] dado el tiempo fatal en Andalucía este invierno (HNN 181)

Como se desprende de los ejemplos, los errores en el uso de esta estructura en los dos subcorpus ([5], [6], [7]) se producen siempre por el uso invariable del masculino singular *dado*. Es decir, cuando no se conoce profundamente esta construcción, se inserta, por defecto, el participio en masculino singular, posiblemente porque se cree invariable en la estructura –así ocurre con los HNN, en cuyas muestras el 100 % de las estructuras se construyen con *dado*, cuando el 66,7 % no deberían tener esta forma– o porque, debido a su semejanza formal y significativa, se confunda *dado algo* por *debido a algo*, como parece ocurrir en el ejemplo nativo [3].

Se observa, una vez más, una estructura poco frecuente en español, compleja desde una perspectiva sintáctica, y que, por lo tanto, se presenta como foco de error para los HNN.

d) Pronominales

	HNN		HN	
	FN	%	FN	%
Pronominales	1*	0,5*	12	3,3

Tabla 21. Frecuencia de *dar* en las pronominales

Hemos hallado dos tipos de estructuras pronominales en los corpus, y hay muestras de ambos en el corpus nativo. En él se dan 11 casos de *darse* + SN, con el significado de “suceder, existir”. Con este sentido solo se usa con pronombre, como en *Lo único que hacen es legalizar una situación que ya se da de todos modos* (27 HN). La muestra [12] de los usos pronominales corresponde a *darse* + SN, con el sentido de “entregarse”, estructura en la que el verbo se combina con sustantivos que denotan “exceso en una actividad cualquiera, de manera puntual”, como en *Realmente satisfecha y contenta de haberme dado lo que se llama: un verdadero banquete cultural* (203 HN).

En el corpus no nativo solo aparece una construcción pronominal, que se corresponde con el primer tipo descrito: *Un problema que se ha dado en Inglaterra y en Escocia, creo, es el tema de que les cuesta a veces a los dueños de bares hacer que se respete la ley* (99 HNN). De nuevo, nos encontramos con un uso que escasea en el corpus no nativo y que solo aparece en una redacción con un nivel de dominio por encima de la media. Esta diferencia estadísticamente significativa entre los dos corpus sugiere que la mayoría de los HNN no son conscientes de este uso (*darse* con el significado de suceder).

e) Construcciones preposicionales

	HNN		HN	
	FN	%	FN	%
Construcción preposicional	2	0,9	12	3,3

Tabla 22. Frecuencia de *dar* en las construcciones preposicionales

La descripción de la producción de estas construcciones es muy semejante a la anterior, a la de las pronominales. Los HN utilizan las construcciones preposicionales con una frecuencia baja, pero más alta que en los HNN, y con mayor variedad en las construcciones empleadas. Así, los HN utilizan la construcción *dar por* + adjetivo o participio, con el significado de “considerar cierta cosa decididamente como ese adjetivo expresa” ([1], [2], [3], [6] y [12]); *dar por* + infinitivo, con el significado de “iniciar una acción inesperada” ([4], [8] y [9]); *dar contra, en o a* + SN ([5] y [10]), con el sentido de “chocar o impactar”; y *dar con* + SN con el significado de “encontrar a alguien o algo” ([7] y [11]). A continuación se incluyen todos los usos de las construcciones preposicionales en los HN para establecer más abajo la comparación con los usos no nativos:

- [1] Durante su estancia allí, tiene un accidente y se da por muerta (3 HN)
- [2] Este absurdo que consiste en dar por hecho lo que no está hecho, en dar por sucedido lo que no ha sucedido (23 HN)
- [3] Este absurdo que consiste en dar por hecho lo que no está hecho, en dar por sucedido lo que no ha sucedido (24 HN)
- [4] Estábamos aterrizando, que se veían las casas desde el avión, y me dio por reír, no me podía creer que de verdad estaba en la India (180 HN)
- [5] Luego sintió que caía y con un golpe seco su cuerpo fue a dar contra el suelo polvoriento (221 HN)
- [6] nunca me he dado por vencida (227 HN)
- [7] por fin, conseguimos dar con la fecha en la que a los tres nos venía bien (241 HN)
- [8] Menos mal que no les ha dado por construir en las mejilloneras (252 HN)
- [9] A lo mejor me da por reconocer en mi psicopatías extremas (264 HN)
- [10] En Madrid la gente no sabe ni por dónde le da el aire (267 HN)
- [11] Al principio resultó un poco complicado pero dimos con él (299 HN)
- [12] Esta le dice que la había visto por el acantilado y la daba por muerta (302 HN)

Los HNN solo usan una de las 4 anteriores construcciones preposicionales, y en un número muy bajo: *dar contra, en o a* + SN [2], con el sentido de chocar o impactar, y añaden una nueva a la lista que no aparece en el corpus nativo [1]:

- [1] el sexto piso al pie del ascensor tenía una tremenda ventana que daba justo al Guggenheim (122 HNN)
- [2] Al africano casi le da en el ojo con el paraguas (198 HNN)

De nuevo, estos dos ejemplos pertenecen también a redacciones que reflejan un nivel de dominio muy alto; se está haciendo uso de estructuras que los HN utilizan con poca frecuencia, pero el hecho de que solo se utilicen dos, y que ninguna de ellas sea de las más repetidas en el corpus nativo, sugiere que el español productivo de este grupo en lo que se refiere a las construcciones preposicionales con *dar* no es muy amplio. Por otro lado, el hecho de que los HN empleen en un porcentaje bajo cada una de las formas de este tipo de construcción (aunque la gama sea relativamente amplia), sugiere que los HNN no encontrarán muchas situaciones para producirlas o que no tendrán gran necesidad de utilizarlas. Precisamente por eso, es probable que gran parte de los aprendientes no conozca dichas formas (*dar por* + infinitivo; *dar con* + SN; *dar por* + adjetivo).

f) Esquemas fraseológicos semiproductivos

	HNN		HN	
	FN	%	FN	%
Esquema fraseológico semiproductivo	2	0,9	8	2,2

Tabla 23. Frecuencia de *dar* en los esquemas fraseológicos semiproductivos

Los esquemas fraseológicos semiproductivos son combinaciones de un verbo en forma personal y un verbo en forma no personal en las que este último está restringido léxicamente, es decir, limitado a unos pocos verbos de determinada clase semántica (NGLE. *Manual*, 2010: 534-535).

Como sucedía en las anteriores estructuras, en los esquemas fraseológicos nos encontramos con la misma situación en la relación de frecuencias: mientras que los HN las utilizan con una frecuencia baja, los HNN las usan todavía menos y en menor variedad, por lo que parece que su conocimiento productivo de estas construcciones tampoco es muy abundante.

A continuación incluimos todos los usos nativos de esquemas fraseológicos semiproductivos:

[1] quien la moldea, quien con sus hechos y dichos la da a conocer a los demás (33 HN)

[2] las palomas, presentes en cualquier ciudad con personas piadosas que las [sic] dan de comer, no es raro ver golondrinas o vencejos en plena ciudad (45 HN)

[3] y por lo tanto dejan paso libre a los fumadores, lo que nos da a entender que, nos guste o no, vivimos ante una sociedad (46 HN)

- [4] por otro, la trayectoria vital de una de las mujeres que más ha dado que hablar en la España que va de los 60 a los 80: Isabel Preysler (97 HN)
- [5] Mi futuro profesional también me da mucho que pensar, ya que soy trabajadora autónoma (247 HN)
- [6] a contrincantes de categorías superiores a la suya aunque cuando se dio a conocer para el gran público español y por extensión, extranjero (265 HN)
- [7] la productora no ha tenido suficiente dinero para darla a conocer (342 HN)
- [8] un día habían soñado como perfecto, capaz de darles de comer, de darles la felicidad que tanto añoran (347 HN)

Los HN usan tres tipos de esquemas fraseológicos: *dar a* + infinitivo con los sentidos de “ejecutar acciones que denotan conocimiento o difusión de informaciones” ([1], [6] y [7]), “ejecutar acciones que denotan percepción e intelección” [3] y “ejecutar acciones que denotan elección (como en *dar a elegir* o *a escoger*) o ingestión (*dar a comer* o *a probar*)”; *dar de* + infinitivo, en el sentido de “ejecutar acciones que denotan ingestión” ([2], [8]); y *dar que* + infinitivo con el sentido de “provocar o dar motivo para que se haga una acción” (lo que el infinitivo expresa) ([4], [5]). Los infinitivos que se combinan con *dar* en este esquema no se agrupan fácilmente en una o varias clases semánticas, como los anteriores. En *Redes* este esquema aparece completado con *comentar*, *hablar* y *pensar*.

Los HNN emplean tan solo el esquema *dar a* + infinitivo:

- [1] Cuando yo era niño, tenía que dar a comer a los caballos cada mañana y cada noche (55 HNN)
- [2] Pero primero debimos dar a comer a las niñas, y después buscar sus juegos de playa (213 HNN)

Llama la atención que los HNN utilizan justamente este esquema de *dar a* + infinitivo con la única clase semántica que no se usa en el corpus nativo, la de los infinitivos de verbos que denotan “ingestión”; pese a ser frecuente *dar a*, lo es más *dar de* (véase *Redes* [2005: 642] al respecto y recuérdense los resultados de nuestra búsqueda en el CREA y en el Corpus del español). En el segundo caso, la participante tiene un nivel avanzado de francés, por lo que podría darse una interferencia de esa lengua, en la que es frecuente la construcción sustantivo + a + infinitivo; aunque se trata de una construcción diferente, existen claras semejanzas formales. En el primer caso, el participante ha pasado una temporada larga en Chile, y, de acuerdo con las búsquedas de corpus realizadas, el esquema con *a* es más frecuente en Sudamérica que en la Península; de hecho, las dos

apariciones de *dar de + comer* en el corpus nativo son producciones de nativos de España.

Así, de este análisis se deduce que los HNN posiblemente desconocen estas estructuras –pues los HNN no usan estos esquemas tanto como lo hacen los HN (ni en cuanto a la variedad de estructuras ni en cuanto a la frecuencia)–, ya que las posibilidades de adquirirlas a partir del input y contrastar sus hipótesis de uso a través del *output* son probablemente mucho menores que las estructuras más frecuentes.

7.1.2.3.3. No frecuencia en HNN *vs.* baja frecuencia en HN

Dos son las estructuras que responden a este esquema, a saber: la construcción introducida por *como* y la del uso del adjetivo o participio:

	HNN FN	%	HN FN	%
Construcción introducida por <i>como</i>	0	0	3	0,8
Adjetivo	0	0	3	0,8

Tabla 24. Frecuencia de *dar* en la construcción introducida por *como* y en la construcción con adjetivo

Tres son los usos nativos de *dar + construcción introducida por como*, con el sentido de “clasificar, asociar, equiparar algo” ([1], [2] y [3]), y 3 son los casos del uso del adjetivo *dado*, que proviene del participio latino *datus* ([4], [5], y [6]):

- [1] Concretando los principales aspectos de la inmigración, podríamos dar como factores clave (138 HN)
- [2] el siglos XX es un siglo de guerras y de crisis de todo tipo que dio como resultado la creación de grandes brechas o abismos insuperables (194 HN)
- [3] abismos insuperables entre los pueblos ricos y los pobres, dando como resultado la dependencia de uno sobre otro, los países ricos (195 HN)
- [4] la película me ha sorprendido muy gratamente aun cuando no he sido muy dada a este tipo de películas (66 HN)
- [5] en un momento dado yo solicité a una persona nativa del lugar que me hiciera un masaje relajante (311 HN)
- [6] [el movimiento migratorio] también viene dado por razones políticas, etc (244 HN)

El primero de los adjetivos se emplea con el sentido de “inclinado”; el segundo, con el sentido de concreto; y el último está precedido del verbo *venir*, que, cuando se construye como semicopulativo, acepta algunos atributos, como *dado*, para expresar las nociones de causa o vínculo necesario (NGLÉ, 2009: 2863).

Estas dos son las estructuras que aparecen con menos frecuencia en el corpus nativo, por lo que no es de extrañar que no haya ninguna ocurrencia en el no nativo. Este porcentaje tan bajo remite, como en casos anteriores, a la posibilidad de que los HNN encuentren escasas ocasiones para producir estos usos, y resulta probable que no los conozcan.

7.1.2.3.4. Síntesis del análisis de las relaciones de frecuencias

Hasta aquí la interpretación de las relaciones de frecuencia en las estructuras sintácticas entre el corpus nativo y el no nativo. No hay ninguna estructura que encaje en las categorías *alta frecuencia vs. baja frecuencia* y *alta frecuencia vs. no existencia*, lo que parece indicar que el español de los HNN del CEDEL2 no es drásticamente diferente del de los HN.

En síntesis, el análisis de las relaciones de frecuencias entre los datos de ambos grupos ha revelado varias cuestiones. En primer lugar, las estructuras que son utilizadas con frecuencia entre los HN también son usadas con frecuencia por los HNN; y aquellas estructuras usadas con poca frecuencia por los HN, son también usadas con poca frecuencia por los HNN. A simple vista, parece que el análisis de frecuencias indica que el uso de *dar* en los dos corpus es similar. Los esquemas de relación de frecuencias utilizados y el test de significatividad Chi-cuadrado permiten profundizar en los índices de frecuencias, cuya aplicación a nuestros datos pone de manifiesto que dicha semejanza no es tal.

Las únicas estructuras que sobreutilizan los HNN son las de V + N; son estructuras que, desde una perspectiva formal, no son complejas y que sintácticamente son equivalentes en la L1. El resto de las estructuras las utilizan con menor frecuencia que los HN, al mismo tiempo que usan un menor número de estructuras. Estos resultados no resultan sorprendentes, dado que el resto de las estructuras son menos frecuentes en el español general –lo que provoca que su adquisición sea más lenta y costosa– y son sintácticamente más complejas.

Las estructuras ditransitivas, transitivas, las locuciones y las impersonales se usan con frecuencia en ambos corpus. El uso de las transitivas es similar en ambos grupos; en cambio, las ditransitivas se sobreutilizan, lo que indica que se

desconocen otros recursos expresivos más adecuados en ciertos contextos, como sucede en aquellos casos en los que se debería o podría emplear un solo verbo en lugar del uso del verbo *dar* + sustantivo deverbal. En cuanto a las impersonales, los HNN usan más el infinitivo, posiblemente por la influencia de la L1 y porque es un uso muy neutro, adecuado, por lo tanto, al tipo de texto analizado. Usan menos que los nativos la 3.^a persona del plural y la 2.^a del singular, lo que, a nuestro juicio, es apropiado, dado que estos usos resultan demasiado informales o coloquiales; en cambio, en su lugar utilizan más la 1.^a persona del plural, lo que demuestra que se involucran más personalmente en el texto. Según lo argumentado, esta sobreutilización y la infrautilización de la impersonal con *se* – más adecuada al tipo de texto examinado– conforman un área de desviaciones en este estadio de la interlengua.

En lo referente a las locuciones, parece que el español productivo de los HNN tiene menor alcance: estos emplean menos tipos de locuciones y las repiten más, por lo que esto constituye otro foco de desviaciones. El uso de *darse cuenta* es muy similar.

Las intransitivas, las pasivas, las construcciones absolutas, las pronominales, las preposicionales y los esquemas fraseológicos semiproductivos son empleados por los dos grupos en una frecuencia baja. En las intransitivas hay divergencias a pesar de que la frecuencia de uso es muy similar a simple vista. Mientras que los HN solo usan intransitivas con el verbo de apoyo, esto es, con “emociones y estados anímicos”, los HN lo utilizan también (en un 43 %) con el verbo en sentido pleno, sobre todo con la expresión *dar tiempo*, que se repite en numerosas ocasiones; así, es probable que los HNN no sean conscientes de este uso. En este segundo uso intransitivo se localiza otra fuente de desviaciones para este grupo en este estadio de la interlengua.

En relación con las pasivas, hemos visto que, posiblemente por influencia de la L1, se utiliza más la estructura con el verbo *ser* que el *se* pasivo, todo lo contrario de lo que sucede con los HN. La pasiva con *se* es compleja, por lo que previsiblemente puede ocasionar desviaciones en la adquisición del español. Lo mismo sucede con la construcción de participio absoluto, muy natural en español pero menos frecuente que otras estructuras, cuyo uso es probablemente poco conocido por los HNN, según se desprende de los errores observados. El español productivo del HNN en las pronominales, en las construcciones preposicionales y en los esquemas fraseológicos no parece ser muy amplio, pues hay poca variedad de usos y son menos frecuentes, por lo que también se presenta como un foco de desviaciones en este estadio de la interlengua. La construcción con *como* y el adjetivo no aparecen en el corpus no nativo, y aunque

no son tan frecuentes como otras estructuras, son naturales y deben conocerse en niveles avanzados.

Para concluir, algunos de los datos de frecuencia obtenidos en el análisis son demasiado bajos para hablar de significatividad estadística; no permiten afirmar con certeza cuál es el conocimiento de una determinada estructura por parte de toda la comunidad de aprendientes. Sería un error afirmar que una ocurrencia en un corpus de aprendientes indica dominio en todo el grupo de HNN. Cuando se intenta establecer hasta qué punto los HNN dominan un uso o hasta qué punto no son conscientes de este, hablamos de probabilidad, no de certeza. No obstante, estos datos sí que nos proporcionan material para poner en marcha futuras investigaciones; los resultados obtenidos en este corpus pueden ser contrastados en corpus más extensos, y deben servir de punto de partida para realizar otros estudios experimentales que nos permitan triangular los resultados obtenidos (estas ideas son desarrolladas en el capítulo 8).

7.1.3. Palabras y tipos de palabras con los que se combina dar

Ya hemos señalado (apartado 7.1.2.1) que el verbo *dar* ocupa una de las primeras posiciones en la relación de los verbos más frecuentes del español por la asiduidad con la que se utiliza la construcción V + N, que se corresponde con las estructuras en las que el verbo actualiza los sentidos de “proporcionar” y “producir”; en ellas, el nombre es su argumento interno y tema; el tema es el CD en las estructuras transitivas, ditransitivas e impersonales y es el sujeto en las pasivas y en las intransitivas.

Para poder construir correctamente estas estructuras, no basta con saber que el verbo exige un CD o un sujeto con el que debe concordar, sino que también se ha de conocer el contenido semántico que desempeña la función semántica *tema*, sea el CD o sea el sujeto. Dado que este verbo es el que tiene mayores posibilidades combinatorias (Lozano Zahonero, 2011: 89), el nombre que lo acompaña puede pertenecer a muchas clases léxicas; ya vimos en el apartado 3.2 que los predicados seleccionan sus argumentos restringiendo el conjunto de piezas léxicas que pueden denotar de acuerdo con sus rasgos semánticos, por lo que las unidades seleccionadas no son piezas aisladas –como en el caso de las combinaciones asistemáticas (recuérdese el ejemplo de *hacer una llave*)–, sino nociones muy diversas que pueden ser denotadas por grupos de sustantivos que son seleccionados por el verbo. Así, atendemos a estas restricciones no como combinaciones en las que la relación entre los elementos

es biunívoca, sino como casos de selección léxica y de concordancia de rasgos. Si la restricción es extralingüística –está fundamentada únicamente en nociones de naturaleza física o corpórea–, el verbo está actualizando la acepción básica. Por el contrario, si la restricción no está basada únicamente en nociones físicas, se tiene que recurrir al conocimiento del idioma para obtener las piezas léxicas que conforman la extensión, de ahí que sean denominadas *restricciones sistemáticas intralingüísticas*. A diferencia de las anteriores, la restricción viene dada en función de las propiedades del predicado, de su estructura interna. No nos basta con nuestro conocimiento del mundo para usar el verbo, y necesitamos un conocimiento profundo de este para obtener el paradigma extensional.

Este conocimiento de las restricciones extralingüísticas e intralingüísticas es el que hemos denominado *léxico-sintáctico*, más concretamente *conocimiento léxico-combinatorio* (véase la tabla 1 en el capítulo 3). En el presente apartado describiremos la extensión de este conocimiento léxico-combinatorio en el grupo de HNN atendiendo a qué clases léxicas emplean, con qué frecuencia y cuáles son las semejanzas y disparidades entre el español de los HNN y de los HN en relación con la combinatoria léxica de *dar*.

Ya hemos señalado que los datos revelan un uso no nativo significativamente mayor ($p < 0,05$) que el nativo de la construcción V + N (75,8 %* de los HNN frente al 62,2 % de los HN). De los 224 usos nativos, 3 no se analizaron porque contenían un error en el sintagma que impedía su interpretación (se analizan en el apartado 7.2). De los 221 restantes, 123 (55,7 %) se corresponden con el uso pleno del verbo, con el sentido de “proporcionar”, y 98 con el uso del verbo como verbo de apoyo, esto es, con el sentido de “producir(se)” (44,3 %). Por otro lado, de los 166 usos no nativos de V + N, 9 no pudieron clasificarse porque contenían errores que impedían integrarlos en una de las clases léxicas. De los 157 restantes, el verbo se emplea como “proporcionar” en 93 ocasiones (59,2 %) y como “producir(se)”, en 64 (40,8 %). Así, *dar* se utiliza más en ambos corpus –en torno a un 10 % más– con el sentido de “proporcionar” que con el sentido de “producir(se)”. Los porcentajes, además, indican que ambos grupos distribuyen los usos de *dar* como “proporcionar” y como “producir(se)” en frecuencias semejantes (la diferencia es menor al 5 %). Pese a esta semejanza, los HNN usan esta construcción V + N en un 14 % más que los HN (como se ha visto en el apartado 7.1.2.1) , diferencia que es estadísticamente significativa ($p < 0,05$); partiendo de estas cifras, se espera que existan coincidencias en ciertos usos pero también diferencias que expliquen la sobreutilización.

Para poder comparar las clases léxicas y los sustantivos con los que coaparece el verbo, se han creado dos tablas (tabla 24 y tabla 26) que incluyen todas las clases léxicas empleadas y los sustantivos utilizados de cada clase: una para los usos de *dar* con el sentido de “proporcionar” y otra para los usos de *dar* con el sentido de “producir(se)”; con ellas se pretende que las semejanzas y diferencias se revelen con claridad. En cada tabla hay tres columnas para cada corpus. Las tres de la izquierda están reservadas para los HN y las tres de la derecha, para los HNN. La primera de las tres columnas –comenzando por la izquierda– corresponde a los NS (números de serie), que indican la posición que ocupa cada clase léxica en la serie según su frecuencia, de mayor frecuencia a menor frecuencia. Junto con el NS, se incluye el porcentaje de frecuencia de uso de la clase léxica. En la segunda columna se coloca la clase léxica (en el capítulo 4, en la tabla 2, pueden verse todas) y en la tercera, los sustantivos empleados de cada clase. Las clases léxicas utilizadas por los HNN y por los HN se colocan en el mismo nivel para facilitar la comparación, precedida del NS correspondiente no nativo. Así, resulta también más fácil ver qué clases léxicas y qué sustantivos no están presentes en el corpus de HNN y cuáles no lo están en el de HN.

7.1.3.1. El uso de *dar* como verbo de transferencia

7.1.3.1.1. Análisis cuantitativo

He aquí la primera de las dos tablas mencionadas, que recoge los datos correspondientes a los usos de *dar* con el sentido de “proporcionar”.

HN			HNN		
NS	Clase léxica ¹⁸⁵	Sustantivos	NS	Clase léxica	Sustantivos
1 (14,6 %)	Asistencia, sustento o protección	Apoyo (2), consejo, beca (4), protección, habitación (2), atención médica, abrigo (2), sueldo, bungalow, amparo, alimentación, salud	7 (2,1 %)	Ídem	Hogar, consejos
2 (12,2 %)	Entidades materiales	Billete, visa, dinero (5), comida (2), vitaminas, champú, raciones, cena, maleta, boleto	2 (17,2 %)	Ídem	Billete, tenedores (2), obras, mate, bebidas, comida (2), maíz, peluche, regalos (3), medicinas, bocadillos, cosa
3 (9,8 %)	Propiedades de las cosas tenidas por fundamentales: contenido, denominación	Forma (2), vida (5), nombre (3), rumbo (2)	3 (7,5 %)	Ídem	Forma (2), vida (4), nombre, apellido
4 (8,1 %)	Valores y cualidades de las personas o cosas (presencia, duración, comportamiento)	Atractivo, educación (4), color, realismo, autenticidad, prestigio, politización	8 (1,1 %)	Ídem	Sabor
5 (7,3 %)	Inclinaciones y sensaciones afectivas que se transmiten	Amor (6), cariño (2), ternura	6 (3,2 %)	Ídem	Amor (3)
5 (7,3 %)	Autorización, aceptación, ausencia de trabas	El alta (2), derecho (5), luz verde, paso	3 (7,5 %)	Ídem	Derecho (3), permiso (3), paso
6 (5,7 %)	Forma en que algo se manifiesta o aspecto que adquiere	Imagen (2), aspecto (2), característica, dejo, toque	7 (2,1 %)	Ídem	Carácter, toque
7 (4,9 %)	Periodos o intervalos	Tiempo (6)	8 (1,1 %)	Ídem	Años
7 (4,9 %)	Información	Chivatazos, argumentos, opinión, declaraciones (2), diagnóstico	4 (6,4 %)	Ídem	Detalles, n.º de teléfono, mensaje, explicaciones, referencia, informaciones
7 (4,9 %)	Cuantificadores	Más, todo (5)	5 (4,3 %)	Ídem	El 120 %, mucho, algo, tanto
8 (3,2 %)	Ocasión, oportunidad o situación favorable	Oportunidad(es) (3), facilidades	1 (22,6 %)	Ídem	Oportunidad(es) (19), opción, ventaja
8 (3,2 %)	Lo que ilustra algo o da razón de su existencia	Ejemplo (3), ideas	3 (7,5 %)	Ídem	Idea (3), perspectiva (2), muestra, ejemplos
9 (2,4 %)	Relevancia, prominencia y otras propiedades	Importancia (3)	8 (1,1 %)	Ídem	Importancia

¹⁸⁵ En la tabla 2 del capítulo 4 se ofrece una definición más completa de cada clase léxica. Aquí aparecen enunciadas de manera más esquemática por razones de espacio.

	relativas al papel que desempeñan las cosas o personas				
9 (2,4 %)	Garantía o confianza; estabilidad, firmeza, ausencia de perturbación	Seguridad (2), paz	6 (3,2 %)	Ídem	Crédito, confianza (2)
10 (1,6 %)	Estímulo, ánimo o fuerza	Esperanza, aliento	6 (3,2 %)	Ídem	Poder (2), coraje
10 (1,6 %)	Capacidad de actuar, difusión, expansión	Libertad (2)	8 (1,1 %)	Ídem	Libertad
10 (1,6 %)	Resultado de algo si es beneficioso. Forma en que se valora algo	Calificaciones, puntaje	7 (2,1 %)	Ídem	Nota, estrellas
10 (1,6 %)	Tiempo, actividad u objetos susceptibles de ser asignados	Hora (2)			
11 (0,8 %)	Instrucción o aviso	El alto			
11 (0,8 %)	Adjetivos sustantivados que denotan propiedades positivas o negativas	Lo bueno			
			7 (2,1 %)	Reconocimiento	Título, regalo
			8 (1,1 %)	Espacio o ubicación	Posición

Tabla 25. Clases léxicas y sustantivos con los que se combina *dar* con el sentido de “proporcionar” (en CVP) en el CEDEL2

Los HN emplean 20 clases léxicas y los HNN, 19, pero no coinciden exactamente las mismas clases. Como se observa en la tabla, los HNN no utilizan la clase de sustantivos que denotan “instrucción o aviso”, la de los adjetivos sustantivados que denotan “propiedades positivas y negativas” ni la de los sustantivos de “tiempo, actividad u objetos susceptibles de ser asignados”. Estas son las clases empleadas con menos frecuencia entre los HN (menos de un 2 % cada una), por lo que esta diferencia no hace que ambos corpus difieran drásticamente. Por otro lado, hay dos clases que los HN no utilizan: la de los sustantivos que denotan “espacio o ubicación” y la de los sustantivos que aluden a un “reconocimiento por una labor u otra circunstancia”; estas son, de igual modo, las menos usadas por los HNN por lo que no supone una brecha en el uso de *dar* en ambos grupos.

Antes de profundizar en el uso que cada grupo hace de las clases léxicas – más concretamente, en los sustantivos que utiliza cada grupo –, observamos a simple vista que la mayoría de las clases son utilizadas por ambos grupos con una frecuencia semejante (diferencia menor del 5 %), y que hay dos que utilizan más los HN y otras dos que utilizan más los HNN. En esta tabla presentamos las

clases léxicas que se usan con una frecuencia diferente en un 5 % o más. El porcentaje entre paréntesis se refiere a la diferencia que se da entre los porcentajes de ambos grupos, y el asterisco representa la significatividad estadística ($p < 0,05$):

+ Frecuencia HN	+ Frecuencia HNN
Asistencia, sustento o protección (12,5 %)*	Ocasión, oportunidad o situación favorable (19,4 %)*
Valores y cualidades de las personas o cosas (presencia, duración, comportamiento) (7 %)*	Entidades materiales (5 %)

Tabla 26. Las clases léxicas con las que se combina *dar* con el sentido de “proporcionar” (en las CVP) en las que difieren los porcentajes de uso nativo y no nativo en más de un 5 %

Esta tabla ilustra la razón de parte de esa diferencia que existe en la frecuencia de uso de las estructuras V + N en ambos grupos, que, como ya se ha señalado, es de un 14 % (los HNN las usan más). Llama la atención la alta frecuencia con la que los HNN utilizan la combinación *dar una oportunidad* frente al bajo porcentaje de uso de los HN (19 usos frente a los 3 usos nativos). Dada la baja frecuencia de este uso en el corpus nativo, no podemos estar seguros acerca de la significatividad de su alta presencia en el corpus no nativo. Cuando las concordancias se examinan, parece que la mayoría de ellas son aceptables. Así, parece que esta coaparición disfruta de una popularidad y homogeneidad entre este grupo de HNN.

7.1.3.1.2. Análisis cualitativo

Cuando la selección de una clase léxica o de una palabra no existe en el corpus nativo o se utiliza con muy poca frecuencia (por debajo del 2 %), menos seguridad tenemos sobre la importancia de la ausencia o presencia en el corpus de HNN. Lo que sugieren estos datos es que hay algunas clases léxicas –en concreto algunas combinaciones– que posiblemente no han sido adquiridas, pues el hecho de que sean utilizadas con poca frecuencia entre los HN nos lleva a pensar que los HNN no tienen muchas oportunidades o mucha necesidad de escribirlas, por lo que la mayoría de ellos no será muy consciente de su uso. Combinaciones como *dar una educación*, *dar una imagen* o *un aspecto*, *dar seguridad*, *dar cariño*, *dar el alta*, *dar rumbo*, *dar una beca*, *dar apoyo* y *dar abrigo* son naturales aunque aparecen con menos frecuencia en comparación con otras (como *dar vida*, *dar*

tiempo, dar amor o darlo todo); no obstante, no son utilizadas por los HNN. El hecho de que no aparezcan con frecuencia en el corpus nativo del CEDEL2 no significa que sean usos extraños o no naturales; en un corpus más extenso (CREA y Corpus del español) comprobamos que sí son usadas, aunque con menos frecuencia que otras combinaciones, por lo que requieren más atención para comprobar si hay desviaciones en su uso. Del hecho de que los HN las usen con menos frecuencia se deduce que los HNN tienen pocas oportunidades de contrastar su producción y confirmar sus hipótesis sobre estos usos. Esta reflexión la aplicamos también al caso de las dos clases léxicas usadas solo por los HNN.

Los sustantivos diferentes identificados en ambos corpus suman un total de 107; solo una proporción pequeña de estos (15) son compartidos por los dos grupos: *consejo, billete, comida, forma, vida, nombre, amor, derecho, paso, toque, oportunidad, idea, ejemplos, importancia, libertad*. HN y HNN solo coinciden en el 14 % de los usos. Los HN usan 68 sustantivos diferentes y los HNN, 54; estos datos absolutos convertidos en porcentajes resultan en unas cifras muy equivalentes: el 55,2 % en los HN y el 58,1 % en los HNN de las combinaciones de *dar* son diferentes en cada corpus. El porcentaje restante representa usos repetidos de una misma combinación, por lo que el uso productivo del verbo *dar* en lo que se refiere a los sustantivos con los que se combina en la construcción V + N cuando el verbo actualiza el significado de “proporcionar” es prácticamente igual de amplio, aunque los sustantivos con los que cada grupo combina *dar* no coincide en la mayoría de casos (86 %), y pese a que se dan diferencias significativas en el porcentaje de uso de algunas clases léxicas o en el de algunas combinaciones léxicas –como en el sobreuso de *dar la oportunidad* que podría ser evitado si los HNN fueran conscientes de él y de algunas combinaciones alternativas (las diferencias en el uso de *dar tiempo*, con seis coapariciones en el corpus nativo frente a ninguna aparición en el no nativo, también son destacables)–.

7.1.3.2. El uso de *dar* como verbo de creación

7.1.3.2.1. Análisis cuantitativo

A continuación analizamos los usos de *dar* cuando funciona como verbo de apoyo, esto es, con el sentido de “producir(se)”, sintetizados en esta tabla.

NATIVO			NO NATIVO		
NS	Clase léxica	Sustantivos	NS	Clase léxica	Sustantivos
1 (28,4 %)	Desplazamiento, frecuentemente breve y repentino	Vuelta(s) (9), paso(s) (11), paseo (3), salto (2), giro (2)	1 (20,6 %)	Ídem	Vuelta (3), paseo (9), volteleta
2 (17,9 %)	Eventos que se interpretan como unidades informativas y que se ofrecen a la contemplación pública	Clase(s) (10), tour turístico, cátedra, conferencia, discurso, concierto (2), lecciones de piano	3 (14,3 %)	Ídem	Clase(s) (4), fiesta, curso, ponencia, conferencia, charla
3 (12,6 %)	Movimientos impulsivos o resultantes de algún contacto.	Vuelta de tuerca, patada, golpe(s)-golpecillos (4), paliza(s) (2), trompazos, beso, empujoncito, caricia	5 (6,3 %)	Ídem	Besito(s) (2), palmas, golpecitos
4 (7,4 %)	Estados de aflicción y sensaciones de infelicidad, enojo o turbación	Lástima, asco, rabia, pena (3), pereza	4 (7,9 %)	Ídem	Pena (2), asco (2), coraje
5 (6,3 %)	Formas de trato social	Bienvenida, gracias (5)	5 (6,3 %)	Ídem	Bienvenida (3), gracias
5 (6,3 %)	Miedo, inquietud, zozobra y estados anímicos similares	Miedo (2), reparo, pánico, cosita, sorpresas	2 (19 %)	Ídem	Vergüenza, miedo (9), sorpresa
6 (4,2 %)	Sensación e impresión y reacciones físicas adversas y repentinas	Sensación (3), impresión	7 (3,2 %)	Ídem	Sensación, ataque de nervios
6 (4,2 %)	Sensaciones y emociones placenteras	Gusto (2), felicidad (2)	3 (14,3 %)	Ídem	Gusto (3), placer (2), alegría (2), risa, satisfacción
7 (2,1 %)	Término, desenlace y otras nociones relativas al fin de las acciones o procesos	Caza, respuesta			
7 (2,1 %)	Sustantivos derivados de verbos	Conversación, cita			
7 (2,1 %)	Deseo o necesidad	Ganas (2)	6 (4,8 %)	Ídem	Ganas (2), hambre
8 (1 %)	Comienzo y final	Origen			
8 (1 %)	Molestia	Lata			
8 (1 %)	Dificultad	Problema			
8 (1 %)	Mejora, limpieza o aseo	Limpiadita			
8 (1 %)	Horas	Doce			
8 (1 %)	Fenómenos atmosféricos o relacionados con el clima	Temperatura	8 (1,5 %)	Ídem	Sombra
			8 (1,5 %)	Sonidos	Respiración

Tabla 27. Clases léxicas y sustantivos con los que se combina *dar* con el sentido de “producir(se)” (en CVA) en el CEDEL2

Ya hemos señalado que los usos de *dar* como verbo de apoyo son 98 en el corpus nativo y 64 en el corpus no nativo. Basándonos en la organización combinatoria de *dar* en *Redes* (2005: 639-641 [2004]), hemos distinguido dos usos del verbo de apoyo muy próximos: cuando se usa como *producir* en el sentido de “ejecutar”, y cuando se utiliza como *producir* en el sentido de “suscitar(se) en alguien”. Así, el primer uso se corresponde con el de *Le dieron una patada*, y el segundo con el de *Le dieron miedo*.

Con el sentido de “ejecutar”, *dar* se combina con sustantivos que designan “movimientos resultantes de algún contacto” (*abrazo, empujón, toque*), sustantivos que denotan “comienzo y final”, sustantivos que denotan “desenlace y nociones relativas al fin de las acciones o los procesos” (*respuesta, solución, cumplimiento*), sustantivos que denotan “enfrentamiento” (*batalla, pelea*); los que denotan “molestia” (*lata, tostón, tabarra*); los que se interpretan como “unidades informativas y por extensión los que expresan eventos que se ofrecen a la contemplación pública” (*clase, lección, concierto, rueda de prensa*); los sustantivos que designan “formas de trato social y otros que se refieren a los sacramentos y a diversos signos que forman parte de ceremonias o actos religiosos” (*buenos días, bienvenida, pésame*); los sustantivos que denotan “desplazamiento breve” (*paseo, salto, paso*); y, por último, los que designan “sonidos” (*grito, soplido*).

Con el sentido de “suscitar(se) en alguien”, *dar* se combina con los sustantivos *sensación* e *impresión*, así como con varios otros que designan “alteraciones y reacciones físicas adversas, repentinas” (*calambre, mareo, retortijón*); sustantivos que denotan “deseo o necesidad” (*gana, hambre, envidia*); los que designan “sensaciones y emociones percibidos como placenteros” (*alegría, gusto, felicidad*); los que denotan “miedo, zozobra y estados anímicos similares” (*miedo, susto, terror, apuro*); y los que designan “otros estados de aflicción relacionados con la infelicidad, el enojo o el rechazo” (*asco, tristeza, pena*)¹⁸⁶.

Partiendo de esta información, los HN usan *dar* en las CVA con el sentido de “ejecutar” tres veces más que con el sentido de “suscitar(se)”, lo cual supone el 76 % frente al 24 %. En cambio, los HNN usan los dos sentidos del verbo prácticamente por igual, un 50,8 %* de las ocasiones con el de “ejecutar” y un 49,2%* con el de “suscitar(se)”. Esta diferencia entre el uso nativo y no nativo es estadísticamente significativa ($p < 0,05$).

¹⁸⁶ En la tabla 2, en el capítulo 4, se presenta el listado de todas estas clases léxicas aquí enumeradas, pero en ella no se marca esta distinción dentro del sentido de *dar* en las CVA; en este apartado esta clasificación sí es relevante.

Al igual que hemos hecho con el uso de *dar* en el sentido de “proporcionar”, antes de profundizar en el uso que cada grupo hace de las clases léxicas, se observa en la anterior tabla que hay 7 clases léxicas no usadas por los HNN. Estas son las clases empleadas con menos frecuencia por los HN (por debajo de 2,1 %), por lo que esta diferencia tampoco hace que ambos corpus difieran drásticamente, pues podría tratarse de usos individuales más que representativos del grupo. Del mismo modo que los HNN no utilizan todas las clases léxicas que emplean los HN, estos no utilizan una clase que sí que emplean los HNN, la de los sustantivos que designan “sonidos”; esta es, de igual modo, la clase menos usada por los HNN –con tan solo una aparición en el corpus– por lo que tampoco supone una gran distancia en el uso de *dar* en ambos grupos.

Por otro lado, la mayoría de las clases léxicas más utilizadas por ambos grupos son empleadas con una frecuencia semejante (diferencia menor del 5 %); pero de nuevo hay dos que utilizan más los HN y otras dos que utilizan más los HNN. En esta tabla presentamos las clases léxicas que se usan con una frecuencia diferente a un 5 %. El porcentaje entre paréntesis se refiere a la diferencia que se da entre los porcentajes de ambos grupos. La diferencia es únicamente significativa desde una perspectiva estadística ($p < 0,05$) en la sobreutilización de los HNN:

+ Frecuencia HN	+ Frecuencia HNN
Desplazamiento, frecuentemente breve y repentino (7,8 %)	Miedo, inquietud, zozobra y estados anímicos similares (12,7 %)*
Movimientos impulsivos o resultantes de algún contacto (6,3 %)	Sensaciones y emociones placenteras (10,1 %)*

Tabla 28. Las clases léxicas con las que se combina *dar* con el sentido de “producir(se)” (CVA) en las que difieren los porcentajes de uso nativo y no nativo en más de un 5%.

7.1.3.2.2. Análisis cualitativo

Por lo que se refiere a la baja frecuencia de uso –o a la no utilización– de algunas clases léxicas por parte de las HN, como ya hemos mencionado a raíz de los casos relativos al uso de *dar* como “proporcionar”, lo que sugieren estos datos es que, dada la baja frecuencia de estas combinaciones, es posible que la mayoría de los HNN no conozca estos usos del verbo con los sustantivos que designan “horas”, “comienzo o final”, “molestia”, “dificultad” o “sonidos”. Si esta combinatoria no es usada apenas por los HN (ya que no se documenta en

nuestro corpus), posiblemente tampoco sean conscientes de este uso la mayoría de los HNN.

Por lo que respecta a la mencionada sobreutilización del sentido de “suscitarse” (49,2 % de los usos no nativos de *dar* en las CVA frente al 24% de los usos nativos) y a la consiguiente infrautilización del de “ejecutar”, ya señalamos –a propósito del sobreuso de la impersonal de 1.^a persona del plural– que los HNN, partiendo de géneros textuales equivalentes, muestran más implicación personal que los HN (Cobb, 2003), lo cual explicaría esta sobreutilización de *dar* con el sentido de “suscitar(se)”, que se traduce en frecuentes referencias al estado mental, físico y anímico del autor de la redacción, como se refleja en la tabla 27. Esta, unida a la tabla equivalente sobre uso de *dar* como “proporcionar” incluida más arriba (tabla 25), muestra las causas de esa diferencia del 14 % que existe en la frecuencia de uso de las estructuras V + N en ambos grupos. Vemos, pues, que gran parte de esa sobreutilización de la construcción V + N se debe a la tendencia de los HNN de implicarse personalmente en la narración, a referirse a su estado anímico.

Por otro lado, llama la atención que precisamente en las dos clases léxicas más frecuentes entre los HNN –la de los “desplazamientos” y la de los sustantivos que denotan “miedo o zozobra”–, la variedad de sustantivos utilizados es menor que en otras clases léxicas menos frecuentes. Los HNN usan menos sustantivos en estas clases pero los repiten más; así, mientras que los HN usan *miedo*, *reparo* o *pánico*, los HNN, dentro de esta gama significativa, solo usan *miedo*; y mientras los HN *dan paseos*, *vueltas*, *giros*, *pasos* y *saltos*, usados en la mayoría de las ocasiones en sentido figurado (*dar un paso en la vida*, *un giro a la historia*, *vueltas a la cabeza*, *dar el gran salto*, *etcétera*), los HNN, en cambio, solo *dan vueltas* y *paseos* en sentido literal.

Cabe relacionar la infrautilización que hacen los HNN de esta última clase léxica (la de los “desplazamientos”) –pese a ser precisamente la más empleada en el corpus, tanto no nativo como nativo–, con el hecho de que lo que convierte esta combinatoria en la más frecuente del corpus nativo sea el uso de los sustantivos correspondientes en sentido figurado; este conocimiento es muy específico, más resbaladizo que el de los sentidos literales, y es posible que la mayoría de los HNN no sean muy conscientes del empleo figurado de sustantivos como *paso*, *salto*, *giro* o *vuelta*; y, por otro lado, el inglés selecciona para esta clase léxica el verbo *to make*, principalmente, lo que también puede contribuir a la infrautilización.

Por lo tanto, parece que, si profundizamos dentro de una clase léxica, tras las aparentes semejanzas de uso se perfilan ciertas diferencias, fundamentadas en

la presencia frente a la ausencia de sustantivos cuasi-sinónimos (*reparo, miedo, pánico*) y en la ausencia frente a la presencia de usos figurados.

Los sustantivos diferentes identificados en ambos corpus suman un total de 63; al igual que ocurre con el uso en el sentido de “proporcionar”, 15 son los sustantivos compartidos por los dos grupos: *vuelta, paseo, clase, conferencia, golpe* (aunque las formas empleadas son *golpecitos* y *golpecillos*), *beso* (aunque los HNN solo se refieren a *besitos*), *asco, pena, bienvenida, gracias, miedo, sorpresa, sensación, gusto y ganas*. HN y HNN coinciden en el 23,8 % de los usos. Los HN usan 47 sustantivos diferentes y los HNN, 31. Estos datos absolutos expresados en porcentajes resultan también en unas cifras muy equivalentes: el 47,9 % en los HN y el 48,4 % en los HNN de las combinaciones de *dar* son diferentes en cada corpus. El porcentaje restante se refiere a usos repetidos de una misma combinación, por lo que el uso productivo de *dar* en lo que respecta a los sustantivos con los que se combina el verbo en la construcción V + N cuando el verbo actualiza el significado de “producir(se)” es prácticamente igual de amplio.

No obstante, pese a las semejanzas en los dos corpus en la frecuencia de uso de la mayoría de las clases léxicas más utilizadas, se dan diferencias significativas en el porcentaje de uso de algunas clases semánticas, como se ha visto, y de algunas combinaciones léxicas; y los sustantivos con los que cada grupo combina *dar* no coincide en la mayoría de los casos (77 %).

En suma, si los HN usan 115 sustantivos diferentes dentro de la construcción V + N, los HNN usan 85. Además, de estos, 15 sustantivos que emplean los HN no forman parte de las 5000 palabras más frecuentes del español, según la información de *A Frequency dictionary of Spanish* (Davies, 2006), a saber: *bungalow, visa, vitaminas, champú, raciones, boleto, autenticidad, politización, el alta, el alto, luz verde, dejo, chivatazo, puntaje y golpe (atraco)*; los HNN usan 5 palabras no frecuentes: *voltereta, ponencia, coraje (irritación), tenedor y peluche*; esto es, el español nativo de este grupo de hablantes utiliza un 87 % de palabras pertenecientes a las 5000 palabras más frecuentes del español, mientras que el grupo de HNN utiliza un 94,1 % de palabras frecuentes. Estos datos demuestran que, en línea con el estudio de Cobb (2003: 403), los aprendientes, cuando han de producir una tarea escrita sobre un tema para un lector anónimo, utilizan el vocabulario más frecuente, restringido de la oralidad,

Al igual que sucede en los demás apartados, algunos de los datos de frecuencia obtenidos en el análisis son demasiado bajos para hablar de significatividad estadística, por lo que en esos casos hemos aportado predicciones sobre el dominio del grupo de aprendientes.

7.2. ACEPTABILIDAD E INACEPTABILIDAD DE LOS DATOS NO NATIVOS. ANÁLISIS CUANTITATIVO Y CUALITATIVO

La frecuencia en el corpus de HNN solo revela cuántas veces ocurre un elemento determinado en el corpus. Aunque puede indicar cierto grado de adquisición en la producción del español por los HNN, no refleja necesariamente dominio, pues la lengua del aprendiente está cargada de rasgos problemáticos y expresiones artificiales. En palabras de Guo (2006: 183-84), “Large frequency in a particular pattern (or phrase) indicates only that this pattern (or phrase) is used fairly often by the group of writers. It has nothing to say about its appropriateness *per se*”. Tras la frecuencia se ocultan los errores.

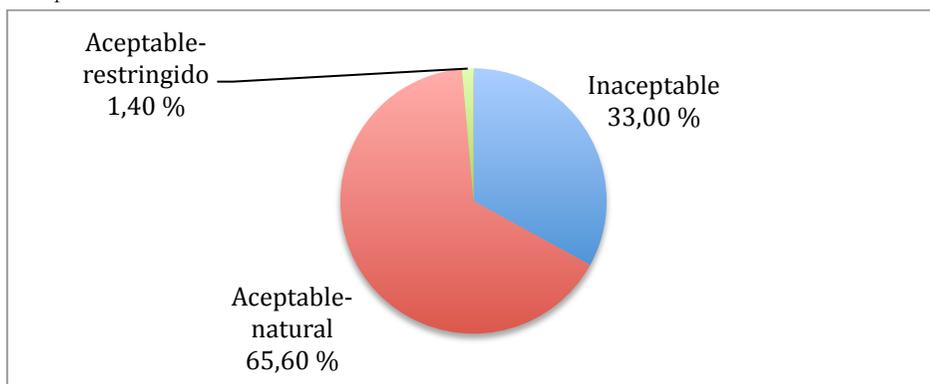
A continuación, presentamos la frecuencia de los errores y de los aciertos, en general, y ofrecemos el porcentaje de error y de acierto en cada estructura sintáctica utilizada en los corpus. Se señala el índice de frecuencias en relación con la localización del error en el sintagma (verbo, complementos, sujeto o totalidad del sintagma); se señala también el índice de frecuencias al respecto de la naturaleza –léxica o computacional– del error en el verbo y, dentro de esta categorización, el índice de frecuencias de cada tipo de error léxico y computacional, de acuerdo con la taxonomía diseñada en el capítulo 6. Estos datos de frecuencia están respaldados por un análisis detallado de las concordancias.

7.2.1. Frecuencia de los datos aceptable-naturales, aceptable-restringidos y de los inaceptables

De los 218 usos de *dar* empleados por los HNN, 143 son aceptable-naturales, esto es, el 65,6 %; 71 son inaceptables¹⁸⁷, lo que equivale al 32,6 %; y tan solo 3 son aceptable-restringidos, lo que se corresponde con el 1,4 %. En el siguiente gráfico se representan estas frecuencias:

¹⁸⁷ Por motivos metodológicos, solo nos es posible cuantificar un error por enunciado, pese a que hay enunciados que contienen más de un error. El criterio seguido para seleccionar un error de entre dos o más en un mismo enunciado es el de la relación directa con el verbo *dar*. Si solo uno de los problemas está en el núcleo verbal, mientras que el resto se localiza en los complementos o el sujeto, hemos dado prioridad al primero por ser más relevante para nuestros objetivos.

Gráfico 4. Índices de frecuencia de los usos aceptable-naturales, aceptable-restringidos e inaceptables



Los cuatro informantes juzgaron los usos siguientes como aceptable-restringidos:

- [1] Para mí corrupción no sólo es lo que esté penado por la ley, también es enchufar a allegados y amigos, dar contratos sin un concurso previo (22 HNN)
- [2] Adri llegó tarde pero luego estaba dando volteletas por la hierba y al final se hicieron las 18:30 (113 HNN)
- [3] A la una de la madrugada dio su última respiración. Cogí el teléfono para llamar a mi padre (166 HNN)

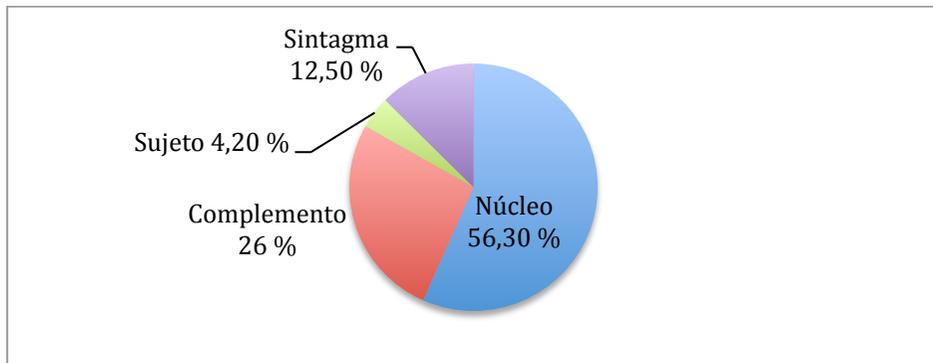
En el ejemplo [1] se propusieron como opciones preferidas “hacer contratos” y “contratar”. Ambas posibilidades responden a dos estrategias mencionadas que pueden estar tras parte de la sobreutilización de V + N: el uso de *dar* en lugar de otro verbo frecuente (Chi Man-lai *et al.*, 1994) o el uso de *dar* en lugar del verbo del que procede el sustantivo con el que coaparece. El [2] parece ser opción natural únicamente en Argentina, mientras que en Méjico y en España no lo es; por ello, “dar volteletas” es un uso considerado restringido a una zona geográfica. En el último ejemplo, los informantes prefirieron por razones estilísticas “exhaló su último aliento” y “respiró por última vez”.

Como se desprende de los datos del gráfico anterior, no podemos decir que el verbo *dar* sea un verbo seguro para este grupo de HNN, dado que se cometen errores en un tercio de sus usos.

7.2.2. Localización del error

Más concretamente, de los 71 errores identificados, 40 se localizan en el núcleo verbal; esto es, más de la mitad (56,3 %) de los errores contenidos en el sintagma proyectado por *dar* se debe a alguna desviación en el uso de *dar* (frente a los errores originados en alguno de los argumentos del verbo). 19 errores se deben a un problema localizado en un complemento exigido por el verbo (26,4 %); 3 errores se encuentran en el sujeto (4,2 %); y, finalmente, 9 errores se extienden por todo el sintagma (12,5 %).

Gráfico 5. Localización del error

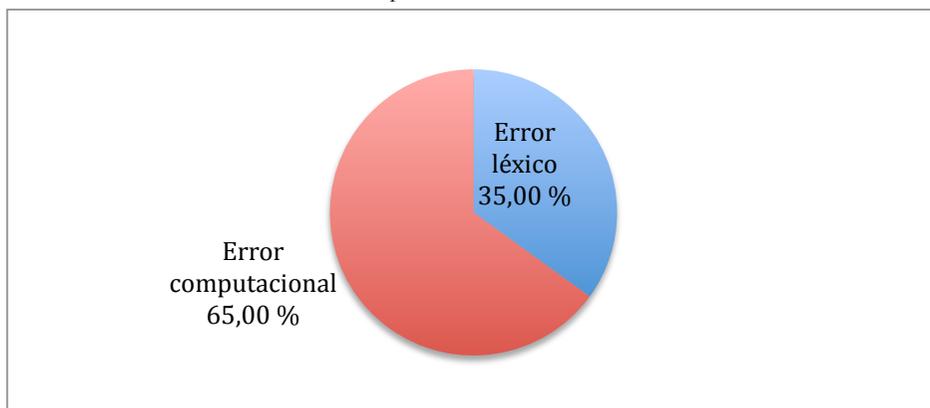


7.2.3. Descripción y explicación del error en el núcleo verbal

Los errores producidos por algún tipo de desviación en el núcleo verbal son los más relevantes en nuestro análisis, que pretende determinar el conocimiento productivo que este grupo concreto de HNN tiene del verbo *dar*.

Así, de esos 40 errores en el núcleo, 14 son errores de tipo léxico y 26 de tipo computacional, de acuerdo con la diferencia entre los tipos de error que establecimos en el capítulo 6.

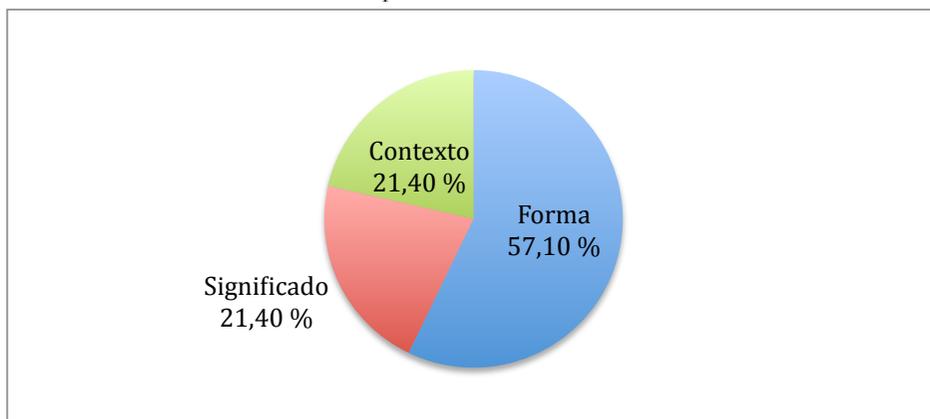
Gráfico 6. Errores léxicos vs. errores computacionales en el uso de *dar*



Estos datos muestran que en prácticamente 2 de cada tres errores (65 %) relacionados con el uso del verbo *dar* el problema está conectado con la selección sintagmática del verbo, con la interfaz léxico-sintáctica –esto es, la red argumental del verbo– y con los vínculos que establece el verbo con las demás piezas léxicas por medio de sus marcas formales. Así, la competencia gramatical y léxico-sintáctica del verbo presentan más problemas que la morfológica, semántica y sociolingüística.

Los errores léxicos, que constituyen el 35 %, tienen que ver con desviaciones en la selección paradigmática, concretamente con la forma del verbo, con el significado del verbo y con el contexto de uso del verbo. De los 14 errores léxicos, 8 se relacionan con problemas en la forma verbal, 3 con el significado y 3 con factores sociolingüísticos. Las principales dificultades léxicas son, por lo tanto, formales.

Gráfico 7. Frecuencia de los diferentes tipos de error léxico del verbo *dar*



A continuación, se incluyen todos los errores léxicos identificados en el núcleo verbal en el corpus no nativo:

- [1] Aprendí que el museo fue establecido antes de su muerte y después Miró le dió muchas obras (43 HNN)
- [2] Me pregunto, ¿porqué no da [por dar] la oportunidad a un pobre niño sin ningún pariente? (52 HNN)
- [3] por la mañana hay camioncitos que vienen a dar cosas a las tiendas (67 HNN)
- [4] Esta vez, uno de estos no funciona bien, y tenían que dar las cosas mas tarde que regular (68 HNN)
- [5] correr los riesgos de estar en un lugar lleno de humo, o no, según nos da la gana (100 HNN)
- [6] Dando clases de quinto año bilingüe en una primaria siempre me presenta con un gran reto (128 HNN)
- [7] pasar el tiempo que quisieramos haciendo lo que nos dio la gana (136 HNN)
- [8] pero la gente va a dar cuenta que es demasiado cara pelear con una planta. (146 HNN)
- [9] Hay mucha mas gente que usa la planta que damos cuenta (147 HNN)
- [10] que el va a hacer milagros y sanaciones para que todo el mundo pueda dar cuenta de que el es el Hijo prometido para salvar el mundo (152 HNN)
- [11] A aquel tiempo, mi tía acababa de dar luz a mi prima nueva, se llama Kira (171 HNN)
- [12] Hay que darse cuenta de que el pueblo achuar es una cultura que les da mucha importancia a sus sueños (178 HNN)
- [13] Manchester daba la luz a muchos grupos musicales, como 'Stone Roses',

'Charlatans', (212 HNN)

[14] se embarazan sin querer y ni están listos para darle a luz a un chico o una chica. (229 HNN)

Los casos [2], [6], [8], [9], [10], [11], [13] y [14] presentan errores léxicos de forma.

En [2] y [6] los HNN no reparan en que el verbo debería estar en infinitivo –y no en la 3.^a persona del singular ni en gerundio–, pues es la forma que se selecciona en español cuando se interpreta que no existe controlador dentro de la oración principal, por lo que el sujeto recibe una lectura no específica, no referencial, del tipo a la que posee el sujeto en una oración impersonal. En ambos casos la causa del error es interlingüística, pues se produce una importación de la forma. En inglés *dar* es *to give*, pero también *give* (dos maneras de poner el infinitivo); de hecho, en el caso de [2], en inglés se diría *why not give?*, estructura que el aprendiente está trasladando al español pero con la 3.^a persona del singular, dado que no reconoce ese *give* como infinitivo en su propia lengua. Probablemente este error se debe a lo que se conoce como “transfer of training”; es una regla muy extendida en inglés que el infinitivo lleve *to* delante, por lo que el aprendiente no percibe que este ejemplo es una excepción a esta regla. El [6] también es un caso de transferencia, dado que en inglés el gerundio puede recibir una lectura no específica, como la del infinitivo.

En [8], [9] y [10] el alumno no domina la forma de la locución *darse cuenta*; es una estructura pronominal y el pronombre está ausente cuando se produce *dar cuenta*. Las causas del error son intralingüísticas: se lleva a cabo una elección léxica errónea del lexicón. Por último, en [11], [13] y [14] sucede lo mismo que en el caso anterior: los HNN no dominan la forma de la locución *dar a luz*; es una construcción invariable y formalmente fija, por lo que no permite que se introduzcan ni se omitan en ella otros elementos, como sucede en *dar luz*, *dar la luz* y *darle a luz*. Las causas del error son también intralingüísticas, por una elección léxica errónea a partir del lexicón.

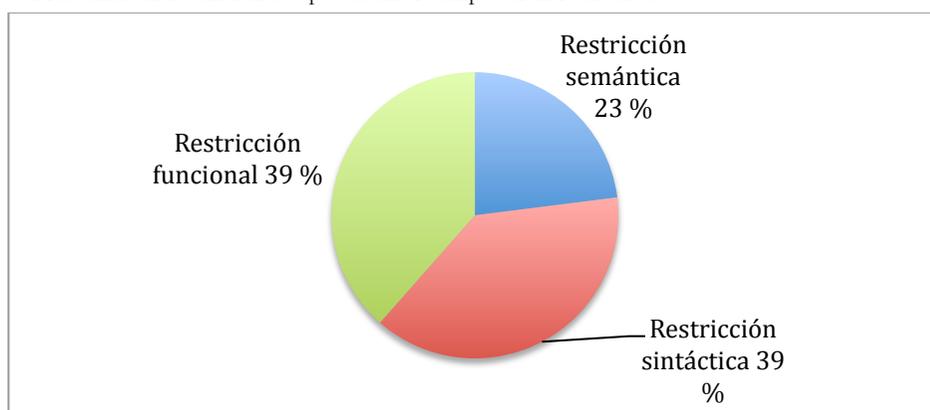
Por otro lado, en los usos [3], [4] y [12] el error léxico es de significado, pues consiste en una confusión en el significado de la pieza léxica seleccionada: la pieza léxica extraída existe pero tiene un significado diferente al que se le ortorga. Así, en [3] y [4] el aprendiente confunde *dar* por *traer*; se trata de un caso de impropiedad semántica. Y en [12] se confunde *darse cuenta* por *tener en cuenta*, construcciones próximas en la forma y en el significado, pero no intercambiables. Estos tres últimos errores son de carácter intralingüístico, debidos a una selección errónea de la pieza léxica.

Por último, los usos de *dar* en [1], [5] y [7] son errores en el contexto comunicativo en el que se encuentran, debido a una desviación de la norma panhispánica en los factores extralingüísticos (geográficos, sociales y situacionales). En [1] se produce un caso de pobreza léxica. La poca extensión de vocabulario del aprendiente causa la ocurrencia de este verbo frecuente en un contexto en el que un nativo –con el mismo perfil– no lo utilizaría; su grado de especialización semántica se halla por debajo de lo que un HN consideraría natural. En [5] y [7] se produce un error de registro, pues *dar la gana* es una locución coloquial (cf. *DRAE*) que no tiene cabida en una redacción de clase que se caracteriza por un registro neutro. Estos tres casos son errores intralinguales, por una desviación de carácter diastrático y diafásico.

Así, en el 57,1 % de los errores léxicos el error está localizado en una locución (*darse cuenta, dar la gana, dar a luz*), y solo el 14,3 % de los errores léxicos son interlinguales, frente a los 85,7 % intralinguales.

Por otro lado, de los 26 errores computacionales, en 10 se presentan problemas relacionados con las exigencias selectivas sintácticas del verbo, esto es, con la preposición que rige el verbo, con el número del sustantivo al que pertenece la clase léxica seleccionada por el verbo o con los argumentos exigidos por el verbo. Otros 10 sintagmas en los que participa *dar* presentan problemas en las exigencias selectivas funcionales del verbo, esto es, en la concordancia de rasgos; y, por último, 6 tienen problemas con las restricciones semánticas del verbo.

Gráfico 8. Frecuencia en los diferentes tipos de error computacional del verbo *dar*



No obstante, si tenemos en cuenta que la restricción semántica que ejerce el verbo se manifiesta solo en la construcción V + N, y no en las demás, el porcentaje de este tipo de error asciende al 50 %, ya que el total de los errores computacionales en el núcleo verbal de las estructuras ditransitivas, transitivas, intransitivas, impersonales y pasivas es 12. Así pues, la desviación en la restricción semántica se convierte en el error computacional más frecuente.

A continuación, incluimos los usos de *dar* que presentan desviaciones de carácter computacional:

- [1] Cuando llegué me saludó en castellano y me dió cuenta de que entendía muchísimo más (11 HNN)
- [2] Los judíos de España, llamados sefardíes, y los moros no nos daban el respeto que nos debieron (15 HNN)
- [3] hace siempre frío durante el invierno dado las bajas temperaturas al exterior (31 HNN)
- [4] Los viernes, siempre tuvimos cena, donde dimos bienvenidas a los nuevos estudiantes (44 HNN)
- [5] Ellos hablaron y hablaron, y se dieron cuenta que todavía no habían cambiado los sentimientos de ellos (46 HNN)
- [6] Su éxito y la creación de una carrera que carece de sustancia da un mensaje negativo a otros (58 HNN)
- [7] Carlos intenta encontrar una mujer de por vida, pero se da cuenta al final que, aunque un poco rara, Julia la es (66 HNN)
- [8] Así que me quejaba muchas veces, pero al final me di cuenta que encontré la cosa más importante (75 HNN)
- [9] Después de mucho tiempo y gran luchas, los derechos necesarios fue dados, primero en la Guerra Civil (77 HNN)
- [10] el pasar tiempo juntos nos daba mucha emoción aun cuando planeábamos (125 HNN)
- [11] me di cuenta que los italianos son tan aficionados a la ópera (127 HNN)
- [12] no sé que idiomas hablaré, en realidad no sé nada y eso me doy un poco miedo (130 HNN)
- [13] No esperaba mucho de la ciudad de Bilbao dado su fama de ser un centro industrial (138 HNN)
- [14] Los pecados nos dan el muerte. !!!Jesus no da VIDA ETERNA!!! (153 HNN)
- [15] es demasiado importante a dar algo a la comunidad que me has dado tanto en mi vida (157 HNN)
- [16] tras pasar las últimas 13 años en la profesión de hostelería, me ha dado cuenta de que el corazón de mi vida existe en la educación (164 HNN)
- [17] Por ejemplo, cuando fui al sur del país, me di cuenta que mucha de la gente me esperaba (180 HNN)

- [18] Fuimos a tomar algo y horas y horas después, nos dimos cuenta que todavía estábamos hablando sin parar en el mismo bar (189 HNN)
- [19] Se da cuenta que no está dispuesta a ser madre (191 HNN)
- [20] un recuerdo que siempre tendré y siempre me dará una sonrisa (220 HNN)
- [21] Cuando mi novio llegó del trabajo, le dio su regalo y nos preparamos (223 HNN)
- [22] en el año pasado me di cuenta que ya no me interesa y por eso yo cambie mi plan (224 HNN)
- [23] Emery y yo se dieron besitos cuando era medianoche porque es una tradición (225 HNN)
- [24] y por eso tengo planes cuidarla bien, y darle todos los deseos de su corazón (234 HNN)
- [25] También me di cuenta que quiero trabajar con latinos, y con los países en Sudamérica (241 HNN)
- [26] Una semana no dio abasto para descubrir todos los misterios de Kashmir (197 HNN)

A partir de estos datos no nativos, describimos los errores encontrados y tratamos de aportar las causas de dichas desviaciones. En primer lugar, nos centramos en los errores en las exigencias selectivas sintácticas, que se dan en el [4], [5], [7], [8], [11], [17], [18], [19], [22] y [25]. De estos, todos salvo el que se produce en [4] se corresponden con el mismo error relacionado con la locución verbal *darse cuenta*. Los HNN desconocen que esta locución va siempre seguida de un complemento precedido de la preposición *de*, y que debe evitarse, en la producción esmerada, la supresión de la preposición *de* cuando el complemento es una oración subordinada introducida por la conjunción *que*, como es el caso de todos los usos anteriores (cf. *DPD* y el *DRAE* [2001 y 2014]). Es un error intralingüístico que puede haber sido transferido a los HNN durante la instrucción formal o durante los periodos de inmersión en los países hispanohablantes, ya que el *queísmo* es un fenómeno relativamente frecuente entre los HN¹⁸⁸.

En [4], se presenta una estructura ditransitiva formada por *dar*, con el sentido de “producir”, esto es, como verbo de apoyo, y un sustantivo que designa “recibimiento”, como es *bienvenida*. Este sustantivo pertenece a la clase de los sustantivos que designan “formas de trato social” (como son los saludos, los recibimientos y las despedidas), junto con otros que se refieren a “diversos signos que forman parte de ceremonias”. Estos SN, cuando coaparecen con *dar*,

¹⁸⁸ En el corpus nativo encontramos este fenómeno en un nada desdeñable 19,2 % del total de los usos de la locución *darse cuenta*.

están precedidos por un determinante (*dar el trato que se merece, dar el último adiós, darle la despedida, darle el recibimiento o la acogida merecida, el pésame, etcétera*), y el sustantivo aparece normalmente en singular –salvo interjecciones fosilizadas (*las gracias, los buenos días, las buenas tardes o las buenas noches*)–, dado que la acción se refiere genéricamente al tipo de recibimiento, excepto cuando se antepone una estructura comparativa, como en *la más calurosa de las bienvenidas* o *la peor de las bienvenidas*. Este error también es intralingüístico y no puede atribuirse a la L1.

Estos datos revelan que los HNN apenas cometen errores productivos en relación con las restricciones sintácticas de *dar*¹⁸⁹; en cambio, sí que tienen importantes problemas con las restricciones sintácticas de la unidad multipalabra *darse cuenta*.

Por lo que respecta a la restricción funcional del verbo, la flexión verbal representa una categoría funcional que establece lazos con las categorías léxicas que también poseen esas marcas formales, como el número y la persona en los pronombres personales que desempeñan la función de sujeto. Así, en el uso [1] y en el [16] los aprendientes no hacen concordar el sujeto y el verbo en la locución *darse cuenta: me dio cuenta y me ha dado cuenta*. En el primer caso es posible que se desconozcan algunas de las formas irregulares del verbo; en el segundo caso, si no es un lapsus, el aprendiente no conoce el presente de *haber*, necesario para construir los tiempos perfectos. Este problema es intralingüístico.

En [3] y en [13] los HNN no demuestran un conocimiento productivo de la construcción de participio absoluto. En estas construcciones, el participio desarrolla una estructura pasiva, pues toma como su sujeto el tema correspondiente del verbo inacusativo: *Dadas las bajas temperaturas; dada su fama*. Así pues, se debe dar una concordancia en el participio (predicado verbal) y su sujeto; como formas verbales, los participios verbales seleccionan a sus argumentos. En todos los casos, el sujeto aparece pospuesto (Rodríguez Ramalle, 2006: 324). Este error es también intralingüístico¹⁹⁰.

¹⁸⁹ En el corpus no nativo hay otros tres usos del verbo *dar* con sustantivos pertenecientes a esta clase semántica: 2 de *dar la bienvenida* y 1 de *dar las gracias*. Los tres usos son adecuados, por lo que no se puede afirmar que este grupo tenga un problema con esta restricción en concreto.

¹⁹⁰ En el corpus nativo, el 40 % de las construcciones de participio absoluto también presentan un error en la concordancia. De acuerdo con Thagg Fisher (1985: 191), aquellos errores producidos por los HNN que también son cometidos por los HN –como es el caso de esta estructura o el del queísmo al que ya nos hemos referido– son valorados por los HN como menos negativos. Se refiere a aquellos casos en los que el uso nativo difiere de la norma prescriptiva establecida, los cuales, según la autora, no deberían ser un foco principal en la enseñanza. Nosotros no establecemos esta diferencia en los errores; todos suponen una desviación de la norma establecida.

En [9] nos encontramos con una construcción pasiva con *ser* en la que el verbo no concuerda en número con el sujeto: *los derechos necesarios fue dados*. Este error es también intralingüístico.

En [6] se demuestra que el aprendiente desconoce que los grupos nominales coordinados mediante la conjunción *y* que desempeñan la función de sujeto concuerdan con el verbo en plural: *Su éxito y la creación de una carrera que carece de sustancia da (...)*. Este no es un caso en el que las dos ideas que se asocian se identifican en la conciencia lingüística del hablante como si designaran una sola entidad compleja, lo que justificaría el singular. El error es intralingüístico.

Los usos [12], [15], [21] y [23] son ditransitivos. En [12] y [21] parece darse el mismo patrón: el aprendiente flexiona el verbo en la persona y número del CI, en lugar de los del sujeto (como en los usos [1] y [16] señalados anteriormente): *Eso me doy un poco de miedo* y *[yo] le dio su regalo*. En [23], *Emery yo se dieron besitos*, el aprendiente confunde igualmente la 1.^a persona del plural con la 3.^a persona del plural. En [15] tampoco coinciden el sujeto (3.^a persona del singular) y el verbo (2.^a persona del singular).

Estos errores, intralingüísticos, son ciertamente básicos, y coinciden en redacciones de un nivel de dominio lingüístico general bastante bajo dentro del nivel del corpus.

Por último, destacan los errores en la restricción semántica del verbo, que se localizan en los usos [2], [10], [14], [20], [24] y [26].

Con referencia a [2], en español, la palabra *respeto*, con el sentido de veneración o consideración (no con el sentido de miedo o reparo) no se da, sino que se muestra, se manifiesta, se profesa o se tiene. En cambio, en inglés sí se da el respeto, por lo que la fuente de este error es interlingüística, pues se está importando la restricción semántica. Lo mismo sucede en los usos [20] y [24]. En español, la sonrisa no se da, pero en inglés sí (aunque tampoco se usaría en este contexto). En español sería más conveniente *provocar* para este contexto. Del mismo modo, tampoco se dan los deseos, sino que se conceden, excepto en la fórmula “le doy mis mejores deseos”. En inglés existe la expresión *to give a desire*, por lo que el aprendiente está importando esta misma restricción al uso el verbo español.

La fuente del error en [10] y [14] no proviene de la L1, por lo que el error es intralingüístico. La *emoción*, con el sentido de conmovir el ánimo, no se da, sino que se provoca, se despierta, se causa. Por otro lado, con respecto al ejemplo [14] *–los pecados dan muerte–*, existe en español *dar muerte a alguien*, como *acabar con*; pero en este caso *muerte* nunca va precedido de determinante y el sujeto se tiene que poder personificar, lo que no es posible en el caso de *los*

pecados. Además, se produce cierta incoherencia semántica, pues *dar muerte* se utiliza como sinónimo de *matar*, y los pecados no matan, sino que conducen a, causan o provocan la muerte.

Por último, en el uso [26], *Una semana no dio abasto para descubrir todos los misterios de Kashmir*, el argumento sujeto de la locución *dar abasto* tiene que ser +humano. Se trata, de nuevo, de un error intralingüístico que demuestra un desconocimiento de la locución, de su restricción semántica.

En suma, dentro de los errores computacionales que se producen en *dar*, se distinguen los sintácticos, los funcionales y los semánticos; el 90 % de los sintácticos está relacionado con la locución *darse cuenta* y la ausencia de la proposición *de*; de los funcionales, el 50 % tiene que ver con una confusión que se produce entre la 1.^a y la 3.^a persona del verbo; y, por último, el 50 % de los errores semánticos son interlingüísticos, pues se transfiere en ellos la restricción semántica del verbo equivalente en la L1. No obstante, la restricción interlingüística en los errores computacionales en general es de un 25 %.

7.2.4. *El error y el acierto en las estructuras analizadas*

En los 106 usos de la estructura ditransitiva hay 36 errores; de las 39 transitivas, 4 contienen error; de las 5 intransitivas, 2 son usos erróneos; 4 son las desviaciones en las 12 impersonales; en las 3 pasivas, se localiza 1 error; de las 45 locuciones, 23 presentan algún error; y, por último, de los 3 usos de la construcción de participio absoluto, 2 son erróneos. Por otro lado, en el único uso pronominal no hay error, ni tampoco lo hay en los dos esquemas fraseológicos semiproductivos utilizados. A continuación, se incluyen dos gráficos: el primero refleja el índice de los porcentajes de error de cada estructura utilizada por los HNN; el segundo, el índice de error en el núcleo verbal producido en cada estructura.

Gráfico 9. Porcentaje de error de cada estructura

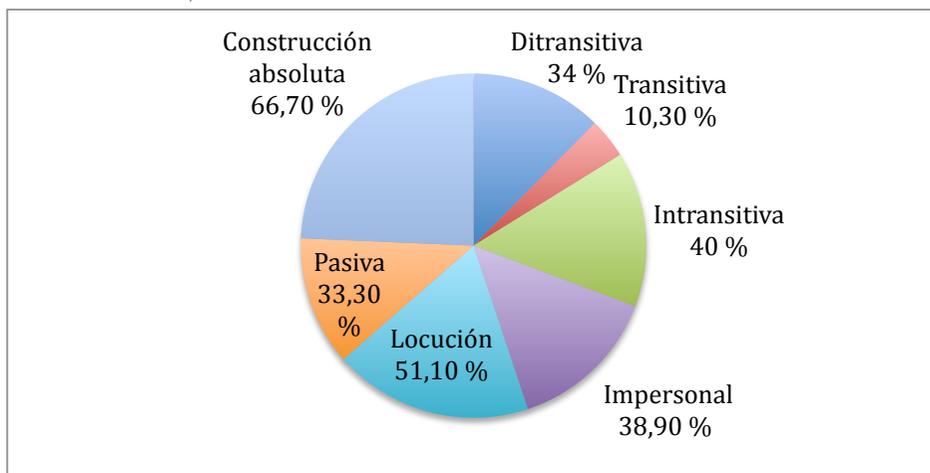
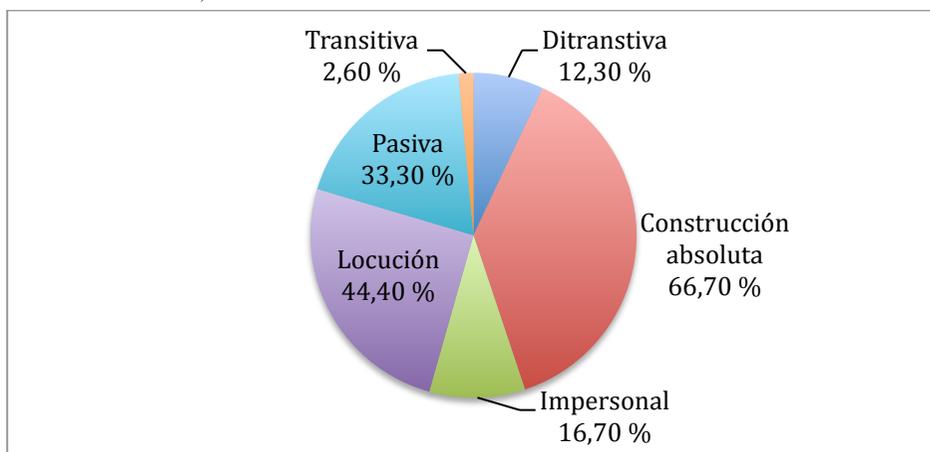


Gráfico 10. Porcentaje de error en el núcleo verbal de cada estructura



Aunque algunos de estos resultados no son significativos, pues las ocurrencias de ciertas estructuras son demasiado bajas como para que se puedan obtener resultados generalizables (13 son los errores en el núcleo verbal de las ditransitivas y 20 en el de las locuciones, pero 2 en las impersonales y en las construcciones de participio absoluto, 1 en las transitivas y 1 también en las pasivas), resultan útiles como indicación a partir de la que comenzar una investigación más a fondo; por ejemplo, con una muestra de datos más amplia, como señalan McEnery y Hardie (2012: 51). No obstante, estos resultados sí nos proporcionan algunos indicios. Por ejemplo, el hecho de que en 2 de los 3 usos

que implican construcciones absolutas no concuerde el participio, predicado verbal, con el sujeto puede indicar desconocimiento en la mayor parte de los HNN. Lo mismo sucede con los usos de la pasiva con *ser*: de los 2 que aparecen en el corpus no nativo, 1 es incorrecto, debido a la falta de concordancia entre el verbo *ser* y el sujeto; habría que investigar si este problema se reproduce en un corpus más extenso.

En cuanto a las locuciones, excepto en el único uso que se hace de *dar igual*, hay errores en porcentajes muy altos. De los 5 usos de *dar a luz*, en 3 casos se desconoce la forma; los 2 usos de *dar la gana* demuestran que los HNN desconocen que es una locución coloquial, y, por tanto, inadecuada en una redacción de clase; en *dar abasto* no se es consciente de que el núcleo verbal no solo selecciona un argumento externo que es el sujeto, sino que además tiene que ser +humano. En 15 de los 36 usos de *darse cuenta* hay un error en el núcleo verbal, ya sea por la forma (confusión entre *darse cuenta* y *dar cuenta*) o por la restricción sintáctica de la locución al exigir un complemento introducido por la preposición *de*.

En el caso de las impersonales, el problema viene dado por la desviación en el uso del infinitivo *dar* cuando se busca que un sujeto reciba una lectura no específica, no referencial.

Por último, en las ditransitivas el error es fundamentalmente computacional. Los errores de concordancia entre el verbo y el sujeto se dan también en otras estructuras, no es algo exclusivo de las ditransitivas; lo que sí es más propio de estas construcciones es el error en la restricción semántica del tema. De los errores computacionales vinculados al núcleo verbal, 5 se deben a la restricción semántica errónea; es decir, el 41,7 % de los errores relacionados con el núcleo del verbo en las oraciones ditransitivas es de esta naturaleza. De nuevo, convendría contrastar estos resultados con un corpus más extenso. Llama la atención que no haya ningún error en la selección semántica de las estructuras intransitivas, en las que se mantiene el tema pero en el papel sintáctico de sujeto. Como tampoco son tantas las intransitivas utilizadas, solo 5, también sería de interés contrastar estos datos con un corpus mayor.

Como se ha visto, la mayor parte de los errores en el núcleo verbal *dar* son de naturaleza intralingüística; son, concretamente, 35 errores, lo que equivale al 87,5 %; los que se deben a la interferencia de la L1 son 5 (12,5 %), pero el 60 % se da en la restricción semántica del verbo (3). De nuevo, convendría contrastar estos resultados con un corpus más extenso para que los resultados sean generalizables.

7.2.5. Descripción y explicación del error en el complemento

Se han localizado 19 errores en el complemento seleccionado por el verbo:

- [1] En cuanto llegué, mis primos, tíos y amigos me dieron la bienvenida
- [2] Si los gays pudieran adoptar, podrían dar buenas hogares a estos huérfanos
- [3] cuando oigo canciones en inglés aquí en España. Me da el sentido de que haya una pérdida de autenticidad de la cultura española
- [4] aunque me gustó esa película, y doy una marca de 6.5 de 10 por la, no pienso en verlo otra vez (38 HNN)
- [5] El amor puede hacer muchísimas cosas, y los homosexuales puedan dar el amor tan bien como los demás (40 HNN)
- [6] Espero que le he dado un paisaje de lo que es Seattle (51 HNN)
- [7] Nico, mi estudiante favorito era argentino y antes de salir, me dio los restos de su mate, la bebida tradicional de Argentina (71 HNN)
- [8] Me da curage ver que a los estudiantes de Penn State nada más les interesa el alcohol (94 HNN)
- [9] que va siempre en aumento el número de personas que les da asco el cigarro (97 HNN)
- [10] no se de donde saca fuerzas y siempre da 120% en el escenario. (107 HNN)
- [11] Para Adri le dimos un peluche, un libro con nuestras fotos firmado (109 HNN)
- [12] [a la película] La daría dos estrellas de cinco (177 HNN)
- [13] Era muy interesante pero me daba miedo un poco porque eran subterráneos. (204 HNN)
- [14] Yo me doy cuenta de la religión es un gran parte de la vida latina también (218 HNN)
- [15] A repasar la implementación y darla otra oportunidad, a lo mejor nos ayudará en salva (230 HNN)
- [16] Tenemos el deber de darles a nuestros hijos la habilidad de tener un mundo me (231 HNN)
- [17] si la pareja puede darle una vida agradable y darle oportunidad (236 HNN)
- [18] ella decidió que tenía que darle a él unos golpecitos si él tomó las galletas sin (238 HNN)
- [19] ayudarles a mejorar su español y darles mucha oportunidad de practicarlo (239 HNN)

De los 19 errores en el complemento, 15 son léxicos y 4 tienen que ver con las restricciones del nombre. Nos centramos a continuación en los errores léxicos.

De los 15 errores léxicos, 6 ([1], [2], [8], [12], [16] y [20]) presentan un problema en la forma; así, se confunde *bienvenida* por *bienvida*; *hogar* es interpretada como sustantivo femenino, dado el género del adjetivo *buenas* que la acompaña; *coraje* se confunde con *curage*; en *la daría* y *darla otra oportunidad* se confunde la forma del pronombre clítico que desempeña la función de CI, que ha de ser *le* (casos de láismo); y, por último, en *darles mucha oportunidad* se interpreta el sustantivo *oportunidad* como no contable, cuando es contable, por lo que se *da una oportunidad* y se *dan muchas oportunidades*. En todos estos errores se produce una elección léxica errónea, y son todos intralinguales, a excepción de *curage*. En inglés existe la palabra *courage*, y en español *coraje*, con el sentido en ambas lenguas de “valor y bravura”; ahora bien, en español tenemos otra acepción de esta misma palabra, la de “irritación o ira”; esta acepción en inglés no existe, y, aunque el aprendiente la conoce en español, interfiere la forma de la voz inglesa en este significado.

Los errores en el significado son 6 ([3], [4], [6], [7], [17] y [19]); en estos usos la pieza léxica extraída existe pero tiene otro significado. Así, el aprendiente confunde *sentido* con *sensación*, *marca* con *nota*, *paisaje* con *idea*, *restos* con *resto*, *habilidad* con *posibilidades* y *golpecitos* con *cachetes* o *azotes*. Los cuatro primeros (66,7 % de los errores de significado) –*sentido*, *marca*, *paisaje* y *restos*– son interferencia del inglés; en los tres primeros se importa el significado, y en el último, la forma. En *sentido*, el aprendiente extiende el significado de una palabra de la L2, *sentido*, porque su equivalente *sense* en la L1 comparte con *sentido* varias acepciones, pero no la que se actualiza en este caso; así, la voz inglesa sí que abarca el contenido aquí traído, el de “impresión que las cosas producen por medio de los sentidos” (DRAE, 2001), pero en español este significado corresponde a *sensación*. En inglés, *nota* es *mark*, y *marca* también es *mark*, pero no con el sentido de “calificación”, sino únicamente con el sentido de “signo visible”, por lo que la interferencia aquí es evidente. El caso de *dar un paisaje* es una transferencia menos directa de *give a picture*, pero igualmente clara. Esta expresión la traduciríamos en español por *dar una idea*. El aprendiente se inspira en *picture* cuando dice *paisaje*, pero no conoce el equivalente de *picture* en español (cuadro), por lo que puede haberse dado cierta relación semántica entre *picture*, *painting* y *paisaje*, unido ello a la coincidencia en los primeros sonidos de estas palabras. En cuanto a *Me dio los restos de su mate*, en español, *restos*, en plural, significa “residuos, sobra de comida”; *resto*, en singular, significa “parte que queda de un todo”, que es a lo que se

refiere el aprendiente. El aprendiente probablemente desconoce estos dos sentidos diferentes; al mismo tiempo, en inglés existe *rest* y *remains*; este último es un plural invariable que el aprendiente transfiere a *resto*, en lugar de hacer la transferencia directamente de su cognado *rest*.

Por otro lado, en los otros dos errores descartamos una influencia de la L1. No hay una explicación a por qué se confunde *habilidad* por *posibilidades*; *golpecitos* y *cachetes* o *azotes* sí que son semánticamente próximos, pero, dado el contexto, se requiere el empleo de un término específico para designar “el golpe [que da una madre a su hijo] en las nalgas con la mano”.

Por último, están los errores léxicos producidos por ausencia del núcleo del complemento, que es el elemento que manifiesta y controla las propiedades de todo el sintagma. Este es el caso de [9], [11] y [15]. En [9], *el número de personas que les da asco el cigarrillo*, el CI debe tener su núcleo, que es la preposición *a* que funciona como marca del sintagma que garantiza la identificación de la función oracional CI. En este caso, el SN *término* tiene un pronombre relativo que ha de ir precedido por un actualizador (*las* que les da), y, al estar en posición preverbal, el CI se reduplica con el pronombre *les*. En este caso, se da una interferencia del inglés: el alumno importa la ausencia de *a* en la estructura equivalente. En [11], el aprendiente produce *Para Adri*, en lugar de *A Adri*; al igual que en el uso anterior, no aparece la preposición *a* que funciona como el núcleo del CI; como puede suceder en español, en la L1 del aprendiente se están mezclando dos estructuras: *Compré un regalo para* (*I got a present for*) y *di un regalo a* (*I gave a present to*); así, ambas se fusionan erróneamente en *Para Adri, le dimos un peluche*, que en inglés sería *For Adry I gave this*, también erróneo. En el último uso, el [15], se suprime la conjunción *que* encargada de introducir la oración subordinada que funciona como suplemento de *darse cuenta*. Esta conjunción es suprimible en inglés, por lo que la L1 interfiere en la construcción.

Por lo que respecta a los 4 errores computacionales ([5], [10], [14] y [18]), son causados todos ellos por desviaciones en la restricción sintáctica del sustantivo. En *los homosexuales pueden dar el amor tan bien como los demás*, el HNN no conoce, al menos desde una perspectiva productiva, las restricciones sintácticas de los sustantivos continuos. En español, los nombres continuos, como *amor*, y el plural de los nombres discontinuos pueden denotar la extensión de los predicados correspondientes, esto es, pueden constituir por sí solos expresiones referenciales (Laca, 1999: 895-896). Así, *amor* denota la muestra de materia que clasificamos como amor –a este caso pertenece el uso no nativo del corpus–, y no se refiere genéricamente al tipo de materia en cuestión (como sí ocurre al comienzo de la oración), en cuyo caso el español sí requiere artículo; este es un

error intralingüístico. En [10], *siempre da 120 %*, se da una ausencia del actualizador (determinante). Las expresiones que designan porcentajes se forman con el número cardinal, que es un sustantivo con el complemento *por ciento*, y un actualizador, como los numerales fraccionarios a los que se asimila (MNGE. *Manual*, 2010: 403). Este error es interlingüístico, dado que en la L1 la construcción con porcentaje no lleva determinante actualizador. En [14], *me dio miedo un poco*, el aprendiente no conoce las construcciones pseudopartitivas, que se ajustan a la estructura general “expresión cuantificativa + de + grupo nominal escueto” –entendiendo por *escueto* sin determinantes ni cuantificadores–. Así, la construcción debería ser *me dio un poco de miedo*. La fuente de error más probable es la L1 del aprendiente, pues existe la construcción equivalente *it scared me a bit*. En el uso [18], *oportunidad* es un sustantivo contable, por lo que necesita un actualizador cuando va en singular, esto es, *la pareja puede darle una oportunidad*; este error es intralingüístico.

7.2.6. Descripción y explicación del error en el sujeto

Son solo 3 los errores en el sujeto, y los 3 son interlingüísticos:

[1] Por lo menos como profesora me ha dado la oportunidad de pasar los veranos en España (76 HNN)

[2] Otra vez me dió mucho gusto realizar que podía entender muy bien! (187 HNN)

[3] me dio mucho placer a verla con las niñas (214 HNN)

Los tres, además, son errores léxicos. En [1] se da la ausencia del sujeto, argumento externo exigido por el verbo. El aprendiente importa la forma de la L1 *As a teacher it has given me*; conoce la regla de que en español no aparece el sujeto en muchos contextos, por lo que no transfiere el *it* a la L2, aunque en este contexto sí se necesita. En [2], hay un error léxico de significado, pues *realizar* significa en español algo diferente a *darse cuenta*, que es lo que el aprendiente busca cuando traduce *realise* como *realizar*; se importa, por lo tanto, la forma. Por último, en [3] el aprendiente no demuestra saber que el infinitivo en español no necesita *a*, pese a que en inglés el infinitivo se construya con *to*; se importa la forma del sujeto en inglés.

7.2.7. *El error en el sintagma*

Son 9 los errores que afectan a toda la construcción y no solo a uno de sus elementos; es decir, en estos usos resulta imposible aislar los elementos concretos desviados. He aquí tales errores:

[1] no es difícil decidir que los niños lo merecen [el amor] de la gente quien se lo dará (47 HNN)

[2] los niños que tienen tanta imaginación y energía que te pueden dar una cosa muy sencilla como una hoja de papel y juegan con ella (54 HNN)

[3] Lo que no se dan cuenta de es que la Biblia tiene muchas otras leyes que nos parecen ridículas (79 HNN)

[4] Decidimos quedarnos en Roma, para conocerla a fondo. El resultado nos dio mucha alegría por dos razones (124 HNN)

[5] Un pareja homosexual puede durar tantos años como una pareja heterosexual. Si la sociedad les da el permiso de existir, como en España o en otros países (143 HNN)

[6] el diálogo dio a la película una forma muy buena, (182 HNN)

[7] que conociendo un idioma le da [a los niños] más valor en este mundo (184 HNN)

[8] nos miraron con unos ojos que no nos dijeron nada pero nos pidieron todo que no pudimos dar (195 HNN)

[9] el gasto y las preocupaciones por la inmigración ilegal dan un sabor amargo a lo que fue anteriormente la fuerza de Estados Unidos (209 HNN)

De entre estos errores que afectan al sintagma, se han clasificado 4 errores léxicos, en los que la construcción tiene un significado diferente al buscado por el usuario. Este es el caso de [4], [6], [7] y [9].

En [4] *El resultado nos dio mucha alegría*, no hay un elemento que pueda identificarse con el resultado (¿las vacaciones?). Cualquiera de las siguientes opciones, u otras similares, habría resultado adecuada: *Nos alegramos de haber tomado esa decisión; El viaje resultó estupendo; La decisión fue muy acertada* o *La decisión nos trajo muchas alegrías*.

Por lo que respecta al caso [6], tanto en inglés como en español se puede dar forma a objetos no materiales, pero no tiene sentido que un diálogo dé forma a una película.

En [7], *valor*, aplicado a personas, significa “coraje” u “osadía”. Cuando significa “grado de utilidad o aptitud de las cosas, para satisfacer necesidades”,

no se aplica a personas, ni en inglés tampoco. Por eso, en *dar valor a un niño* da la impresión que se cosifica *niño*; mejor *da oportunidades a un niño* o *lo hace más valioso*.

En [9], *sabor amargo* no está bien traído, ni tampoco el CI, porque *sabor amargo* no se relaciona con el desprestigio, sino con la amargura y el sufrimiento.

Aparte de estos errores léxicos en el sintagma, se han localizado 5 errores computacionales ([1], [2], [3], [5] y [8]), es decir, errores en los que los elementos se encuentran presentes pero desordenados ([3] y [5]); en algunos casos, el sintagma es incomprensible debido al modo en el que las piezas léxicas han sido unidas en el eje sintagmático (y no se puede corregir intercambiando elementos aislados) ([1] y [8]) o a que los elementos no se fusionan o coordinan de la manera adecuada, esto es, cuando las relaciones de concordancia de rasgos no son las adecuadas dentro del sintagma [2]).

En [1], una opción más adecuada sería *los niños merecen amor, sea quien sea quien se lo dé*; y en [8], *nos pedían aquello que no podíamos darles*. En [5], se da un error en la coaparición del verbo *existir* y *permiso*; no se puede dar un permiso para existir. En su lugar cabrían otras opciones, como *debería dar el derecho de declararse* o *de reconocerse*. En [2], en lugar de *te pueden dar una cosa*, debería decirse *les puedes dar una cosa*, es decir, que las marcas funcionales de concordancia del verbo con el sujeto aparecen en el pronombre cuya función es de CI, y la persona y número que debería tener el CI (3.^a del plural) aparece en la flexión del verbo. Y, por último en [3], se presenta una oración hendida o escindida, que contiene la cópula, el constituyente escindido y la relativa, en este caso en un proceso de expansión que consiste en lo siguiente: separar o escindir el constituyente que desempeña la función de suplemento, poner en su lugar el relativo que desempeña la misma función y luego añadir el constituyente escindido al principio precedido o seguido de la cópula o al final precedido de la cópula. Hay varios tipos de estructuras hendidas; aquí tenemos la relativa en primer lugar, luego la cópula y en último lugar el constituyente escindido. Como señala Moreno Cabrera (1999: 4247-4280), una característica esencial de estas estructuras es “la existencia de la equivalencia funcional entre el constituyente escindido y el pronombre relativo de la correspondiente relativa. Ambos deben tener exactamente los mismos índices de función, si han de presentarlos”. De esta manera, el error se produce porque el suplemento ha de ir introducido por la preposición regida por el verbo, la preposición *de*, tanto en la relativa como al final, cuando recuperamos el constituyente escindido tras la cópula.

7.3. CONCLUSIONES SOBRE EL USO DE DAR

En este capítulo se han examinado las diferencias y semejanzas entre el uso del verbo *dar* que hacen los HNN y el que hacen los HN a partir de las muestras obtenidas en sus respectivos corpus. Los resultados de esta investigación cuantitativa y cualitativa sugieren que, aunque este uso no difiere drásticamente en ambos grupos, no puede decirse que los aprendientes de nivel avanzando dominen este verbo frecuente del todo.

7.3.1. *Diferencias y semejanzas por estructuras*

Para empezar, el verbo, en general, es menos utilizado por los HNN que por los HN, como preveía Sinclair (1991) y como –a propósito del verbo *to make*– demuestran también los resultados en Altenberg y Granger (2001). No obstante, el estudio contrastivo de las estructuras sintácticas revela que los aprendientes sobreutilizan las ditransitivas, al tiempo que usan por igual las transitivas y usan menos el resto de las estructuras, ya sea por desconocimiento o porque, conscientes de su complejidad, las evitan. En relación con esto, Cobb (2003: 403) mantiene que cuando se requiere del aprendiente la producción de una tarea escrita sobre un tema para un lector anónimo, este tiende a emplear el léxico restringido de la oralidad; esto explicaría que se sobreutilicen las ditransitivas y que no se recurra a una gama más amplia de estructuras, característica de la escritura nativa. De acuerdo con el estudio de Cobb (2003), no podemos descartar que los HNN conozcan estas estructuras; ya se ha señalado que, cuando las frecuencias son bajas en el corpus nativo, es muy difícil alcanzar conclusiones indiscutibles sobre el conocimiento de una estructura dada por parte de toda la comunidad de HNN. El hecho de que las estructuras pasivas, las pronominales, las construcciones absolutas, las preposicionales, las introducidas por *como*, los adjetivos o los esquemas fraseológicos semiproductivos no sean utilizadas con mucha frecuencia por los HN implica que posiblemente no sean tan necesarias para los HNN a la hora de escribir; no obstante, si estos fueran capaces de producirlas, su español sería mucho más expresivo de lo que lo es ahora, sin ellas. Algunos de los usos de estas estructuras son más bien individuales¹⁹¹ y no característicos del grupo, por lo que estas

¹⁹¹ Ya se ha señalado que los usos de estas estructuras que son menos frecuentes también entre los HN se corresponden con usos individuales de aprendientes cuya actuación demuestra un

estructuras poco empleadas constituyen un foco de desviaciones entre la interlengua y la L2.

Nos referimos concretamente al uso de la pasiva con *se* y al de las intransitivas con el verbo empleado con el sentido pleno de “proporcionar”; el 100 % de las intransitivas de los HNN son utilizadas con el verbo de apoyo combinado con sustantivos que denotan “emociones y estados anímicos” –lo que refuerza la idea de que los HNN se implican personalmente más en sus textos, presentan más referencias al estado mental, físico y anímico–, mientras que los HN utilizan este tipo en un 57 % de los casos, y en el 43 % restante usan *dar* como “proporcionar”, principalmente en la combinación *dar tiempo*. Asimismo, tampoco se puede decir que el grupo de HNN utilice las pronominales; en concreto, el tipo de pronominal más empleado por los HN es el de *darse* + SN –con el significado de “suceder o existir”–. Ni tampoco hacen uso de las diferentes construcciones preposicionales en las que participa *dar*; así, los nativos utilizan la construcción *dar por* + adjetivo o participio, *dar por* + infinitivo, *dar contra*, *en* o *a* + SN y *dar con* + SN. Lo mismo sucede con los esquemas fraseológicos semiproductivos; los nativos emplean *dar a* + infinitivo con el sentido de “ejecutar acciones que denotan conocimiento o difusión de informaciones y percepción e intelección”; usan también el esquema *dar de* + infinitivo, en el sentido de “ejecutar acciones que denotan ingestión”, y *dar que* + infinitivo con el sentido de “provocar o dar motivo para que se haga una acción” (lo que el infinitivo expresa), usados normalmente con *comentar*, *hablar* y *pensar*; por el contrario, los HNN solo emplean un esquema que no es ni siquiera el preferente entre los HN: *dar a* + infinitivo de verbos que denotan ingestión, en lugar de *dar de*.

Otras estructuras que sí son utilizadas en una frecuencia alta, como son las locuciones, se usan con menor variedad, esto es, se usan menos tipos de locuciones diferentes (mientras que los HN emplean 10, los HNN usan la mitad), lo que demuestra que el español de los HNN es menos amplio en lo que se refiere a esta estructura.

En otras estructuras usadas también con frecuencia (como las impersonales o las ditransitivas) también se observan divergencias al analizar las concordancias. Las impersonales no se emplean de manera similar. Ya vimos que los HNN usan más la 1.^a persona del plural –por esa tendencia a involucrarse más

dominio lingüístico por encima de la media, lo que no resulta sorprendente, dado que se trata de estructuras formalmente más complejas y más infrecuentes, por lo que las posibilidades de adquirirlas a partir del *input* y contrastar sus hipótesis de uso a través del *output* son mucho menores.

personalmente en la narración– y el infinitivo; en cambio, los HN usan más la 3.^a del plural y la 2.^a del singular. Mientras que esta última resulta demasiado informal para el tipo de texto analizado –lo que confirma que el corpus nativo no puede ser utilizado como única norma para juzgar los usos no nativos–, estos resultados muestran que para que no haya desviaciones entre la interlengua y la L2 entorno a este aspecto, los HNN deberían usar más la 3.^a persona del plural o la impersonal con *se* en lugar de la 1.^a persona.

7.3.2. Diferencias y semejanzas por usos de *dar* en sus dos significados

Si nos adentramos en las ditransitivas y transitivas, comprobamos, en primer lugar, que se utilizan prácticamente las mismas clases léxicas, sobre todo cuando el verbo actualiza el sentido de “proporcionar”. Con el sentido de “producir”, los no nativos utilizan 6 clases menos, pero, como estas son las clases menos utilizadas por los HN –de hecho, se corresponden con usos individuales–, no representan el español de la comunidad de HN. Así, estas ausencias en el corpus no nativo no provocan una diferencia drástica entre el español de los dos grupos.

No obstante, pese a esta semejanza, algunas clases semánticas no se emplean de igual modo. Con respecto al sentido de “proporcionar”, los HNN sobreutilizan *dar la oportunidad*, lo que podría evitarse si conocieran alternativas como *ofrecer la posibilidad*. Ambos grupos utilizan la misma cantidad de combinaciones diferentes con *dar* con el sentido de “proporcionar” (en torno al 55 % de los usos), pero solo coinciden en el 14 % de los sustantivos empleados. Así, el uso de *dar* como “proporcionar” es igual de amplio en ambos grupos, pero al mismo tiempo hay diferencias: los HNN utilizan menos que los HN los sustantivos en sentido figurado: mientras que los HNN usan en sentido figurado *vida, toque, sabor* (el uso figurado que se hace de este sustantivo es erróneo) y *forma*, los HN usan *vida, toque, deajo* (por sabor) y *forma*, y, además, *rumbo, luz verde* y *aliento*. No obstante, esta diferencia no es significativa. Estos resultados son aplicables también al uso de *dar* con el sentido de “producir”, pero las diferencias resultantes en el uso de *dar* como verbo de apoyo sí son estadísticamente significativas. Aunque el conocimiento productivo de los HNN con respecto al sentido de “producir” es, en general, igual de extensivo (los HN emplean un 47,9 % de usos diferentes y los HNN un 48,4 %), solo coinciden en el 14 % de los sustantivos que utilizan. Varios aspectos llaman la atención. En primer lugar, los HN emplean el verbo con el sentido de “producir” con una gran cantidad de

sustantivos en sentido figurado, en concreto, 23 usos (*vuelta de tuerca, paso(s), vueltas, patada, golpe, giro, salto, caricia*), mientras que en el corpus no nativo tan solo se documenta una aparición (*ataque de nervios*); en porcentajes supone un 21,4 % frente a un 1,6 % de usos figurados. Como vemos, la mayoría de los sustantivos que se interpretan figuradamente en el corpus nativo pertenecen a la clase léxica de los que denotan “desplazamiento”: 19 de los 27 usos que pertenecen a esta clase son figurados, esto es, el 70,4 %, mientras que de los 13 usos no nativos correspondientes a dicha clase, ninguno se emplea figuradamente –pese a que los sustantivos utilizados son prácticamente los mismos–. En segundo lugar, los HN emplean el verbo de apoyo con el sentido de “ejecutar” en un 76 % de los casos, mientras que los HNN lo usan en un 50,8 %* ($p < 0,05$). Esto quiere decir que los HNN sobreutilizan el verbo de apoyo con el sentido de “suscitar(se)”, pues lo usan un 49,2 %* ($p < 0,05$) frente al 24 % nativo. Con respecto a las clases léxicas seleccionadas por el verbo, los HNN utilizan significativamente más el grupo de sustantivos que denotan “miedo, inquietud, zozobra y otros estados anímicos similares”, aunque el conocimiento extensivo de estos sustantivos es menor (menos sustantivos, más repetidos); asimismo, también sobreutilizan la clase léxica de los sustantivos que designan “diversas sensaciones, emociones y estados anímicos generalmente percibidos como placenteros”, lo que concuerda con la idea reflejada en el trabajo de Cobb (2003) sobre el mayor grado de implicación personal –referencias al estado mental físico y anímico– de los aprendientes en las producciones nativas cuando se comparan géneros equivalentes.

7.3.3. El papel de la transferencia de la L1

Para terminar, nos referimos brevemente al papel de la L1 en las desviaciones del uso del verbo *dar*. Un análisis como el presente no nos permite asegurar si la influencia de la L1 tiene, en efecto, lugar. De hecho, como observa Nesselhauf (2005: 197), determinar el peso de la influencia de la L2 y de la L1 en la producción de los usos desviados del verbo *dar* es difícil porque la influencia de L2 es directa y la de la L1 es indirecta. La semejanza entre L1 y L2 se considera en este análisis como una indicación de que la influencia es probable.

Los resultados acerca del papel de la L1 en la adquisición del léxico son contradictorios. Hay quienes subrayan que la influencia de la L1 no es muy importante en el ámbito del léxico (Martin, 1984); en cambio, otros afirman que el léxico es de todos los niveles lingüísticos el más susceptible de ser afectado por

la transferencia (Ellis, 1994). En este estudio empírico se da la razón a los dos argumentos, pues la influencia de la L1 parece depender del tipo de léxico investigado. De los errores analizados a propósito del verbo *dar*, algunos son el resultado de la transferencia de la L1 (12,5 %), pero la mayoría son intralingüísticos, debidos a una confusión en la forma, significado, contexto o combinatoria de la palabra en la L2. Por el contrario, en las desviaciones del complemento, el porcentaje asciende considerablemente, hasta alcanzar el 47,4 %; y en las del sujeto, la L1 está presente en el 66,7 % de los casos.

En las desviaciones centradas en el propio verbo, llama la atención que la L1 está presente en muy pocos tipos de error (mientras que en las desviaciones del complemento se localiza en prácticamente todos los tipos de error), tan solo en los usos inadecuados por la forma y por la restricción semántica. Así pues, de los 14 errores léxicos, solo 2 son interlingüísticos: en ambos se produce una importación de la forma del infinitivo inglés y del gerundio cuando el sujeto de estas formas verbales no se interpreta de forma específica. Así pues, el 14,3 % de los errores léxicos son interlingüísticos, frente al 85,7 % que corresponde a los intralingüísticos.

De los errores computacionales, el 11,5% (3 errores) es interlingüístico, frente al 88,5 % de errores intralingüísticos. Llama la atención que los tres errores computacionales se deben a una selección semántica del verbo equivocada; esos tres errores conforman el 50 % de los usos de la selección semántica, lo que quiere decir que en el uso sintagmático del verbo *dar* (a la hora de proyectarse un sintagma) la L1 está detrás solo de los errores de selección semántica que se producen en el verbo *dar*; ahora bien, su interferencia es importante, pues es la causa de la mitad de estos errores (en el trabajo de Nesselhauf [2005] la influencia de la L1 se encuentra en el 50 % de las colocaciones inapropiadas, mientras que los errores intralingüísticos suman el 70 % de los casos).

Dicho esto, son varios los casos en los que se produce en la lengua no nativa algo similar a elementos de la L1 y algo similar a elementos de la L2; resultaría sorprendente que se tratara de una coincidencia. Esto sucede, por ejemplo, en la construcción de participio absoluto, la cual se construye en inglés con el participio *given*, siendo este invariable; por ello, al mismo tiempo que el hablante desconoce las características formales de esta construcción en español (concordancia sujeto y predicado verbal), de manera indirecta se puede estar transfiriendo también la L1 (y el sustantivo invariable no marcado en español es el masculino).

8. DISCUSIÓN

En este capítulo se lleva a cabo una revisión metodológica a través de la cual se ponen en relación los resultados del análisis con las hipótesis de partida de esta investigación (8.1); se exponen las principales limitaciones (8.2) y se apuntan varios estudios que sería interesante realizar para completar la presente investigación (8.3).

8.1. REVISIÓN DE LA METODOLOGÍA DESARROLLADA EN ESTA INVESTIGACIÓN

8.1.1. El enfoque cuantitativo y cualitativo. Hipótesis de partida contrastadas y conclusiones sobre los resultados obtenidos

En esta investigación de CAI hemos utilizado un enfoque cuantitativo y cualitativo de los datos. La relación de los índices de frecuencia en los dos corpus ha constituido la primera parte de cada uno de los apartados que conforman el capítulo 7 y, al mismo tiempo, ha supuesto un punto de partida para análisis posteriores cuantitativos y cualitativos: por un lado, una frecuencia similar no garantiza un uso detallado similar; además, tras los índices de frecuencia a menudo se esconden errores; por otro lado, este análisis nace de la necesidad de contrastar una serie de hipótesis –consideradas a partir de los datos observados en el aula, de nuestro razonamiento lógico y de nuestra intuición– y concluye con una serie de generalizaciones sobre la adquisición de *dar* por parte de un grupo concreto de aprendientes de español establecidas a partir de los resultados que se han obtenido.

Como señalábamos en el apartado 2.1.3, la contribución de la investigación de corpus de aprendientes ha sido más substancial en la descripción que en la interpretación de los datos referidos a la ASL (Granger, 2004), más centrada en un acercamiento pedagógico a la ASL, con pocas referencias a los debates, hipótesis y teorías actuales sobre la ASL y sus implicaciones en el desarrollo de la lengua del aprendiente (Myles, 2005, citado en Lozano y Mendikoetxea, 2013: 1-2). Pues bien, esta investigación no sigue esta tendencia, pues se originaba a partir de una serie de explicaciones plausibles en torno a los problemas de uso percibidos en los HNN o intuídos con respecto al verbo *dar* en relación con la interfaz léxico-sintaxis, que es uno de los centros de interés reciente en la ASL, pues es considerada como causa potencial de desviaciones en el desarrollo de la interlengua de los aprendientes (Lozano y Mendikoetxea, 2008

y 2010)¹⁹². Estas hipótesis han sido planteadas a lo largo de la investigación y eran contrastadas empíricamente en el capítulo 7. Nos referimos a ellas a continuación, y también a su confirmación o refutación.

1) Si atendemos a la estructura del sintagma verbal, los verbos establecen relaciones más estrechas con sus complementos (normalmente el tema) que con sus sujetos (según se explica en los trabajos de Haegemanm [1992] y Escandell [2004], y también a partir de nuestras inferencias del Esquema de Imagen del CAMINO, de Lakoff y Johnson [1999]), por lo que el CD está sujeto a más restricciones semánticas y sintácticas, lo que potencialmente convierte al CD en un elemento de más difícil manejo para un aprendiente de L2 (esta hipótesis se ponía de manifiesto en el capítulo 6, en el apartado sobre la localización del error [6.2.5.1]). Los datos analizados confirman esta hipótesis; de los 11 errores localizados en dar en las estructuras transitivas y ditransitivas –que son las que al mismo tiempo tienen sujeto y CD–, 5 errores se producen en la restricción semántica del CD y 1 en las restricciones sintácticas del CD, por lo que suman 6 –más de la mitad– los errores de dar relacionados con el CD. Con todo, todavía son 5 los errores combinatorios producidos en la unión entre el verbo y el sujeto, si bien todos los casos se deben a la falta de concordancia del número y persona entre ambos elementos –lo que resulta llamativo si tenemos en cuenta que esta concordancia se contempla en el inventario gramatical del nivel A1 del PCIC–. No obstante, lo que aquí nos interesa es que no hay ningún error que esté provocado por las restricciones semánticas o sintácticas entre el verbo y el sujeto. Aun así, si hubiéramos podido recuperar todas las combinaciones léxicas en las que los HNN han utilizado erróneamente un verbo diferente a *dar* en lugar de este, el porcentaje de error en el CD habría sido probablemente más elevado.

2) En relación con la anterior hipótesis, al comienzo de la investigación (en el capítulo 1 y, de nuevo, en el 4) exponíamos que los errores en el verbo *dar* producidos al fusionarse el verbo con el complemento (en las estructuras V + N) serían posiblemente semánticos en su mayoría, puesto que la red argumental del verbo no es compleja desde una perspectiva sintáctica (tanto intralingüísticamente como interlingüísticamente), pero sí que lo es desde una perspectiva semántica (téngase en cuenta la gran cantidad de clases léxicas que selecciona este verbo). Así, al hablante, para dominar el verbo *dar*, no le basta con saber que selecciona un sustantivo como CD precedido o no de un determinante que lo actualice (restricción sintáctica), sino que además ese

¹⁹² En numerosos estudios se discute sobre las combinaciones léxicas (en concreto se refieren a las colocaciones) como foco de problemas para los aprendientes de lenguas (Granger, 1998; Lewis, 2000; Nesselhauf, 2004, 2005; o Wanner *et al.*, 2013, entre otros).

sustantivo ha de tener determinados rasgos semánticos que se coordinan con ciertos rasgos subléxicos del verbo (restricción semántica). Como se ha expuesto con respecto a los resultados obtenidos en la anterior hipótesis [en la 1], 5 son los problemas semánticos del complemento, frente a 1 problema sintáctico. Queda así, también, confirmada esta hipótesis.

3) El verbo *dar* participa en muchas estructuras sintácticas diferentes a la prototípica V + N, para la mayoría de las cuales no existe un equivalente exacto en la L1 de los aprendientes. Esta complejidad, que es tanto intralingüística como interlingüística, debería de causar desviaciones en el grupo nativo (como anunciábamos en el capítulo 1 y exponíamos con mayor detalle en las tablas del capítulo 4). De nuevo, esta hipótesis se confirma tras el análisis; por un lado, se ha demostrado que los HNN utilizan significativamente menos las estructuras que no se corresponden con la construcción V + N; esto es, las locuciones, construcciones absolutas, pronominales, preposicionales, esquema fraseológicos semiproductivos y el adjetivo son usadas por los HN en un 12,2 %, mientras que los HNN apenas las usan, tan solo en un 3,7 % ($p < 0,05$). Por otro lado, el porcentaje de error en estas estructuras es muy alto, si bien los datos obtenidos son demasiado bajos para considerarlos significativos. Por todo ello, los resultados obtenidos (en cuanto a las frecuencias de uso y el índice de errores) sugieren que los HNN podrían estar evitando usar estas estructuras.

4) Como ya se ha señalado, la construcción V + N con *dar* no implica dificultad desde una perspectiva sintáctica, pero sí que es compleja desde una perspectiva semántica, dado que se trata de un verbo con una capacidad extensional muy alta. Dentro de esta construcción, considerábamos que las CVA deberían de ser más problemáticas porque son más opacas que las construcciones en las que participa el verbo con un sentido pleno, aunque figurado, motivo por el cual consideramos que hay más posibilidades que en la L1 se seleccione un verbo de apoyo diferente en estas CVA (como avanzábamos en el capítulo 1 y desarrollábamos en el 4). Los resultados obtenidos confirman que hay más divergencias entre el uso nativo y no nativo en las CVA que en las construcciones en las que el verbo actualiza el sentido pleno de “proporcionar” (las CVP). No obstante, el porcentaje de error es muy similar en ambos casos. Aun así, no hemos podido determinar si hay más casos de error en la restricción semántica de las CVA que en las restricciones de las CVP (lo que era el fundamento de nuestra hipótesis de partida), dado que el corpus solo nos permite analizar los usos de *dar* que han sido utilizados en él, y no los casos en los que debiera aparecer el verbo *dar* pero en su lugar ha sido seleccionado otro verbo (como señalábamos al respecto de la hipótesis [1]); es decir, solo contamos

con los problemas en la restricción semántica de dar por su presencia, y no por su ausencia. Por lo tanto, los casos de selección semántica errónea detectados no se pueden categorizar como errores semánticos de las CVA ni como errores semánticos de las CVP –pues precisamente se trata de sustantivos que no son seleccionados por el verbo *dar-*, ni podemos conocer los errores semánticos que sí se han producido en el corpus por la ausencia del verbo *dar*. Para poder confirmar esta hipótesis habría que realizar un análisis manual de los datos, y solo así podríamos determinar si la L1 está detrás de los errores en la restricción semántica de las CVA.

Como se puede observar, todas estas hipótesis –inspiradas en los datos conseguidos en el aula, en nuestro razonamiento lógico o intuición– surgen en torno a la interfaz léxico-sintaxis del verbo, que, como hemos señalado, es uno de los asuntos más debatidos en ASL. Una parte importante de los capítulos 3, 4 y 6 está dedicada a esta interfaz; en el 3 se presenta dicha interfaz, se describe y se contextualiza dentro de la competencia léxica del verbo junto a los demás contenidos que la integran; en el capítulo 4 se profundiza en la naturaleza léxico-sintáctica del verbo dar, se proporcionan unas tablas que recogen todas las estructuras sintácticas en las que participa dar utilizadas por los participantes en el corpus y todas las clases semánticas seleccionadas por el verbo; y en el 6 diseñamos una tipología del error que da cuenta de todas las especificaciones de carácter léxico-sintáctico observadas en el capítulo 4. Nuestros resultados confirman que esta interfaz sí que es fuente de importantes déficits en la producción de este grupo de HNN.

Por otro lado, y en esto sí coincidimos con la tendencia general de los estudios de interlengua, este análisis del verbo sirve en el intento de construir hipótesis en la investigación sobre la ASL. Las más destacadas de esta investigación son las siguientes:

- 1) Los HNN infrautilizan el verbo *dar* acaso por el desconocimiento de muchas de las estructuras en las que aparecen –lo que puede implicar que erróneamente se use otro verbo, frecuente (Chi Man-lai et al., 1994) o no, donde debería emplearse *dar-*, y posiblemente también porque las eviten, conscientes de la compleja combinatoria de este verbo. Al mismo tiempo sobreutilizan las estructuras ditransitivas, lo que se podría explicar a partir de la frecuencia con la que aparece esta estructura en el input junto con el hecho de que existe una tendencia entre los HNN a usar en la escritura libre el lexicón restringido propio de la oralidad (Cobb, 2003), en lugar del extensivo característico de las argumentaciones escritas.

2) Sobreutilizan el verbo de apoyo cuando este actualiza el sentido de “suscitar(se)”, concretamente en el uso que hacen de *dar* combinado con sustantivos que denotan “miedo, inquietud y otros estados anímicos similares” y con sustantivos que designan “diversas sensaciones, emociones y estados anímicos percibidos como placenteros”; esto puede explicarse porque los HNN tienden a implicarse más personalmente en sus textos que los HN (Cobb, 2003).

3) Aunque el conocimiento productivo de la estructura V + N es igual de amplio en ambos grupos, resulta que los HN utilizan el verbo de apoyo en un 21 % con sustantivos figurados mientras que los HNN solo lo utilizan un 1,6 %. La producción de sustantivos figurados con *dar* es posiblemente más compleja para los aprendientes, dado que la proyección metafórica que hace una comunidad de hablantes a partir de un sentido básico no tiene por qué coincidir con la de otra comunidad; aunque la comprensión resulte lógica y natural, la producción puede presentar más problemas. Por medio de un análisis manual se podría comprobar si los hablantes directamente evitan los usos de sustantivos figurados o si los usan en coaparición con otro verbo de apoyo, en lugar de con *dar*; esto permitiría determinar la influencia de la L1 en estas CVA, como ya se ha señalado.

Estas interpretaciones acerca de los resultados obtenidos –que se fundamentan en nuestros conocimientos en otras ramas de la lingüística aplicada, como la ASL, la psicolingüística y la experiencia docente en ELE– abren nuevas vías para futuras investigaciones en los estudios de interlengua.

8.1.2. La identificación de las necesidades de los HNN a partir de las semejanzas y diferencias en el español nativo y el español no nativo. La necesidad de una norma prescriptiva y de una matización de los sobregeneralizadores conceptos de infrantilización y sobreutilización

El CEDEL2, que está bien diseñado de acuerdo con los criterios establecidos (Sinclair, 2005), proporciona una base empírica sólida para la descripción de la interlengua. La comparación es la esencia de este estudio, dado que el principal objetivo de esta investigación de interlengua era identificar las semejanzas y diferencias entre el grupo nativo y el grupo no nativo del CEDEL2 para descubrir las necesidades de los HNN; por ello, este estudio se enmarca en el enfoque metodológico del ACI. Así, resulta de enorme utilidad disponer de un corpus de control nativo de iguales o muy similares características al no nativo, que sirve de norma descriptiva, pues permite, al contrastar las frecuencias de los usos de ambos corpus, determinar qué elementos se utilizan con igual frecuencia

en los HN y en cuáles hay desviaciones. Como hemos visto en el capítulo 5, además de esta norma descriptiva es necesario la aplicación de una norma reguladora que precise el grado de adecuación de los usos no nativos (recuérdese que el corpus no nativo presenta usos que oscilan entre los aceptables y los no aceptables), en tanto que el contraste con el corpus de control, como se ha dicho, no siempre resulta suficiente. Piénsese, por ejemplo, en el uso de la 2.^a persona del singular para la construcción de oraciones impersonales, nada representativo de los textos argumentativos propios de la escritura nativa especializada, y que, en cambio, aparece con relativa frecuencia entre los HN del CEDEL2 (en el 20 % de los usos impersonales). Otro ejemplo es el del *queísmo*; aunque su presencia en el corpus de control sea relativamente amplia, tanto los focos especializados como un buen número de hablantes lo señalarían como no adecuado.

Para un corpus como el CEDEL2, en el que tanto los HN como los HNN manejan referentes diatópicos distintos, la unión de los estándares peninsular y americano en un estándar panhispánico constituye un referente lingüístico operativo para categorizar los datos de la interlengua. En los casos en los que el investigador alberga dudas sobre la estandarización de algún elemento y sienta la necesidad de contrastar sus intuiciones, puede recurrir a la combinación de varias herramientas, como son los corpus de referencia más amplios (CREA y el Corpus del español), diccionarios como *Redes*, *DRAE* o el *DPD*, o la *NGLE* (en el capítulo 6 se ha abordado con profundidad cada una de las fuentes seleccionadas para contrastar los datos, y los criterios de uso establecidos para el análisis).

Por otro lado, los estudios del ACI pueden proporcionar una descripción mejorada de la interlengua de este grupo de HNN si se consideran los siguientes procedimientos: 1) la aplicación a los sobregeneralizadores conceptos de infrautilización y sobreutilización de los diferentes esquemas de relaciones de frecuencias entre los dos corpus (adaptados de Guo [2006]) contribuye a describir la lengua del aprendiente con mayor precisión; 2) la aplicación a estos esquemas de un test de significatividad (Chi-cuadrado) permite conocer con exactitud si las diferencias entre los resultados de HN y HNN son estadísticamente significativas; y 3) la aplicación a este análisis de las frecuencias de un análisis del error que lo complementa permite determinar con precisión el grado de acierto y de error en el uso del verbo *dar* en los HNN; además, gracias a la tipología del error propuesta, es posible identificar, localizar, describir y explicar cada una de las desviaciones.

8.2. LIMITACIONES DE LA INVESTIGACIÓN

A lo largo de esta investigación se ha demostrado que analizar la lengua del aprendiente por medio de un enfoque comparativo basado en corpus produce importantes resultados en la investigación sobre ASL. No obstante, existen ciertas limitaciones en este enfoque metodológico.

1) En primer lugar, es evidente que solo podemos contestar de forma científica a las preguntas sobre la interlengua del español si disponemos de corpus de HNN de español bien diseñados. Son todavía escasas las bases de datos de interlengua del español fiables, contrastables, que estén diseñadas de acuerdo con los criterios de calidad propuestos por Sinclair (2005) que aseguran el control de las variables que subyacen a la comparación. El CEDEL2 reúne estos criterios; no obstante, la experiencia del análisis nos lleva a concluir que no es un corpus lo suficientemente amplio. La muestra de CEDEL2 a la que tuvimos acceso en otoño de 2010 era de 370000 palabras, de las cuales unas 172000 palabras pertenecían al corpus de aprendientes de nivel avanzado y aproximadamente 198000, al corpus nativo. Esta cantidad prácticamente coincide con el tamaño de cada uno de los dieciséis subcorpus del ICLE –el corpus de aprendientes del inglés más empleado en los estudios de interlengua–, que constan de 200000 palabras, cantidad considerada en el ACI como suficiente para realizar análisis válidos (Lozano y Mendikoetxea, 2013: 16). El CEDEL2 estaba entonces (y sigue estando) en proceso de compilación de datos; se pretende alcanzar las 750000 palabras no nativas –por lo que cada subcorpus tendrá 250000 palabras (correspondientes al nivel inicial, intermedio y avanzado)– lo cual, según los expertos, constituye, como se acaba de señalar, un tamaño suficiente como para aportar un panorama fiable de cualquier uso lingüístico. Posiblemente, esta asunción deba ser matizada. A juzgar por nuestra experiencia, el tamaño adecuado de un corpus está determinado, en última instancia, por el tipo de análisis que se vaya a realizar, esto es, por el uso lingüístico que vaya a ser examinado. Si, como es nuestro caso, se realiza un análisis en el que se requieren numerosas segmentaciones de los datos obtenidos para su clasificación en diferentes categorías –como es el caso de los usos de *dar* en cada una de las estructuras sintácticas, o de *dar* combinado con las diferentes clases léxicas, o incluso en lo que se refiere a *dar* en relación con la localización del error y el tipo de error–, en ocasiones se extraen conclusiones a partir de números demasiado bajos, con lo que no es posible hablar de resultados representativos, significativos o, en definitiva, generalizables. Es imprescindible,

por lo tanto, que el CEDEL2 –y los demás corpus de aprendientes que se están desarrollando y se desarrollarán en los años venideros– reúna más de 1000000 palabras para que se pueda responder científicamente a cualquier pregunta que nos podamos formular sobre la interlengua del español. Por lo tanto, en estos primeros tiempos de desarrollo de los estudios de la interlengua del español basados en corpus, hemos de asegurarnos en el campo de la ASL de que las reivindicaciones y limitaciones metodológicas sentidas y manifestadas por los investigadores en relación con los corpus que ahora mismo están siendo diseñados y compilados sean atendidas desde el principio para iniciar esta andadura correctamente.

2) Un análisis de corpus proporciona información sobre la interlengua y el proceso desarrollado en la producción del verbo; no obstante, las inferencias sobre este proceso fundamentadas únicamente en datos producidos de manera natural deben ser interpretadas como posibilidades, pues los investigadores dependemos de los datos de producción, por lo que estamos restringidos a una cantidad de datos limitada, en la que solo se atiende a aquellos datos que pueden ser contados y se deja fuera los que no son contados; por ejemplo, el hecho de que no hayan podido ser contabilizados o recuperados aquellos usos donde debiera haber sido empleado el verbo *dar* y en su lugar aparece otro verbo nos ha impedido determinar si hay más casos de error en la restricción semántica de las CVA que en las restricciones de las CVP, lo que constituía una de nuestras hipótesis de partida cuya confirmación o refutación solo se obtiene a partir de un análisis manual de los datos.

Además, la ASL siempre requerirá otras fuentes de datos, como los experimentales, los metalingüísticos y los de la introspección (que se usan tradicionalmente en la SLA), para contrastar los resultados obtenidos del análisis de corpus y para triangular los resultados y obtener así resultados más convincentes (Guilquin y Gries, 2009; Mendikoetxea y Lozano, 2012 y 2013):

The triangulation of corpus methods with other research methodologies will be an important further step in enhancing both the rigour of corpus linguistics and its incorporation into all kinds of research (...) To put it another way, the way ahead is methodological pluralism (McEnery y Hardie, 2012: 227).

En algunos trabajos recientes sobre la interlengua del español esta triangulación metodológica ya se está aplicando; en estos estudios se combina el análisis de corpus y el análisis experimental psicolingüístico, como ocurre en

Domínguez *et al.* (2013). Es preciso que el resto de los estudios sobre la interlengua del español terminen de girar en esa dirección¹⁹³.

3) Otra de las limitaciones sentidas en esta investigación radica en el hecho de que el CEDEL2 no está todavía anotado. Si el corpus tuviera anotados los errores ortográficos y estuviera lematizado, tendríamos garantías de que hemos recuperado en el análisis todos los usos de *dar* contenidos en el corpus, y no hubiéramos tenido que dedicarle tiempo a la tarea de descartar los usos no relevantes obtenidos de la búsqueda de *dar* (la preposición *de*, la conjunción *dado que* o el sustantivo *dado*, entre otros).

En cambio, nos resulta beneficioso que el corpus no estuviera etiquetado con los errores de tipo léxico. Ya se ha señalado que no existe un estándar comúnmente aceptado que nos permita determinar con precisión lo que es un error y lo que es aceptable, ni lo que es natural frente a lo que es poco natural o restringido, por lo que, en definitiva, existe una falta de sistematicidad a la hora de usar en el análisis de interlengua un concepto de *norma* (excluimos la ortografía de esta consideración, pues sus normas, sencillamente, pueden ser conocidas o desconocidas por los aprendientes). Esta falta de rigor limita la eficacia de los corpus etiquetados con errores. No nos oponemos a la práctica de la etiquetación del error, pero esta debería reevaluarse, pues se ha de ser cauto si se trabaja con un corpus etiquetado con errores. De hecho, si el CEDEL2 hubiera presentado los errores léxicos etiquetados, habríamos tenido que realizar nuestra propia propuesta metodológica en relación con la identificación del error para evitar las limitaciones descritas.

4) Aunque el nivel lingüístico en el que se clasifican los HNN del CEDEL2 –inicial, intermedio o avanzado– es fruto de la aplicación de una prueba de diagnóstico estandarizada (de la Universidad de Wisconsin) –a diferencia de otros corpus, como el SPLLOC, en los que se clasifican los aprendientes de acuerdo con el nivel educativo en el que se encuentran–, el nivel de dominio no está basado en los niveles marcados por el *MCER*, por lo que desconocemos los criterios y los parámetros establecidos para evaluar las redacciones; la fiabilidad es, por lo tanto, menor. De hecho, nos ha sorprendido que algunas redacciones que hemos analizado hayan sido clasificadas como nivel

¹⁹³ De alguna manera, este planteamiento ya existía cuando Vázquez (1991) y Fernández (1997) llevan a cabo –tras el análisis de los errores extraídos de los textos– una entrevista posterior con los aprendientes para constatar que las hipótesis formuladas una vez realizado el análisis sobre el origen de los errores eran ciertas. Algunos años antes, se publicaba la investigación de Linnarud (1986), uno de los primeros trabajos de interlengua inglesa en el que se recurre a una combinación del análisis de redacciones, esto es, lengua natural, con los datos procedentes de test experimentales de elicitación.

avanzado y no intermedio. Esta situación es aplicable a los corpus de interlengua del inglés, como señalan Tono (2003: 801) o Guo (2006: 43-44), entre otros:

In most studies, the criteria are not specified for what counts as “advanced level” or “intermediate level” or “novice level”. Among the few studies in which the criteria are mentioned, the parameters are all external and bear no relation whatsoever to the internal linguistic proficiency (...) No matter whether theoretically or empirically, there is a need to establish a relative norm so that when someone mentions “advanced learners”, it will be explicitly understood as the same (or approximately the same) thing with the same parameters in the measurement of the learners’ English (Guo, 2006: 43-44).

Así pues, mientras no se logre combinar corpus de HNN de español bien diseñados, buenas herramientas de búsqueda y anotaciones en el corpus, no será posible explorar todas las preguntas de investigación imaginables. Asimismo, tampoco se producirán importantes avances en el análisis de la interlengua del español si no buscamos una solución a las limitaciones ya referidas en torno a la metodología de análisis de corpus.

8.3. FUTURAS INVESTIGACIONES

La naturaleza de la interlengua del español no ha sido suficientemente estudiada. Esta investigación ha examinado el uso de un solo verbo en profundidad y en relación con un grupo de aprendientes concreto. Los resultados obtenidos sugieren características estables en la interlengua de este grupo; no obstante, estos resultados deben ser triangulados (Guilquin y Gries, 2009) y contrastados en corpus más extensos –en el caso de aquellos resultados no significativos que consisten en cifras demasiado bajas–. Asimismo, ciertos datos relacionados con la influencia de la L1 en la producción del verbo *dar* podrían ser contrastados con un corpus comparable de otra interlengua, como el WriCLE, de aprendientes de inglés de L1 español. Esta comparación permitiría saber si ciertos rasgos que caracterizan la interlengua de nuestro corpus resultan de la influencia de la L1 verdaderamente, o si, por el contrario, son consecuencia del *input* o de patrones de desarrollo universales (Lozano y Mendikoetxea, 2013: 6).

Asimismo, queda pendiente realizar un análisis manual del corpus que permita descubrir aquellos sintagmas en los que se debería usar *dar* y, en su

lugar, aparece otro verbo. Ese análisis nos permitiría confirmar tres de nuestras hipótesis que no han podido ser contrastadas en este trabajo: 1) si en las CVA hay más errores en la selección del verbo que en la construcción con el verbo usado en sentido recto; 2) si hay más problemas con las restricciones léxico-semánticas entre el verbo y el CD que entre el verbo y el sujeto; y 3) si los HNN están usando sustantivos figurados con otros verbos diferentes, cuando deberían estar empleando el verbo *dar*.

Aunque hay ciertas semejanzas entre la lengua escrita y la lengua oral, los estudios de corpus del inglés muestran importantes diferencias entre las dos modalidades de comunicación (Carter y McCarthy, 2006). Para conocer las diferencias más a fondo, es necesario comparar también la producción oral nativa y la no nativa; por esto, sería importante contrastar estos resultados con los de un corpus oral comparable. En español hay muy pocos corpus que permitan hacer un estudio contrastivo de la interlengua del español oral (posiblemente por la dificultad para compilar y transcribir); el más adecuado para este análisis podría ser el SPLLOC, concretamente las muestras de lengua de nivel avanzado, si bien el CEDEL2 se corresponde con muestras de lengua natural y el SPLLOC con muestras semi-naturales.

Por otro lado, dado que este estudio es sincrónico, queda pendiente para un futuro trabajo la realización de un análisis diacrónico o longitudinal del verbo *dar* a partir del corpus quasilongitudinal CEDEL2. Este nos permitirá describir y explicar la evolución del desarrollo del conocimiento de este verbo, y nos posibilitará responder a preguntas como ¿en qué estadio del aprendizaje se comienza a producir cierta estructura o combinación de *dar* en el español escrito de un grupo de aprendientes? o ¿cuáles son las dificultades compartidas por el grupo de aprendientes en los diferentes estadios de adquisición? Las respuestas a estas preguntas nos ayudarán a enriquecer nuestro conocimiento del aprendizaje del español.

8.4. CONCLUSIÓN

En definitiva, por medio de la revisión metodológica llevada a cabo en este capítulo se han puesto en relación los resultados del análisis con las hipótesis de partida de este trabajo, formuladas al comienzo de la investigación (véase capítulo 1); a su vez, los datos obtenidos nos han permitido construir nuevas hipótesis en la investigación sobre ASL, fundamentalmente referidas a la interfaz léxica-sintaxis como fuente de déficits; no obstante, se ha expuesto la necesidad

de contrastar algunos de estos resultados en corpus más extensos, lo que guarda relación con el tamaño del corpus, insuficiente para poder establecer la significatividad de todos los resultados obtenidos en esta investigación. Esta limitación se unía a otras expuestas en este capítulo; por último, basándonos en esta investigación, hemos apuntado varios estudios que sería interesante realizar para completar el presente trabajo.

9. CONCLUSIÓN

A continuación, nos referimos a las implicaciones de los hallazgos de esta investigación en los estudios de interlengua del español (9.1) y se proponen posibles aplicaciones didácticas que pueden tenerse en cuenta en la creación de materiales de ELE y en el aula de ELE, partiendo de la asunción –comúnmente aceptada en el campo de estudio de la lengua del aprendiente– de que el español nativo representa el estado de lengua ideal al que los HNN aspiran llegar (9.2). Por último, se indican las perspectivas de futuro de los estudios de la interlengua del español en relación con los corpus de aprendientes (9.3).

9.1. IMPLICACIONES EN LOS ESTUDIOS DE INTERLENGUA DEL ESPAÑOL

En esta investigación se ha descubierto que existe una importante diferencia entre las estructuras y combinaciones usadas por los HN y por los HNN en lo que se refiere al uso de este verbo. Una conclusión general de este análisis es que no podemos asumir que una palabra tan frecuente y aparentemente simple como es *dar* sea dominada por los aprendientes de nivel avanzado. A continuación nos referimos a la contribución de esta investigación a los estudios de la lengua del aprendiente por medio de las dos funciones principales del ACI señaladas por Guo (2006: 233), la de diagnóstico y la evaluativa, que analizamos en detalle.

9.1.1. *La función de diagnóstico*

La función de diagnóstico nos permite descubrir qué usan y que no usan los HNN en comparación con el uso nativo; esto nos ofrece la posibilidad de conocer los rasgos de la interlengua del español colectivo y diagnosticar las posibles necesidades de los HNN. Por medio del empleo de los ocho esquemas de relación de frecuencias adaptado de Guo (2006: 176) y de la aplicación del test de significatividad Chi-cuadrado, hemos podido describir la lengua del aprendiente con más precisión que con los extendidos conceptos de *sobreutilización* e *infrautilización*, a través de los cuales se cuantifican las diferencias lingüísticas pero no implican necesariamente juicios cualitativos acerca de la actuación en la interlengua.

Así, hemos podido comprobar que los HN, al igual que los HNN, usan con una frecuencia alta (por encima de un 5 %) las estructuras ditransitivas y las transitivas, las impersonales y las locuciones; que los HN y los HNN usan con baja frecuencia (por debajo de un 5 %) las intransitivas, las pasivas, las construcciones absolutas, las pronominales, las construcciones preposicionales y los esquemas fraseológicos; y por último, que los HN usan en una muy baja frecuencia (podríamos hablar de usos individuales en estos casos) las construcciones con *como* y las de *dar* como adjetivo, estructuras estas no usadas por los HNN. Esto nos demuestra que el español de los HNN y el español nativo del CEDEL2 no son drásticamente diferentes, dado que utilizan con frecuencia y con poca frecuencia las mismas estructuras, y ninguna de las estructuras se corresponde con los esquemas de “alta frecuencia *vs.* baja frecuencia” o con el de “alta frecuencia *vs.* no frecuencia” (véase la tabla 9, en el capítulo 7). De acuerdo con las asunciones sobre estos esquemas basadas en la intuición, las estructuras clasificadas en el esquema de “alta frecuencia *vs.* alta frecuencia” representan la parte del español colectivo del aprendiente más compartida –aunque no está garantizado que el uso nativo y no nativo sea similar–; y las que se clasifican en el esquema de “baja frecuencia *vs.* baja frecuencia” se corresponden con usos que los HNN no tienen muchas posibilidades de producir ni de recibir del *input* –dado que no son muy frecuentes en los HN–, por lo que es muy posible que la mayoría de HNN no sea del todo consciente de estos usos. Recuérdese que cuando estamos ante frecuencias bajas en los HN, tenemos menos certeza a la hora de interpretar la baja frecuencia entre los HNN.

Estos esquemas, en realidad, nos permiten describir e interpretar la interlengua desde una perspectiva muy general; tanto el test de significatividad como el análisis de la adecuación de los datos nos permiten realizar un diagnóstico más completo. Al aplicar el test de significatividad a los datos hemos comprobado que dentro de un mismo esquema se pueden dar sobreutilizaciones o infrautilizaciones (entendiendo aquí por estos conceptos un mayor o menor uso estadísticamente significativo); por ejemplo, las ditransitivas se sobreutilizan y las pronominales se infrautilizan. Además, por medio de un análisis más pormenorizado, hemos comprobado que incluso esta semejanza no garantiza un uso detallado similar, como en el caso de las locuciones, de las construcciones preposicionales o de los esquemas fraseológicos semiproductivos, en los que hemos visto que el español productivo de los HNN es menos amplio, independientemente de que el tipo de estructura sea usado con mucha frecuencia (locuciones) o con poca (el resto).

Asimismo, gracias a que en esta investigación los errores constituyen uno de los intereses centrales del análisis, se ha podido comprobar que una frecuencia similar no es sinónimo de una adecuación en los usos –como se ha visto, por poner un ejemplo, en las construcciones absolutas, donde dos de cada tres usos estaban mal contruidos–.

En las desviaciones centradas en el propio verbo, se ha diagnosticado que la transferencia de la L1 no es muy importante –solo un 12,5 % de los errores son interlingüísticos– y está presente en muy pocos tipos de error, tan solo en los usos inadecuados por la forma –el 14,3 % de los errores léxicos son interlingüísticos– y por la restricción semántica. En relación con esta, pese a que únicamente el 11,5 % de los errores computacionales son interlingüísticos, todos ellos conforman el 50 % de los usos que contienen un error en la selección semántica. Esto quiere decir que la L1 está detrás de la mitad de los errores de selección semántica que se producen en el verbo *dar*.

El análisis de cada concordancia también nos ha permitido diagnosticar otras características del español de los HNN; la agrupación de las unidades léxicas combinadas con *dar* en clases léxicas ha sido imprescindible para realizar este análisis. Así, un hallazgo importante es el hecho de que la construcción preposicional *dar por* + participio, que es la más utilizada entre los HN de entre todas las construcciones preposicionales, es utilizada con determinados participios: *dar por muerto/a*, *por hecho* y *por vencido/a*; de igual manera, de todos los esquemas fraseológicos, los HN utilizan con mayor frecuencia *dar a* + infinitivo, concretamente con infinitivos de verbos que denotan “conocimiento y percepción” –más específicamente, los infinitivos *conocer* y *entender* son los repetidos–; y, por otro lado, en *dar que* + infinitivo, los infinitivos utilizados son *hablar* o *pensar*. Los HNN no usan ninguna de estas estructuras; si las usaran, ganarían en expresividad; y, si además utilizaran los participios e infinitivos más usados por los HN, ganarían en naturalidad. Esta situación es aplicable a otros usos: los HN usan únicamente la pasiva con *se*, y los HNN, en cambio, prefieren la pasiva con *ser*; los HN prefieren utilizar el infinitivo *comer* combinado con *dar de* mientras que los HNN han elegido *dar a*.

Otra característica importante que ha sido diagnosticada por medio de esta investigación comparativa de las combinaciones léxicas de *dar* es que los HNN utilizan más y con más variedad sustantivos que pertenecen a ciertas clases léxicas, como la formada por sustantivos que designan “valores y cualidades de las personas o las cosas que suelen aludir a alguna característica esencial de su presencia, su duración o su forma de actuar o comportarse”, como *atractivo*, *autenticidad*, *prestigio*, *color*, *realismo* o *educación*. Estos sustantivos aportan una gran

riqueza expresiva de la que carece el texto no nativo, que solo registra un uso de estos sustantivos: *sabor*. Lo mismo sucede con otras clases léxicas como las siguientes:

1) Los HN también utilizan repetidamente varios sustantivos que designan “diversas inclinaciones y sensaciones afectivas que se caracterizan por transmitirse a los demás”, como *amor*, *cariño* y *ternura*; de nuevo, los HNN tan solo utilizan uno de estos sustantivos repetidamente: *amor*. La frecuencia con la que se emplea esta clase léxica es similar en ambos grupos, pero la variedad léxica solo la encontramos en los HN.

2) Los HN utilizan repetidamente varios sustantivos que denotan “la forma en que algo se manifiesta o el aspecto que adquiere”, como *imagen*, *aspecto*, *característica*, *dejo* y *toque*. Los HNN usan tan solo *deje* y *toque*.

3) Los HN usan diferentes sustantivos que designan “lo que ilustra algo o da razón de su existencia”, como *idea*, *perspectiva*, *muestra* y *ejemplo*; los HNN utilizan tan solo *idea* y *ejemplos*.

4) Los HN emplean una variación amplia en los sustantivos que denotan “desplazamiento, más frecuentemente si es breve o repentino”, y los interpretan figuradamente (*vuelta*, *paso*, *paseo*, *salto* y *giro*); los HNN, en cambio, muestran menos variación: *vuelta*, *paseo* y *voltereta*, y no hacen un uso figurado de estos sustantivos.

5) Los HN utilizan también diferentes sustantivos que designan “ciertos movimientos, reales pero también figurados, con frecuencia bruscos, vehementes, impulsivos o resultantes de algún contacto”, como *patada* (figurado), *golpe(s)* (figurado principalmente) y *golpecillos*, *paliza*, *trompazo*, *empujoncito* (figurado) y *vuelta de tuerca* (figurado). Los HNN solo utilizan *golpecitos* y además para significar *cachete*.

6) Aunque ambos grupos utilizan con la misma frecuencia *dar* con sustantivos que designan “reacciones y sensaciones anímicas relacionadas con la infelicidad, el enojo o el rechazo”, los HN utilizan, de nuevo, más variedad léxica: *lástima*, *asco*, *rabia*, *pena* y *pereza*. Los HNN usan *pena*, *asco* y *coraje*.

7) Lo mismo sucede con los sustantivos que denotan “miedo o inquietud”; mientras los HN usan *miedo*, *reparo* y *pánico*, los HNN solo usan *miedo*.

Por último, hay solo una clase léxica en la que sucede lo contrario: los HN utilizan únicamente dos sustantivos que designan “sensaciones y estados anímicos placenteros”, que son *gusto* y *felicidad*, y los HNN utilizan *placer*, *alegría*, *risa*, *satisfacción* y *gusto*. Resulta llamativo que varios de estos sustantivos que no son utilizados en el corpus nativo fueron especialmente difíciles de clasificar debido a su dudoso grado de adecuación. Esto sucedió en *quiero expresar el placer*

que me dieron estas dos películas, en me dio mucho placer verla [a mi hermana] con las niñas; en en las últimas escenas, lo que me dio más risa era [...]; o en El resultado nos dio mucha alegría, en lugar de Nos alegramos de haber tomado esa decisión. Esto demuestra la necesidad de analizar con detalle aquellos usos que son más frecuentes en los HNN que en los HN, pues normalmente es más natural que un ítem aparezca en el español nativo y no aparezca en el no nativo, por lo que puede esconder inadecuaciones.

En suma, si los HNN utilizaran los sustantivos más usados por los HN, ganarían en naturalidad. A menudo, se dice de la producción no nativa que es bastante correcta gramaticalmente, pero que adolece de una falta de naturalidad, que no es fácil de descubrir sin la aplicación de esta metodología.

De esta manera, la función de diagnóstico del ACI está muy presente en nuestra investigación. Nos ha permitido identificar las necesidades de los HNN en el uso de *dar* a través de la aplicación de varias técnicas y procedimientos combinados –los cuales se complementan entre sí–, a saber, 1) los ocho esquemas de la relación de frecuencias –con las respectivas asunciones basadas en la intuición– (adaptado de Guo, 2006); 2) el test de significatividad Chi-cuadrado; 3) el análisis pormenorizado de las concordancias en el que se examinan las palabras y tipos de palabras con las que se combina *dar* –lo que exige una agrupación de las coapariciones de *dar* en clases léxicas (basada en los presupuestos teóricos de Bosque [2005 (2004), 2001b] sobre combinatoria léxica) que permite extraer generalizaciones acerca de las palabras y tipos de palabras con las que unos y otros combinan este verbo–; y, por último, 4) una tipología del error que pretende ser consistente, informativa, flexible y reusable (calidades destacadas por Granger [2003]) que permite identificar, localizar, describir y explicar los errores (basada en el tratamiento del error propuesto por Corder, 1971).

Con un conocimiento de los errores y de las semejanzas y diferencias de uso entre el español de los HNN y el de los HN, podemos comprender y describir mejor el estadio de interlengua en el que se encuentran los HNN, y, en última instancia, sus necesidades¹⁹⁴. La identificación de estas necesidades puede

¹⁹⁴ No hemos detectado que haya determinadas formas verbales usadas únicamente por los HN; de haber habido algunas, estas formas deberían convertirse en una prioridad para el aprendizaje. No obstante, sí que hemos detectado que el 50 % de las desviaciones del verbo *dar* en las restricciones funcionales se debe a una confusión entre la 1.^a y la 3.^a persona (*[yo] me dio cuenta; [eso] me doy un poco de miedo; [me] ha dado cuenta de que; [yo] le dio su regalo; o Emery y yo se dieron besito*), posiblemente porque al ser las personas que con más frecuencia aparecen en toda narración básica (el hablante y de quien se habla) se entremezclan.

convertirse en punto de partida para aplicaciones en ELE (como se esboza en el apartado 9.2).

9.1.2. *La función evaluativa*

La segunda de las funciones que tiene un corpus de HNN es la evaluativa, dado que permite encontrar indicadores de una actuación de nivel de dominio alto y bajo en el español colectivo (Guo, 2006: 231). El hecho de que los HNN no utilicen o utilicen con muy poca frecuencia o utilicen mal las estructuras y combinaciones más usadas por los HN indicaría que el español colectivo de los HNN no muestra un nivel de dominio lingüístico muy alto. Esta situación no está muy extendida en nuestro corpus, aunque sí hay algunos casos; por ejemplo, las locuciones, que son usadas con frecuencia, presentan errores en casi un 50 % de los usos; o los sintagmas de *dar* combinado con la clase léxica de los sustantivos que denotan “asistencia, sustento o protección” son usados por los HN en un 14,2 % (de los usos de *dar* cuando actualiza el sentido de “proporcionar”), mientras que los HNN usan esta combinatoria léxica en un 2,1 %; lo mismo se aplica a la clase léxica de los sustantivos que designan “valores y cualidades de las personas o las cosas, frecuentemente si aluden a alguna característica esencial de su presencia, su duración o su forma de actuar”, pues es utilizada por los HN en un 8,1 % de los usos de *dar*, cuando actualiza el sentido de “proporcionar”, y en un 1,1% por los HNN.

De igual modo, el hecho de que los HNN no utilicen o utilicen significativamente menos las estructuras y combinaciones léxicas usadas con menos frecuencia por los HN (pero que no se corresponden con usos individuales) parece reflejar un dominio colectivo del español relativamente bajo; estos casos tampoco están muy extendidos en nuestro corpus, pero los hay. Un ejemplo es el del bajo índice de frecuencia con el que se utilizan las estructuras pronominales entre los HNN; asimismo, el conjunto de las estructuras que no se corresponden con la construcción V + N se utiliza en un porcentaje significativamente más bajo en los HNN (12,2 % frente al 3,7 %), aunque, si se examina con detalle cada estructura, solo hay diferencia significativa en las pronominales, y el resto simplemente se utiliza menos (pese a que, como se acaba de señalar, el uso total de todas ellas sí sea significativamente más bajo). Esto sí reflejaría un español menos amplio en la escritura y, por lo tanto, menos expresivo. No obstante, hay que tener en cuenta que es más difícil llegar a una

conclusión sobre el conocimiento colectivo de los HNN si las frecuencias no son suficientemente elevadas.

De igual modo que el análisis de corpus nos permite evaluar el grado de dominio del español colectivo, la alta o baja frecuencia de un elemento determinado en el corpus nos puede servir para evaluar a aprendientes individuales. Como sugiere Guo (2006: 232), aquellos HNN que no producen o no producen bien usos comunes o frecuentes entre los HN estarían en un estadio de adquisición inferior, por lo que su nivel se correspondería con un dominio lingüístico general más bajo que el de aquellos que sí los producen y lo hacen correctamente. Parece improbable que aquellos que no usan o no usan bien estas estructuras comunes y frecuentes tuvieran una puntuación alta en la calificación de la redacción. Asimismo, aquellos HNN que utilizan bien ciertos rasgos que no son muy frecuentes entre los HN (como las construcciones preposicionales, los esquemas fraseológicos semiproductivos o ciertas locuciones), parecen mostrar en la redacción un nivel de dominio más alto, por encima de la media, lo cual no implica que aquellos que los hayan usado, pero mal, no tengan necesariamente un dominio más alto. Para poder contrastar estas intuiciones y, por tanto, aceptar esta función evaluativa del ACI, sería conveniente tener la calificación obtenida en la redacción de cada uno de los participante del corpus.

En suma, la función de diagnóstico del corpus de aprendientes, cuando este se compara con el corpus nativo, es explícita. El diagnóstico llevado a cabo en este análisis nos lleva a reivindicar la importancia de que los HNN se familiaricen con la combinatoria de las palabras “pequeñas” y frecuentes como *dar*, que es usada a diario en numerosas ocasiones. Dominar el uso de estas palabras tan frecuentes hará que la producción no nativa se vaya aproximando gradualmente hacia el español nativo. En esta tarea tanto los profesores de ELE como los creadores de materiales tienen mucho que decir, por lo que se espera que extraigan también el máximo beneficio de esta función de diagnóstico¹⁹⁵. Los resultados también permiten concienciar a los aprendientes de los problemas que tienen como grupo y les da la posibilidad de atajarlos. Precisamente la

¹⁹⁵ Recuérdese que en el capítulo 1 defendíamos que las descripciones de las interlenguas obtenidas de los análisis de corpus de aprendientes pueden ser utilizadas en la enseñanza de las L2 (Granger, 2002: 4; Pastor Cesteros, 2004: 98), lo cual se enlazaba directamente con nuestros intereses en el ámbito de la actuación en el aula; ya hemos aludido (capítulo 1) al hecho de que cuando nos referimos a las posibilidades de aplicación de los resultados de este estudio de adquisición a la enseñanza lingüística como nuestro objetivo último, no lo hacemos simplemente desde una perspectiva investigadora, sino también –y fundamentalmente– sobre la base de nuestra trayectoria en la enseñanza de ELE.

identificación de las desviaciones en el uso de *dar*, tanto en las estructuras como en las combinaciones léxicas, pueden constituir el punto de partida de aplicaciones en la enseñanza de ELE.

9.2. ALGUNAS POSIBLES APLICACIONES DIDÁCTICAS DE ESTA INVESTIGACIÓN

Nuestro análisis de *dar* demuestra que este grupo de HNN de nivel avanzado no domina este verbo del todo. Está extendida la idea de que cuanto más vocabulario tienen los HNN más avanzado es su nivel, lo cual se traduce en la intención y en el deseo de aquellos de querer aumentar su vocabulario. No obstante, no sabemos con certeza cómo amplían los HNN el conocimiento de una palabra una vez que adquieren su uso básico, por lo que no se tiene total seguridad sobre la manera en la que se puede ayudar a expandir –a través de los materiales didácticos y actuaciones en el aula– toda esa gama de usos. En los siguientes apartados exponemos algunas ideas encaminadas a ampliar el conocimiento de *dar*, fundamentadas en los progresos de la teoría lingüística y psicolingüística (a los que nos hemos referido en el capítulo 3).

9.2.1. Materiales didácticos

La primera implicación que resulta de un estudio de corpus de aprendientes es la elaboración de materiales didácticos dirigidos a aquellos HNN que tienen el mismo perfil que el de los HNN del corpus; en nuestro caso, estudiantes o graduados anglófonos con un nivel avanzado de español, según el test de la Universidad de Wisconsin (para actuar con más fiabilidad en las posibilidades de aplicación de los resultados, sería útil conocer la equivalencia de dicho test con respecto a los niveles del *MCER*).

Los materiales de enseñanza de ELE que habitualmente se utilizan en el aula suelen estar basados en la experiencia de profesores, o bien proceden de manuales cuyas muestras de lengua se fundamentan principalmente en la intuición de sus autores (Vangehuchten, 2004); así, no tenemos manera de medir hasta qué punto los materiales reflejan las necesidades de los alumnos¹⁹⁶, aunque esto puede empezar a cambiar gracias, precisamente, a los estudios de interlengua

¹⁹⁶ El manual *Cumbre*, sin embargo, sí ha sido elaborado a partir del corpus del mismo nombre (sobre el corpus *Cumbre*, véase Sánchez Pérez [1996]).

basados en corpus. La comparación de los corpus ha revelado algunas necesidades del grupo de HNN en relación con el uso del verbo *dar* que podrían ser consideradas en el proceso de elaboración de materiales.

El objetivo no ha de ser solo aprender nuevos usos, sino mejorar el uso de los elementos conocidos. Mejorar los usos conocidos exige también prestar atención a los usos inaceptables (recuérdese que un 32,6 % de los usos de *dar* es considerado inaceptable de acuerdo con la norma descriptiva y prescriptiva fijada; esto es, 1 de cada 3 usos presenta algún error; y un 56,3 % de los errores se localiza en el núcleo verbal, en *dar*; es decir, en 1 de cada 2 errores hallados en el sintagma, el problema se localiza en el verbo). Mientras que algunos errores son más bien individuales, otros son ciertamente característicos del grupo. De esta manera, donde hay una tendencia general compartida por un grupo de hablantes, un enfoque lingüístico de corpus ayuda a descubrirla.

Por ejemplo, los HNN utilizan con demasiada frecuencia las estructuras ditransitivas; al respecto, los manuales pueden presentar alternativas, como la del empleo de un solo verbo (del que proviene el sustantivo utilizado en la combinación con *dar*). Las actividades tendrían que especificar los contextos en los que las estructuras son intercambiables, lo cual parece suceder en aquellos casos en los que el foco comunicativo está más bien en la acción que se desarrolla; pese a ser intercambiables, de acuerdo con nuestro análisis, los HN tienden más que los HNN a utilizar en estos casos un solo verbo (recuérdese el ejemplo de [*en Año Nuevo*] *Emery yo nos dimos besos a media noche porque es la tradición*, en lugar de *nos besamos*); en cambio, si el foco está más en el referente del sustantivo, probablemente no omitiríamos la estructura *dar* + N, como ocurriría, por ejemplo, si el foco estuviera en *el beso*, porque ese beso hubiera sido significativo y esto se quisiera transmitir al interlocutor.

En esta misma línea, los materiales pueden trabajar la sustitución de combinaciones sobreutilizadas, como es el caso de *dar la oportunidad* en lugar de otras coapariciones también naturales y frecuentes, como *ofrecer la posibilidad*.

Por otro lado, los HN utilizan con más frecuencia y más variedad las estructuras que no se corresponden con la construcción V + N. Si se busca que el español de los HNN sea igual de expresivo y amplio que el de los HN, los materiales deberían recoger, en primer lugar, las coapariciones léxicas que aparecen solo en el corpus nativo y no se corresponden con usos individuales – es decir, son naturales –, como por ejemplo *dar tiempo*; en segundo lugar, deberían introducirse los sustantivos de algunas clases léxicas que son muy empleados por los HN con su sentido figurado (*rumbo, golpe, aliento, vuelta, salto, paso, giro, empujón, patada*, etc.); y debería atenderse a la selección de *dar* de una clase léxica

determinada haciendo hincapié en la variedad de sustantivos próximos –desde una perspectiva semántica– contenidos en esa clase que son empleados con frecuencia por los HN (*miedo, pánico, reparo; o lástima, pena, rabia, etcétera*).

Asimismo, locuciones tan comunes como *dar igual* o *dar lugar* deben considerarse en los materiales, y estas, al igual que las demás unidades léxicas, deben trabajarse atendiéndose a su comportamiento sintáctico. En relación con ello, más énfasis se debe otorgar a la restricción sintáctica del núcleo verbal *darse cuenta*, que, como hemos visto en nuestro análisis, ocasiona el 90 % de los problemas sintácticos del verbo a raíz de la ausencia de la proposición *de*. Y también en estas, al igual que en las demás construcciones, tendría que prestarse atención al registro (recuérdese el caso de *dar la gana* o *dar cosa*), aspecto que con demasiada frecuencia se desatiende en los materiales didácticos¹⁹⁷.

El análisis realizado también revela que este grupo de HNN confunde la 1.^a con la 3.^a persona del verbo *dar*, sobre todo cuando el verbo está precedido por un pronombre (tenga esta función sintáctica o no); más actividades deberían tener como objetivo practicar estas formas verbales en contexto, dado que son estas las que ocasionan problemas. En relación con ello, si los aprendientes quieren aproximar su producción todo lo posible a la de los HNN, los materiales deberían plantear la composición escrita sobre narraciones más centradas en el contenido del mensaje, en las que las referencias al estado mental, físico y anímico propio no tengan cabida fácilmente, y con un mayor empleo de sustantivos usados en sentido figurado (como se acaba de señalar). De esta manera (y si, como se acaba de señalar, se logra un mayor empleo de sustantivos usados en sentido figurado) el uso de *dar* como verbo de apoyo sería más semejante en los dos grupos.

Por otra parte, como se desprende del trabajo de Cobb (2003), los HNN tienden a no usar en las tareas escritas de carácter libre aquellas estructuras con las que se sienten menos seguros; por ello, deben desarrollarse materiales que obliguen a los alumnos a irse familiarizando –por medio de su aparición frecuente en el *input* y su recuperación en el *output*, y con la práctica previa del *foco en la forma* [Long, 1991]¹⁹⁸– con aquellas estructuras y combinaciones que son naturales, ya sean más frecuentes o menos frecuentes (recuérdese que se trata de

¹⁹⁷ En Sánchez Rufat (2014) reseñamos algunos manuales para la enseñanza del léxico de reciente publicación que, pese a estar en mayor o menor medida centrados en la enseñanza de la palabra a través de su combinatoria léxica, prescinden por completo de indicaciones sobre la frecuencia y el registro de uso de cada una de esas combinaciones, lo que, sin duda, es una información imprescindible para el uso completo de una palabra.

¹⁹⁸ *Foco en la forma*, del inglés *focus on form*, implica prestar atención a la forma y al uso de la forma en contexto en el momento necesario para la realización de la tarea o actividad.

aprendientes de nivel avanzado), si bien las estructuras o combinaciones más usadas por los HN son las que deben tener prioridad a la hora de ser aprendidas.

Si las estructuras o combinaciones nuevas o no del todo conocidas por los HNN no se trabajan explícitamente y no se introducen con frecuencia en las muestras de *input*, no se podrá ampliar el lexicon de los HNN. Estas combinaciones se convertirán en fuente de errores o en construcciones cuyo uso se tratará de evitar.

A continuación, presentamos propuestas pedagógicas para el aula de ELE, centradas en algunas de las necesidades que han sido reveladas en este análisis contrastivo basado en corpus.

9.2.2. *El aula de ELE*

En la última década ha surgido un interés por la adquisición del español como lengua extranjera (Liceras, 2009; Geeslin, 2014). Cada vez son más los estudios sobre la interlengua del español, en general, y sobre aspectos léxicos, en particular (véase el capítulo 2 para una enumeración de estos trabajos), pero siguen siendo escasos los materiales y las actuaciones de aula que incorporan los hallazgos de estos estudios. Al mismo tiempo, también son pocos los materiales de ELE que incorporan los nuevos enfoques teóricos y que hacen un tratamiento del léxico coherente y sistemático, pese a la fuerte revalorización que ha experimentado el componente léxico en los últimos 30 años en las teorías lingüísticas, psicolingüísticas y pedagógicas (capítulo 3); del mismo modo, resultan escasas las actuaciones en el aula acordes con los nuevos planteamientos surgidos en torno al léxico.

Los procedimientos o técnicas de enseñanza desarrollados en este apartado (terminología de Richard y Rodgers, 2009 [1986]) se fundamentan en varias ideas comúnmente aceptadas acerca del aprendizaje de una lengua, en general, que son aplicables al aprendizaje del léxico, en particular: 1) aprendemos de forma personal, según nuestras motivaciones y desde nuestra identidad —de ahí el beneficio de que los alumnos completen en el aula en los primeros días de curso una ficha y encuesta para conocer las características, así como las necesidades e intereses del grupo, lo cual ayudará al profesor en la tarea de seleccionar el vocabulario que se va a presentar en el aula—; por eso, 2) el *input* recibido a partir de los textos —que es imprescindible para contextualizar el léxico y para la construcción de hipótesis sobre su funcionamiento— no es indiferente, esto es, debe interesar al aprendiente; los textos que contienen los ejemplos

contextualizados de *dar* habrían de tratar temas sobre los que el propio aprendiente quisiera leer o escuchar en su L1; 3) aunque aprendemos vocabulario participando en interacciones significativas, es imprescindible también observar y reflexionar sobre el sistema, ensayar a partir de modelos, de recursos o secuencias lingüísticas con *dar* a partir de los cuales el aprendiente crea su propio mensaje; 4) se aprende el vocabulario también desarrollando estrategias y siendo aprendientes autónomos; 5) se aprende en espiral; es decir, aprendemos lo nuevo a partir de lo conocido; por eso, es importante contar con unidades didácticas o con actividades *escala*, que retoman los usos de *dar* ya trabajados en unidades anteriores antes de proseguir con los nuevos usos que serán aprehendidos más fácilmente gracias a los conocimientos previos adquiridos sobre el verbo; y, por último, 6) dado que, según muchos investigadores, nuestro lexicón mental se encuentra altamente organizado en redes interrelacionadas, conviene aprender las palabras por medio de redes. La rapidez con que se recuperan las palabras ha llevado a algunos lingüistas a pensar que la mente organiza las palabras de muchas maneras. Suelen distinguirse las asociaciones fonológicas, morfológicas (derivación y composición), discursivas, semánticas (hiponimia e hiperonimia, sinonimia y antonimia, metonimia, polisemia, campos semánticos y asociativos), categoriales, léxicas y léxico-sintácticas (cognados, expresiones idiomáticas, coapariciones frecuentes), pragmáticas (elección léxica determinada por las variables de la situación comunicativa), psicológicas (elementos que se organizan en torno al prototipo), personales o afectivas, socioculturales, intralingüísticas e interlingüísticas. La búsqueda de una palabra en el lexicón mental se ha comparado con un viaje a través de una red compleja donde ciertas asociaciones son más fuertes que otras. Diversos test de asociación de palabras demuestran que las redes colocaciones y de coordinación (léxicas y léxico-sintácticas) son de las más fuertes (Nagini, 2008: 184). Partiendo de estas premisas, sería del todo contraproducente enseñar las palabras aisladas del contexto en el que aparecen.

Sin perder de vista estos seis aspectos generales acerca del aprendizaje del léxico, nos trasladamos a los inventarios de contenidos del *Plan Curricular del Instituto Cervantes (PCIC)* –como concreción de las ideas y planteamientos más generales del *MCER* a la realidad del español–, en concreto, al apartado de *nociones específicas*, donde se recogen las orientaciones en torno a los contenidos léxicos; allí podemos ver la manera en la que el contenido de *dar* está distribuido por los seis niveles de dominio lingüístico. Constatamos que el *PCIC* no hace ninguna referencia a usos figurados ni literales (distinción de importancia demostrada en nuestro análisis); no recoge esquemas fraseológicos abiertos, ni las construcciones preposicionales utilizadas en el corpus analizado (*dar a*, *dar de*,

dar por); sí se recogen algunas locuciones, pero no las más frecuentes del análisis, como *darse cuenta* (locución más empleada y con un 90 % de error debido al régimen preposicional) o *dar lugar a*. Las coapariciones léxicas de *dar* no se agrupan en clases léxicas, sino que nos encontramos con combinaciones que pertenecen a una misma selección semántica diseminadas por los niveles B1, B2, C1 y C2, como se observa en el siguiente ejemplo: en el inventario de contenidos del nivel B1 aparecen las combinaciones *dar la mano*, *dar un abrazo* y *dar un beso*, y, por otro lado, *darse un golpe*; en el B2, se recoge *dar una patada*, *una torta*, *un puñetazo* y, por otro lado, *dar un golpe de Estado*; en el C1, *dar una palmada en el hombro* y *darse un apretón de manos*; y en el C2, *dar una paliza*.

Aunque sin duda la combinación *dar una palmada en el hombro* es más compleja –por el número de elementos que la constituyen– y menos frecuente que *dar un abrazo* (lo que justifica que la primera sea propuesta como contenido léxico de nivel C1 y la segunda, de nivel B1), consideramos que supone una pérdida de eficacia en el proceso de enseñanza-aprendizaje de *dar* el hecho de que no se introduzca la clase léxica en el momento en el que se trate por primera vez en el aula el verbo combinado con un sustantivo que pertenezca a dicha clase. Es decir, cuando aparece por primera vez en un nivel B1 *dar un abrazo*, en lugar de tratar de que el aprendiente fije y almacene en la memoria esta expresión como relación biunívoca, se debe aprovechar la oportunidad para considerar esta expresión como una muestra de una restricción intralingüística del verbo concreta, a saber, la de *dar* combinado con sustantivos que designan “ciertos movimientos, reales o figurados, con frecuencia bruscos, vehementes, impulsivos o resultantes de algún contacto”. De esta manera, se obtienen varias combinaciones léxicas a partir de una misma relación léxico-sintáctica, como todas las mencionadas anteriormente (desde *dar un beso* hasta *dar una paliza*), lo que resulta muy rentable para el aprendiente. No solo porque este conocimiento le permite codificar otras combinaciones a partir de lo conocido (como *dar un empujón*, *dar un toque*, *dar un bofetón* o *un cachete*, por poner algunos ejemplos), sino porque también podría, a partir de ese conocimiento adquirido en el B1, descodificar correctamente todas las demás combinaciones que van apareciendo en los demás niveles, infiriendo la lógica en la combinatoria composicional que hay tras ellas¹⁹⁹. Esta práctica promueve la adquisición del léxico a través de la elaboración de redes léxicas de gran potencial combinatorio.

Es cierto que el *PCIC* tan solo aporta “una base de orientación general que podrá servir de referencia a quien lo utilice a la hora de seleccionar y

¹⁹⁹ Dedicamos un trabajo a la productividad de las clases léxicas (Sánchez Rufat, 2011).

distribuir por niveles las unidades léxicas que precise para sus propios fines”. Por lo tanto, no tiene por qué haber coincidencias totales en la manera en la que los distintos manuales de ELE distribuyen los contenidos de *dar* a lo largo de los diferentes niveles de dominio; el problema es que no hay en absoluto coincidencias, lo cual demuestra una falta de sistematicidad y de coherencia en el tratamiento del léxico en el panorama de ELE que no se da en el tratamiento de ninguno de los demás niveles lingüísticos. Si, como hemos señalado, el *PCIC* no contempla muchos de los usos de *dar* más utilizados por los nativos (de nuestro corpus) ni tampoco los usos figurados de los sustantivos con los que frecuentemente se combina el verbo, y dado que los creadores de materiales y profesores tienen en el *PCIC* una guía orientativa acerca del léxico que se puede presentar en cada nivel, no es sorprendente que ni en los manuales ni en el aula de ELE se trabajen muchas de las construcciones preposicionales, muchos de los esquemas fraseológicos semiproductivos, los usos figurados de ciertos sustantivos frecuentes o ciertas coapariciones frecuentes. Los resultados de investigaciones como esta son claves para homogeneizar los contenidos léxicos en los manuales de ELE y para orientar las actuaciones en el aula atendiendo a las necesidades reales de los aprendientes; si son los profesores los que han de decidir en cada curso exactamente qué elementos léxicos tendrá que aprender a reconocer y utilizar su grupo de aprendientes y cómo se les capacitará para ello; y si de aquellos también depende cómo se eligen –normalmente en función de la frecuencia y de los intereses o necesidades del grupo– y cómo se ordenan los elementos léxicos que va a introducir en el aula, el potencial de una investigación como esta es evidente, pues destapa ante los profesores las necesidades que tiene el grupo para llegar a dominar el verbo *dar* de la misma manera en la que lo hacen los nativos.

A lo largo de este análisis hemos visto que, de todos los aspectos que abarca la competencia léxica de *dar*, las restricciones gramaticales y léxico semánticas son las que causan más problemas en este grupo de aprendientes (el 65 % de los errores son de esta naturaleza²⁰⁰); por ello, estos contenidos son los que más se deben trabajar en el aula, sin que por ello perdamos de vista el

²⁰⁰ Recuérdese que entre las desviaciones computacionales el 39 % son sintácticas (en el 90 % de estas, el foco de error está en la ausencia de la preposición regida por *darse cuenta*) y el 39 % son funcionales (en el 50 % de estas el error se debe a una confusión entre la 1.^a y la 3.^a persona del verbo). Si tenemos en cuenta que la selección semántica particularmente restrictiva en la relación entre el predicado y argumento-tema se aplica únicamente a las estructuras ditransitivas, transitivas, pasivas, impersonales e intransitivas, la desviación semántica suma el 50 % de los errores computacionales del núcleo.

contexto de aparición de la estructura o de la combinación léxica, lo que implica una atención al registro y a la frecuencia de uso.

En el apartado 6.4.7.1 del *MCER* se recogen diferentes técnicas para presentar el vocabulario en el aula, pero solo algunas de ellas facilitan precisamente la creación de redes léxicas de manera más o menos efectiva. Se debe tener en cuenta que, de acuerdo con el paradigma de enseñanza vigente comunicativo (al que le incorporamos las contribuciones del enfoque léxico, de Lewis [1993, 1997 y 2000]), al aprendiente se le debe facilitar a través del *input*, en primer lugar, la observación y la percepción del léxico nuevo en contexto –lo que supone la *entrada* de la unidad en el lexicón mental–, para después poder pasar a experimentar con ese vocabulario nuevo, en un segundo estadio del proceso de adquisición, por medio de actividades guiadas –a través de las cuales se logra *almacenar* la palabra en el lexicón–, antes de poder ser capaz de usarlo correctamente en una tarea de producción libre –lo que significa poder *recuperar* la palabra en el acto comunicativo–, último estadio en la adquisición. A continuación, esbozamos algunas propuestas didácticas –centradas en algunas de las necesidades identificadas en el análisis– que se refieren a los distintos pasos o estadios mencionados del proceso de enseñanza-aprendizaje del léxico.

1) Explorar campos semánticos y construir mapas conceptuales, al tiempo que se reflexiona sobre la distinta distribución de los rasgos semánticos de *dar* en la L1 y en la L2 (semántica contrastiva). Consistiría en adaptar la tabla 2 del capítulo 4 al formato del mapa conceptual (seleccionando de la tabla los contenidos en función del nivel del grupo). El mapa conceptual de *dar* se puede ir completando en el cuaderno del alumno a lo largo del curso, y a medida que se van descubriendo nuevas restricciones intralingüísticas del verbo.

2) Enseñar a los alumnos a usar diccionarios combinatorios (*Redes y Práctico*), diccionarios de sinónimos y otras obras de consulta para que exploren las relaciones paradigmáticas y sintagmáticas del verbo. En torno al diccionario *Práctico*, de más fácil manejo que *Redes* para los aprendientes, pueden desarrollarse actividades en las que el aprendiente formula con sus palabras la clase léxica a partir de los sustantivos con los que se combina *dar*. En este diccionario, una cadena de sustantivos que comparten cierto rasgo semántico que es seleccionado por *dar* aparece entre dos barras; por ejemplo, aparecen entre barras *grito*, *alarido*, *resoplido*, *ladrido*, *aullido*, *pitido* y *do de pecho*; se les puede preguntar a los aprendientes qué es lo que tienen en común todos esos sustantivos (designan sonidos). De esta manera, ellos mismos van elaborando las clases léxicas que selecciona *dar*, y, al ser ellos quienes hacen este descubrimiento, aumentan las posibilidades de que empleen esta combinatoria en sus producciones lingüísticas

y que sean capaces de producir combinaciones nuevas originadas a partir de ese conocimiento o restricción intralingüística, como *bufido*, *suspiro* o *chillido*. Por otro lado, el diccionario *Redes* también puede servir para examinar los ejemplos de usos figurados de los sustantivos de ciertas clases léxicas que son usados con frecuencia (como en los ejemplos de *salto*, *paso*, *vuelco*, *marcha atrás*, *giro*, *toque*, *empujón*, etcétera) y para reflexionar sobre la coincidencia o no de esta proyección metafórica en la L1 del aprendiente (semántica contrastiva).

3) Ayudar a los alumnos a elaborar las clases léxicas que selecciona un predicado (ya sea el predicado *dar* [+ N], o *dar que*, *dar de* y *dar a* [+ infinitivo]), pero esta vez con la ayuda de un corpus nativo²⁰¹. Este enfoque, que pone al alumno en contacto con textos auténticos en el aula, es conocido como *data-driven learning* (DDL) y fue iniciado por Tim Johns, quien lo define del siguiente modo: “the use in the classroom of computer-generated concordances to get students to explore regularities of patterning in the target language and the development of activities and exercises based on concordance output” (1991: iii, citado en Granger y Tribble [1998]). El corpus de control nativo puede utilizarse, además, para descubrir la restricción sintáctica de *dar* en las estructuras diferentes a la construcción V + N, que son más desconocidas por los aprendientes (reflexionar sobre los diferentes usos preposicionales, esquemas fraseológicos semiproductivos, construcciones de participio absoluto...); y puede servir para tratar de *elaborar*²⁰² (Barcroft, 2002), a partir de las concordancias, el significado de la estructura en contexto a partir del significado básico y de los demás usos de

²⁰¹ En Sánchez Rufat y Jiménez Calderón (2013) se desarrolla una propuesta de aprendizaje de las combinaciones léxicas (para su descubrimiento o entrada en el lexicón y su fijación o almacenamiento) basada en la restricción intralingüística de los predicados, mediante la aplicación de los corpus textuales nativos informatizados. Lo que se pretende es que el aprendiente identifique en el *input* no solamente ciertas combinaciones léxicas (objetivo de la enseñanza), sino que, utilizando el *input* que los corpus suponen, sea capaz de producir otras combinaciones, toda vez que haya descifrado la relación que se establece entre un predicado y sus argumentos. Para esto nos servimos del concepto de restricción intralingüística, según el cual los predicados seleccionan sus argumentos restringiendo el conjunto de piezas léxicas que pueden denotar de acuerdo con sus rasgos semánticos; y también de los enfoques de la lingüística cognitiva, mediante los que el alumno concebirá la lengua, no como un conjunto de relaciones azarosas, sino como el resultado de relaciones lógicas desde los planos semántico y léxico; todo ello facilita la elaboración de la información léxica, es decir, se prevé que ayudará al alumno a producir combinaciones léxicas basadas en el conocimiento de las propiedades combinatorias de los predicados y de su carácter restrictivo. Estos corpus funcionan como *input* muy provechoso que los alumnos pueden manejar autónomamente; tan solo hay que estimular su curiosidad. No obstante, para obtener el máximo beneficio de trabajar con corpus en el aula, el profesor tiene que adecuarlo en ocasiones y tomar control de las actividades que se hagan con las concordancias en el corpus. Como señala Cruz Piñol (2012: 104), en todo planteamiento de la enseñanza que otorga un papel activo al alumno, la función del profesor es fundamental. Este profesor puede servirse de los textos aportados en los manuales y abordar el léxico de manera sistemática a partir de propuestas como la descrita aquí.

²⁰² Véase n. 10.

dar conocidos. Si descubren en el aula cuáles son esas nuevas formas y sus usos (incluso si llegaran a ser capaces de construir las clases léxicas con las que se combinan *dar a* + infinitivo, *dar de* + infinitivo o *dar que* + infinitivo), hay más posibilidades de que ellos empiecen a utilizarlas en su propia producción²⁰³. En relación con los datos del corpus, hemos visto que algunos errores formales de la construcción *darse cuenta de* se deben a la confusión con *dar cuenta de*; considerando esto, las concordancias sirven para que los aprendientes dilucidan si existen diferencias de significado entre ambas expresiones.

4) Si se informa a los aprendientes de que sobreutilizan las estructuras ditransitivas, se les da la oportunidad de que intenten usar los verbos de los que proceden los sustantivos con los que *dar* se combina. Para ayudar a los alumnos a concienciarse de este sobreuso, se les puede pedir que comparen en un corpus nativo *dar* + N *vs.* verbo, por ejemplo, *dar un beso vs. besar*. Se espera que los aprendientes perciban que los hablantes nativos tienden a usar mucho más el verbo *besar*. En una búsqueda del CREA se obtienen 149 casos de *da* + N y 1006 casos de *besa*. Pueden detenerse a observar el contexto que rodea a *da* + N en los ejemplos, para comprobar que suele haber un complemento a la derecha de *beso* (*rápido, fuerte, en la mejilla, en la mano, en la frente, en la boca, en los labios, de buenos días, de despedida, un beso que él prolonga, etcétera*), mientras que esto no sucede con frecuencia en las concordancias de *besa*. Una vez examinados estos usos, pueden ver sus propias concordancias y reemplazar *dar* + *beso* por el sustantivo *besar* en aquellos ejemplos en los que el contexto no aporta información sobre el beso para que, de esta manera, la producción no nativa se aproxime más a la nativa.

5) Los aprendientes pueden observar las peculiaridades sintácticas de sus propios usos de *dar* al compararlos con los usos del CREA; por ejemplo, la comparación entre corpus puede servir para detectar los casos de *queísmo* en la locución *darse cuenta*, o la falta de concordancia entre el participio y el sustantivo en las construcciones de participio absoluto. Se prevé que los aprendientes aprenderán pronto a usar las estructuras del mismo modo en el que lo hacen los nativos. Para subrayar el efecto de haber hecho este descubrimiento por sí mismos, se puede pedir a los aprendientes que directamente practiquen el uso nativo.

6) Hemos visto en el capítulo 4 que, por medio de la definición sistémica de *dar* en la estructura “A da B a C”, es posible interpretar que tras ese cambio

²⁰³ No obstante, basándonos en el trabajo de Schmitt y Redwood (2011) sobre el aprendizaje de los *phrasal verbs* del inglés, no descartamos el papel facilitador que puede desempeñar en el proceso de adquisición de los usos de *dar* multipalabra (*dar con, dar de, dar por, dar a, dar que...*) la exposición fuera del aula a estos usos.

que se produce en *B* –de no existir a existir, o de no existir en *C* a existir en *C*– subyace un contenido de trayectoria implícita presente a lo largo de todo el Esquema de Imagen del Camino propuesto por Lakoff y Johnson (1999). A partir de este esquema se observa que el Camino surge de nuestra percepción de entidades que se mueven a lo largo de un trayecto, de una línea descrita en el espacio, en el que hay un punto inicial (lugar de origen), un destino y un recorrido. Según este esquema, conceptualizamos directamente recurriendo a las categorías perceptivas, a una experiencia física que tiene que ver con dominios básicos –en este caso, el espacio–, como ocurre en *dar un caramelo a alguien*, o mediante proyecciones metafóricas que toman como dominio de origen el Esquema de Imagen, como en *dar una explicación a alguien*. En el aula, podemos trabajar con este esquema de imagen para ir estableciendo los lazos que existen entre los usos físicos y figurados de *dar* –que tienen en común la *trayectoria* y la *meta*– y también para establecer la distinción entre los usos de *dar* en las CVA (en los que se destaca la noción de *creación*) y los usos de *dar* en sentido recto aunque figurado (en los que la noción subrayada es la *transferencia*).

7) El análisis revela que los aprendientes, pese a usar las locuciones con frecuencia, utilizan muy poca variedad. Además, veíamos que el 57,1 % de los errores léxicos se debe a errores en alguna locución (*darse cuenta, dar la gana, dar a luz*), ya sea por la forma, por el significado o por el contexto de uso. Al igual que proponíamos explicar el uso de *dar* en las combinaciones composicionales V + N por medio del Esquema de Imagen del Camino, las locuciones nuevas pueden explicarse en el aula atendiendo también a varios principios cognitivos (teoría del código dual [Clark y Pavio, 1991], la del rastro [Baddeley, 1997] y la de los niveles de procesamiento en relación con el concepto de *elaboración*²⁰⁴ de la información léxica [Barcroft, 2002]), que permiten al aprendiente beneficiarse del potencial mnemotécnico de estas expresiones²⁰⁵.

²⁰⁴ La operación mental que exige la elaboración se caracteriza por su complejidad, y se opone al bajo nivel de procesamiento cognitivo que requieren muchas veces las actividades sobre léxico comunicativas.

²⁰⁵ El proceso de enseñanza-aprendizaje de estas combinaciones se inicia estableciendo una clasificación de dominios o campos semánticos como los siguientes: deportes y juegos, flora y fauna, comida y cocina, transporte, religión y superstición, caza y pesca, etcétera. Esta clasificación resulta muy útil por su productividad: todas las locuciones trabajadas en el aula pertenecen a uno de estos dominios; a medida que el alumno se vaya encontrando con nuevas locuciones y elaborando su significado, podrá clasificarlas y ordenarlas de acuerdo con su significado originario en el campo correspondiente. Conocer el significado originario de la combinación supone, por lo tanto, incluir la locución en un campo de dominio; y acceder al significado composicional sirve de gran ayuda al aprendiente en el proceso de retención de la locución, ya que le permite concebir el sentido figurado y, por tanto, no composicional que tiene esa combinación de palabras en la actualidad como una extensión natural del primero. Boers y Lindstromberg (2011) recogen los resultados de una serie de análisis experimentales sobre la interlengua del inglés que demuestran

8) Hay estudios (Horst *et al.*, 2005) que muestran el éxito de algunas aplicaciones del diseño de ejercicios interactivos en línea para la enseñanza del inglés. El *Compleat Lexical Tutor* (<http://www.lextutor.ca>) es una base de datos para el aprendizaje colaborativo de vocabulario en línea diseñada para aprendientes de inglés universitarios. El *software* tiene diferentes aplicaciones y permite a los aprendientes crear sus concursos de vocabulario y ejercicios de huecos. Este estudio demuestra que las palabras que son introducidas en el *software* por los aprendientes fueron aprendidas significativamente mejor que las que no entraron; y en una encuesta realizada posteriormente a los aprendientes, estos afirmaron que valoraban mucho –como herramientas útiles para el aprendizaje del vocabulario– la aplicación del diccionario en línea y el programa para la preparación del concurso. Así pues, a nuestro juicio, una vez que está creado el *software*, no resultaría complicado emplear estas aplicaciones para trabajar las necesidades extraídas de los resultados de los análisis de corpus de aprendientes. Cruz Piñol (2012: 154-155) reconoce la escasez de prácticas interactivas para ELE basadas en corpus; destaca las que se proponen en la *web* del *Diccionario de Colocaciones del Español (DICE)* y las creadas a partir de *FrameNet Español* (difíciles, estas últimas, para un aprendiente de español). A estas les podemos añadir las actividades basadas en los datos del corpus CORELE²⁰⁶ diseñadas, en este caso, para formar a profesores de ELE. Más esfuerzos pueden ir encaminados en esta dirección.

9) Por último, dado que, como hemos señalado, se aprende a partir de modelos sobre los cuales el aprendiente crea su propio mensaje, consideramos una práctica útil que los aprendientes, antes de redactar una composición de aula de nivel universitario, pudieran reflexionar y discutir con el profesor las características comunes a una serie de textos modelo (en este caso, buenas composiciones escritas por nativos), el tono y el registro general del texto. De esta manera, en el aula se puede hacer hincapié en las pocas referencias que hay en los textos nativos al estado mental, físico o psíquico del autor en una composición de este nivel, en contraposición con las composiciones escritas por

que la retención y comprensión de las locuciones se incrementan cuando se ayuda a asociar cada locución con el dominio de experiencia al que pertenecía cuando se originó. En Sánchez Rufat (2013) y en <<http://es.slideshare.net/EncuentroPracticoELE/la-claboracin-semntica-en-el-tratamiento-del-lxico>> (panel presentado en el XIX Encuentro práctico de profesores de ELE, en Barcelona [2010]) aplicamos al español estos presupuestos a través de los cuales proponemos la enseñanza de las locuciones *dar en el clavo* y *dar el brazo a torcer*, respectivamente.

²⁰⁶ Las actividades están disponibles en línea en <http://cartago.llf.uam.es/corele/home_en.html#activities> (07-08-14).

no nativos. Esto ayudará a los aprendientes a replantearse el sobreuso de *dar* con sustantivos que denotan ciertos estados anímicos y psíquicos.

En definitiva, la relevancia de las palabras más frecuentes de una lengua, entre las que se incluye *dar*, implica que los profesores y aprendientes deben dedicar un tiempo suficiente para asegurar su adquisición. Esta atención puede prestarse en forma de enseñanza directa, aprendizaje directo, aprendizaje incidental y a través de encuentros planeados con la palabra (Nation, 2001: 16). En este apartado se han presentado varias técnicas para el aprendizaje y la enseñanza directos de algunos aspectos de *dar*, considerados –de acuerdo con los resultados de este análisis de corpus– conflictivos para este grupo de aprendientes. Todo el tiempo que se dedique a *dar* en el proceso de enseñanza-aprendizaje está sobradamente justificado por la frecuencia con la que se emplea este verbo, por su alcance y extensión. En relación con ello, no podemos olvidar, como ya señalara Lennon (1990), que el reto para los aprendientes de nivel avanzado en lo que se refiere al léxico reside en ampliar su almacén de palabras (*wordstore*) al tiempo que aumentar el dominio de uso de las palabras que ya se conocen.

9.2.3. *Diccionarios*

Nos ocupamos aquí de las implicaciones de los hallazgos de nuestro análisis en relación con los diccionarios de uso para aprendientes de ELE. La mayor parte de los diccionarios no sirven de mucha ayuda a los aprendientes que necesitan conocer más detalles sobre la manera en la que se usa una palabra en un contexto específico; no suelen incluir ejemplos, y, si lo hacen, son muy pocos los que aparecen, a causa de las limitaciones de espacio. Sería muy útil poder contar con un diccionario basado en la actuación de los nativos en un tipo de texto particular, como, por ejemplo, la escritura argumentativa; no obstante, como señala Ädel (2006: 206), si consideramos que los aprendientes buscan poder redactar buenas argumentaciones, tal vez no deberíamos usar como modelo la escritura de los estudiantes nativos (ya hemos hecho referencia a sus carencias y limitaciones), sino la escritura profesional nativa.

Dicho diccionario no tendría que incluir una gran cantidad de entradas, dado que hay un número limitado de vocabulario activo que los aprendientes aprenden a usar; pero es deseable que cubra las palabras que ocurren a menudo en el corpus considerado de referencia. De las 5000 palabras más frecuentes del español, 1072 son verbos, y, si extraemos de la lista algunos que no podrían

aparecer en una argumentación escrita como *joder*, hablamos de un diccionario que incluiría en torno a 1000 verbos, con sus especificaciones de uso tanto sintácticas como semánticas. *Dar*, sin duda, sería uno de ellos. El diccionario recogería las construcciones sintácticas en las que aparece *dar* (desde la construcción V + N hasta las locuciones), pero no se deberían incluir las estructuras con *dar* que no se usan o que, al menos, no aparecen en las argumentaciones de los nativos. Junto a cada estructura, aparecerían las palabras o tipos de palabras que coaparecen con el predicado; si este selecciona las palabras por determinado rasgo semántico que estas contienen, todas aquellas que compartan dicho rasgo se agrupan en la misma clase léxica.

En español contamos con los diccionarios *Práctico* y *Redes*, que contienen información muy rica en torno a la restricción léxico-sintáctica de *dar*. Por comodidad para el manejo por parte de los aprendientes, convendría informatizarlos y añadir a la información contenida en ellos, por un lado, las nuevas clases léxicas propuestas en esta investigación y, por otro lado, las estructuras sintácticas obtenidas del análisis de corpus realizado pero que no son contempladas en estos dos diccionarios. Al mismo tiempo, sería útil ampliar las clases léxicas contenidas en los diccionarios con los sustantivos hallados en el corpus que no están incluidos todavía; dado que el verbo no selecciona de manera exhaustiva todos los sustantivos que pudieran integrarse en una misma clase léxica, sí que conviene introducir todos los que, al menos, sabemos que coaparecen con el verbo. Como se acaba de señalar, si la información lexicográfica estuviera informatizada, resultaría más sencillo actualizarla a medida que se realizaran más estudios de corpus que nos fueran dando nuevos resultados.

9.3. PERSPECTIVAS DE FUTURO DE LOS ESTUDIOS DE LA INTERLENGUA DEL ESPAÑOL BASADOS EN CORPUS

El estudio de la interlengua del español basado en corpus está comenzando su desarrollo. En unos años, se espera que haya un aumento en la cantidad de los corpus contrastivos del español y en su tamaño, gracias a los avances informáticos. En relación con ello, habrá también más corpus etiquetados y *software* disponible para realizar búsquedas muy avanzadas.

Este desarrollo permitirá realizar investigaciones sobre la interlengua que abarquen aspectos más generales (de los que aquí se examinan) que posibiliten sistematizar mejor los rasgos de la lengua de los aprendientes en cada estadio.

Así, por ejemplo, una vez que el corpus CEDEL2 sea anotado por categorías de palabras, como nombres, adjetivos, adverbios, preposiciones, conjunciones y verbos, podrá analizarse el uso que se hace de cada categoría gramatical; podrá hacerse un listado de lemas utilizados y lemas ausentes, y extraer las diferencias en la distribución de las formas utilizadas de un lema²⁰⁷. Tras realizar el análisis de cada categoría en el corpus de un grupo de aprendientes determinado, será posible conocer las desviaciones de los aprendientes en cada una de las categorías, lo que a su vez permitirá describir de manera más sistemática la interlengua de ese grupo. A partir de ahí, se debería pasar a analizar la profundidad del conocimiento de elementos o rasgos comunes pero más concretos, de un modo similar al análisis que hemos realizado del uso del verbo *dar*. Sería interesante continuar con el estudio de los verbos frecuentes del español (como *hacer* y *tomar*) para poder determinar las tendencias generales de este grupo con respecto a esta área particular del español. Las perspectivas tomadas en este estudio pueden servir de referencia para el análisis de los verbos de la interlengua y para otras categorías de palabras, aunque dada la diferente naturaleza léxica de cada categoría, posiblemente se requerirán nuevos diseños y metodologías.

También se espera que se vayan reforzando los lazos entre los lingüistas de corpus de aprendientes y los investigadores de ASL, de tal modo que los análisis sirvan para contrastar las hipótesis formuladas en el ámbito de la ASL²⁰⁸.

Se prevé, asimismo, que poco a poco se vayan desarrollando los corpus de aprendientes de español con fines específicos (español académico, español de los negocios, jurídico, etcétera), y que, asimismo, se creen corpus multilingües (diferentes corpus de aprendientes de español comparables con diferentes L1) y corpus comparables de diferentes interlenguas (como el WriCLE y el CEDEL2), que permitirán explorar el papel de la transferencia en ambas direcciones (como es el caso del trabajo [tesis doctoral en preparación] de Ferrandis-Estrella titulado *Crosslinguistic interference at the interfaces: subjects in L2 grammars*) e investigar si los déficits de la interlengua resultan de la transferencia de la L1 o del *input* recibido, o son consecuencia de patrones de desarrollo universales, como también observan Lozano y Mendikoetxea (2013: 7). Con el tiempo, si se compilan más corpus longitudinales, como el CEDEL2, o SPLLOC, también

²⁰⁷ A esta previsión también se refería Guo (2006) a propósito de la interlengua del inglés hace casi una década, cuando concluía su análisis de los verbos en el inglés escrito de aprendientes chinos; esto demuestra, por tanto, el trabajo que todavía tenemos por delante en el desarrollo de los estudios de interlengua del español.

²⁰⁸ Granger (2009: 14) y Lozano y Mendikoetxea (2013: 17) se refieren a la falta de relaciones entre estos dos ámbitos.

habrá más estudios de interlengua del español que compararán la interlengua de varios niveles: la interlengua de nivel inicial con la del nivel intermedio, o la del intermedio con la del avanzado.

Aunque, hasta ahora, los análisis de la interlengua del español basados en corpus o en otras fuentes de datos se localizan de forma dispersa en diferentes publicaciones²⁰⁹, nos encontramos ante un panorama investigador en el que se están comenzando a abrir cauces estables de publicación, lo que permitirá construir un cuerpo manejable de datos sobre la interlengua del español, esto es, sobre el aprendizaje/adquisición del español como LE/L2. Merece una atención especial el lanzamiento del *International Journal of Learner Corpus Research*, cuyo primer número verá la luz en 2015; esta revista está planteada para investigadores que compilan, anotan o analizan corpus de aprendientes de cualquier lengua, por lo que supondrá un punto de referencia para la investigación de esta rama de estudio. Asimismo, indica el auge de este ámbito la reciente creación de la Asociación de Corpus del Aprendiente o Learner Corpus Association (LCA), que ha organizado las jornadas y congresos sobre estudios de corpus de aprendientes que tuvieron lugar en Lovaina la Nueva (2011) y en Bergen (2013), y que se llevarán a cabo en la universidad holandesa de Radboud en septiembre de 2015. Está prevista también la creación de una asociación sobre la lengua del aprendiente de español que suponga un punto de referencia y de diálogo en los estudios de este ámbito, a cuyo cargo esté la organización de eventos en torno a este tema y la creación de una revista especializada.

De esta manera, en los años venideros se prevén avances más rápidos en la descripción de la interlengua del español, y se espera que los hallazgos de estos estudios se encuentren más accesibles a los profesores, creadores de materiales y otros investigadores.

En conclusión, se han alcanzado grandes logros desde el comienzo de la era de la investigación de corpus de aprendientes informatizados, a principios de los años noventa. El estudio del corpus de aprendientes es ahora un campo

²⁰⁹ En su gran mayoría, se trata de revistas con un enfoque amplio que también comprenden trabajos de índole no “adquisicionalista”, como la literaria o la puramente didáctica; este es el caso de *MarcoELE* o *RedELE*. La *Revista Nebrija de Lingüística Aplicada a la Enseñanza* acepta contribuciones de cualquier ámbito teórico y metodológico relacionado con la lingüística aplicada; y la de más reciente aparición, *Journal of Spanish Language Teaching* (primer número del 2014), publicada por Routledge, recoge investigaciones centradas exclusivamente en el terreno de la enseñanza del español –con un claro equilibrio entre el aspecto más práctico de la enseñanza y el discurso teórico y académico– y posee vocación interdisciplinar. Aparte de estas revistas, estudios de corpus de la interlengua del español pueden ubicarse diseminados también en cualquiera de las revistas especializadas en Lingüística de Corpus, de carácter general, o en revistas de Humanidades, esto es, no relacionadas con el español como lengua extranjera. Esta dispersión hace que los análisis de corpus realizados para estudiar la interlengua del español sean difíciles de rastrear.

consolidado de la Lingüística Aplicada, un campo que está en continua evolución. Aunque el ACI se ha convertido en un paradigma fructífero, en el caso de la interlengua del español se está iniciando ahora su desarrollo; poco a poco, por medio de diferentes modos de acceder a los datos, vamos descubriendo algunos de los rasgos más significativos del español de los aprendientes. Los lazos entre la investigación de corpus y la SLA son todavía débiles; si se logran reforzar, el futuro para la enseñanza de ELE y para el entendimiento de la interlengua y los procesos de adquisición de la L2 es prometedor. No obstante, todavía quedan importantes cuestiones metodológicas por resolver para que los investigadores de ASL se beneficien plenamente de los estudios de CAI.

REFERENCIAS BIBLIOGRÁFICAS

- ÄDEL, Annelie (2006): *Metadiscourse in L1 and L2 English*. Amsterdam: John Benjamins.
- AITCHISON, Jean (1987): "Reproductive furniture and extinguished professors". En *Language Topics. Essays in Honour of Michael Halliday*, editado por R. Steele y T. Threadgold, Volumen 2, 3-14. Amsterdam: John Benjamins.
- AITCHISON, Jean (1994): *Words in the Mind: an Introduction to the Mental Lexicon*. Oxford: Blackwell.
- ALLEN, Virginia F. (1983): *Techniques in Teaching Vocabulary*. Oxford: Oxford University Press.
- ALONSO RAMOS, Margarita (1994-1995): "Hacia una definición del concepto de colocación: de J.R. Firth a I. A. Mel'cuk". *Revista de Lexicografía* I: 9-28.
- ALONSO RAMOS, Margarita (1997): "Coocurrencia léxica y descripción lexicográfica del verbo 'dar'. Hacia un tratamiento de los verbos soportes". *Zeitschrift für Romanische Philologie* 113: 380-417.
- ALONSO RAMOS, Margarita (2004): *Las construcciones con verbo de apoyo*. Madrid: Visor Libros.
- ALONSO RAMOS, Margarita, dir.: *Diccionario de colocaciones del español* [en línea] <http://www.dicesp.com/paginas>
- ALONSO RAMOS, Margarita, (2010a): "Tagging collocations for learners". En *eLexicography in the 21st Century: New Challenges, New Applications: Proceedings of ELEX2009, Cahiers du CENTAL 7*, editado por S. Granger y M. Paquot. Lovaina la Nueva: Presses universitaires de Louvain.
- ALONSO RAMOS, Margarita, (2010b). "Towards a motivated annotation schema of collocation errors in learner corpora". En *Proceedings of the Seventh Conference on International Language Resources and Evaluation (LREC'10)*, editado por N. Calzolari. Valetta, Malta: Language Resources Evaluation.
- ALTENBERG, Bengt y Silviene GRANGER (2001): "The grammatical and lexical patterning of MAKE in native and non-native student writing". *Applied Linguistics* 22 (2): 173-195.
- ALTENBERG, Bengt (2011): "Preface". En *A Taste for Corpora. In honour of Sylvianne Granger*, editado por F. Meunier et al., xiii-xv. Amsterdam: Benjamins.
- ÁLVAREZ CAVANILLAS, J. Luis (2008): *Algunas aplicaciones del enfoque léxico al aula de ELE. Memoria de Máster, Universidad de Barcelona* [en línea]: <

- <http://www.mepsyd.es/redele/Biblioteca2008/JL.Cavanillas/memoriaJose%20Luis%20cavanillas.pdf> > [consulta: 21-03-2013].
- BÁEZ, Valerio (1995): *Bibliografía de Lingüística general y española (1964-1990)*, Tomo II. Alcalá de Henares: Servicio de Publicaciones de la Universidad de Alcalá.
- BARALO, Marta (2001): “El lexicon no nativo y las reglas de la gramática”. En *Tendencias y líneas de investigación en adquisición de segundas lenguas*, editado por V. Salazar y S. Pastor, 23-38. Alicante: Universidad de Alicante.
- BARALO, Marta (2004): *La adquisición del español como segunda lengua*. Madrid: ArcoLibros
- BARALO, Marta (2005): “Aspectos de la adquisición del léxico y su aplicación en el aula”. FIAPE. I Congreso internacional: El español, lengua del futuro [en línea] <<http://www.mepsyd.es/redele/biblioteca2005/fiape/baralo.pdf>> (15-12-2013).
- BARALO, Marta (2006): “Cómo crear redes entre palabras en el aula de ELE”. Conferencia presentada al III Encuentro Práctico de Profesores de ELE, Würzburg, 19-20 de mayo de 2006 [en línea]: <<http://www.encuentro-practico.com/pdfw06/baralo.pdf>> [consulta: 15-02-2014].
- BARALO Marta (2007): “Adquisición de palabras: redes semánticas y léxicas”. Actas del Foro de español internacional: Aprender y enseñar léxico [en línea] <http://www.cervantes-muenchen.de/es/05_lehrerfortb/Actas06-07/3MBaralo.pdf> [consulta: 29-04-2011].
- BARALO Marta (2010): “La investigación en español como lengua segunda: necesidad de un corpus de español no nativo”. Ponencia impartida en el V Congreso Internacional de la Lengua Española (CILE) [en línea] <http://congresosdelengua.es/valparaiso/ponencias/lengua_educacion/baralo_marta.htm> [consulta: 29-04-2011].
- BARLOW Michael Y Suzanne KEMMER, eds. (2000): *Usage Based Models of Language*. Palo Alto: CSLI Publications.
- BENINCÀ, Paola (1993): “Sintassi”. En *Introduzione al italiano contemporaneo. Le strutture*, editado por A. A. Sobrero, 247-288. Roma-Bari: Laterza.
- BIBER, Douglas et al. (1999): *Longman Grammar of Spoken and Written English*. Harlow: Pearson Education.
- BLANCO ESCODA, Xavier (2000): “Verbos soporte y clases de predicados en español”. LEA XXII: 99-117.
- BOERS, Frank (2009): “Optimizing a Lexical Approach in TEFL”. Conferencia impartida en la Universidad de Extremadura, Cáceres, 15 de noviembre de 2009.

- BOERS, Frank y Seth LINDSTROMBERG (2008): "How cognitive linguistics can foster effective vocabulary teaching". En *Cognitive Linguistic Approaches to Teaching Vocabulary and Phraseology*, editado por F. Boers y S. Lindstromberg, 1-61. Berlin: Mouton de Gruyter.
- BOERS, Frank y Seth LINDSTROMBERG (2009): *Optimizing a Lexical Approach to Instructed Second Language Acquisition*. Basingstoke: Palgrave Macmillan.
- BOGAARDS, Paul (1996): "Lexicon and grammar in second language learning". En *Investigating Second Language Acquisition*, editado por P. Jordens y J. Lalleman, 357-379. Berlin: Mouton de Gruyter.
- BOGAARDS, Paul (2001): "Lexical units and the learning of foreign language vocabulary". *Studies in Second Language Acquisition* 23(3): 321-344.
- BOLINGER, Dwight (1976): "Meaning and memory". *Forum Linguisticum* 1:1-14.
- BORREGO NIETO, Julio (1999): "El español de Castilla León. ¿Modelo lingüístico o complejo dialectal?". En *La lengua española patrimonio de todos*, coordinado por A. Álvarez Tejedor, 13-40. Burgos: Caja de Burgos.
- BORREGO NIETO, Julio (2001) "EL concepto de norma regional y su aplicación a las hablas castellano-leonesas". II Congreso Internacional de la Lengua Española". RAE / Instituto Cervantes, Valladolid, octubre de 2001. Publicado en: <http://cvc.cervantes.es/obref/congresos/valladolid/>
- BOSQUE, Ignacio (2001a): "Sobre el concepto de colocación y sus límites". *LEA* 23 (1): 8-35.
- BOSQUE, Ignacio (2001b): "On the Weight of Light Predicates". En *Features and Interfaces in Romance. Essays in honor of Heles Contreras*, editado por Julia Herschensohn et al., 23-38. Amsterdam/Philadelphia: Benjamins.
- BOSQUE, Ignacio (2004): "La direccionalidad en los diccionarios combinatorios y el problema de la selección léxica". En *Lingüística teòrica: anàlisi i perspectives*, editado por Teresa Cabré, 13-58. Bellaterra: Universitat Autònoma de Barcelona.
- BOSQUE, Ignacio (2005 [2004]): "Combinatoria y significación. Algunas reflexiones", En *Redes. Diccionario combinatorio del español contemporáneo*, dirigido por I. Bosque, LXXVII-CLXXIV. Madrid: SM.
- BOSQUE, Ignacio, dir. (2005): *Redes. Diccionario combinatorio del español contemporáneo*. Madrid: SM.
- BOSQUE, Ignacio, dir. (2006): *Práctico. Diccionario combinatorio del español contemporáneo*. Madrid: SM.

- BOSQUE, Ignacio y Javier GUTIÉRREZ-REXACH (2009): *Fundamentos de Sintaxis Formal*. Madrid: Akal.
- BRIEÑO GARCÍA, Claudia (2013): *El léxico del mundo animal y el acceso léxico sinonímico en ELE: estudio semántico y didáctico*. Tesis doctoral, Universidad Complutense de Madrid [en línea]: <<http://eprints.ucm.es/22358/>> [consulta: 26-03-2014].
- BYBEE, Joan y Paul HOOPER, eds. (2001): *Frequency and the Emergence of Linguistic Structure*. Amsterdam: Benjamins.
- BUTT, Miriam y Wilhelm GOUDER (2001): “On the (Semi)lexical Status of Light Verbs”. En *Semi-lexical Categories. The Functions of Content Words and the Content of Function Words*, editado por N. Corver y H. V Riemsdijk, 1-19. Berlin/Nueva York: Mouton de Gruyter.
- CALVO, Julio (2005): “Los sentidos del lenguaje”. En *Lenguaje y comunicación*, editado por A. López y B. Gallardo, 155-215. Valencia: Universitat de València.
- CANALE, Michael y Merrill SWAIN (1980): “Theoretical bases of communicative approaches to second language teaching and testing”. *Applied Linguistics* 1: 1-47.
- CANDLIN, Christopher N. (1990): “Hacia la enseñanza de lenguas basada en tareas”. *Comunicación, lenguaje y educación*, 7-8: 33-53.
- CARTER, Ron y Michael M. MCCARTHY (1988): *Vocabulary and Language Teaching*. London: Longman.
- CASTILLO CARBALLO, M.^aAuxiliadora (2001): “Colocaciones léxicas y variación lingüística: implicaciones didácticas”. *Lingüística española actual* XXIII/1: 133-144.
- CERVERO, M.^a Jesús y Francisca PICHARDO (2000): *Aprender y enseñar vocabulario*. Madrid: Edelsa.
- CESTERO MANCERA, Ana M.^a et al. (2001): *Corpus para el análisis de errores de aprendices de E/LE (CORANE)*. Actas del XII Congreso Internacional de ASELE: tecnologías de la información y de las comunicaciones en la enseñanza de la E/LE, [en línea]: <http://cvc.cervantes.es/ensenanza/biblioteca_ele/asele/pdf/12/12_0527.pdf> [consulta: 25-06-2013].
- CHI, Man-lai, et al. (1994). “Collocational problems amongst ESL learners: A corpus-based study.” En *Entering Text*, editado por L. Flowerdew y A. K. Tong, 157–165. Hong Kong: University of Science and Technology.
- CHOMSKY, Noam (1995): *The Minimalist Program*. Cambridge MA: MIT Press

- CONSEJO DE EUROPA (2002): Marco común Europeo de referencia para las lenguas. Aprendizaje, enseñanza y evaluación. Madrid: Ministerio de Educación, Cultura y Deporte, y Anaya.
- CORDER, Stephen Pit (1967): "The significance of learners' errors". IRAL 5: 161-170.
- CORPAS PASTOR, Gloria (2001): "Apuntes para el estudio de la colocación". Lingüística española actual, XXIII (1): 41-57.
- COSERIU, Eugenio (1967): Teoría del lenguaje y lingüística general. Madrid: Gredos.
- COSERIU, Eugenio (1977): Principios de semántica estructural. Madrid: Gredos.
- COSERIU, Eugenio (1981): Lecciones de lingüística general. Madrid: Gredos.
- COSERIU, Eugenio (1990): "Semántica estructural y semántica cognitiva". I Jornadas de Filología. Homenaje al profesor Francisco Marsá 239-282. Barcelona: Universidad de Barcelona.
- CRAIK, Fergus y Robert S. LOCKHART (1972): "Levels of processing: a framework for memory research". Journal of Verbal Learning and Verbal Behavior 11: 671-684.
- CRUTTENDEN, Alan (1981): "Item-learning and system-learning". Journal of Psycholinguistic Research 10: 79-88.
- CRYSTAL, David (1995): The Cambridge Encyclopedia of the English Language. Cambridge: Cambridge University Press.
- DAVIES, Mark (2002) Corpus del Español: 100 million words, 1200s-1900s [en línea] <<http://www.corpusdelespanol.org>> [consulta: 28-03-2014].
- DAVIES, Mark y Dee Gardner (2010) A Frequency Dictionary of Contemporary American English. Londres: Routledge.
- DE ALBA, Virginia (2009): "El análisis de errores en el campo de ELE. Algunas cuestiones metodológicas". Revista Nebrija de Lingüística aplicada a la enseñanza de lenguas, 5.
- DE COCK, Sylvie et al. (2011): "Putting corpora to good uses". En A Taste for Corpora. In honour of Sylviane Granger, editado por F. Meunier et al., 1-6. Amsterdam: Benjamins.
- DE MIGUEL, Elena (1999): "El aspecto léxico". En Gramática descriptiva de la lengua española, dirigido por I. Bosque y V. Demonte, 2977-3061. Madrid: Espasa.
- DE MIGUEL, Elena (2008): "Construcciones con verbo de apoyo en español. De cómo entran los nombres en la órbita de los verbos". Actas de XXXVII Simposio Internacional de la Sociedad Española de Lingüística [en línea]: <<http://www.unav.es/linguis/simposiosel/actas/>>

- DE MIGUEL, Elena (2011): "En qué consiste ser verbo de apoyo". En 60 problemas de gramática, editado por M.V. Escandell Vidal et. al., 139-146. Madrid: Akal.
- DEMONTE, Violeta (2003): "Lengua estándar, norma y normas en la difusión actual de la lengua española". Circunstancia [en línea]: <<http://digital.csic.es/bitstream/10261/13074/1/Documento1.pdf>> [consulta: 15-06-2013].
- DIK, Simon C. (1997): *The Theory of Functional Grammar*. Berlin: Mouton de Gruyter.
- DOMÍNGUEZ, Laura et al. (2013): "The role of dynamic contrasts in the L2 acquisition of Spanish past tense morphology". *Bilingualism: Language and Cognition*, 16 (3): 558-577
- DULAY, Heidi et al. (1982): *Language two*. Oxford: Oxford University Press.
- ELLIS, Rod (1994): *The Study of Second Language Acquisition*. Oxford: Oxford University Press.
- ELLIS, Nick C. (2011): "Implicit and Explicit SLA and Their Interface". En *Implicit and Explicit Language Learning*, editado por C. Sanz y R.L. Leow, 35-47. Washington: Georgetown University Press.
- ENKVIST, Nils Erik (1973): "Should we count error or measure success?". En *Errata: Papers in error analysis*, editado por J. Svartvik, 16-23. Lund: Gleerup/Liber.
- ERMAN, Britt y Beatrice WARREN (2000): "The idiom principle and the open choice principle". *Text* 20: 87-120.
- ESCANDELL, M. Victoria et al. eds (2011): *60 Problemas de Gramática*. Madrid: Akal.
- ESCUTIA, Marciano (2012). "Expletives and Unaccusative Predicates in L2". *Higher Education of Social Science*, 2 (3): 1-14.
- FERNÁNDEZ, Sonsoles (1997): *Interlengua y análisis de errores*. Madrid: Edelsa.
- FERRARO Gabriela, et al. (2011): "Collocations: A Challenge in Computer Assisted Language Learning". En *Proceedings of the Fifth International Conference on Meaning-Text Theory*, editado por I. Boguslavsky y L. Wanner, 69-79.
- FERRARO Gabriela, et al. (2013) "Towards advanced collocation error correction in Spanish learner corpora". *Language Resources and Evaluation*.
- FILLMORE, Charles J. (1976). "The Need for a Frame Semantics within Linguistics". *Statistical Methods in Linguistics*, 5-29. Estocolmo: Skriptor.
- FILLMORE, Charles J. et al. (1988): "Regularity and Idiomaticity in Grammatical Constructions: The Case of let alone". *Language* 64: 501-38.
- FIRTH, John R. (1957): *Papers in Linguistics 1934-1951*. Londres: Oxford University Press.

- FITCH, Roxana (2011): *Diccionario de coloquialismos y términos dialectales del español*. Madrid: Arco Libros.
- GARCÍA-PAGE, Mario (2008): *Introducción a la fraseología española. Estudio de las locuciones*. Anthropos: Barcelona.
- GILQUIN, Gaetanelle y Stephan Grice (2009): "Corpora and experimental methods: a state-of-the-art review". *Corpus Linguistics and Linguistic Theory* 5 (1): 1-26.
- GILQUIN, Gaetanelle, et al. (2010): *The Louvain International Database of Spoken English Interlanguage. Handbook and CD-ROM*. Lovaina la Nueva: Presses universitaires de Louvain.
- GÓMEZ MOLINA, J. Ramón (1997): "El léxico y su didáctica: una propuesta metodológica". *Reale* 8: 69-93.
- GÓMEZ TORREGO, Leonardo (1999): "Los verbos auxiliares. Las perífrasis verbales de infinitivo". En *Gramática descriptiva de la lengua española*, dirigida por I. Bosque y S. Varela, 3323-3389. Madrid: Espasa.
- GRANGER, Sylviane (1996a): "From CA to CIA and back: An integrated approach to computerized bilingual and learner corpora". En *Languages in Contrast. Papers from a Symposium on Text-based Cross-linguistic Studies*, editado por K. Aijmer et al., 37-51. Lund: Lund University Press.
- GRANGER, Sylviane (1996b): "Romance words in English: From history to pedagogy". *KVHAA Konferenser* 36, 105-121
- GRANGER, Sylviane (1998a): "Introduction". En *Learner English on Computer*, editado por S. Granger, xxi-xxii. Londres: Longman.
- GRANGER, Sylviane (1998b): "The computer learner corpus: a versatile new source of data for SLA research". En *Learner English on Computer*, editado por S. Granger, 3-18. Londres: Longman
- GRANGER, Sylviane (2002): "A bird's-eye view of learner corpus research". En *Computer Corpora, Second Language Acquisition and Foreign Language Teaching*, editado por S. Granger et al., 3-33. Amsterdam: Benjamins.
- GRANGER, Sylviane (2003): "Error-tagged learner corpora and CALL: a promising synergy". *CALICO (Special issue on Error Analysis and Error Correction in Computer-Assisted Language Learning)* 20 (3): 465-480.
- GRANGER, Sylviane (2004): "Computer learner corpus research: current status and future prospects". En *Applied corpus Linguistics: A Multidimensional Perspective*, editado por U. Connor y T. A. Upton. Amsterdam: Rodopi.

- GRANGER, Sylviane (2009): "The contribution of learner corpora to second language acquisition and foreign language teaching: A critical evaluation". En *Corpora and Language Teaching*, editado por K. Aijmer, 13-32. Amsterdam: Benjamins.
- GRANGER, Sylviane et al., eds. (2002): *Computer Corpora, Second Language Acquisition and Foreign Language Teaching*. Amsterdam: Benjamins.
- GRANGER, Sylviane et al., eds. (2009): *International Corpus of Learner English. Version 2. Lovaina la Nueva: Presses universitaires de Lovaina*.
- GROSS, Maurice (1975): *Méthodes en syntaxe*. París: Hermann.
- GROSS, Maurice (1989): *Les constructions converses du français*. Genève: Droz.
- GUO, Xiaotian (2006): *Verbs in the Written English of Chinese Learners: A Corpus-based Comparison between Non-native Speakers and Native Speakers*. Tesis doctoral, Universidad de Birmingham. En *The linguistics journal* [en línea]: <http://www.linguistics-journal.com/thesis_Guo.pdf> [consulta: 20-06-2013].
- GUTIÉRREZ CUADRADO, Juan (1996): *Diccionario Salamanca de la lengua española*. Madrid: Santillana, Universidad de Salamanca
- HALLIDAY, Michael A.K. (1961): "Categories of the theory of grammar". *Word* 17(3): 241-292.
- HALLIDAY, Michael A.K. (1966): "Lexis as a linguistic level". En *In Memory of John Firth*, editado por C. E. Bazell et al., 148-162. London: Longman.
- HALLIDAY, Michael A.K. (1977): "Estructura y función del lenguaje". En *Nuevos horizontes de la lingüística*, compilado por J. Lyons, 145-173. Madrid: Alianza.
- HAMMARBERG, B (1973): *The insufficiency of error analysis*. En *Errata: Papers in error analysis*, editado por J. Svartvik, 29-36. Lund: Gleerup/Liber.
- HASSELGARD, Hilde y Stig JOHANSSON (2011): "Learner corpora and contrastive interlanguage analysis". En *A Taste for Corpora. In honour of Sylviane Granger*, editado por F. Meunier et al., 33-61. Amsterdam: Benjamins.
- HAUSMANN, Franz J. (1979): "Un dictionnaire des collocations est-il possible?". *Travaux de Linguistique et de Littérature* 17 (1): 187-195.
- HERRERO INGELMO, JOSÉ L. (2002): "Los verbos soporte: el verbo dar en español". En *Léxico y Gramática* [en línea] <http://diarium.usal.es/joluin/files/2013/12/darsoporte.pdf> [consulta 12-05-2014].
- HIGUERAS GARCÍA, Marta (1996): "Aprender y enseñar léxico". En *Didáctica del español como lengua extranjera* 3, coordinado por L. Miquel y N. Sans, 111-126. Madrid: Fundación Actilibre.

- HIGUERAS GARCÍA, Marta (2006): *Las colocaciones y su enseñanza en la clase de ELE*. Madrid: Arcolibros.
- HILL, Jimmie (1999): "Collocational competence". *English Teaching Professional* 11: 3-6.
- HOFLAND, Knut y Stig JOHANSON (1982): *Word Frequencies in British and American English*. Bergen: Norwegian Computing Centre for the Humanities.
- HOWATT, Anthony P. R. (1984): *A History of English Language Teaching*. Oxford: Oxford University Press.
- HUNSTON, Susan (2002): *Corpora in Applied Linguistics*. Cambridge: Cambridge University Press.
- HUNSTON, Susan y Gill FRANCIS (2000): *Pattern Grammar: Corpus-driven Approach to the Lexical Grammar of English*. Amsterdam: John Benjamins.
- HYMES, Dell (1972): "On communicative competence". En *Sociolinguistics*, editado por J.B. Pride y J. Holmes, 269-293. New York: Penguin.
- INSTITUTO CERVANTES (1997-2014): *Diccionario de términos clave de ELE* [en línea]: <http://cvc.cervantes.es/ensenanza/biblioteca_ele/diccionario_ele/indice.htm>
- JONES, Susan y John SINCLAIR (1974): "English Lexical Collocations. A study in computational Linguistics". *Cahiers de Lexicology* 24: 15-61.
- JACKENDOF, Ray (2003): *Foundations of language: Brain, Meaning, grammar, evolution*. Oxford: Oxford University Press
- KASZUBSKI, Przemek (2000): *Selected aspects of lexicon, phraseology and style in the writing of Polish advanced learners of English: a contrastive, corpus-based approach*. Tesis doctoral, Universidad de Adam Mickiewicz [en línea]: <<http://www.staff.amu.edu.pl/~przemka/rsearch.html#PhD>> [consulta: 21-06-2013].
- KOIKE, Kazumi (2008): "Colocaciones atípicas". *Lingüística española actual* XXX (1): 87-109.
- KOERNER, Konrad E.F. (1982): *Ferdinand de Saussure. Génesis y evolución de su pensamiento en el marco de la lingüística occidental*. Madrid: Gredos.
- KRASHEN, Stephen (1977): "Some issues relating to the monitor model". En *TESOL* 77: *teaching and learning English as a second language*. Trends in research and practice, editado por H. D. Brown, et al., 144-158. Washington: TESOL.
- KRASHEN, Stephen (1982): *Principles and Practice in Second Language Acquisition*. Oxford: Pergamon Press.

- KRASHEN, Stephen et al. (1978): "Two studies in language acquisition and language learning". *ITL Review of applied linguistics* 39/40: 73-92.
- KRASHEN, Stephen y Tracy TERRELL (1983): *The Natural Approach: language acquisition in the classroom*. Oxford: Pergamon.
- KUCERA, Henry y W. Nelson FRANCIS (1967): *Computational Analysis of Present-day American English*. Providence RD: Brown University Press.
- LACA, Brenda (1999): "Presencia y ausencia de determinante". En *Gramática descriptiva de la lengua española*, dirigido por I. Bosque y V. Demonte, 891-928. Madrid: Espasa Calpe.
- LADO, Robert (1957 [1971]): *Linguistics across Cultures: Applied Linguistics for Language Teachers*. Ann Arbor MI: University of Michigan Press.
- LAKOFF, George (1987): *Women, Fire and Dangerous Things: What Categories Reveal about the Mind*. Chicago: Chicago University Press.
- LAKOFF, George y JOHNSON, Mark (1999): *Philosophy in the Flesh: The Embodied Mind and Its Challenge to Western Thought*. Nueva York: Basic Books.
- LANGACKER, Ronald (1987): *Foundations of Cognitive Grammar, Volume 1: Theoretical Prerequisites*. Stanford: Stanford University Press.
- LARDIERE, Donna (1998): "Dissociating syntax from morphology in a divergent L2 end-state grammar". *Second Language Research* 14 (4): 359-375.
- LARSEN-FREEMAN, Diane y Michael H. LONG (1994): *Introducción al estudio de la adquisición de segundas lenguas*. Madrid: Gredos.
- LAUFER, Batia (1990): "Words you know: How they affect the words you learn". En *Further Insights into Contrastive Analysis*, editado por J. Fisiak, 573-593. Amsterdam: John Benjamins.
- LAUFER, Batia (1994): "Appropriation du vocabulaire: mots faciles, mots difficiles, mots impossibles". En *Acquisition et interaction en Langue Etrangere* 3: 97-113.
- LAUFER, Batia (2005): "Focus on Form in second language vocabulary acquisition". En *EUROSLA Yearbook* 5, editado por S. Fister-Cohen, 223-250. Amsterdam: John Benjamins.
- LEECH, Geoffrey (1997): "Introducing corpus annotation". En *Corpus annotation: Linguistic information from computer text corpora*, editado por R. Garside et al., 1-18. Londres: Longman.
- LEECH, Geoffrey (1998): "Preface". En *Learner English on Computer*, editado por S. Granger, XIV-XX. London: Longman.

- LEECH, Geoffrey (2001): "The role of frequency in ELT: new corpus evidence brings re-appraisal". *Foreign Language Teaching and Research* 33 (5): 328-339.
- LEECH, Geoffrey (2011): "Frequency, corpora and language learning". En *A Taste for Corpora*. In honour of Sylvianne Granger, editado por F. Meunier et al., 7-31. Amsterdam: Benjamins.
- LEWIS, Michael (1993): *The Lexical Approach: The State of ELT and a Way Forward*. Hove: Language Teaching Publications.
- LEWIS, Michael (1997): *Implementing the Lexical Approach: Putting Theory Into Practice*. Hove: Language Teaching Publications.
- LEWIS, Michael (2000): *Teaching collocation: Further Developments in the Lexical Approach*. Hove: Language Teaching Publications.
- LEZCANO, Emma (1994-95): "Una aproximación a la gramática de valencias". *Lenguaje y textos* 6-7: 161-174.
- LI, Wen-Zong(2003): "A CLEC-based analysis of key words and associates". *Modern Foreign Languages* 26 (3): 284-293.
- LICERAS, Juana (2009): "La interlengua del español en el siglo XXI". *Revista Nebrija de Lingüística Aplicada a la Enseñanza de Lenguas* 5.
- LINDSTROMBERG, Seth y Frank BOERS (2008): *Teaching Chunks of Language. From noticing to remembering*. Helbling Languages.
- LINNARUD, Moira (1986): *Lexis in Composition: A Performance Analysis of Swedish Learners' Written English*. Lund: Gleerup/Liber.
- LONG, Michael H. (1991): "Focus on form: A design feature in language teaching methodology". En *Foreign Language Research in Cross-Cultural Perspective*, editado por K. De Bott et al., 39-52. Amsterdam: John Benjamins.
- LONG, Michael H. (1997): "Focus on form in task-based language teaching". En *The Annual McGraw- Hill Teleconference in Second Language Teaching* [en línea]: <<http://www.mhhe.com/socscience/foreignlang/top.htm>> [consulta: 20-08-2010].
- LÓPEZ-MEZQUITA, M. Teresa (2007): *La evaluación de la competencia léxica: Test de vocabulario. Su fiabilidad y validez*. Madrid: Secretaría General Técnica
- LOZANO, Cristóbal (2009a): "CEDEL2: Corpus Escrito del Español L2". En *Applied Linguistics Now: Understanding Language and Mind*, editado por Carmen M. Bretones Callejas et al., 197-212. Almería: Universidad de Almería.
- LOZANO, Cristóbal (2009b). "Selective deficits at the syntax-discourse interface: Evidence from the CEDEL2 corpus". En *Representational Deficits in Second*

- Language Acquisition, editado por Y. Leung et al., pp. 127-166. Amsterdam: John Benjamins.
- LOZANO, Cristóbal y Amaya Mendikoetxea (2013): "Learner corpora and SLA: the design and collection of CEDEL2". En *Automatic Treatment and Analysis of Learner Corpus Data*, editado por A. Díaz-Negrillo et al., 65-100. Amsterdam: Benjamins.
- LOZANO ZAHONERO, María (2011): *Gramática de perfeccionamiento de la lengua española*. Milán: Urlico Hoepli.
- LUQUE DURÁN, Juan de Dios (2002): "Sobre el papel del lexicón y la evolución de los lenguajes naturales". *Language Design*, 4: 111-147.
- LYONS, John (1980): *Semántica*. Barcelona: Teide.
- MAIRAL USÓN, Ricardo (2010): *Teoría Lingüística: Métodos, Herramientas y Paradigmas*. Madrid: Fundación Ramón Areces/ UNED.
- MAKKAI, Adam (1975): "Introduction". En *A Dictionary of American Idioms*, editado por A. Makkai et al., VI-VIII. New York: Barrons's Educational Series.
- MARTÍ, María A., coord. (1996): *Gran diccionario de la lengua española*. Barcelona: Larousse.
- MARTÍN ARISTA, Javier ed. (1999) *Estudios de Gramática Funcional*. Zaragoza: Mira Editores.
- MCENERY, Tony y Andrew HARDIE (2012): *Corpus Linguistics*. Cambridge: University Press.
- MEARA, Paul (1996): The dimensions of lexical competence. En *Performance and competence in second language acquisition*, editado por G. Brown et al. Cambridge: Cambridge University Press.
- MEL'ČUK, Igor (1996): "Lexical Functions in Lexicography and NLP". En *Lexical Functions: A Tool for the Description of Lexical relations in a Lexicon*, editado por L. Wannet, 37-102. Amsterdam: John Benjamins.
- MEL'ČUK, Igor et al. (1984-1999): *Dictionnaire explicatif et combinatoire du français contemporain*. Universidad de Montreal.
- MENDÍVIL, José L. (1999): *Las palabras disgregadas. Sintaxis de las expresiones idiomáticas y los predicados complejos*. Zaragoza: Prensas Universitarias de Zaragoza.
- MEUNIER, Fanny (1998): "Computer tools for the analysis of learner corpora". En *Learner English on Computer*, editado por S. Granger, 19-37. Londres: Longman

- MITATOU, Zafeira (2011): “Las construcciones del verbo dar en la enseñanza del español como lengua extranjera (E/LE)”. RedELE, 23 [en línea] http://www.mecd.gob.es/dctm/redele/Material-RedEle/Revista/2011_23/2011_redELE_23_15Reina%20Mitatu.pdf?documentId=0901e72b8101eb7a [consulta 12-05-2014]
- MITCHELL Rosamond, et al. (2008): “SPLLOC: A new corpus for Spanish second language acquisition research”. En *EUROSLA Yearbook 8*, editado por L. Roberts et al., 287-304. Amsterdam: Benjamins.
- MONTORO DEL ARCO, Esteban T. (2010): “Norma y uso de las colocaciones léxicas en español”. En *Norma y usos correctos en el español actual*, coordinado por M. Aleza Izquierdo et al., 249-266. Valencia: Tirant lo Blanch.
- MORANTE VALLEJO, Roser (2005): *El desarrollo del conocimiento léxico en segundas lenguas*. Madrid: Arco libros.
- MORENO CABRERA, Juan C. (1999): “Las funciones informativas: las perífrasis de relativo y otras construcciones perifrásticas”. En *Gramática descriptiva de la lengua española*, dirigido por I. Bosque y V. Demonte, 4245-4302. Madrid: Espasa Calpe.
- MORIMOTO, Yuko (1998): *El aspecto léxico: delimitación*. Madrid: Arco libros.
- MUÑOZ-BASOLS, Javier (2014): “Hacia una internacionalización del discurso sobre la enseñanza del español como lengua extranjera”. *Journal of Spanish Language Teaching 1* (1): 1-14.
- MUÑOZ GUTIÉRREZ, Carlos: *La mente inteligente* [en línea]: <<http://pendientedemigracion.ucm.es/info/pslogica/mente/>> [consulta: 28-03-2014].
- MYLES, Florence (2007). “Using electronic corpora in SLA research”. En *Handbook of French Applied Linguistics*, editado por D. Ayoun, 377-400. Amsterdam: Benjamins.
- NAGINI, Pushpa (2008): “The Mental Lexicon”. *Journal of the Indian Academy of Applied Psychology 34* (1): 181-186.
- NATION, Paul (1990): *Teaching and Learning Vocabulary*. New York: Newbury House.
- NATION, Paul (2001): *Learning Vocabulary in Another Language*. Cambridge: Cambridge University Press.
- NATTINGER, James R. (1980): “A lexical phrase grammar for ESL”. *TESOL Quarterly*: 337-344.

- NATTINGER, James R. y Jeanette S. De Carrico (1992): *Lexical Phrases and Language Teaching*. Oxford: Oxford University Press.
- NEMSER, William (1971): "Approximative systems of foreign language learner". *IRAL* 9 (2): 115-123.
- NESSSELHAUF, Nadja (2004): "Learner corpora and their potential for language teaching". En *How to Use Corpora in Language Teaching*, editado por J. Sinclair, 125-152. Amsterdam: Benjamins.
- NESSSELHAUF, Nadja (2005): *Collocations in a Learner Corpus*. Amsterdam: Benjamins.
- OSTER, Ulrique (2009): "La adquisición del vocabulario en una lengua extranjera; de la teoría a la aplicación didáctica". *Porta Linguarum* 11: 33-50.
- PASTOR CESTEROS, Susana (2005): "La enseñanza de segundas lenguas". En *Conocimiento y lenguaje*, editado por A. López y B. Gallardo, 361-399. Valencia: Universitat de València
- PAWLEY, Andrew y Frances SYDER (1983): "Two puzzles for linguistic theory: Nativelike selection and nativelike fluency". En *Language and Communication*, editado por J. Richards y R. Schmidt, 191-226. London: Longman.
- PENADÉS, Inmaculada (1999): *Lingüística contrastiva y análisis de errores*. Madrid: Edinumen.
- PENADÉS, Inmaculada (2003): "Las clasificaciones de errores lingüísticos en el marco del análisis de errores". *LinRed* 1: 1-29.
- PEREZ BASANTA, Carmen (2003): "Psycholinguistics and Second Language Vocabulary Teaching". *Indian Journal of Applied Linguistics* 29 (2): 35-47.
- PETERS, Ann M. (1983): *Units of Language Acquisition*. Cambridge: Cambridge University Press.
- PIERA, Carlos y Soledad VARELA (1999): "Relaciones entre morfología y sintaxis", en *Gramática descriptiva de la lengua española*, dirigida por I. Bosque y S. Varela, 4367-4422. Madrid: Espasa.
- PINO, Aymé (2012): *El uso de combinaciones de palabras con que en un corpus de aprendices suecos de español como lengua extranjera*. Tesis doctoral, University of Gothenburg.
- POTTIER, Bernard (1963) *Recherches sur l'analyse sémantique en linguistique et en traduction mécanique*. France: Nancy Université.
- PRAGGLEJAZ GROUP (2007): "MIP: A Method for Identifying Metaphorically Used Words in Discourse". *Methaphor y Symbol* 22(1): 1-39.

- PRENDERGAST, Thomas (1864): *The mastery of languages; or, The art of speaking foreign tongues idiomatically*. Londres: Richard Bentley.
- PRIETO GONZÁLEZ, Sabela et al. (2009): “Córpora y enseñanza de lenguas: se buscan colocaciones”. En *A survey on corpus-based research*, editado por P. Cantos Gómez y A. Sánchez Pérez, 336-373. Murcia: AELINCO.
- QUIRK, Randolph (1995): *Grammatical and lexical variance in English*. London: Longman.
- RAYSON, Paul y Alistair BARON (2011): “Automatic error tagging of spelling mistakes in learner corpora”. En *A Taste for Corpora. In honour of Sylvianne Granger*, editado por F. Meunier et al., 109-126. Amsterdam: Benjamins.
- REAL ACADEMIA ESPAÑOLA (2001): *Diccionario de la lengua española*. Madrid: Espasa.
- REAL ACADEMIA ESPAÑOLA: Banco de datos (CREA). Corpus de referencia del español actual [en línea]: <<http://www.rae.es>> [consulta: 20-06-2013].
- REAL ACADEMIA ESPAÑOLA: Diccionario panhispánico de dudas [en línea]: <<http://www.rae.es/dpd/>> [consulta: 26-06-2013].
- REAL ACADEMIA ESPAÑOLA y ASOCIACIÓN DE ACADEMIAS DE LA LENGUA ESPAÑOLA (2009): *Nueva gramática de la lengua española*. Madrid: Espasa.
- REAL ACADEMIA ESPAÑOLA y ASOCIACIÓN DE ACADEMIAS DE LA LENGUA ESPAÑOLA (2009): *Nueva gramática de la lengua española. Manual*. Madrid: Espasa.
- RICHARDS, Jack. C. (1976): “The role of vocabulary teaching”. *TESOL Quarterly* 10, 1: 77-89.
- RICHARDS, Jack C. (1985): *The Context of Language Teaching*. Cambridge: Cambridge University Press.
- RICHARDS, Jack C. y Theodore S. RODGERS (1998): *Enfoques y métodos en la enseñanza de idiomas*, Cambridge: Cambridge University Press.
- RICHARDS, Jack C. y Theodore S. RODGERS (2009): *Enfoques y métodos en la enseñanza de idiomas*, Madrid: Edinumen.
- RINGBOM, H (1998a): “High frequency verbs in the ICLE corpus”. En *Explorations in Corpus Linguistics*, editado por A. Renouf, 191-200. Amsterdam: Rodopi
- RINGBOM, H (1998b): “Vocabulary frequencies in advanced learner English: A cross-linguistic approach”. En *Learner English on computer*, editado por S. Granger, 41-52. Londres: Longman.
- RODRÍGUEZ RAMALLE, Teresa M. (2005): *Manual de Sintaxis del Español*. Madrid: Castalia.

- ROLLINSON, Paul y Amaya Mendikoetxea (2010): "Learner corpora and second language acquisition: Introducing WriCLE". En *Analizar datos > Describir variación / Analysing Data > Describing Variation*, editado por J. L. Bueno Alonso et al., 1-12. Vigo: Universidad de Vigo (Servizo de Publicacións).
- ROSCH, Eleanor (1973): "On the internal structure of perceptual and semantic categories". En *Cognitive Development and the Acquisition of Language*, editado por T. E. Moore, 111-144. Londres: Academic Press.
- RUHL, Charles (1989): *On Monosemy: A Study on Linguistic Semantics*. Albany: State University of New York Press.
- SÁNCHEZ RUFAT, Anna (2010): "Apuntes sobre las combinaciones léxicas y el concepto de colocación". *Anuario de Estudios Filológicos XXIII*: 291-306.
- SÁNCHEZ RUFAT, Anna (2011): "Léxico gramaticalizado y lengua formulaica: algunas precisiones al enfoque léxico". *Sintagma* 23: 85-98.
- SÁNCHEZ RUFAT, Anna (2014): "Uso interactivo del vocabulario". En *Journal of Spanish Language Teaching* 1: 123-125 [en línea]: <http://www.tandfonline.com/doi/full/10.1080/23247797.2014.898527>
- SÁNCHEZ RUFAT, Anna (en prensa): "La coma en oraciones subordinadas antepuestas: un estudio de corpus contrastivo". En *Ortografía e Imagen Personal*, coordinado por E. T. Montoro del Arco. Granada: Universidad de Granada.
- SANTOS GARGALLO, Isabel (1993): *Análisis contrastivo, análisis de errores e interlengua en el marco de la lingüística contrastiva*. Madrid: Síntesis.
- SCHACHTER, Jacquelyn y Marianne CELCE-MURCIA (1977): "Some reservations concerning error analysis". *TESOL* 11(4): 441-451.
- SCHMIDT, Richard (1990): "The Role of Consciousness in Second Language Learning". *Applied Linguistics* 11: 129-158.
- SCHMITT, Norbert y Paul MEARA (1997): "Researching Vocabulary Through a Word Knowledge Framework. Word Associations and Verbal Suffixes". *Studies in Second Language Acquisition* 19 (1): 17-36.
- SCHMITT, Norbert (2000). *Vocabulary in Language Teaching*. Cambridge: Cambridge University Press.
- SCHMITT, Norbert y Ronald CARTER (2000): "Lexical Phrases in Language Learning". *The Language Teacher Online*: <http://jalt-publications.org/old_tlt/articles/2000/08/schmitt>
- SCHUMANN, John (1976): "Second language acquisition: The pidginization hypothesis". *Language Learning* 26: 391-408.

- SELINKER, Larry (1972): "Interlanguage". *IRAL* 10 (2): 209-231.
- SINCLAIR, John M. (1966): "Beginning the study of lexis". En *In Memory of John Firth*, editado por C. E. Bazell et. al., 410-30: London: Longmans.
- SINCLAIR, John M. (1987): *Look Up: An Account of the COBUILD Project in Lexical Computing*. London: Collins
- SINCLAIR, John M. (1991): *Corpus, Concordance, Collocation*. Oxford: Oxford University Press.
- SINCLAIR, John M., ed. (2004): *How to use corpora in language teaching*. Amsterdam: Benjamins.
- SINCLAIR, John M. et al. (2005). *English Collocation Studies: The OSTI Report*. Londres: Continuum.
- SINCLAIR, John M. y Anna MAURANEN (2006): *Linear Unit Grammar: Integrating Speech and Writing*. Amsterdam: John Benjamins.
- SINGLETON, David (2000): "Connectivité interlinguistique dans l'acquisition et l'exploitation du lexique de la langue seconde". En *Le Français dans le Monde. Actas del coloquio internacional de Crapel, Nancy*, mayo 56-69.
- SKEHAN, Peter (1998): *A Cognitive Approach to Language Learning*. Oxford: Oxford University Press.
- SOBRERO, Alberto A., ed. (1993): *Introduzione al italiano contemporaneo. Le strutture*. Roma-Bari: Laterza.
- SONBUL, Suhad (2012): "Approaches to Second Language Vocabulary Teaching". *The Encyclopedia of Applied Linguistics*, [en línea]: <<http://onlinelibrary.wiley.com/doi/10.1002/9781405198431.wbeal0034/abstract>>
- STEIN, Gabriele (1991): "The Phrasal Verb Type to have a look in Modern English". *International Review of Applied Linguistics* 19 (1): 1-29.
- STENGERS, Hélène (2009): *The Idiom Principle put to the test: An exercise in Applied Comparative linguistics*. Tesis doctoral, Vrije Universiteit Brussel.
- STOCKWELL, Robert et al. (1965): *The Grammatical Structures of English and Spanish*. Chicago: University of Chicago Press.
- SVARTVIK, Jan, ed. (1973): *Errata: Papers in Error Analysis*. Lund: Gleerup/Liber.
- TERRÁDEZ GURREA, Marcial (2001): *Frecuencias léxicas del español coloquial: análisis cuantitativo y cualitativo*. Valencia: Universidad de Valencia.
- THAGG FISHER, Ulla (1985): *The Sweet Sound of Concord: A Study of Swedish learners' Concord Problems in English*. Lund: Gleerup/Liber

- THORNBURY, Scott (2002): *How to teach vocabulary*. Harlow: Longman.
- TOMASELLO, Michael (2003): *Constructing a Language: a Usage-Based Theory of Language Acquisition*. Cambridge MA: Harvard University Press.
- TONO, Yukio (2003): "Learner corpora: Design, development and applications". En *Proceedings of the 2003 Corpus Linguistics Conference*, editado por D. Archer et al., 800-809. UCREL, Lancaster University: UCREL Technical Paper number 16.
- TRÉVILLE, Marie-Claude (2000): *Vocabulaire et apprentissage d'une langue seconde. Recherches et théories*. Québec: Les éditions Logiques.
- TRIBBLE, Christopher (2011): *Revisiting apprentice texts*. En *A Taste for Corpora*. In honour of Sylvianne Granger, editado por F. Meunier et al., 85-108. Amsterdam: Benjamins.
- TRUJILLO, Ramón (1976): *Elementos de semántica lingüística*. Madrid: Cátedra.
- VÁZQUEZ, Graciela (1991): *Análisis de errores y aprendizaje de español / lengua extranjera*. Frankfurt am Main: Peter Lang.
- VÁZQUEZ, Graciela (2008): "¿Qué español enseñar? Preguntas frecuentes". Teleconferencia pronunciada en las III Jornadas de Español como Lengua Extranjera, I Congreso Internacional de Enseñanza e Investigación en ELSE, en la Universidad Nacional de Córdoba, Argentina [en línea]: <http://www.lenguas.unc.edu.ar/elsecongreso/teleconferencia_vazquez.pdf> [consulta: 24-06-2013].
- WANNER, Leo et al. (2013a): "Writing assistants and automatic lexical error correction: word combinatorics". *Proceedings of eLex 2013*: 472-487.
- WANNER, Leo et al. (2013b): "Annotation of Collocations in a Learner Corpus for Building a Learning Environment". En *Twenty Years of Learner Corpus Research: Looking back, Moving ahead*, editado por S. Granger et al., Lovaina: Presses universitaires de Louvain.
- WIDDOWSON, Henry (1978): *Teaching Language as communication*. Oxford. Oxford University Press.
- WIDDOWSON, Henry (1998): "Skills, abilities and contexts of reality". *Annual Review of Applied Linguistics* 18: 323-333.
- WILKINS, David A. (1972): *Linguistics and Language Teaching*. London: Edward Arnold.
- WILKINS, David A. (1976): *Notional Syllabuses*. Oxford: Oxford University Press.
- WILLIS, Dave (1990): *The Lexical Syllabus: A New Approach to Language Teaching*. London: Collins ELT.

WRAY, Alison (2002): *Formulaic Language and the Lexicon*. Cambridge: Cambridge University Press.

ANEXO

dar	dariais	dalo
daban	darían	dala
dában	darian	dale
daría	di	dalo
daria	dí	dalas
darías	diste	dales
darías	díste	dáselo
daríamos	dio	dáselos
daríamos	dió	dásela
dado	dimos	dáselas
dados	dímos	daselo
dada	dísteis	daselos
dadas	distéis	dasela
dando	dísteis	daselas
doy	dieron	dárselo
das	dé	dárselos
da	des	dársela
dá	demos	dárselas
damos	démos	darselo
dámos	déis	darsela
daís	de	darselos
dáis	deis	darselas
daís	deís	darlo
dan	den	darlos
daré	diera	darla
dare	diéra	darlas
darás	diese	darle
daras	diése	darles
dará	dieras	dándoselo
dara	diéras	dándosela
daremos	dieses	dándoselos
darémos	diéramos	dándoselas
daréis	dieramos	dandoselo
dareis	diésemos	dandosela
darán	diesemos	dandoselos
darán	dierais	dandoselas
daba	diérais	
dába	dieséis	
dabas	diéseis	
dábas	dieran	
daba	diéran	
dábamos	diesen	
dabamos	diésen	
dabais	dad	
dábais	darse	
dabáis	dándose	
dariais	dandose	

Tesis doctoral

El verbo *dar* en el español escrito de aprendientes de L1 inglés: estudio comparativo entre hablantes no nativos y hablantes nativos basado en corpus

Anna Sánchez Rufat

DEPARTAMENTO DE FILOLOGÍA HISPÁNICA Y LINGÜÍSTICA GENERAL
2015

