

LA INVESTIGACIÓN DE CORPUS DE  
APRENDIENTES BASADA EN  
EL ANÁLISIS CONTRASTIVO DE  
LA INTERLENGUA: EL CASO DE *DAR*

ANNA RUFAT SÁNCHEZ



# ÍNDICE

	Pág.
INTRODUCCIÓN .....	XIII
1. LA INVESTIGACIÓN SOBRE LA LENGUA DEL APRENDIENTE BASADA EN CORPUS .....	1
1.1. <i>Estudios de la lengua del aprendiente anteriores al uso de corpus</i> .....	2
1.1.1. El análisis contrastivo .....	2
1.1.2. El análisis de errores .....	3
1.1.3. La ASL y el problema de los datos .....	7
1.2. <i>La investigación de corpus de aprendientes</i> .....	8
1.2.1. El desarrollo de los corpus de aprendientes .....	11
1.3. <i>Estudios sobre la lengua del aprendiente del español: propósitos y alcance de este estudio</i> .....	17
2. LA COMPETENCIA LÉXICA Y EL CONOCIMIENTO DEL VERBO .....	23
2.1. <i>Los rasgos morfológicos</i> .....	24
2.2. <i>La naturaleza semántica</i> .....	25
2.3. <i>El significado y sus relaciones paradigmáticas y sintagmáticas</i> .....	28
2.4. <i>La interfaz léxico-sintaxis</i> .....	30
2.5. <i>Los rasgos sociolingüísticos</i> .....	34
2.6. <i>Los rasgos de la competencia léxica del verbo</i> .....	36
3. LA NATURALEZA SEMÁNTICA Y LA PROYECCIÓN SINTÁCTICA DEL VERBO <i>DAR</i> .....	39
3.1. <i>Consideraciones sobre la naturaleza semántica de dar</i> .....	40
3.1.1. <i>Hacia una definición de dar que se visualice en la sintaxis</i> .....	40
3.1.1.1. <i>Los contenidos aspectuales y las idiosincrasias léxicas en el significado de dar: la trayectoria y la meta en la transferencia y en la creación.</i> ..	41
3.1.1.2. <i>La concordancia léxica y la redundancia</i> .....	45
3.1.2. <i>Descripción de los sentidos de dar</i> .....	49
3.2. <i>Usos y estructuras del verbo dar</i> .....	54
4. LOS PROCEDIMIENTOS PARA EL ANÁLISIS .....	61
4.1. <i>El aspecto de la norma</i> .....	61
4.2. <i>La obtención de los datos</i> .....	64
4.2.1. <i>El CEDEL2</i> .....	64
4.2.2. <i>Antconc</i> .....	68
4.3. <i>Procedimientos para el análisis de los datos</i> .....	72
4.3.1. <i>Cuestiones preliminares</i> .....	72

4.3.2. Estructuras consideradas en el análisis .....	73
4.3.3. Fuentes para contrastar los datos .....	74
4.3.3.1. Introspección .....	75
4.3.3.2. Otros corpus textuales .....	76
4.3.3.3. Diccionarios combinatorios.....	77
4.3.3.4. Obras de consulta generales de carácter panhispánico.....	78
4.3.4. Criterios para determinar el grado de aceptabilidad de los datos.....	78
4.3.5. Criterios para determinar la inaceptabilidad de los datos .....	83
4.3.5.1. Localización del error.....	84
4.3.5.1.1. Error relevante frente a error no relevante .....	86
4.3.5.2. Descripción del error .....	88
4.3.5.2.1. El modelo de funcionamiento del lexicón utilizado en la descripción del error: el lexicón y el sistema computacional. ....	88
4.3.5.2.2. Error léxico y error computacional .....	90
4.3.5.3. Explicación del error .....	94
5. EL USO DEL VERBO <i>DAR</i> EN EL CORPUS .....	97
5.1. <i>Relaciones del índice de frecuencia entre los dos corpus. Análisis cuantitativo y cualitativo</i> .....	98
5.1.1. Frecuencia total del lema <i>dar</i> : ¿tienden los HNN a utilizar este lema más o menos que los HN?.....	100
5.1.1.1. Análisis cuantitativo .....	100
5.1.1.2. Análisis cualitativo .....	101
5.1.2. Construcciones sintácticas del verbo y su frecuencia .....	102
5.1.2.1. Análisis cuantitativo .....	103
5.1.2.2. Análisis cualitativo .....	105
5.1.2.3. Interpretación de las relaciones de frecuencia.....	106
5.1.2.3.1. Alta frecuencia en HNN vs. alta frecuencia en HN.....	107
5.1.2.3.2. Baja frecuencia en los HNN vs. baja frecuencia en los HN.....	112
5.1.2.3.3. No frecuencia en HNN vs. baja frecuencia en HN.....	119
5.1.2.3.4. Síntesis del análisis de las relaciones de frecuencias.....	120
5.1.3. Sustantivos y clases léxicas con las que se combina <i>dar</i> .....	121
5.1.3.1. El uso de <i>dar</i> como verbo de transferencia .....	123
5.1.3.1.1. Análisis cuantitativo.....	123
5.1.3.1.2. Análisis cualitativo.....	125
5.1.3.2. El uso de <i>dar</i> como verbo de creación .....	126
5.1.3.2.1. Análisis cuantitativo.....	126
5.1.3.2.2. Análisis cualitativo.....	129
5.2. <i>Aceptabilidad e inaceptabilidad de los datos no nativos</i> .....	130
5.2.1. Frecuencia de los datos aceptable-naturales, aceptable-restringidos e inaceptables .....	131
5.2.2. Localización del error.....	132
5.2.3. Descripción y explicación del error en el núcleo verbal .....	132
5.2.4. El error y el acierto en las estructuras analizadas.....	138
5.2.5. Descripción y explicación del error en el complemento .....	140

## ÍNDICE

5.2.6. Descripción y explicación del error en el sujeto .....	143
5.2.7. El error en el sintagma.....	143
5.3. <i>Observaciones sobre el uso de dar</i> .....	145
5.3.1. Diferencias y semejanzas por estructuras.....	145
5.3.2. Diferencias y semejanzas de los usos de dar (V + N) en sus dos significados	146
5.3.3. El papel de la transferencia de la L1 .....	147
6. CONCLUSIONES .....	149
6.1. <i>Sobre la dificultad de uso de dar</i> .....	149
6.2. <i>Implicaciones del análisis y sus resultados</i> .....	151
6.2.1. Aplicaciones didácticas .....	153
6.2.1.1. Materiales didácticos .....	154
6.2.1.2. Contenidos léxicos.....	156
6.2.1.3. Procedimientos de enseñanza .....	157
6.3. <i>Los estudios de la interlengua del español en el futuro</i> .....	162
REFERENCIAS BIBLIOGRÁFICAS .....	165

## INTRODUCCIÓN

Este libro constituye un análisis del uso del verbo *dar* en la interlengua del español de un grupo representativo de aprendientes anglófonos a través de una investigación de corpus de aprendientes basada en el método del análisis contrastivo de la interlengua (Granger, 1996a, 2015). Combina, por lo tanto, perspectivas y planteamientos de la adquisición de segundas lenguas, la lingüística aplicada y la lengua española.

El propósito del libro es, por un lado, familiarizar a investigadores y profesores de español como segunda lengua con los principios fundamentales de la investigación de corpus informatizados de aprendientes a través de una metodología diseñada expresamente para el análisis de la interlengua que, al mismo tiempo, es consistente, informativa, flexible y reusable; y, por otro lado, mostrar que nuestro conocimiento acerca de cómo se adquiere el español como L2 puede verse ampliado gracias a este enfoque metodológico.

Para este análisis, se ha seleccionado un elemento léxico porque son todavía muy pocos los estudios de interlengua del español que tienen como objeto de estudio este componente, pese a la revalorización que ha experimentado en los últimos cuarenta años en las teorías lingüísticas, psicolingüísticas y pedagógicas. Dentro del léxico, el verbo es un área muy importante de la estructura de cualquier lengua, ya que contiene concreciones argumentales que provocan que el significado de la oración se asimile al del verbo que la proyecta. Esto lo convierte en un área fundamental de la estructura de cualquier lengua, susceptible de generar errores (Housen, 2002: 78).

En concreto, el verbo *dar* es uno de los más representativos del grupo de verbos frecuentes del español y, al mismo tiempo, forma parte del grupo de palabras “pequeñas” (del inglés *smallwords*), que parecen desempeñar un papel central en el éxito comunicativo (Sinclair, 1991; Hasselgren, 2002), lo que lo sitúa como un buen punto de partida para evaluar la lengua de este grupo de hablantes. Parece evidente que el hecho de que un elemento sea frecuente en la lengua conlleva que el aprendiente esté muy expuesto a él y, por lo tanto, resulte más fácil de recordar (Nation, 2001; Nagini, 2008; Schmitt, 2010; Schmitt y Redwood, 2011). No obstante, las características atribuidas a los verbos frecuentes (Viberg, 1996), así como los datos observados en el aula, sumados a los debates actuales sobre la ASL, conducen a pensar que este verbo es un elemento difícil de manejar y que así lo perciben los profesores y los alumnos.

Algunas de las hipótesis de partida que explicarían esta dificultad sentida son las siguientes. En primer lugar, este verbo polisémico participa en construcciones sintácticas

muy variadas; una variación que es tanto intralingüística (piénsese en *dar por*, *dar en*, *dar a*, *dar con*, *dar contra*, *dar que*, seguidos de nombres, adjetivos, infinitivos, conjunciones, entre otras muchas posibilidades) como interlingüística, ya que las estructuras difieren de una lengua a otra; esto es, no hay una correspondencia plena, en el caso que nos ocupa, entre los usos de *dar* y los de *give*. En segundo lugar, la construcción verbo + nombre (V + N), que desde una perspectiva formal es sencilla, puede resultar compleja desde una perspectiva semántica, pues el verbo coaparece con una gran variedad de nombres. En tercer lugar, si nos detenemos en esta estructura V + N, llaman nuestra atención las *construcciones con verbo de apoyo* (como *dar un salto* o *dar un paseo*) por su complejidad. Estas combinaciones son opacas, comparadas con las construcciones en las que *dar* actualiza un significado pleno —aunque figurado—, próximo a “proporcionar” o “hacer llegar” (como *dar un consejo*); las primeras se consideran más opacas porque es difícil rastrear en ellas el vínculo que hay entre el significado del verbo de apoyo —próximo a “hacer” o “ejecutar”— y el significado básico o físico a partir del cual se proyectan todos los figurados (el dar, de *dar un lápiz*).

Estas tres hipótesis acerca de la dificultad del aprendizaje de este verbo están planteadas en torno a la interfaz léxico-sintaxis, que es uno de los centros de interés reciente en la ASL por considerarse causa potencial de desviaciones en el desarrollo de la interlengua. Así pues, partir de unas hipótesis que establecen un vínculo con los debates actuales sobre la ASL y sus implicaciones en el desarrollo de la lengua de los aprendientes constituye un buen motor para esta investigación que arranca de las siguientes preguntas cuyas respuestas nos han permitido contrastar las hipótesis anteriores:

1. ¿Cuáles son las semejanzas y diferencias más destacadas entre el español de los hablantes no nativos y el español de los hablantes nativos con respecto al uso del verbo *dar*? En relación con ello:

1.1. ¿Tienden los no nativos a utilizar este lema más o menos que los nativos?

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 nativos y los no nativos el verbo *dar*?

2. ¿Es un verbo seguro para el no nativo o, por el contrario, es una fuente de problemas fruto de las diferencias interlingüísticas y complejidades intralingüísticas?

3. ¿Qué papel desempeña la L1 en el uso de este verbo? ¿Y qué presencia tiene el factor intralingüístico en la adquisición de este verbo?

En definitiva, como se desprende de las cuestiones anteriores, en este análisis se tienen en cuenta las cuestiones ortográficas, morfológicas, gramaticales, léxicas (asociaciones y colocaciones) y sociolingüísticas del verbo; y en concreto, se presta especial atención a la naturaleza semántica de *dar* y a sus proyecciones sintácticas.

La necesidad de este estudio se ve respaldada por la escasez de trabajos que se centran en este aspecto de la interlengua del español, además de por las discrepancias entre los investigadores sobre verbos frecuentes de la interlengua del inglés: unos hablan de sobreutilización y otros de infrautilización del verbo; unos de verbo seguro y otros de foco de errores; unos subrayan que la influencia de la L1 no es muy importante y, en cambio, otros afirman que el léxico es el nivel más susceptible de ser afectado por la transferencia. Las

conclusiones, además, proporcionarán nuevos caminos para la enseñanza y el aprendizaje del léxico del español basados en las investigaciones de corpus a través de las implicaciones pedagógicas del análisis de interlengua propuesto.

El libro se organiza en cuatro partes principales para cubrir cada uno de los objetivos planteados anteriormente. En la primera sección (capítulo 1) se lleva a cabo una revisión de los estudios sobre la lengua del aprendiente de español que se centra en los corpus de aprendientes informatizados; se revisan los corpus más destacados y los estudios de corpus de español L2 sobre la competencia léxica, y se subrayan los aspectos metodológicos más característicos y las limitaciones de las investigaciones. La segunda sección (capítulos 2 y 3) está dedicada al conocimiento del verbo, en general, y al examen del verbo *dar*, en particular, a través de su naturaleza semántica y su proyección sintáctica. Se propone una descripción de los diferentes significados que actualiza el verbo y de las diferentes estructuras sintácticas en las que participa; por último, con la intención de desarrollar herramientas eficaces para la categorización de los datos analizados del corpus y para facilitar su explicación, se recoge una propuesta formalizada de los aspectos que integran la competencia léxica de este verbo. En la tercera sección (capítulo 4) se describe la metodología de la investigación; tras referir las herramientas empleadas y describir en profundidad los dos corpus comparados y analizados en este trabajo (un corpus nativo y un corpus no nativo de L1 inglés, procedentes del CEDEL2), se presentan los procedimientos seguidos en el análisis de frecuencias y en el de errores; para ello se plantea la cuestión de la selección de la norma en los estudios de interlengua y se 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). La última sección (capítulo 5), en fin, está dedicada al análisis de la producción de *dar* en el corpus nativo y en el no nativo, que se comparan desde un enfoque cualitativo y cuantitativo. Se atiende, asimismo, al papel de la transferencia de la lengua materna de los aprendientes. En la conclusión (capítulo 6) se discuten los retos y las oportunidades de la futura investigación de corpus de aprendientes de español, así como de la práctica pedagógica basada en los resultados de los análisis de corpus.