

La evaluación de la competencia colocacional en aprendices de español como lengua extranjera

Autora: Ana Orol González

Tesis doctoral UDC / 2024

Directora: Margarita Alonso Ramos

Tutora: María Dolores Sánchez Palomino

Programa de doctorado en Estudios Lingüísticos



UNIVERSIDADE DA CORUÑA

AGRADECIMIENTOS

Allá por 2012 inicié este largo camino en el que no he estado sola. En primer lugar, quiero mostrar mi agradecimiento a la directora de esta tesis, Margarita Alonso Ramos, por su esfuerzo, exigencia y, sobre todo, por mantenerse a la espera durante tantos años. En segundo lugar, me gustaría agradecer a los centros de enseñanza que me han permitido administrar el TCE (Centro de Linguas da Universidade da Coruña, da Universidade de Santiago de Compostela, al Centro de Linguas Modernas, la Universidad de Nápoles y, muy especialmente, a la Escuela Oficial de Idiomas de A Coruña). No obstante, el más cálido de los agradecimientos es para los aprendices de ELE que dedicaron su tiempo, su conocimiento y competencia a realizar un test de colocaciones. Esta investigación no hubiera sido posible sin su colaboración.

Quisiera hacer extensibles mis agradecimientos a aquellos que en otra época también formaron parte de este proyecto: al profesorado y personal de la Facultad de Filoloxía, a los miembros del Grupo LYS y, sobre todo, a Marcos y Estela por acompañarme, ayudarme y enseñarme. También a la Universidade da Coruña por apoyar esta investigación con la concesión del Contrato predoctoral. Mis agradecimientos quedarían injustamente incompletos sin mencionar al profesor James Milton, que me acogió en la universidad de Swansea durante aquella estancia de investigación y me permitió usar X-LEX en este proyecto.

El desarrollo de esta tesis es también deudor de las personas que forman parte de mi círculo cercano. Antes de seguir mostrando mi agradecimiento, me gustaría dedicar esto a dos personas:

Para ti, que me enseñaste que ser coherente es obligatorio, que tener carácter es necesario y que la bondad es imprescindible. Gracias, sé que ver esto te hubiera hecho mucha ilusión. No es una novela, pero he escrito mucho. Para ti, que me enseñaste lo que es la resistencia, la generosidad y la sonrisa (pese a que no siempre se consigue). Gracias por crear un hogar, por compartirlo y mostrarnos que eso es indispensable. Gracias eternas a los dos.

Por último, daros las gracias a vosotros:

A ti, que lo aprendiste de ellos y me lo has enseñado a mí (aunque no lo haya aprendido todo), por confiar y no dejar de creer. A ti, que respetaste mis silencios, que aguardaste (con lo impaciente que tú eres) sin esperar mi respuesta. A ti, que me ayudaste a tomar grandes decisiones, aunque quizás no lo sepas, y que me has hecho reír tantas y tantas veces. Y a ti, por estar siempre ahí para mí con una sonrisa.

A todos los que ya no preguntabais... y lo agradecía. A ti, que me escuchaste, me entendiste (o eso me parecía) y me animaste a no dejarlo. A todos los que me ayudasteis en innumerables ocasiones con la estadística, los idiomas u otras cuestiones académicas, administrativas o emocionales. Y a todos aquellos y aquellas que en algún momento me acompañaron a lo largo de este dilatado encargo que muchos años atrás me hice.

Burela, 8 de mayo de 2024

RESUMEN

La búsqueda de un método adecuado para medir la competencia colocacional de aprendices de español como lengua extranjera es el principal propósito de esta investigación. Esto supuso la elaboración de un test de diagnóstico y de competencia, el Test de colocaciones del español (TCE), desde la creación a la administración. Para llevar a cabo esta tarea, propusimos la definición operativa de colocación y la descripción de un modelo de conocimiento de colocaciones, así como la selección de una muestra de colocaciones *testables*, que implicó la creación de una lista de colocaciones niveladas y permitió proponer algunos criterios para su nivelación, así como el diseño del test. El TCE consta de una batería de test que tratan de mostrar una perspectiva panorámica del conocimiento que los aprendices de español como lengua extranjera tienen de las colocaciones del español. La batería está compuesta por cuatro modalidades que miden la producción y la comprensión desde el reconocimiento y la recuperación. Los resultados muestran que el aprendizaje de colocaciones no siempre es un reto para los aprendices de ELE, puesto que solo en el caso de la producción desde la recuperación les supone un desafío, a diferencia de lo que muestran los resultados relacionados con la comprensión, que son positivos.

RESUMO

A busca dun método adecuado para medir a competencia colocacional de aprendices de español como lingua estranxeira é o principal propósito desta investigación. Isto supuxo a elaboración dun test de diagnóstico e de competencia, o Test de colocacións do español (TCE), dende a creación ata a administración. Para levar a cabo esta tarefa, propuxemos a definición operativa de colocación e a descrición dun modelo de coñecemento das colocacións, así como a selección dunha mostra de colocacións testables, o que implicou a creación dunha lista de colocacións niveladas e permitiu propoñer algúns criterios para a súa nivelación, así como o deseño do test. O TCE consta dunha batería de test que trata de mostrar unha perspectiva panorámica do coñecemento que os aprendices do español como lingua estranxeira teñen das colocacións do español. A batería está composta por catro modalidades que miden a produción e a comprensión dende o recoñecemento e a recuperación. Os resultados mostran que a aprendizaxe de colocacións non sempre é un reto para os aprendices de ELE, posto que unicamente no caso da produción dende o recoñecemento supón un desafío, a diferenza do que mostran os resultados relacionados coa comprensión, que son positivos.

ABSTRACT

The search for an adequate method to measure the collocational competence of learners of Spanish as a foreign language is the main purpose of this research. This involved the development of a diagnostic and competency test, the Spanish Collocation Test (TCE), from creation to administration. To carry out this task, we proposed the operational definition of collocation and the description of a collocation knowledge model, as well as the selection of a sample of testable collocations, which entailed the creation of a leveled collocation list and allowed us to propose some criteria for their leveling, as well as the design of the test itself. The TCE consists of a battery of tests that seek to provide a panoramic perspective of the knowledge that learners of Spanish as a foreign language have of Spanish collocations. The battery is composed of four modalities that measure production and comprehension from recognition and retrieval. The results show that learning collocations is not always a challenge for learners of Spanish as a foreign language, since only in the case of production from retrieval does it pose a challenge for them, unlike what the results related to comprehension show, which are positive.

ÍNDICE DE TABLAS

Tabla 1. Recopilación de ejemplos de los cuatro formatos de CATSS	17
Tabla 2. Enfoque para definir el conocimiento reconocido y recuperado (Schmitt & Schmitt, 2020, p. 38)	18
Tabla 3. Variantes en la organización de las dimensiones de la competencia léxica.....	24
Tabla 4. Comparación entre factores que influyen en la carga de aprendizaje	39
Tabla 5. Comparación entre varios modelos que explican la dificultad de la palabra	42
Tabla 6. Diferente terminología para los componentes de la colocación en el enfoque fraseológico	46
Tabla 7. Componente colocacional del conocimiento de las palabras. Extraído de Nation (2001, p. 27)....	56
Tabla 8. Componentes que se necesitan dominar para conocer una unidad multipalabra (Nation & Webb, 2011)	58
Tabla 9. Comparación entre el conocimiento componencial de la palabra (Nation, 2013) y de las unidades multipalabra (Nation & Webb, 2011)	60
Tabla 10. Recopilación de listas de colocaciones frecuentes.....	64
Tabla 11. Ejemplo de escala del MCER (Consejo de Europa, 2002, p. 109)	69
Tabla 12. Componentes del test de diagnóstico. Adaptación (Jang, 2017).....	79
Tabla 13. Comparación terminológica entre el ABA y el AUA.	81
Tabla 14. Comparación entre inferencias en el Enfoque basado en argumentos y el Argumento de uso de la evaluación	83
Tabla 15. Comparación entre diferentes propuestas de los principios de la evaluación	85
Tabla 16. Usos del MCER.....	92
Tabla 17. Muestra de ítems en CATSS, versión bilingüe (Laufer & Goldstein, 2004).....	115
Tabla 18. Muestra de ítems de CATSS, versión monolingüe (Aviad-Levitzky et al., 2019)	116
Tabla 19. Clasificación de los test de la primera generación según sus características	154
Tabla 20. Clasificación de los test de la primera generación según sus características	156
Tabla 21. Conocimiento de la base.....	166
Tabla 22. Conocimiento del colocativo.....	170
Tabla 23. Conocimiento (propio) de la colocación	172
Tabla 24. Estándar del conocimiento colocacional en su grado óptimo	173
Tabla 25. Lista de aspectos del conocimiento de la colocación mínimos para comprender una colocación	176
Tabla 26. Aspectos imprescindibles del conocimiento de la colocación para su producción	176
Tabla 27. Ciclo del TCE.....	178
Tabla 28. Estructura del TCE	183
Tabla 29. Ejemplo de ítem en todas las modalidades en TCE	191
Tabla 31. Guía del TCE	192
Tabla 34: Rúbrica básica para graduar el conocimiento colocacional	193

Tabla 35. Escala del MCER.....	197
Tabla 37. Correspondencia entre la escala numérica a porcentual del TCE	199
Tabla 38. Primera escala del conocimiento colocacional.....	201
Tabla 39. Segunda escala del conocimiento colocacional.....	201
Tabla 40. Correspondencia entre la gradación del conocimiento colocacional y el TCE	202
Tabla 41. Orden y duración en la administración del TCE de la Modalidad A.....	203
Tabla 42. Relación entre el nivel y las bandas de frecuencia (Meara & Milton, 2003).....	212
Tabla 43 Esquema de colocaciones de un nivel para el TCEp	217
Tabla 44. Estadística descriptiva de la frecuencia de las combinaciones	219
Tabla 45. Esquema de colocaciones de un nivel para el TCE	219
Tabla 46. Comparación de las tres muestras en las tres versiones del TCE	221
Tabla 47. Ejemplo de ítem en las cuatro modalidades del TCE.....	222
Tabla 48. Datos de participación en las tres versiones del TCE	230
Tabla 49. Número de participantes por modalidad y nivel del preTCEp	231
Tabla 50. Número de participantes por modalidad y nivel del TCEp.....	232
Tabla 51. Número de participantes por modalidad y nivel del TCE	232
Tabla 52. Estadística descriptiva de preTCEp.....	234
Tabla 53. Resultados de la Modalidad A en el preTCEp	236
Tabla 54. Resultados de la Modalidad B en el preTCEp.....	236
Tabla 55. Índice de facilidad de un ítem catalogado como de baja dificultad	237
Tabla 56. Índice de facilidad de un ítem catalogado como de alta dificultad	237
Tabla 57. Media de resultados de X-LEX	239
Tabla 58. Media de resultados de X-LEX según el nivel de competencia lingüística	240
Tabla 59. Resultados de X-LEX en TCEp.....	240
Tabla 60. Porcentaje de participantes asignado a un nivel según los resultados de X-LEX.....	240
Tabla 61. Coeficiente Pearson entre el valor de la CL y el valor asignado de la Competencia lingüística ..	241
Tabla 62. Estadística descriptiva TCEp.....	241
Tabla 63. Índice de facilidad del ítem del TCEp.....	242
Tabla 64. Asignación del nivel de la CCol según los criterios de clasificación preestablecidos	243
Tabla 65. Estadística descriptiva en la Modalidad A del TCEp.....	243
Tabla 66. Estadística descriptiva en la Modalidad B del TCEp.....	243
Tabla 67. Matriz de correlación de la Modalidad A (CC-A), PRECU, CRECO, la Competencia léxica (CL) y la competencia lingüística (CLing)	245
Tabla 68. Matriz de correlación de la Modalidad B (CC-B), PRECU, CRECO, la Competencia léxica (CL) y la competencia lingüística (CLing)	245
Tabla 69. Análisis descriptivo de la influencia de la categoría gramatical de las colocaciones en las distintas modalidades.....	247
Tabla 70. Estadística descriptiva de los resultados de X-LEX en la Modalidad A	249

Tabla 71. Estadística descriptiva de los resultados de X-LEX en la Modalidad B.....	249
Tabla 72. Relación entre los niveles del MCER y los resultados de XLEX (Milton & Alexiou, 2010)	250
Tabla 73. Clasificación en niveles del MCER según los resultados de X-LEX	251
Tabla 74. Resultados del TCE en la Modalidad A	252
Tabla 75: Resultados del TCE en la Modalidad B	252
Tabla 76. Análisis de la normalidad de las cuatro modalidades del TCE	253
Tabla 77: Estadística descriptiva de la Modalidad A del TCE.....	254
Tabla 78. Estadística descriptiva de la Modalidad B del TCE	254
Tabla 79. Índice de facilidad (IF) en cada una de las modalidades del TCE	256
Tabla 80: IF por niveles de cada una de las modalidades del TCE	256
Tabla 81. Fiabilidad de cada una de las opciones	258
Tabla 82. Porcentaje de aprendices clasificados en un nivel del MCER según su CCOL.....	258
Tabla 83. Porcentaje de aprendices clasificados en un nivel del MCER según cada una de las modalidades del TCE.....	259
Tabla 84. Tabla de contingencia de las dos modalidades de la Modalidad A	260
Tabla 85. Tabla de contingencia de las dos modalidades de la Modalidad B.....	261
Tabla 86. Tabla de contingencia	261
Tabla 87. Matriz de correlaciones en la Modalidad A del TCE	263
Tabla 88. Matriz de correlaciones en la Modalidad B del TCE	264
Tabla 89. Medidas de ajuste del Modelo de regresión logística binomial	265
Tabla 90. Comparaciones Entre Parejas (Durbin-Conover)	269
Tabla 91. Modelo de gradación del conocimiento colocacional.....	270
Tabla 92. Porcentaje de grado de conocimiento general en la Modalidad A por niveles del MCER	271
Tabla 93. Porcentaje de grado de conocimiento general en la Modalidad B por niveles del MCER	271
Tabla 94. Porcentaje de grado de conocimiento concreto en la Opción A por niveles del MCER	272
Tabla 95. Porcentaje de grado de conocimiento general en la Opción B por niveles del MCER	272
Tabla 96. Comparación de los grados del conocimiento colocacional en las dos Modalidades del TCE	273
Tabla 97. Influencia de la categoría gramatical en la competencia y conocimiento colocacional.....	274
Tabla 98. Prueba T para Muestras Independientes.....	275
Tabla 99. Orden jerárquico de las cuatro modalidades del TCE a lo largo de todas sus versiones.....	276
Tabla 100. Fases del MACE en relación con el grado y las dimensiones del conocimiento colocacional...281	

ÍNDICE DE GRÁFICAS

Gráfica 1. Porcentaje de participantes según su L1 en el TCE	231
Gráfica 2. Histogramas y gráficos de densidad de los resultados de X-LEX.....	249
Gráfica 3. Gráficos de dispersión de las cuatro modalidades: PRECU, CRECO, PRECO y CRECU	253
Gráfica 4. Comparación de los niveles del MCER en las diferentes modalidades.....	260
Gráfica 5. de dispersión de la CL con la CC-A y con cada una de las modalidades (PRECU y CRECO).....	264
Gráfica 6. dispersión de la CL con la CC-B y con cada una de las modalidades (PRECO y CRECU).....	265
Gráfica 7 descriptiva de las dimensiones de la competencia colocacional del TCE.....	269
Gráfica 8. Proporción del grado de conocimiento en cada Opción del TCE	270
Gráfica 9. Proporción del grado de conocimiento en cada Modalidad del TCE: CN, conocimiento nulo; CP, conocimiento parcial; CT, conocimiento total:	271
Gráfica 10. de barras que muestra el número de aciertos del esquema gramatical de la colocación en cada modalidad del TCE	274

ÍNDICE DE FIGURAS

Figura 1. Conceptualización de la profundidad de Doczi & Cormos (2016)	15
Figura 2. Grados de solidez del conocimiento (Laufer & Goldstein, 2004)	17
Figura 3. Enfoque componencial del conocimiento léxico (Nation, 2020).....	22
Figura 4. Perfil del vocabulario de un aprendiz típico basado en frecuencias (Meara, 1992, p. 4)	28
Figura 5. Representación de bandas de frecuencias (Schmitt, Schmitt, et al., 2020)	32
Figura 6. Modelo de dificultad (Dunn, 2013)	41
Figura 7. Representación del grado de transparencia y frecuencia de la colocación (Martinez & Murphy, 2011)	51
Figura 8. Clasificación de los frasemas o unidades fraeológicas (Mel'čuk & Mel'čuk, 2012)	54
Figura 9. Modelo longitudinal de los aspectos del conocimiento de la palabra (Dóczi & Kormos, 2016)	57
Figura 10. Tipos de listas de combinaciones frecuentes según su finalidad	64
Figura 11. Clasificación de los test atendiendo a diferentes criterios	75
Figura 12. Diferencia entre test referidos a la norma (TRN) y test referidos al criterio (TRC) (Barrios & Cosculluela, 2013).....	77
Figura 13. Argumento de uso de la evaluación (Schmidgall, 2017)	82
Figura 14. Modelo efectivo y útil (Green, 2014) Figura 15. Modelo de buenas prácticas (ALTE, 2020)..	86
Figura 16. Adaptación (Green & Fulcher, 2021) del ciclo propuesto por Fulcher (2010).....	89
Figura 17. Escala tipo de tres niveles.....	92
Figura 18. Ejemplo de ítem de VLT (Schmitt et al., 2001, p. 58)	105
Figura 19. Ejemplo de ítem del PVL (Laufer y Nation, 1999).....	106
Figura 20. Ejemplo de ítem de VocabLab (Peters et al., 2019)	107
Figura 21. Ejemplo de ítem de VKS (Wesche & Paribakht, 1996)	110
Figura 22 Ejemplo de ítem de SPVLT	119
Figura 23. Ejemplo de ítem de PVL	119
Figura 24. Ejemplo de ítem de COLLEX 5 (Gyllstad, 2009)	126
Figura 25. Ejemplo de ítem de COLLMATCH 3 (Gyllstad, 2009).....	126
Figura 26. Muestra de ítem (Lee Revier, 2009, p. 129)	128
Figura 27. Ejemplo de ítem de LexCombi (Barfield, 2009)	131
Figura 28. Ejemplo de ítem del primer formato de ADELEX (Moreno-Jaén, 2007b, p. 301).....	133
Figura 29. Ejemplo de ítem del segundo formato de ADELEX (Moreno-Jaén, 2007, p. 302)	134
Figura 30. Ejemplo de ítem del tercer formato de ADELEX (Moreno-Jaén, 2007b, p. 303)	134
Figura 31. Ejemplo de ítem del cuarto formato de ADELEX (Moreno-Jaén, 2007, p. 303)	134
Figura 32. Ejemplo de formato de LexCombi 2 (Brown, 2018, p. 201).....	137
Figura 33. Ejemplo de ítem de producción (Nizonkiza, 2012).....	139
Figura 34. Ejemplo de ítem (González Fernández & Schmitt, 2015)	141
Figura 35. Ejemplo de ítem (Nguyen & Webb, 2017)	143

Figura 36. Ejemplo de ítem del CPCT académico (Frankenberg-García, 2018).....	144
Figura 37. Ejemplo de ítem del CPCT no académico (Sonbul et al., 2022, p. 12).....	146
Figura 38. Ejemplo de ítem de recuperación (González-Fernández & Schmitt, 2020).....	147
Figura 39. Ejemplo de ítem de reconocimiento (González-Fernández & Schmitt, 2020).	147
Figura 40. Ejemplo de ítem de reconocimiento (Nontasee & Sukying, 2021)	150
Figura 41. Ejemplo de ítem de recuperación (Nontasee & Sukying, 2021).....	150
Figura 42. . Opción C del test que mide las colocaciones.....	151
Figura 43. Ejemplo de ítem de test productivo de colocaciones en francés (Forsberg Lundell et al., 2018)	152
Figura 44. Componentes básicos del modelo de conocimiento de la colocación	165
Figura 45. Representación de los grados del conocimiento componencial de una palabra (Barclay & Schmitt, 2019)	200
Figura 46. Evolución del test de colocaciones del español	206
Figura 47. Organización de las Nociones del PCIC (Instituto Cervantes, 2006).....	218
Figura 48. Relación entre niveles del MCER y resultados de XLex (Milton & Treffers-Daller, 2013)	224
Figura 49. Asignación niveles del MCER y los resultados de X-LEX en inglés, francés y griego (Milton & Alexiou, 2010)	225
Figura 50. Jerarquía del conocimiento colocacional en preTCEp	235
Figura 51. Jerarquía del conocimiento colocacional en el TCEp	246
Figura 52. Imagen global y escalar de la competencia colocacional.....	268
Figura 53: PreTCEp Figura 54: TCEp Figura 55: TCE	276

LISTA DE ABREVIATURAS

CCOL: Competencia colocacional

CLéx: Competencia léxica

CLing: Competencia léxica

CN: Conocimiento nulo

CP: Conocimiento parcial

CRECO: Comprensión desde el reconocimiento

CRECU: Comprensión desde la recuperación

CT: Conocimiento total

DiCE: Diccionario de colocaciones del español

ELE: Español como lengua extranjera

MACE: Modelo de aprendizaje de colocaciones del español

MCER: Marco Común Europeo de Referencia

PCIC: Plan Curricular del Instituto Cervantes

PRECO: Producción desde el reconocimiento

PRECU: Producción desde la producción

TAI: Test adaptativo informatizado

TCE: Test de colocaciones del español

TRC: Test referido al criterio

TRN: Test referido a la norma

ÍNDICE

1.	INTRODUCCIÓN	1
1.1.	Ámbito de estudio y justificación	1
1.2.	Objetivos y preguntas de investigación	4
1.3.	Planificación y desarrollo	6
1.4.	Estructura	7
2.	EL APRENDIZAJE DE VOCABULARIO	9
2.1.	Hacer operativa la definición del conocimiento léxico	9
2.1.1.	Destrezas y procesos cognitivos	10
2.1.2.	Extensión y profundidad	13
2.1.3.	Solidez	16
2.2.	El léxico	18
2.2.1.	El concepto de léxico	18
2.2.2.	El conocimiento léxico	20
2.2.3.	Relación entre léxico y frecuencia	26
2.2.4.	Aprendizaje del léxico	36
2.3.	Las colocaciones	42
2.3.1.	El concepto de colocación	42
2.3.2.	El conocimiento colocacional	55
2.3.3.	Relación entre colocaciones y frecuencia	61
2.3.4.	Aprendizaje de colocaciones	68
2.4.	Conclusión	71
3.	LA EVALUACIÓN DE LENGUAS: MEDIR EL CONOCIMIENTO LÉXICO Y EL CONOCIMIENTO COLOCACIONAL	73
3.1.	La evaluación de lenguas	73
3.1.1.	Clasificación de los test	74
3.1.2.	Principios de la evaluación de lenguas	79
3.1.3.	Ciclo y especificaciones	88
3.1.4.	Estandarización	91

3.1.5.	Estadística para la evaluación de lenguas	94
3.2.	La evaluación del léxico	101
3.2.1.	Tipos de test de vocabulario	102
3.2.2.	Test de vocabulario en inglés.....	104
3.2.3.	Test de vocabulario en español	117
3.3.	La evaluación de las colocaciones	119
3.3.1.	Clasificación de los test de colocaciones.....	121
3.3.2.	Descripción de una selección de test de colocaciones	122
3.3.3.	Recapitulación y comparación	153
4.	DISEÑO Y CONSTRUCCIÓN DEL TCE.....	157
4.1.	Planificación del TCE	158
4.1.1.	Condiciones: test de diagnóstico y argumento de validez	158
4.1.2.	Propuesta de un modelo de conocimiento colocacional	160
4.2.	Ciclo del TCE	177
4.3.	Diseño del TCE.....	180
4.3.1.	Adaptación de CATSS	181
4.3.2.	Estructura y forma del TCE	182
4.3.3.	Especificaciones.....	191
4.4.	Evolución del TCE	205
4.4.1.	Del pre-TCEp al TCEp.....	207
4.4.2.	Del TCEp al TCE: reconstrucción	208
4.5.	Selección de la muestra del TCE.....	210
4.5.1.	Muestra para el <i>preTCEp</i>	211
4.5.2.	Muestra para el TCEp	215
4.5.3.	Muestra para el TCE	219
4.6.	Construcción del TCE	221
4.7.	X-LEX en español	222
4.8.	Conclusión	226

5.	ADMINISTRACIÓN, ANÁLISIS E INTERPRETACIÓN ... DEL TCE.....	229
5.1.	Población y administración.....	229
5.2.	Análisis de los resultados del TCE	232
5.2.1.	Análisis del <i>preTCEp</i>	233
5.2.2.	Análisis del TCEp	238
5.2.3.	Análisis del TCE	247
5.2.4.	Consideraciones finales.....	275
5.3.	Aplicación de los resultados del TCE a la didáctica	278
5.3.1.	Problemas y necesidades	278
5.3.2.	Propuestas y soluciones	280
6.	CONCLUSIÓN.....	283
6.1.	Aportaciones e implicaciones.....	283
6.2.	Limitaciones y líneas futuras	288
	REFERENCIAS BIBLIOGRÁFICAS.....	293
	ANEXOS.....	327

1. INTRODUCCIÓN

There is [also] a need for diagnostic test of strategies for dealing with aspects of word knowledge that are relevant for foreign language learners (Schmitt et al., 2020).

1.1. Ámbito de estudio y justificación

Las colocaciones son el objeto de estudio de este trabajo, uno de los componentes clave de la competencia léxica y esencial en la construcción de la competencia lingüística. Tradicionalmente, se ha identificado erróneamente el léxico solo con unidades léxicas monoverbales. La confusión radica en la creencia de que el léxico de una lengua está constituido únicamente por palabras, pero también es necesario incluir todo tipo de *entidades léxicas* (Polguère, 2010); es decir, cualquier entidad que pertenece al léxico de la lengua, tiene propiedades individuales y que, por lo tanto, tiene que ser aprendida y descrita como tal. Bajo ese término, se incluyen también las expresiones fraseológicas (Read, 2011), entre las que se encuentran las colocaciones. Entendemos por *colocaciones* las combinaciones de palabras como *leche entera* o *café cortado*, en las que uno de sus componentes selecciona al otro para expresar un sentido dado (Alonso Ramos, 2002). En estos ejemplos, el sustantivo *leche* selecciona el adjetivo *entera* para expresar el sentido ‘sin eliminar la grasa’ y el sustantivo *café*, el adjetivo *cortado* para expresar el sentido ‘con un poco de leche’. Así, entendemos las colocaciones como la combinación de dos elementos en los que uno de ellos, la *base*, elige léxicamente al otro, el *colocativo*, para expresar un sentido dado (Hausmann, 1998). Seguimos para esta definición la *Lexicología Explicativa Combinatoria* (a partir de ahora LEC), rama de la *Teoría Sentido Texto*, TST (Mel’čuk, 2012; Mel’čuk et al., 1995) .

En los últimos años ha habido una gran proliferación de estudios sobre el aprendizaje de colocaciones, especialmente para inglés como L2 (Gyllstad & Wolter, 2016; Peters, 2016a; Pulido & Dussias, 2020; Sonbul & Schmitt, 2013; Webb et al., 2013). Una de las claves para el auge de estudios sobre el aprendizaje de colocaciones es que se consideran signo de la fluidez necesaria para hablar como un nativo (Henriksen, 2013a; Nation, 2013b); esto es, acercan al estudiante de L2 al nivel más alto de competencia (C2)

del *Marco Común Europeo de Referencia*, a partir de ahora MCER (Consejo de Europa, 2002); por lo tanto, si las colocaciones están ampliamente distribuidas en el discurso, consecuentemente, la competencia debe incluir el conocimiento y dominio de estas secuencias (Schmitt, 2004). Así, para comunicarse eficazmente en una lengua, deben adquirirse, además de palabras o unidades monoverbales, también unidades pluriverbales o expresiones fraseológicas (Gyllstad, 2007b, p. 1). Sin embargo, también se ha identificado el aprendizaje de colocaciones como uno de los retos a los que se deben enfrentar los aprendices (Schmitt, Schmitt, et al., 2020; Zaabalawi & Gould, 2017) debido a su carácter arbitrario.

Hasta ahora no se ha hecho un estudio en profundidad sobre la medición del conocimiento colocacional de los aprendices de español como L2. Dicho estudio tiene que estar enmarcado en el ámbito de la *Evaluación de lenguas* (del inglés *Language Testing*), rama de la lingüística aplicada que se ocupa de proporcionar las bases teóricas a la investigación mediante test. La mayor parte de estas pruebas desempeñan una función social, en la medida en que los resultados de algunos test implican decisiones que, en algunos casos, afectan en mayor o menor medida la vida de las personas. De ahí que a lo largo de los últimos años los estudios sobre *Evaluación de Lenguas* hayan dedicado especial atención a la búsqueda de la calidad de sus métodos y herramientas de evaluación. En concreto, la evaluación de léxico se considera parte esencial de la evaluación de lenguas (Fulcher & Davidson, 2007), ya que los test de vocabulario se utilizan como pruebas de nivel por su capacidad para predecir muchas otras habilidades lingüísticas como la competencia lectora (Hirsh & Nation, 1992).

En este trabajo uniremos el campo de la lexicología o fraseología y el de la evaluación de lenguas, valorando un aspecto en concreto del área del conocimiento lingüístico: las colocaciones. Además, nos acercaremos a estas disciplinas desde la didáctica de la enseñanza de lenguas. De modo que uno de los propósitos de este trabajo de investigación es la creación de un método de evaluación que mida la competencia colocacional de español como lengua extranjera de forma eficaz, así como la aplicación de los resultados en una propuesta metodológica. Mientras en inglés ya se ha trabajado la evaluación de las colocaciones, de algún aspecto de la competencia colocacional (Moreno-Jaén, 2007) o algún aspecto de esta competencia: la comprensión (Gyllstad,

2007b), la producción (Brown, 2018; Nizonkiza, 2012) y pese a la variedad de test existentes que miden la competencia colocacional, todavía no se considera ninguno como estandarizado¹ (Gyllstad, 2007a; Gyllstad & Schmitt, 2019) ni ninguno se adapta por completo a los objetivos de esta investigación. Por lo tanto, se hace imprescindible crear y diseñar un nuevo test para medir la competencia colocacional de los aprendices de español. Pero no solo este vacío nos lleva al diseño de un nuevo test, sino que otras razones lo hacen necesario. De forma ordenada, los motivos que lo justifican se resumen en los siguientes puntos. Primero, no hay ninguna versión en español, es decir, ningún test mide la competencia colocacional del español, lo que obliga a dedicar parte de la investigación a la validación del instrumento de medida. Segundo, la definición de colocación que sostienen los test existentes parte del enfoque basado en frecuencia y no en el enfoque fraseológico, tal y como se hace en esta investigación. Tercero, ninguno de los test existentes mide el conocimiento de una forma global, sino que se centra en algún aspecto en concreto: la comprensión, pero no la producción (Eyckmans, 2009; Gyllstad, 2007); la producción, pero no la comprensión (Lee Revier, 2009; Nizonkiza, 2012), etc. Estas razones determinan las características que definirán el *Test de colocaciones del español* (TCE): un test en español, con una concepción fraseológica de la colocación y que mida tanto la extensión como la profundidad, es decir, el conocimiento de una forma global. Para evaluar el conocimiento global de las colocaciones no será suficiente con un único formato, sino con varios que tengan en cuenta más de un aspecto del conocimiento de las colocaciones (Ishii & Schmitt, 2009).

Asimismo, es necesaria una batería de test para poder obtener una imagen panorámica de la competencia colocacional de los aprendices de español. No solo se busca saber cuántas colocaciones se conocen sino cómo se conocen para atender a las dificultades de aprendizaje buscando un diagnóstico del conocimiento y calibrando el peso de su aprendizaje. Este objetivo condiciona todo el proceso: desde la planificación hasta el análisis, pasando por el diseño y administración.

¹ Con test estandarizado nos referimos a un test aceptado por la comunidad científica al nivel de los test estandarizados de léxico como VLT (Nation, 1990).

1.2. Objetivos y preguntas de investigación

El punto de partida de esta investigación es responder, por un lado, a si la competencia colocacional es índice de la competencia léxica del español y, por extensión, de la competencia lingüística de una L2 y, por el otro, a si el aprendizaje de colocaciones supone un reto para los aprendices de español como lengua extranjera (ELE). De modo que los dos primeros objetivos de investigación se pueden formular como sigue:

- Objetivo 1: Medir la competencia colocacional de aprendices de español y compararla con su competencia léxica y su competencia lingüística.
- Objetivo 2: Identificar las dificultades y fortalezas que entraña el dominio de la competencia colocacional en ELE
- Objetivo 3: Asignar un nivel de competencia colocacional a los aprendices de ELE según los parámetros del MCER.

Elegimos el test como herramienta para cumplir con los tres objetivos anteriores. En el ámbito del inglés como lenguaje extranjera, la competencia léxica se ha medido tanto a partir de test (Laufer & Goldstein, 2004; Laufer & Nation, 1999; Read, 1998) como a partir de análisis de corpus de aprendices (Laufer & Nation, 1995). De igual forma, la competencia colocacional se ha medido a partir de test (Barfield & Gyllstad, 2009), pero sobre todo de corpus (Nesselhauf, 2005b). En el ámbito del español, la mayor parte de los análisis se han realizado a partir del análisis de corpus (Alonso-Ramos, Wanner, Vázquez-Veiga, et al., 2010; García-Salido et al., 2016), pero hasta ahora no contábamos con un test que mida la competencia colocacional de aprendices con español.

Consecuentemente, necesitamos un test de nivel y un test de diagnóstico. Entendemos por *test de nivel* aquel en el que se clasifica la competencia de los aprendices gradualmente según unos niveles de dominio (Stoeckel et al., 2021), en nuestro caso utilizaremos el MCER (Consejo de Europa, 2002) como guía. Un *test de diagnóstico* es aquel que detecta las fortalezas y debilidades en la competencia lingüística de una población (Alderson et al., 2015), en nuestro caso, la colocacional. Consecuentemente, debemos preguntar cuál es el mejor método para evaluar la competencia colocacional. Consecuentemente, debemos preguntar cuál es el mejor método para evaluar la competencia colocacional en español. Así, podemos formular otro de los objetivos de la investigación que se convierte en principal ya que es el eje central de la investigación.

- Objetivo 4: Diseñar un test válido y fiable que mida la competencia colocacional de aprendices de español como lengua extranjera.

Siguiendo las directrices actuales de la evaluación de lenguas y con el propósito de que la herramienta de evaluación diseñada cumpla con los principios de la evaluación, se ha optado por seguir la *validez basada en argumentos* (Schmitt et al., 2019). Con la necesidad de crear un nuevo test justificado y los propósitos definidos, debemos ahora delimitar los aspectos y niveles del vocabulario que van a ser medidos; es decir, hay que definir el constructo del test: aportar una definición de colocación operativa, describir los componentes del conocimiento colocacional y establecer un estándar, además de seleccionar la muestra de las colocaciones que se van a testar. Estos requisitos nos obligan a definir un modelo de conocimiento colocacional y crear una lista de colocaciones niveladas, ya que no existe en el ámbito de ELE una lista de colocaciones niveladas, aunque como explicaremos en el apartado (vid. 2.3.3) sí existen trabajos centrados en la nivelación de las colocaciones (García-Salido et al., 2019; García-Salido & Alonso-Ramos, 2018). Esto nos lleva al siguiente objetivo:

- Objetivo 5: Seleccionar una muestra de colocaciones niveladas

Para completar el proceso del test de diagnóstico, esbozamos brevemente cuál sería la forma más eficaz de aprender colocaciones. El análisis cuantitativo y, sobre todo, cualitativo de los resultados obtenidos con el test propuesto nos permitirá obtener una imagen general del conocimiento colocacional de los aprendices y compararlo con el modelo estándar, es decir, aquel que se espera de los hablantes nativos de una lengua y que se considera como grado óptimo de consecución. El análisis cuantitativo, basado en métodos estadísticos propios de la evaluación de lenguas, permite comparar la diferencia entre la competencia comprensiva y productiva de la competencia colocacional. El análisis cualitativo nos permite centrarnos en los errores cometidos, sobre todo en la parte productiva. Una visión panorámica del conocimiento colocacional de aprendices de español en diferentes niveles permite observar y analizar las debilidades y fortalezas. Esta observación, junto con otras cuestiones relevantes como una definición objetiva y la descripción clara de su procesamiento, permitirán trazar las líneas que definan un modelo eficaz de aprendizaje de colocaciones del español.

En conclusión, el eje central de esta investigación es la creación del *Test de colocaciones del español* (TCE) y su implementación. Se trata de una herramienta que nos va a permitir identificar el nivel de la competencia colocacional de los aprendices de ELE y detectar las debilidades y fortalezas de estos aprendices en relación con su competencia. De este modo, podremos dar respuesta a una de las preguntas de investigación: si la competencia colocacional es índice de la competencia lingüística, pero también si entre la competencia léxica y la colocacional existe correlación o si la colocacional es índice de la competencia léxica.

1.3. Planificación y desarrollo

La creación del TCE es el instrumento para alcanzar los propósitos clave: obtener una imagen global de la competencia colocacional de los aprendices de ELE y comprobar la relación causa-efecto entre el nivel de competencia lingüística y léxica de los aprendices en la L2. El nivel de competencia léxica se obtiene a partir de un test específico, X-LEX (Meara & Milton, 2003), que realiza cada uno de los participantes del TCE; el nivel de competencia lingüística lo obtenemos a partir de la información proporcionada por los centros que colaboran con esta investigación.

Los pasos realizados no siguen una línea cronológica, sino que muchos de los pasos se tuvieron que realizar de forma simultánea y en continua revisión, de modo que el regreso a alguno de los pasos se hizo inevitable. Por lo tanto, la planificación y el desarrollo no siguieron una línea recta sino circular, siguiendo el ciclo clásico de la Evaluación de lenguas para la creación de test. El primer paso en la elaboración del test fue la definición operativa de colocación y la descripción del modelo de conocimiento, cumpliendo así con el cuarto objetivo. Una vez definido el constructo, se pasó al diseño del test, en el que adaptamos el test de léxico CATSS (Aviad-Levitzky et al., 2019). La definición del constructo, es decir, de la competencia lingüística evaluada, es imprescindible en el desarrollo del test. Posteriormente, y tras comprobar que no contábamos con una muestra válida para nuestros propósitos y requisitos, creamos una muestra propia. Este proceso se repitió ya que el pilotaje al que se sometió el TCE nos obligó a realizar mejoras en las sucesivas versiones

El proceso de administración se realizó en tres fases, una por cada versión del TCE. El test pasó por dos triajes o pruebas piloto. La primera versión, preTCEp, fue superada por la segunda versión, TCEp, y esta dio lugar a la versión definitiva, el TCE. La población de cada una de las fases de administración se llevó a cabo gracias a la colaboración de distintas instituciones: Centro Linguas de la UDC, Centro de Linguas de la USC, la EOI de A Coruña y el Centro de Cursos Internacionales de la USC. Previamente, tras conseguir el informe favorable del Comité de Ética de la Investigación y la Docencia de la Universidade da Coruña (vid. ANEXO XXIX), se solicitó el permiso para dicha participación y cada uno de los estudiantes aceptaron su participación en la investigación. En cada una de las versiones se realizó la corrección de forma manual y un posterior análisis estadístico con dos herramientas gratuitas, JASP y JAMOVI.

1.4. Estructura

Este trabajo se estructura en seis grandes capítulos que se inicia con esta introducción y termina con la conclusión y un apartado que recoge el aparato bibliográfico utilizado para la investigación. En la parte final se recogen los anexos en los que se puede acceder al TCE en sus diferentes versiones.

El capítulo dos se dedica al aprendizaje del vocabulario, cediendo un papel preponderante al constructo. Empezamos, por lo tanto, por aclarar cómo entendemos en este trabajo algunos de los conceptos clave que nos van a permitir hacer operativo el constructo de esta investigación. Después, pasamos por definir el conocimiento léxico, la relación con la frecuencia para finalizar con cómo modelar el aprendizaje del léxico. Por último, nos centramos en las colocaciones, también, partimos de la definición de colocación, describimos el conocimiento colocacional, las relacionamos con la frecuencia y, por último, esbozamos algunos de los principios generales en el aprendizaje de las colocaciones.

El capítulo tres recoge el aparato teórico en el que se fundamenta este trabajo. Está dividido a su vez en tres grandes secciones o subapartados: uno dedicado a la evaluación de lenguas; otro, a la evaluación del léxico; y, por último, uno dedicado a la evaluación de colocaciones. En la primera de estas secciones se explican los pasos

necesarios para el diseño, creación y puesta en marcha de un test de lengua, acompañados de la explicación de los conceptos teóricos asociados imprescindibles para este fin. A continuación, nos centramos específicamente en los test que miden el léxico y describimos algunos de los más relevantes y estandarizados. Posteriormente, de forma paralela a la subsección anterior, abordamos las mismas cuestiones desde la perspectiva de la colocación y hacemos un repaso exhaustivo por los test de colocaciones existentes.

Los capítulos cuatro y cinco se dedican a la explicación de todas las fases necesarias para poner en marcha un test de lengua que dividimos en dos capítulos diferentes. En el cuarto nos centramos en las dos primeras fases: diseño y construcción, mientras que en el quinto presentamos las tres fases siguientes: la administración, el análisis y la interpretación de los resultados. En el capítulo cuarto, centrado en el TCE, no solo se abordan las dos primeras fases del ciclo, también se incluye un apartado dedicado a la planificación y diseño del test, así como a la selección de la muestra. Tanto la planificación como las condiciones de la muestra son unos de las aportaciones más relevantes de esta investigación ya que ahí trazamos las líneas del aparato teórico que sustenta el TCE. Entre otras cuestiones, presentamos un modelo de conocimiento colocacional y la adaptación de CATSS a las colocaciones. Además, proponemos un modelo de nivelación de la competencia colocacional y un modelo de gradación del conocimiento colocacional (vid. 2.2.2). En el espacio dedicado a la selección de la muestra se plasman las carencias del español como lengua extranjera ya que no se dispone, por ejemplo, de una lista de colocaciones frecuentes niveladas.

El capítulo quinto se divide también en tres secciones. En la primera, nos ocupamos de las cuestiones relacionadas con la administración y las dificultades que supuso. En el segundo, el más amplio, nos centramos en el análisis de los resultados que organizamos teniendo en cuenta la evolución del test. Así, primero analizamos las pruebas piloto, el preTCEp y el TCEp, para, por último, centrarnos en el análisis de la versión definitiva del TCE de forma más profunda. Finalmente, terminamos con unas consideraciones y acercamiento a la aplicación didáctica de los resultados, así como una serie de soluciones y propuestas a los problemas y necesidades mostrados.

Por último, la conclusión se estructura en dos bloques en el que se enumeran las aportaciones de la investigación, pero también las limitaciones que presenta.

2. EL APRENDIZAJE DE VOCABULARIO

El presente capítulo ofrece una visión general sobre el aprendizaje de vocabulario de segundas lenguas, centrándonos en el aprendizaje de las colocaciones del español como L2. Primero, haremos operativo el constructo definiendo qué entendemos por conocimiento léxico y por conocimiento colocacional, punto determinante en la construcción de un test ya que su diseño va a depender de cómo se defina el conocimiento (Laufer & Goldstein, 2004). Esto nos llevará a dejar claro qué entendemos por alguno de los conceptos clave de la evaluación de lenguas: *producción, comprensión, reconocimiento, recuperación, extensión, profundidad y solidez*. Posteriormente, compararemos diversos aspectos indispensables en relación con el léxico, por una parte, y a las colocaciones, por la otra. Estos aspectos versan entre el conocimiento, la frecuencia y sus implicaciones en el aprendizaje.

2.1. Hacer operativa la definición del conocimiento léxico

Esta sección proporciona un acercamiento a varios conceptos clave en la evaluación del léxico. Algunos de ellos siguen siendo confusos en la actualidad, por la complejidad de su naturaleza, por su proximidad a otros conceptos o por el solapamiento con otros términos. En la evaluación se denomina constructo al conocimiento evaluado por medio de un test, aunque son muchas las definiciones propuestas. Además, el concepto se ha ido desarrollando hasta llegar a la concepción actual, entendiendo el *constructo* como una interpretación significativa del comportamiento observado (Chapelle, 2012b), es decir, la teoría lingüística a partir de la que se interpretan los resultados de un test. No todo el mundo coincide en entender lo mismo por constructo léxico (Read & Chapelle, 2001). En esta investigación seguimos la siguiente definición:

The knowledge, skills, abilities, and other characteristics that are the focus of an assessment are described in terms of a “construct.” A construct is an abstract, theoretical concept, such as “knowledge” or “proficiency,” that has to be explicitly described and specified in test design (National Academies of Sciences, Engineering, 2020, p. 27).

2.1.1. Destrezas y procesos cognitivos

Antes de adentrarnos en el contraste entre *producción* y *comprensión* en relación con el léxico debemos dejar claro que esta dicotomía se mantiene en todos los niveles de lengua y que se engloba bajo la denominación *destrezas de la lengua*. Las destrezas las podemos definir como las formas en las que se activa el uso de la lengua (Martín Peris, 2008). Según el papel que desempeñan en el proceso de comunicación hablamos de *competencias productivas* o *activas* y de *competencias receptivas* o *pasivas*. Las *productivas* se relacionan con las acciones de hablar y escribir y las *receptivas*, con las de escuchar y leer. Por lo tanto, las destrezas se clasifican atendiendo a dos criterios: cómo se activa el uso de la lengua y el medio en el que se realiza. La *comprensión* se relaciona con la capacidad de entender la información que escuchamos o leemos, y la *producción* se relaciona con la idea de producir hablando o escribiendo para intercambiar mensajes con otros (Nation, 2013b). Esta definición resulta inoperativa, sobre todo en la evaluación, ya que no se tienen en cuenta, entre otras cuestiones, los procesos cognitivos que intervienen en dichas acciones. En un intento de hacerlo operativo, Read (2012) propone la distinción de *comprensión* y *uso*, pero en relación con los dos procesos cognitivos, *reconocimiento* y *recuperación*, que trataremos a continuación.

Más enfocado al vocabulario, se puede definir el *conocimiento comprensivo* como la capacidad de comprender la información recibida (*input*) al ser capaz de recibir la forma de una palabra y relacionarla con su significado o significados. Y el *conocimiento productivo* como la capacidad para recuperar la forma de la palabra adecuada del significado que queremos expresar (Laufer, 2012). O dicho de otro modo, conocimiento receptivo en cuanto que se reconoce o se recupera el significado y conocimiento productivo en cuanto que se reconoce y se recupera la forma (Nation & Hunston, 2013), como se explica a continuación de forma más detallada. Además de la dificultad que supone definir de forma operativa la distinción entre comprensión y producción, y su estrecha relación con la recuperación y el reconocimiento, hay otro punto polémico, el límite entre las nociones de comprensión y producción. Este límite, en el campo del aprendizaje del léxico, se ha venido explicando a partir de dos visiones contrapuestas: a) como dos compartimentos estancos del lexicón mental y b) como la gradación del conocimiento léxico (Laufer & Goldstein, 2004; Melka, 1997; Read, 2011). Tomaremos

esta segunda vía según la cual el conocimiento receptivo precede al comprensivo y, además, implica un menor grado de complejidad (Schmitt, 2010).

En relación con las destrezas de la lengua y el vocabulario se han aceptado una serie de generalizaciones que sientan las bases de la concepción actual. Primero, qué se necesita saber para producir una palabra y qué se necesita saber para comprenderla (Henriksen, 1999). Aunque no se ha contestado con certeza a estas preguntas, los estudios al respecto sí han dejado claro que la comprensión es más fácil que la producción y que, en el proceso de aprendizaje, la comprensión es previa a la producción. La producción evidencia solo una parte de lo que se sabe. Se dice, incluso, que lo que alguien produce es una medida pobre de su (extensión de) vocabulario (Nation & Webb, 2011, pp. 197–198).

En conclusión, estos conceptos son indispensables para la evaluación del léxico ya que nos permiten clasificar y ordenar los test de vocabulario y, por extensión, los test de colocaciones. Al diseñar los ítems de los instrumentos de evaluación es imprescindible decantarse por una u otra tendencia. En este estudio, siguiendo la propuesta de desarrollada en CATSS (Aviad-Levitzky et al., 2019; Laufer & Goldstein, 2004), tomaremos el conocimiento como una entidad gradual que maneja dos dualidades. Aquí entenderemos el reconocimiento y la producción como dos procesos cognitivos desde los que se accede a la comprensión o producción de un elemento lingüístico. La rapidez o facilidad de acceso a la información requerida vendrá determinada por los resultados obtenidos y reflejarán los resultados en el ámbito de las colocaciones. Sin embargo, es imposible plantear la evaluación de la producción y la comprensión sin incorporar los procesos cognitivos (reconocimiento y recuperación) a la descripción de los formatos y medidas.

Con la dicotomía *reconocimiento* y *recuperación* nos referimos a los diferentes procesos cognitivos a partir de los que se activa el uso de la lengua, también entendidos como diferentes funciones de la memoria a largo plazo, que están íntimamente relacionados con la conexión entre la forma y el significado. Podemos definir el *reconocimiento* como el acceso a una palabra, o combinación de palabras, a partir de la percepción de su forma. La *recuperación* se define como la reactivación de una información que recuerdas sin la ayuda de la percepción de su forma, sino que se

recuerda su forma a partir de uno de sus componentes, especialmente del sentido. Es decir, con la recuperación se rescata una información de la memoria a partir de un estímulo y con el reconocimiento se asocia un información a una forma a partir de un estímulo dado (González-Fernández & Schmitt, 2020). Al igual que se asume que la producción supone un reto mayor que la comprensión, se presupone que la recuperación también supera en dificultad al reconocimiento.

De este modo, se plantea el reconocimiento como el primer nivel del conocimiento y la recuperación como un segundo nivel que implica un mayor grado de dificultad (Read, 2012). Además, las definiciones, así como las tareas de los test, de *reconocimiento* suelen ir asociadas a la *comprensión* y las de *recuperación* a la *producción* (Read, 2012). De este modo se rechaza la idea de identificar inequívocamente *producción* con *recuperación* y *comprensión* con *reconocimiento*, ya que no hay un simple continuo que vaya desde el conocimiento mínimo comprensivo hasta la competencia productiva avanzada (Read, 2011). Sobre todo para la evaluación, debemos tener en cuenta que tanto la producción como la comprensión pueden relacionarse tanto con el proceso de reconocimiento como con el de recuperación (Nation, 2013a).

A la hora de medir el conocimiento, mientras el reconocimiento se asocia con la selección entre varias opciones, la recuperación se asocia con la idea de satisfacer o proveer una forma a partir de un sentido. Más concretamente, tareas como traducir, rellenar huecos o nombrar son las comúnmente asociadas a la recuperación en las que los participantes generan una respuesta memorística basada en una señal o pista dada. Por el contrario, la elección múltiple, la relación entre elementos o la identificación son tareas que miden el reconocimiento de modo que el aprendiz demuestra la habilidad de reconocer en algún grado el ítem presentado (Sonbul & El-Dakhs, 2020). Se mantiene, por lo tanto, en algún sentido el problema planteado por Read (2011): delimitar el umbral donde el dominio de la comprensión se transforma en dominio productivo, esto es, qué cantidad de conocimiento se requiere antes de poder usar una palabra. Detectar qué hace complejo al conocimiento de las palabras y del léxico es uno de los constantes objetivos de los investigadores. La distinción más importante en cuanto al conocimiento del léxico es relativa a la diferencia entre el reconocimiento y la recuperación más que

los componentes o dimensiones del conocimiento del vocabulario (González-Fernández & Schmitt, 2020).

2.1.2. Extensión y profundidad

Otro de los puntos clave en la configuración del conocimiento léxico es la diferencia entre el concepto de *extensión* y *profundidad*. El origen de estos dos primeros términos está en los estudios de Anderson y Freebody (1981) que contraponen el concepto de profundidad (*depth*)², al de extensión (*breadth*)³. Ambos son imprescindibles tanto en la pedagogía de segundas lenguas como en la evaluación del vocabulario. A grandes rasgos, se entiende la *extensión* como el número de palabras que conoce un hablante y la *profundidad* como el grado en el que se conoce una palabra (su ortografía, su pronunciación, sus significados, etc.). Se han conceptualizado como dimensiones del vocabulario entre otras que varían en función de la propuesta planteada (vid.2.2.2). Por ejemplo, el *espacio léxico* (Daller et al., 2007; Martínez, 2011) además de estar ocupado por la *extensión* y la *profundidad* incorpora otras dimensiones⁴ como la *fluidez* (Daller et al., 2007). Otros enfoques incluyen la *organización* (Meara, 1996) o la *solidez* (Laufer et al., 2004) como dimensiones que completan la difícil configuración del espacio léxico.

La definición de *extensión* no suele ocasionar discrepancias entre los especialistas. Aun así, ciertos aspectos son objeto de discusión. Por ejemplo, la forma en la que se mide el número de palabras que un hablante conoce. En los test de léxico que miden la extensión del vocabulario se han venido utilizando mayoritariamente dos formas: a partir de la identificación (Meara & Milton, 2003), es decir, teniendo en cuenta únicamente el reconocimiento de la forma de la palabra meta; o a partir de la relación entre la forma de una palabra y su significado (Nation, 2012). Los estudios sobre la extensión del vocabulario demuestran que es índice de la competencia de un hablante, es decir que la

² En inglés se alude a profundidad con dos términos: *depth* y *strength*. En otros trabajos se establece una diferencia conceptual entre las dos palabras (Laufer et al., 2004).

³ En inglés se refiere al concepto de extensión con dos términos: *breadth* y *size*.

⁴ Definiremos esta dimensiones al tratar la perspectiva dimensional del conocimiento léxico, a excepción de la solidez, a la que dedicamos la sección siguiente por la relevancia que tiene para el TCE.

extensión y la competencia son correlativos (Schmitt, 2014), lo que otorga un papel esencial al léxico.

Uno de los problemas a la hora de medir la extensión del vocabulario es la selección de los elementos léxicos que se van a testar (Nation, 2012) para lo que es necesario que la muestra sea representativa con lo que se debe atender a dos aspectos: el número de palabras que hay en la lengua y la forma para seleccionar un fragmento representativo para ser testado. Uno de los métodos más utilizados es la lista de palabras frecuentes, que se han convertido en una de las herramientas clave en la evaluación de lenguas. De ahí que la extensión quede indiscutiblemente unida a la idea de frecuencia, partiendo de la idea de que las palabras más frecuentes son más útiles, más necesarias y fáciles de aprender. Esto, además, dota a la extensión de una objetividad con la que siempre han tenido que lidiar los estudios sobre la profundidad.

Al contrario de lo que ocurría con la extensión, la definición de la profundidad no está libre de variaciones y discrepancias, que se suelen recoger como tres enfoques (Dóczy & Kormos, 2016; Read, 2004; Schmitt, 2014). Los dos primeros, centrados en la palabra, y el último, en el sistema. Primero, el enfoque que interpreta la profundidad como *precisión de significado* se centra en la diferencia existente entre tener una idea vaga y limitada de lo que significa una palabra o tener un conocimiento de su significado mucho más elaborado y específico. Segundo, el enfoque que identifica la profundidad con el *conocimiento completo de las palabras*. Una de las definiciones más extendidas es la descripción de lo que implica conocer una palabra (Nation, 2013), que trataremos más en profundidad en apartados sucesivos (vid. 2.2.2). Así el conocimiento de una palabra no solo se adquiere mediante la adquisición de sus rasgos semánticos sino también de los ortográficos, fonológicos, morfológicos, sintácticos, las colocaciones y las características pragmáticas. Tercero, el enfoque que iguala la profundidad con una *red de conocimiento* a la que se incorporan palabras y se relacionan unas con otras, acercándose, así al concepto de lexicón metal. En la siguiente imagen (vid. Figura 1) se puede ver de forma más esquemática la conceptualización de la profundidad que proponen.

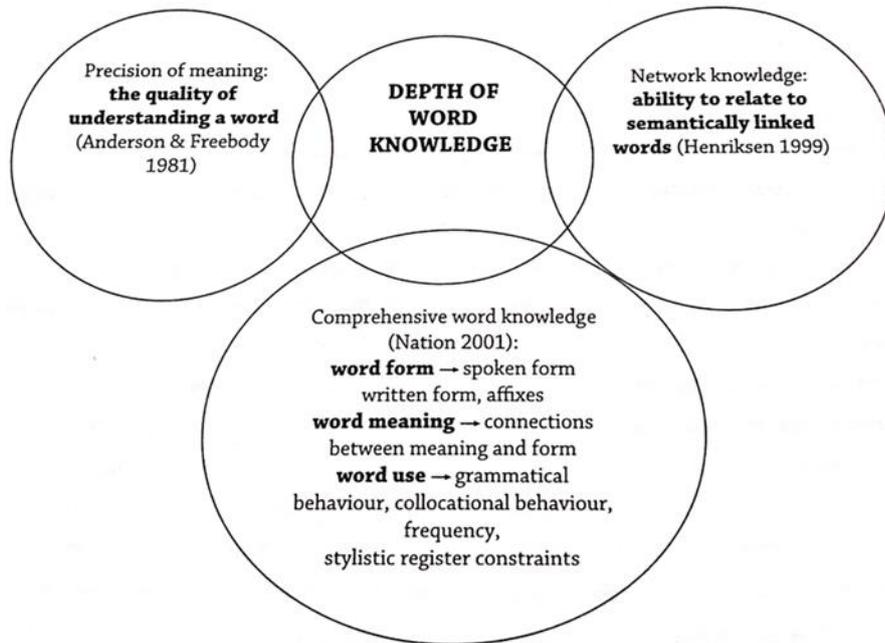


Figura 1. Conceptualización de la profundidad de Doczi & Cormos (2016)

Existe un mayor número de test que miden la *extensión* del vocabulario que de test que miden la *profundidad* (Schmitt, 2010, p. 89). Las formas de medir la profundidad se pueden categorizar en tres grupos: enfoque de desarrollo, enfoque de redes léxicas, enfoque de competencias (Yanagisawa & Webb, 2020), cada uno de los cuales se relaciona con una de las formas en las que se entiende la profundidad. Se entienden por *test de desarrollo* aquellos que miden el grado con el que se conoce una palabra en concreto, es el caso de VKL (Bruton, 2009), *Vocabulary Knowledge Levels*; en los de redes léxicas se evalúa la conexión de la palabra meta con las que está conectado, como el WAF (Wolter, 2002), *Word Associates Format*; y en los test de competencia se realiza una *batería de test* para poder comparar cada aspecto (González-Fernández & Schmitt, 2020).

Uno de los puntos cuestionados sobre la profundidad y la extensión es la relación que se establece entre ellas. En la actualidad parece estar confirmado que se trata de dos conceptos independientes, pero también interrelacionados (Schmitt, 2014). No se puede medir la extensión, sin tener en cuenta la profundidad (Dóczy & Kormos, 2016; Schmitt, 2010) aunque esta no haya sido la tendencia general en los estudios previos. En varios trabajos se encuentran correlaciones positivas entre la extensión y profundidad del vocabulario, incluso entre la extensión y la competencia, no así entre la profundidad y la competencia. Sin embargo, en los estudios que comparan las dos dimensiones suele

haber cierta superposición entre los constructos (Webb, 2012). Por lo tanto, el desarrollo de estos dos conceptos y la relación entre ambos no solo se centra en su definición o descripción sino en la conceptualización y metodología usada para evaluarlas (Schmitt, 2014). Frente a la posición anterior, otras visiones afirman que la distinción entre extensión y profundidad es desafortunada (Meara & Wolter, 2004) y que la distinción entre extensión y *organización* es mucho más práctica o productiva para la investigación del vocabulario. La principal crítica a esta dualidad es que no son comparables porque sus constructos son diferentes (Wolter & Gyllstad, 2013), punto que trataremos en los siguientes apartados.

2.1.3. Solidez

De la carencia de medir el conocimiento léxico a partir de la extensión y de la dificultad para medir todos los aspectos que implica la profundidad, diferentes estudios proponen otras propiedades que puedan captar una imagen global del conocimiento colocacional. Entre otras, se propone la *solidez* (Laufer et al., 2004), que permite medir el conocimiento léxico de forma más global teniendo en cuenta la extensión, número de palabras conocidas, y la profundidad (la conexión entre la forma y el significado). La solidez se define como el grado en el que se conoce una palabra teniendo en cuenta dos de los aspectos más habituales en la descripción del léxico: las destrezas (producción y comprensión) y los procesos cognitivos (reconocimiento y recuperación). Se diferencia de la extensión en que mide el conocimiento de una misma palabra de cuatro formas diferentes y de la profundidad en que solo tiene en cuenta uno de los aspectos del conocimiento, el significado, y deja de lado los otros. Se presenta como una solución a las dificultades que presenta el límite entre producción y comprensión (Read, 2011) y las limitaciones de medir la extensión o la profundidad (Schmitt & Schmitt, 2020).

La gradación que mide la solidez se produce a partir de la combinación de la producción, la comprensión, el reconocimiento y la recuperación, de modo que se establecen cuatro grados del conocimiento léxico que forman una escala jerárquica desde el más complejo al más sencillo. Esta escala se confirmó con el análisis de los resultados de dos test, CATSS (Aviad-Levitzky et al., 2019; Laufer & Goldstein, 2004), en el que se asoció cada uno de los grados con una forma de medida. En el grado de menor

dificultad se mide la comprensión desde el reconocimiento a partir de un formato de opción múltiple en el que se selecciona la traducción en la L1. En el test original se aporta una palabra en la L1 y los participantes debe elegir entre cuatro opciones la traducción correcta al inglés. En el grado de mayor dificultad se mide la producción desde la recuperación a partir de la sustitución de una palabra en la L2 por la traducción de la L1. En este caso, se presenta la palabra en la L1 y se debe aportar la traducción a la L2. Entre ambos se sitúan dos grados que no se han podido distinguir significativamente y por lo tanto encontraríamos la comprensión desde la recuperación, en la que se traduce desde la L2 a la L1, y la producción desde el reconocimiento en la que se selecciona la traducción de una palabra en la L2. En la *Figura 2* podemos ver cómo se disponen las cuatro perspectivas desde las que se observa la imagen global del conocimiento de una palabra que ofrece la solidez. Observamos que no se refiere a las destrezas como comprensión y producción sino como destreza activa en la que se busca la forma y la destreza pasiva en la que se busca el significado. En la Tabla 1 vemos un resumen de los ítems utilizados en el test para medir esta dimensión.

	Recall	Recognition
Active (Productive) (retrieval of form)	Active recall	Active recognition
Passive (Receptive) (retrieval of meaning)	Passive recall	Passive recognition

Figura 2. Grados de solidez del conocimiento (Laufer & Goldstein, 2004)

	Recuperación	Reconocimiento
Producción	Turn into water <i>m</i> _____	<i>Turn into water</i> a. elect b. blame c. melt d. threaten
Comprensión	When something <i>melts</i> it turns into _____	<i>Melt</i> a. choose. b. accuse. c. makes threats d. turn into water

Tabla 1. Recopilación de ejemplos de los cuatro formatos de CATSS

La presentación de cuatro maneras de acceder al conocimiento léxico implica que la solidez sea una propiedad del léxico gradual. Esto es, que se puedan establecer grados

en el conocimiento léxico y que la relación entre los cuatro tipos de vocabulario sea jerárquica, es decir, impliquen una mayor o menor grado de dificultad. Uno de los argumentos para defender esta jerarquía es la diferencia entre la extensión del vocabulario productivo y comprensivo de los aprendices y hablantes de una lengua. Dos hipótesis: la producción es más difícil que la comprensión y la recuperación es más difícil que la comprensión. Según esto, establecer la jerarquía implicaría ordenar los tipos de conocimiento desde el menor grado de solidez hasta el más alto con los siguientes puntos. Es decir, miden el conocimiento del léxico de forma gradual como una escala implicativa. En un enfoque que trata de definir el conocimiento reconocido y recuperado (Schmitt, Schmitt, et al., 2020) desde la evaluación, basado en CATSS (Aviad-Levitzky et al., 2019), presentan una diferencia entre la información de partida, el conocimiento aportado en el test y el conocimiento solicitado al aprendiz que realiza el test. En la *Tabla 2* podemos ver ejemplos de extracción de la información: de qué manera recuperar el significado de una forma o la forma de un significado y de qué manera demandar el reconocimiento de la forma de un significado o el significado de una forma.

Conocimiento aportado	Conocimiento solicitado	
	RECUPERACIÓN	RECONOCIMIENTO
SIGNIFICADO 'líquido blanco que se bebe'	Recuperación de la forma (proveer el elemento de la L2) <i>leche</i>	Reconocimiento de la forma (seleccionar el elemento de la L2) <i>Leche, vino, cerveza</i>
FORMA leche	Recuperación del significado (aportar la definición o traducir a la L1) <i>milk</i>	Reconocimiento del significado (selección de la definición o traducción a la L1) <i>Milk, wine, beer</i>

Tabla 2. Enfoque para definir el conocimiento reconocido y recuperado (Schmitt & Schmitt, 2020, p. 38)

2.2. El léxico

2.2.1. El concepto de léxico

Uno de los conceptos clave de esta investigación es el de *léxico*, así que debemos aclarar cómo entendemos ese concepto y cuáles son las barreras que lo separan de otros términos afines: *vocabulario* y *lexicón*. Si buscamos estos términos (*léxico* y *vocabulario*) en un diccionario de uso común, encontramos que se tratan como sinónimos; sin embargo, en diccionarios especializados, nos encontramos con un resultado diferente, tal y como podemos apreciar en las siguientes definiciones:

Vocabulario: El vocabulario o léxico puede definirse como el conjunto de unidades léxicas de una lengua (Martín Peris, 2008).

Vocabulary: "a set of lexemes, including single words, compound words and idioms" (Richards & Schmidt, 2010)

En el *Diccionario de términos CLAVE* ya se especifica claramente que "en didáctica de la lengua, los términos vocabulario y léxico son equivalentes, aunque en algunas disciplinas lingüísticas se establece una distinción entre ambos términos" (Martín Peris, 2008). Sin embargo, desde la TST (Mel'čuk, 2012) sí se establece una diferencia clara entre léxico y vocabulario. Se entiende el léxico de una lengua como la entidad teórica que se corresponde con el conjunto de unidades o entidades del léxico de una lengua (Polguère, 2010), mientras que el vocabulario como el conjunto de unidades o entidades del léxico de un ámbito concreto, ya sea de un texto o un individuo (Polguère, 2010).

El *lexicón* se suele relacionar con el almacenamiento, más específicamente, con cómo y dónde las palabras se podrían almacenar y representar en la mente (Dóczy, 2019). Aunque se puede explicar comparándolo metafóricamente con un diccionario, puede definirse como el conocimiento que un hablante tiene interiorizado del vocabulario (Martín Peris, 2008) y que equiparan al concepto de *competencia léxica* (vid. 2.2.2). Entre las cuestiones más interesantes en relación con este concepto está la de qué tipo de información se almacena y cómo se organiza cuando se habla más de una lengua. Por lo tanto, la organización, el acceso y la selección del léxico son temas clave de este ámbito. Pese a la dificultad del concepto, los últimos estudios se inclinan por tratarlo como una red integrada por diferentes ítems léxicos y representada mediante rasgos (Vitevitch et al., 2015). Otra de las áreas que se han investigado es la diferencia entre el lexicón de la L1 y de la L2; a este respecto, quedan claros dos aspectos: que el de la L2 es menos extenso que el de la L1 y que el de la L2 está menos organizado, aunque cualquier aspecto formal del lexicón está condicionado por una serie de factores de uso como el contexto o la frecuencia (Dóczy & Kormos, 2016).

Entendemos por *unidades del léxico*, o *entidades léxicas* (Polguère, 2010), todas aquellas palabras y combinaciones de palabras que forman parte del léxico. En diferentes estudios se utiliza el término de *unidad léxica* referido tanto a las palabras (como unidades simples) como a otras unidades mayores con sentido unitario como modismos o locuciones (C. Ainciburu & Santana, 2022). Aunque no mencionan las colocaciones

como unidades léxicas complejas, tomando la definición de forma amplia, entendemos que se encuentran entre las unidades del léxico pese a que como veremos no tienen un significado unitario. Aquí seguiremos el enfoque de la TST según el cual una *unidad léxica* es equivalente a una acepción de una palabra polisémica y es descrita en un artículo lexicográfico completo puesto que toda unidad léxica es una entidad compleja que puede describirse a través de tres componentes: significado, forma y propiedades combinatorias (Alonso Ramos, 2016). En este sentido, una locución como *estirar la pata* es una unidad léxica y tiene un artículo lexicográfico, mientras que una colocación no es una unidad léxica, puesto que es composicional. Con todo, puesto que debe ser aprendida y descrita en el léxico, es una unidad del léxico o como prefiere Polguère, una entidad léxica.

Para evitar este tipo de confusión nos decantaremos por *unidades del léxico* como término clave. Incluimos, por lo tanto, las unidades fraseológicas. En este capítulo hablaremos de léxico, incluyendo todas las unidades o entidades que forman parte de él. Las colocaciones forman parte de este *material léxico* o de la *estructura del léxico* (C. Ainciburu & Santana, 2022). A esta estructura del léxico, además de las palabras, pertenecen las unidades fraseológicas como las locuciones, las colocaciones.

2.2.2. El conocimiento léxico

Una de las bases de los test de vocabulario es la definición de su constructo, es decir, la definición del aspecto de la lengua que se va a medir, en este caso, del conocimiento léxico, que condicionará el diseño del test (Dóczy & Kormos, 2016; Laufer, 1998). En este apartado abordaremos esta cuestión desde una perspectiva amplia porque es un concepto problemático y no es el objeto directo de este test, pero sí indispensable, ya que no podemos hacer operativo el conocimiento colocacional, constructo del TCE, sin previamente haber hecho operativo el conocimiento léxico. Además, trataremos de dejar claro qué entendemos por conocimiento léxico y por competencia léxica, estableciendo así una diferencia clara entre dos conceptos polémicos, utilizados en ocasiones como sinónimos, pero imprescindibles para el diseño y propósito del TCE.

La dificultad a la hora de definir el conocimiento léxico viene, entre otras razones, por la variedad de enfoques con que se ha formulado. Sin embargo, unas propuestas no invalidan a las otras y complementarlas se plantea como una solución adecuada (Leńko-Szymańska, 2020). Incluso, son varias las clasificaciones que se hacen de los diferentes enfoques que tratan de definir el conocimiento o la competencia léxica aportando diferentes términos que agrupan a unos y otros enfoques: *modelo de rasgos separados*, *modelo de rasgos globales* (Nizonkiza & Van den Berg, 2015; Zareva, 2010); *enfoque componencial*, *dimensional* y *enfoque evolutivo* (Dóczy & Kormos, 2016; James Milton & Fitzpatrick, 2014). Tomaremos una clasificación sencilla que ordena los enfoques según estén centrados en la palabras individuales o en el léxico como sistema, que se explican claramente en la siguiente cita, punto clave en el diseño y análisis de los resultados del TCE.

Two approaches to the description of lexical competence have been dominant in the literature. The first one perceives lexical competence as constructed of the knowledge of individual words. This is a word-centred approach. The other approach is not concerted with individual words but with the lexicon as a whole and its functioning as a system. The researches taking the former stand list components and degrees of word knowledge; the proponents of the system-centred approach concentrate on dimensions of vocabulary knowledge (Leńko-Szymańska, 2020, p. 16).

A continuación, explicaremos brevemente en cada una de las propuestas mencionadas incidiendo sobre sus ventajas y limitaciones. Los enfoques centrados en la palabra se suelen relacionar con los test que miden la profundidad y los enfoques centrados en el sistema, con los que miden la extensión. Empezaremos con los enfoques centrados en la palabra para, posteriormente, centrarnos en los enfoques centrados en el léxico.

El *enfoque componencial*, uno de los enfoques centrado en la palabra, que inició con los trabajos de Richards (Richards, 1976), se consolidó con la propuesta de Nation⁵ (Nation, 2013) y explica qué significa conocer una palabra a partir de la división en sus componentes constituyentes. El enfoque componencial es sin duda una de las propuestas de definición y descripción del conocimiento léxico más aceptadas y, sobre todo,

⁵ Mientras el propósito de Richard no era el de explicar de forma sistemática lo que significa conocer una palabra sino describir y explicar el conocimiento léxico, Nation sí propone de forma clara una explicación sistemática de lo que significa conocer una palabra.

principalmente aplicada con un amplio recorrido en la evaluación de lenguas (González-Fernández & Schmitt, 2020; Schmitt, 2019a). Aun así, y como veremos a continuación, se han observado limitaciones que el propio autor comparte y acepta, adaptando su propuesta.

El conocimiento léxico se concibe desde este enfoque como la suma de una serie finita de componentes organizados en torno a tres grupos: el conocimiento de la forma de una palabra, el conocimiento del significado y el conocimiento del uso tal y como podemos ver en la *Figura 3*. El conocimiento de la forma incluye tres componentes: el conocimiento de la forma escrita, el conocimiento de la forma hablada y el conocimiento de las partes de la palabra. El conocimiento del significado incluye otros tres componentes: la relación entre la forma y el significado de la palabra, la relación entre la palabra con el concepto y con el referente, y, por último, las asociaciones. Finalmente, el uso de la palabra incluye los tres últimos componentes del conocimiento de las palabras: las funciones gramaticales, las colocaciones y las restricciones de uso. Una de las grandes aportaciones de esta propuesta es la dualidad de cada componente: el conocimiento léxico tiene una versión de cada uno de los componentes del conocimiento, una para la producción y otra para la comprensión. El aprendizaje de vocabulario, siguiendo este enfoque, se daría a través de sistemas de conocimiento y no de un simple ejercicio de memorización de palabras aisladas; de esta forma muchos de los componentes se podrían derivar de la L1 o de otras palabras ya conocidas en el sistema generalizando el conocimiento previo.

FORM	Spoken	[R] What does the word sound like?
		[P] How is the word pronounced?
	Written	[R] What does the word look like?
		[P] How is the word written and spelled?
Word parts		[R] What parts are recognisable in this word?
		[P] What word parts are needed to express the meaning?
MEANING	Form and meaning	[R] What meaning does this word form signal?
		[P] What word form can be used to express this meaning?
	Concept and referents	[R] What is included in the concept?
		[P] What items can the concept refer to?
Associations		[R] What other words does this make us think of?
		[P] What other words could we use instead of this one?
USE	Grammatical functions	[R] In what patterns does the word occur?
		[P] In what patterns must we use this word?
	Collocations	[R] What words or types of words occur with this one?
		[P] What words or types of words must we use with this one?
	Constraints on use	[R] Where, when and how often would we expect to meet this word?
		[P] Where, when, and how often can we use this word?

Figura 3. Enfoque componencial del conocimiento léxico (Nation, 2020)

Pese a reconocer su exhaustividad, una de las principales críticas es la falta de relación entre los componentes que conlleva diversas cuestiones sin responder: a) la relativa contribución de varios de los componentes al constructo del conocimiento y b) si algunos de los componentes tienden a adquirirse antes que otros (Daller et al., 2007; González-Fernández & Schmitt, 2020). Otra es la dificultad para aplicar esta lista de componentes a la evaluación ya que no es factible evaluar el conocimiento de todos los componentes de forma simultánea (Milton, 2013, p. 16). El propio autor en su último artículo sobre los diferentes aspectos del conocimiento de las palabras señala alguna de estas limitaciones como su carácter estático y la falta de atención a la evolución del aprendizaje (Nation, 2020a, p. 25), por lo que presenta como una de las futuras direcciones de la investigación la creación de un modelo más centrado en el uso y crecimiento del conocimiento. Así, propone una nueva explicación de lo que significa conocer una palabra en la que, sin rechazar sus principales preceptos, matiza algunas cuestiones.

Otro de los enfoques centrado en la palabra es la gradación del conocimiento de la palabra según el cual las palabras se pueden conocer en mayor o menor grado según el número de componentes conocidos o según el proceso de activación del conocimiento demostrado. Por ejemplo, es posible escribir una palabra, pero hacerlo con faltas de ortografía o ser capaz de reconocer el signado de una palabra si la vemos escrita pero no utilizarla en caso de necesitarla. Esta propuesta tampoco está exenta de dificultades. Primero, la asimilación del enfoque componencial que plantea cuestiones como la de que no todos los componentes tienen la misma relevancia y el dominio de todos es esencial en todas las situaciones comunicativas. Segundo, la distinción entre la producción y la comprensión, que ya se plantea desde la propuesta componencial, pero que muestra dificultades en el límite entre las dos ideas y la prioridad de una frente a la otra. Tercero, la frontera entre las etapas de una escala. Una de las soluciones planteadas tiene en cuenta el proceso cognitivo que implica la diferencia entre rescatar la forma (recuperación) o el rescatar el significado (reconocimiento) de una palabra: es la propuesta base de CATSS (Aviad-Levitzky et al., 2019; Laufer & Goldstein, 2004) en la que se propone una escala jerárquica del conocimiento centrada únicamente en uno de los componentes del conocimiento, la unión entre la forma y el significado (vid. 2.1.3 para

una explicación más detallada). Otras propuesta proponen otro tipo de escalas en las que los aprendices tienen que demostrar o no el grado de conocimiento (vid. 3.2.2.3).

Los enfoques centrados en el sistema, o *propuestas dimensionales*, se apartan de nivel de la palabra para centrarse en los conocimientos y habilidades del aprendiz (Daller et al., 2007; Henriksen, 1999; Meara, 1996). Estas propuestas no están unificadas, es decir, no todos los autores proponen las mismas dimensiones para medir la competencia léxica. En el intento de describir la competencia léxica y los factores que contribuyen a su desarrollo se buscan dimensiones que describan el léxico considerado como un todo y no como propiedades de los elementos concretos del léxico. En el siguiente cuadro podemos ver de forma comparada las distintas propuestas dimensionales, que aunque con diferencias, presentan un enfoque general muy similar (Meara & Miralpeix, 2017b).

	Propuesta 1 (Meara, 1996)	Propuesta 2 (Henriksen, 1999)	Propuesta 3: Espacio léxico (Daller et al., 2007)
DIMENSIONES de la competencia léxica	Extensión	Conocimiento parcial o preciso	Amplitud
	<i>Organización</i>	Profundidad	Profundidad
	Accesibilidad	Habilidad receptiva y comprensiva	<i>Fluidez</i>

Tabla 3. Variantes en la organización de las dimensiones de la competencia léxica

En este apartado analizaremos los puntos en común y las diferencias entre algunas propuestas dimensionales que presentamos en la Tabla 3. Además, revisaremos las otras dimensiones como la *organización* y la *accesibilidad* o *fluidez*.

Con la *organización* se pretende especificar la riqueza de la estructura que relaciona las palabras (Meara, 1996). Esta riqueza se plasma en el salto desde las listas de palabras a las redes de vocabulario. Parte de la investigación en este ámbito se ha centrado en el diseño de una herramienta de evaluación que refleje este tipo de relaciones. Los test de asociaciones de palabras han ido evolucionando, sobre todo en relación a su validez y fiabilidad, hasta dar con un modelo adecuado (Meara & Miralpeix, 2017b; Meara & Wolter, 2004). La *fluidez*, que se corresponde con la accesibilidad de Meara, se define como la facilidad y automaticidad de una aprendiz para usar las palabras y conocer toda la información necesaria sobre su uso (Daller et al., 2007). Tiene que ver con la precisión y rapidez a la hora de recuperar una palabra. Q_Lex (Meara & Miralpeix, 2017a) es un programa que mide la facilidad de un aprendiz para acceder a una serie de palabras muy frecuentes. La comparación entre la inmediatez de un nativo para reconocer o recuperar una palabra y las dificultades de procesamiento para los hablantes

de una L2. Para hacer operativa la facilidad para acceder a las palabras necesarias, siguió el método del *Lexical Decision Task* en el que hay que decir si distintas combinaciones de letras, con apariencia de palabras, son palabras reales o pseudopalabras y donde se tienen en cuenta diferentes variables. A estas propuestas se unen otras dimensiones que completan esa visión global del léxico de un aprendiz, o un hablante, y que todavía está incompleta. Desde la solidez (Laufer et al., 2004), hasta la *facilidad léxica*⁶ (Harrington, 2018) o el concepto de *complejidad léxica* (Kalantari & Gholami, 2017), en la línea de lo propuesto para otros sistemas complejos de la lengua.

Hay cierta confusión entre conocimiento léxico y *competencia léxica* o *dominio léxico*⁷ sin una definición clara de ninguno de los dos conceptos en el entorno de la evaluación de lenguas y con una clara tendencia a no equipar el conocimiento de la palabra con la competencia léxica (Milton & Fitzpatrick, 2014). La propuesta de traducción de *lexical proficiency*, que se interpreta como el correlato conductual de la competencia léxica en el nivel teórico cognitivo (Bulté et al., 2008), formaría parte de la competencia léxica: tanto el componente declarativo, formado por el conocimiento léxico (extensión, anchura y profundidad) y el componente procesual (habilidades sobre el conocimiento). A nivel conductual, el dominio léxico permitiría medir la competencia a través de la diversidad, sofisticación, complejidad, productividad o fluidez léxica. Así que el dominio del léxico sería la demostración de la competencia léxica. La principal diferencia entre uno y otro concepto radica en el ámbito de aplicación: el conocimiento léxico se reduce al ámbito de la palabra y la competencia al ámbito del aprendiz. Es decir, los dos enfoques se centran en una idea común, pero cuando se refiere al léxico como sistema se usa el concepto de competencia léxica (propia del hablante nativo y objetivo de aprendiz) y conocimiento léxico para atender a las palabras específicas que componen el sistema.

Por lo tanto, el conocimiento léxico de la palabra entronca con la pregunta de qué significa conocer una palabra o en qué grado se conoce una palabra, es decir, con el enfoque componencial, y la competencia léxica del aprendiz con el enfoque dimensional.

⁶ Capacidad para reconocer palabras rápidamente y combinando el conocimiento del vocabulario, manifestado como extensión del vocabulario, con la habilidad de procesamiento de acceso a las palabras (Harrington, 2018, p. 68).

Es importante en este punto explicar que nos situamos en la línea de los modelos relacionados con la competencia comunicativa⁸ y competencia léxica (Consejo de Europa, 2002) que distingue entre *conocimiento*, *competencia* (como grado de dominio) y *actuación*, distinta de la diferencia dicotómica entre *competencia* y *actuación* de Chomsky y el generativismo (Leńko-Szymańska, 2020, p. 13). Se define la competencia léxica como una combinación de conocimiento y habilidad que permiten la demostración de la lengua. Por lo tanto, la competencia no solo recoge información o conocimiento sino también capacidad para aplicar ambos al uso de la lengua. En la misma línea, se opone *conocimiento léxico* y *uso léxico* (Levitzky-Aviad & Laufer, 2013) en el que entiende el conocimiento como la información sobre una palabra y el uso como la manifestación del conocimiento en la producción a tiempo real.

Entendemos, por lo tanto, el *conocimiento léxico* como la información necesaria para usar una palabra y su grado de conocimiento y que relacionamos directamente con la profundidad y la solidez, pero indirectamente con la extensión. Actúa como sinónimo de conocimiento de la palabra. La *competencia léxica* la entendemos como el conocimiento y uso que un aprendiz hace del conocimiento léxico de todas las palabras y entidades léxicas que forman parte de su vocabulario o lexicón mental, es decir, del sistema. Recordamos que en este caso no estamos usando el concepto de *competencia* como oposición a *actuación* sino en la línea de *competencia comunicativa*, entendida como una habilidad fundamentada en el conocimiento y demostrada en la acción.

2.2.3. Relación entre léxico y frecuencia

En el ámbito de la didáctica, la relación entre frecuencia y vocabulario se ha ido consolidando por su objetividad y facilidad para la generalización. Se ha investigado de forma longitudinal desde varios enfoques y con diferentes propósitos. En este apartado solo atenderemos a aspectos que creemos relevantes para esta investigación, aquellos que nos conducen a la creación del test de colocaciones. Aunque el objeto de estudio no es el léxico en general sino las colocaciones, las bases del conocimiento actual sobre

⁸ Este modelo de competencia comunicativa consta de dos componentes: el conocimiento de la lengua y la habilidad para usarla que tiene como resultado un uso concreto y que siguen diversos autores (L. Bachman & Palmer, 2010b)

aprendizaje de colocaciones tienen como principal fuente el aprendizaje del léxico hasta el punto de considerar la frecuencia como el criterio más objetivo y estandarizado para distintos propósitos: las palabras más frecuentes son más útiles y fáciles que las menos frecuentes (Nation, 2013b; Schmitt & Schmitt, 2020). Por lo tanto, resulta imprescindible esbozar algunas ideas del aprendizaje del léxico en relación con la frecuencia. En primer lugar, revisaremos algunos de los puntos clave del aprendizaje del léxico en relación con la frecuencia, como son el *perfil basado en frecuencias* (Milton, 2013) o la *ley Zipf* (Sorell, 2013). Después abordaremos dos ámbitos de especial relevancia para la enseñanza-aprendizaje de cualquier idioma, y también para la evaluación de lenguas: la *selección* (vid. 2.3.3.1) y *nivelación* (vid. 2.3.3.3) de las palabras u otras unidades del léxico.

Los aspectos tratados en relación con el aprendizaje del léxico se centran en qué palabras, cuántas y cómo se deberían aprender. De la búsqueda de la respuesta a estas preguntas han partido la mayoría de los estudios sobre el aprendizaje del léxico, la inmensa mayoría centrados en el aprendizaje del inglés. La lingüística de corpus ha ejercido una importante influencia en este ámbito, ya que su base cuantitativa permite dar respuesta de forma objetiva a estas cuestiones. Aun así, no es la frecuencia el único factor que se ha tenido en cuenta para dar respuesta a estas preguntas, pero sí uno de las más fiables y estables ya que depende de la propia L1. A pesar del relativo consenso en cuanto a la importancia de la frecuencia en el uso, en la elección y secuenciación del vocabulario, no está del todo claro hasta qué punto la frecuencia léxica está supeditada a otro tipo de criterios (Vincze et al., 2016, p. 640). El desarrollo de la lingüística de corpus ha permitido avanzar en la relación entre frecuencia y aprendizaje. Una mirada a las diferentes respuestas dadas a dos preguntas base (qué palabras y cuándo enseñarlas) nos lleva a dividir dos etapas en los estudios de vocabulario. Por un lado, se relacionan las palabras, la frecuencia y las bandas de frecuencia en corpus, que identificaremos con la selección de palabras. Por otra parte, la correlación entre las palabras, la frecuencia y los niveles propuestos por el MCER.

Al igual que en la enseñanza-aprendizaje de lenguas, en la evaluación, la frecuencia es de gran utilidad en la evaluación del léxico y ha marcado el camino a seguir.

La extensión, propiedad de orden cuantitativo, parte de los principios extraídos de esta relación (frecuencia-léxico) y basa tanto la selección de las palabras meta.

Existe una relación entre la frecuencia de una palabra y la probabilidad alta de que esa palabra se encuentre en un texto y se aprenda (Milton, 2013, p. 25). De este principio se deriva un modelo de aprendizaje testado empíricamente, el *perfil basado en frecuencias (frequency profile)* (Meara, 1992) desde un punto de vista cuantitativo. Se trata de un modelo de uso del vocabulario que representa la cantidad de palabras que un hablante conoce en base a su frecuencia (Milton, 2013, p. 26). Según este perfil, el conocimiento típico de un aprendiz se establece en relación con la frecuencia de las palabras, de modo que conocerá más palabras frecuentes y menos palabras no frecuentes. Se representa en forma de un gráfico de barras, cada una de las cuales representa una banda de frecuencia (vid. Figura 4). Como vemos en la imagen, a medida que la frecuencia de las palabras que se incluyen en cada banda disminuye, también lo hace, de forma correlativa, el número de palabras que se conocen de dicha banda. El gráfico representa en el eje vertical el número de palabras conocidas y en el eje horizontal cada una de las bandas de frecuencia en que se dividen las palabras de la lengua. Así cada barra representa el número de palabras conocidas en cada banda; de modo que, el perfil nos indica el número de palabras conocidas por banda, dando lugar al perfil de frecuencia. Este también, es el principio en el que se basan la mayoría de los test que miden la extensión (*breadth*) de la competencia léxica.

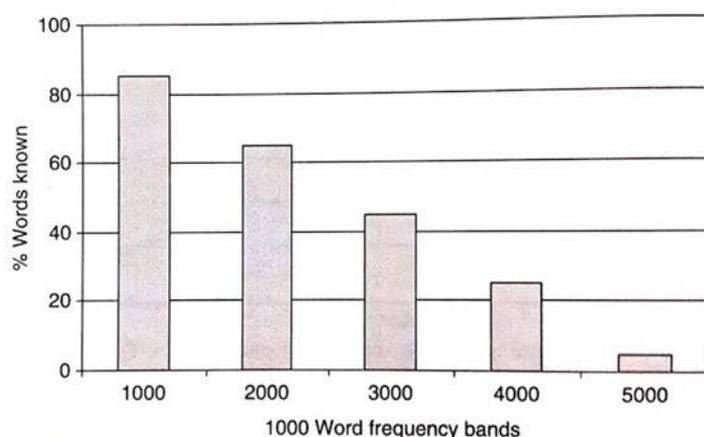


Figura 4. Perfil del vocabulario de un aprendiz típico basado en frecuencias (Meara, 1992, p. 4)

La validez y actualidad de este modelo se ha probado en diferentes lenguas, entre ellas griego y francés, que confirman, con mínimas variaciones cuantitativas en cada banda, la curva del perfil. Esto da lugar a establecer como regla de oro del aprendizaje de L2 que las palabras más frecuentes en una lengua tienden (pero solo tienden) a aprenderse antes que las palabras menos frecuentes.

Además, desde la lingüística de corpus, se observa que, cuando las palabras de una muestra de expresión oral o escrita se cuantifican y ordenan de la más frecuente a la menos frecuente, el puesto del ranking en que está cada palabra de la lista es inversamente proporcional a su frecuencia de ocurrencia (en ese texto o en la lengua) y que se conoce como ley Zipf. En otras palabras, hay una relación inversamente proporcional entre el ranking y la frecuencia de la palabra en un texto (Sorell, 2013). Este descubrimiento es interesante por la explicación que da de este fenómeno: la selección de las palabras no se hace únicamente por motivos de comprensibilidad o información, sino para mantener la atención del receptor (Jarvis & Daller, 2013).

2.2.3.1. Selección de las palabras meta

Seleccionar el vocabulario que se va a aprender en cada etapa no es una tarea sencilla; suelen ser los profesores o los libros de texto los encargados de decir qué elementos del léxico se aprenden, normalmente, según el vocabulario previamente conocido y con la intención de ampliarlo y en qué orden. Parece obvia la idea de que el vocabulario que se debe aprender es el que no se conoce (Nation & Webb, 2013), pero la duda reside en delimitar qué léxico se aprende, en qué orden y qué criterios nos sirven para seleccionar las palabras y demás elementos léxicos que debemos aprender. En un sentido similar, para evaluar el conocimiento léxico se necesita la selección de unas palabras meta que representen aquello que se quiere medir. Esto se realiza a partir de una serie de criterios y unas herramientas propias.

Distintos manuales enumeran y ordenan los criterios de selección del léxico (García & Rull, 2019; Izquierdo Gil, 2004; Salazar García, 2004). De forma sistemática, se clasifican los criterios de selección del léxico según su grado de objetividad. Así, contaríamos con una serie de criterios objetivos entre los que estaría la *frecuencia*, la

disponibilidad léxica, la *rentabilidad comunicativa*, etc. Y, entre los subjetivos, estarían las necesidades del aprendiz y la experiencia del profesor, es decir, los factores contextuales. Los criterios subjetivos parecen ser los criterios elegidos para la selección del léxico en los materiales didácticos, pese a que los instrumentos pedagógicos disponibles dan algunas claves. De hecho, el PCIC (Instituto Cervantes, 2006) expone una breve explicación de los criterios que utilizan para la selección del vocabulario: la frecuencia y la rentabilidad comunicativa. Pese a que la frecuencia es el criterio más recurrente, no ha sido el único criterio que se ha tenido en cuenta para la selección del léxico. Así, se habla de *léxico fundamental*, *léxico disponible* (Ainciburu, 2012; Ávila Muñoz, 2016) y *léxico básico* (Palapanidi, 2012). En /dispoLex/⁹ se define el léxico disponible como el conjunto de palabras que los hablantes tienen en el léxico mental y cuyo uso está condicionado por un tema concreto de comunicación. Se diferencia del léxico básico en que este es el más frecuente de la lengua. El léxico fundamental se define como la suma del léxico básico y el léxico disponible.

Existen varias herramientas que recogen el léxico de una lengua, una de ellas son los diccionarios. Una de las premisas es que el léxico seleccionado sea útil para el aprendiz y en los diccionario se recoge todo el léxico de una lengua, por lo tanto, sería pertinente discernir cuál es útil para el aprendiz (Bartol Hernández, 2010). Una de las cuestiones a trabajar sería la inclusión del nivel en cada forma y más concretamente en cada acepción. Esto es lo que se hace en el Cambridge Learner Dictionary¹⁰, en el que cada acepción está nivelada según la escala el MCER. Otras herramientas como las *listas de palabras frecuentes* o los *recopilatorios nocionales* también son útiles. Debemos, sin embargo, puntualizar el esfuerzo que desde la lexicografía se ha hecho para ampliar la utilidad de los diccionarios al incorporar mejoras a sus ediciones, sobre todo a las electrónicas, en las que han incorporado datos de frecuencia e incluso se ha adjudicado un nivel a cada palabra (Ezquerro, 2011).

Sin duda, la herramienta más habitual en la selección del léxico que se va a testar, son las listas de palabras frecuentes. Las listas de palabras, de larga tradición en el ámbito

⁹ <http://www.dispoplex.com/info/la-disponibilidad-lexica>

¹⁰ Esta función se extrae de *English Vocabulary Profile*, (<https://englishprofile.org/wordlists>). Se trata de una herramienta que con una amplia base de datos que proporciona las palabras que conocen los aprendices que realizan las pruebas de nivel de Cambridge en cada nivel del MCER.

de la lingüística aplicada en inglés (Cobb, 2007; Coxhead, 2000; Gardner & Davies, 2013), delimitan y priorizan el vocabulario que serán útil en el océano de posibles palabras y frases (Schmitt, Schmitt, et al., 2020). Las listas pueden tener dos finalidades: una como herramienta pedagógica y otra como instrumento para la elaboración de test (Martínez, 2011). Las listas se pueden clasificar según los propósitos de aprendizaje o según sean listas generales o especializadas (Ngoc & Dang, 2020). La necesidad, los criterios para crear las listas¹¹ y la evaluación de la propia lista, así como su clasificación (según su propósito y el tipo de unidad del léxico que incluyan) son algunos de los retos que presenta la creación de una nueva listas. La selección de las palabras de una lengua que se deben aprender ha llevado a contabilizar el número de palabras que un hablante necesita para comunicarse, el número de palabras que un aprendiz necesita para entender un discurso de forma eficaz y el número de palabras necesario para producir un discurso.

La frecuencia permite organizar el léxico de una lengua en bandas de frecuencia, proponiendo así un modelo en dos o tres bandas (vid. Figura 5). En un sistema tripartito encontramos palabras de frecuencia alta, palabras de frecuencia media y palabras de frecuencia baja. La principal dificultad de esta clasificación reside en el corte o límite entre las bandas. Las primeras son las palabras frecuentes en la lengua y, por lo tanto, las que tienen una mayor cobertura (Vilkaite & Schmitt, 2019). *General Service List* (GSL) (Garnier & Schmitt, 2015) ha sido la lista de palabras más estandarizado hasta las últimas décadas, cuando a partir de corpus actualizados se crearon listas de palabras representativas de la lengua actual. Entre las listas generales actuales en inglés la *New General Service List* (Brezina & Gablasova, 2015) es una de las más conocidas .

¹¹ Una de las cuestiones más debatidas en la bibliografía disponible es la forma de contar las palabras (o unidades del léxico).

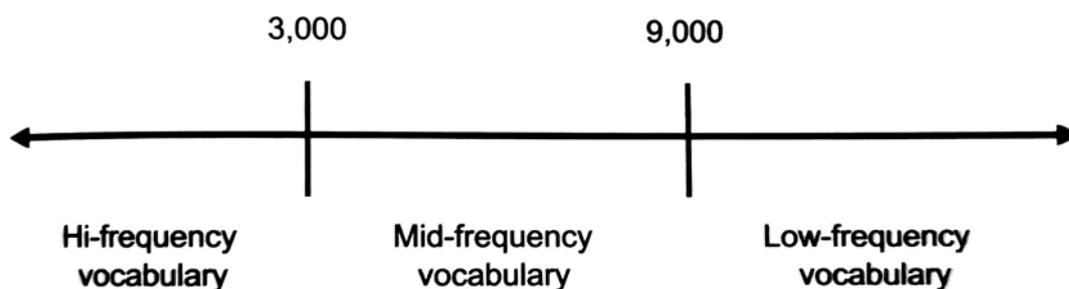


Figura 5. Representación de bandas de frecuencias (Schmitt, Schmitt, et al., 2020)

Antes de pasar a analizar algunas de las claves de las listas de frecuencias y su aplicación en el ámbito del español, debemos centrarnos en la cuestión de la identificación de palabras ya que hay que delimitar qué se entiende por palabra. Entre los problemas que se encuentran están las diferencias entre *types* y *tokens*, la inclusión de las familias de palabras o las unidades pluriverbales. En la actualidad se toma uno u otro criterio en función del propósito de la lista de frecuencia, considerando la inclusión de unidades mayores que la palabra una de las limitaciones de las listas de frecuencia (Martínez, 2011, p. 23). De hecho, como veremos en futuros apartados, la creación de listas de colocaciones frecuentes supone una dificultad aún mayor (vid.2.3.3). Entre las limitaciones de las listas de palabras frecuentes, además de la incorporación de las unidades fraseológicas, está la cuestión de la polisemia (Alonso-Ramos, 2012; Crossley et al., 2010).

Sin discutir la utilidad de las listas de frecuencia, también se reconocen sus limitaciones y se proponen otro tipo de herramientas (Brysbaert et al., 2021; Durrant et al., 2022, p. 157). Por ejemplo, las *Knowledge-Based Vocabulary List*, a partir de ahora KVL (Schmitt et al., 2021) que representan el vocabulario que probablemente conozcan los aprendices de L2. Podemos decir que estas y las listas de frecuencia son listas de palabras inversas: en unas se parte de la frecuencia de ocurrencia para representar lo que comprenden o deben aprender los aprendices y en otros se parte del conocimiento del aprendiz para representar lo que estos pueden producir considerando las palabras más fáciles aquellas que los aprendices pueden escribir con corrección y las más difíciles las que los aprendices no pueden escribir. Al comparar este estudio se comprueba que la frecuencia se correlaciona con la comprensión, pero no con la producción; por lo tanto,

es un criterio útil pero no definitivo. La KVL se compiló a partir de la colaboración de aprendices cuya lengua materna es el español¹², el alemán y el japonés.

En español contamos con listas de palabras frecuentes, pero no tienen el mismo desarrollo que en inglés. Debemos distinguir varias herramientas: los diccionarios con información de frecuencia, los diccionarios de frecuencia y las listas de palabras frecuentes. Incluso corpus, como CORPES XXI, que permite visualizar listas de palabras frecuentes ordenadas en bandas de frecuencia. Algunos ejemplos son el *Diccionario de frecuencia* (Davies, 2006), que reúne el vocabulario nuclear ordenado por frecuencia, o *Cumbre* (Cantos-Gómez & Almela-Sánchez, 2018; Sánchez et al., 2001). La información que ofrecen estos recursos no es suficiente para la selección de una muestra de colocaciones niveladas, de modo que hay que recurrir a otro tipo de recursos en los que el léxico sí esté nivelado como el PCIC, cuyo eje vertebrador de la lista no es la frecuencia. Un estudio reciente analiza los criterios de selección del léxico de manuales de ELE (García & Rull, 2019), recurso fundamental a partir del que se selecciona el vocabulario que se estudia en las aulas. La hipótesis de partida es que en los manuales de ELE ni se tienen en cuenta los datos de frecuencia ni de disponibilidad léxica. Frente a la no aclaración de los criterios, sí se comprueba una amplia analogía entre los elementos seleccionados en los distintos manuales, así como la coincidencia con las nociones del PCIC. Desde el ámbito de la didáctica, se ha estudiado la incidencia de las listas de frecuencia en la presentación del léxico a aprendices de español como lengua extranjera con la confirmación de que este tipo de aprendizaje explícito se muestra más eficaz que el incidental (Hidalgo Gallardo, 2020).

2.2.3.2. *Asignación de nivel del MCER*

La asignación de nivel al léxico, que está ampliamente vinculada con la competencia lingüística, en relación con cualquiera de sus destrezas (Qian & Lin, 2020), se puede ver desde dos perspectivas sucesivas. Primero, la asignación de nivel a cada palabra o elemento del léxico en concreto. Esto quiere decir que se asigna a la palabra

¹² Se accede de forma gratuita a esta lista en el siguiente enlace: <https://www.britishcouncil.org/exam/aptis/research/knowledge-based-vocabulary-lists-kvl>

lluvia el nivel A2 y a la palabra *granizo* un nivel B2, por ejemplo, porque los aprendices deberían conocer *lluvia* en el nivel A2 y *granizo* en el B2. Segundo, la asignación de nivel a los aprendices, es decir, determinar si la competencia léxica de un aprendiz se corresponde con el nivel A2 o con el nivel B2. La nivelación de las palabras u otros elementos del léxico se ha venido correlacionando con bandas de frecuencia, en relación con la teoría de que el léxico más frecuente es el léxico más necesario. La competencia léxica, sin embargo, se ha venido midiendo a partir de diferentes instrumentos de evaluación como rúbricas, aludiendo a criterios más menos objetivos como los que ofrece el MCER.

En esta investigación necesitamos tanto nivelar las colocaciones como la competencia colocacional de los aprendices y, para eso, en este apartado, haremos una breve revisión sobre cómo se asignan los niveles a las palabras y cómo se asigna un nivel de competencia léxica a los aprendices. En otro apartado nos centramos en la nivelación de las colocaciones y de la competencia colocacional (vid. 2.3.3.3). Como habitualmente, la cantidad de recursos e investigaciones del inglés son muy superiores a la realizadas en el ámbito del español como lengua extranjera.

La frecuencia es un recurso objetivo utilizado para asignar nivel a las palabras. De la relación entre utilidad y frecuencia se establece que las palabras más frecuentes se encuentran en un nivel y las palabras menos frecuentes en un nivel alto. Pese a este acuerdo general, la dificultad reside en el establecimiento del umbral de cada uno de los grados de la escala del MCER. Una de las primeras preguntas realizadas fue la de cuántas palabras debía conocer un aprendiz en cada nivel, posteriormente, se estableció un límite que relacionaba cada banda de frecuencia con un nivel.

Así, en test estandarizados como el VLT o VST (Nation & Beglar, 2007b) la competencia de los aprendices se asigna según la baja, media o alta frecuencia de las palabras que forman la muestra (vid. 3.2.2.1 para una descripción de estos test). Uno de los recursos más utilizados es a través de listas de palabras frecuentes organizadas en bandas de frecuencia a las que se asigna un nivel. Es el caso de KELLY Project, (Kilgariff et al., 2013) o CERFLex¹³ (Pintard & François, 2020), dos herramientas que extraen listas

¹³ <https://cental.uclouvain.be/cefrlex/>

de palabras frecuentes de corpus y relacionan, en base a diferentes criterios y umbrales, con los niveles del MCER. Estas herramienta se desarrollan en varias lenguas entre las que está el español ELELex¹⁴.

Otra estrategia es extraer la información a partir de las aportaciones de los aprendices. Es el caso de *English Vocabulary Profile* (Capel, 2010, 2012) que utiliza el corpus de aprendices, *Cambridge Learner Corpus*, para asignar un nivel a cada palabra. Esta herramienta de acceso semi restringido es un recurso extremadamente útil para todos los agentes que intervienen en el proceso de enseñanza-aprendizaje, que lo integran a partir del número de ocurrencias que las palabras y combinaciones tienen en corpus. Pearson creó una herramienta similar, *Global Scale of English*¹⁵ (De Jong & Zheng, 2016) (GSE), que incluye una lista de significados de palabras nivelados. En ambas herramientas se nivelan los sentidos y no solo las formas como ocurre con las listas de palabras frecuentes.

También se está trabajando la nivelación desde los resultados que aplicaciones de test adaptativas con respuestas masivas ofrecen como es el caso de Duolingo¹⁶. Con input y a partir de medidas estadísticas probabilísticas se pueden extraer conclusiones sobre el nivel de cada palabra. En este tipo de trabajos el análisis se centra en los ítems y no en los participantes. Así el número de aciertos de cada ítem según el nivel del participante que acierto o falle permitirá clasificar a la palabra evaluada en cada ítem en un nivel del MCER.

En español contamos con un repositorio de léxico nivelado, el PCIC. Sin embargo, esta herramienta cuenta con varios problemas: no se pueden realizar búsquedas filtradas según el nivel, tópico o noción, categoría gramatical, etc. ni tampoco conocemos los criterios que han elaborado la selección del material léxico ni la asignación de nivel (Contreras-Izquierdo & Martos-Eliche, 2020; García-Salido & Alonso-Ramos, 2018).

¹⁴ <https://cental.uclouvain.be/cefrlex/elelex/>

¹⁵ <https://www.english.com/gse/teacher-toolkit/user/lo>

¹⁶ <https://research.duolingo.com/>

2.2.4. Aprendizaje del léxico

El aprendizaje del léxico es un proceso continuo y sin fin, aún en la L1 (Dóczy & Kormos, 2016) por lo que también lo es en la L2. En este apartado nos centraremos en dos aspectos: la presencia del léxico en los documentos de referencia y la carga de aprendizaje.

2.2.4.1. *El léxico en los documentos de referencia*

En esta sección analizaremos cómo tratan el léxico los documentos de referencia, el MCER (Consejo de Europa, 2002) y el PCIC (Instituto Cervantes, 2006), recordemos que se trata de guías orientativas y no de materiales prescriptivos. En el MCER se sitúa la competencia léxica como uno más de los tipos de competencia lingüística, entre los que también está la gramatical, semántica u ortográfica. Se tratan dos asuntos en relación con la competencia léxica: la lista de elementos léxicos y una serie de escalas orientadas a la evaluación de esta competencia. Enumera los elementos léxicos entre los que se encuentran las fórmulas fijas, los modismos, las estructuras fijas, las frases hechas y el régimen semántico (Consejo de Europa, 2002), este último definido como expresiones que se componen de palabras que habitualmente se utilizan juntas. Por los ejemplos que presentan, *cometer un crimen* o *cometer un error*, comprobamos que se refieren al concepto de colocación con el término *régimen semántico*. En las escalas, que sirven para graduar el conocimiento de vocabulario, se desglosan dos propiedades, la riqueza y el dominio del vocabulario. En ningún caso se define ni una ni otra propiedad, sino que se plantean una serie de descriptores que aluden a la variedad y concreción de las palabras. Tampoco se propone un inventario sobre qué palabras o elementos del léxico se deben enseñar o aprender en cada nivel para alcanzar la competencia léxica deseada, ni tampoco unos criterios objetivos que permitan a los especialistas seleccionar el léxico.

En el PCIC (Instituto Cervantes, 2006) se atiende a cuestiones relacionadas léxico a partir del componente nocional, que en lugar de ajustarse al concepto tradicional de palabra se centra en un concepto más amplio de unidad léxica que atiende a su dimensión combinatoria. Este componente nocional se organiza en nociones generales y nociones específicas. Las nociones *generales* son a las que un aprendiz necesitará recurrir en cualquier contexto y las *específicas* son a las que se recurre en situaciones concretas que requieren usos más específicos; que se organizan en categorías y subcategorías de

carácter temático. En el PCIC se presenta un inventario de las nociones organizado según los niveles del MCER, sin la especificación de los criterios, para la elección y nivelación de estas nociones más que un criterio semántico. Esto supone un problema a la hora de elaborar pruebas objetivas en base a estos recursos.

Los criterios que menciona el PCIC para la selección del léxico son los siguientes: a) registro, b) rentabilidad comunicativa y c) frecuencia léxica¹⁷. La repercusión de los dos primeros criterios sobre el tercero es evidente (García-Salido & Alonso-Ramos, 2018). Aun así, aclaramos que la frecuencia léxica se establece a partir de la introspección de los autores que, aunque suele coincidir, no viene avalada por los datos objetivos de un corpus, de hecho, se encuentra en algunos casos una correlación negativa entre frecuencia léxica y nivel.

2.2.4.2. *La carga de aprendizaje del léxico*

La relación entre conocimiento y dificultad de la palabra hace explícita la vinculación entre el enfoque componencial, que define el conocimiento léxico, y la *carga de aprendizaje* que describe el grado de esfuerzo al adquirir una palabra nueva (Laufer, 2012). Este concepto es esencial en el desarrollo del TCE ya que, al tratarse de un test de diagnóstico, que diagnostica las debilidades y fortalezas, su objetivo es detectar los puntos débiles que se derivan del aprendizaje de colocaciones. En este caso, las debilidades en el aprendizaje se identificarán con factores intrínsecos, que son los que tenemos en cuenta en el TCE, y nos permitirán esbozar la carga de aprendizaje de las colocaciones de ELE.

Cada uno de los aspectos que describen el conocimiento léxico se considera factor potencial de la carga de aprendizaje (Nation, 2020b). Las conclusiones que se extraen de los estudios relacionados con esto son relevantes tanto desde un punto de vista pedagógico, ya que conocer las dificultades permitirá mejorar los procesos de enseñanza aprendizaje del léxico, como desde la evaluación de lenguas, proporcionando una base más sólida para el diseño del formato e interpretación de los resultados de los test. El concepto de dificultad graduable se centra en los aspectos particulares de las palabras concretas que conocemos y no en el léxico como conjunto o lexicón mental. En una

¹⁷ Se especifica en la introducción a los capítulos cinco y nueve del Plan.

primera época los estudios se centraron en la enumeración de los factores que pueden afectar, con mayor o menor precisión, al aprendizaje y su clasificación, según diferentes criterios (Laufer, 1997). En un segundo momento, se dedicaron a la creación un modelo general que describa la dificultad de las palabras a partir de la implicación de distintos factores (Dunn, 2013).

Los factores que condicionan la dificultad de aprendizaje de las palabras se dividen en *intraléxicos* (los relacionados con la palabra en sí misma), como puede ser la longitud de la palabra, e *interlingüísticos* (la influencia de la L1), como la presencia de cognados (Schmitt, Schmitt, et al., 2020); y también los *extralingüísticos* como la motivación, la necesidad, el estado de ánimo, la presión, etc. Los *intrínsecos* se suelen dividir a su vez en *objetivos*, aquellos que se pueden medir, y los *subjetivos*, que no se atienen a medidas concretas. En el ámbito del español, también se enumera los factores de dificultad idiosincráticos en el aprendizaje del léxico tomando como base las aportaciones propuestas en la enseñanza-aprendizaje del inglés (Ainciburu & Santana, 2022, pp. 46-48).

Laufer (1997, 2012) propuso una serie de elementos intrínsecos que siguen vigentes hasta día de hoy y que divide en factores que facilitan el aprendizaje, factores que inducen dificultad y factores sin un claro efecto sobre la dificultad de aprendizaje. En la tabla siguiente se actualizan estos factores que afectan al *reconocimiento*, *procesamiento* y *producción* de una nueva palabra (Schmitt & Schmitt, 2020, p. 165). En ella, los rasgos se adaptan a las investigaciones actuales y se presentan de forma dual, de modo que puede ser un rasgo facilitador o lo contrario. Por ejemplo, la familiaridad de los fonemas de la palabra se considera un elemento facilitador, pero la presencia de fonemas extraños a la L1 en la palabra se considera un rasgo que induce a la dificultad de aprendizaje. También se propone una clasificación que distingue entre los factores que se dan al inicio del aprendizaje, primer contacto, y aquellos que aparecen a lo largo del proceso, o longitudinales.

Pese a que cada estudio propone sus propios factores condicionantes del peso de aprendizaje hay bastante coincidencia entre unos y otros. La razón es que muchos autores entienden los subcomponentes o rasgos del conocimiento como potenciales

dificultades del aprendizaje. En el siguiente cuadro podemos comparar diferentes clasificaciones que aportan pequeñas variaciones con respecto a la propuesta estándar.

LAUFER ¹⁸ (2012)	MILTON ¹⁹ (2009)	SCHMITT (2010)
Pronunciación: forma hablada	Diferencia de sonido y apariencia	
Ortografía: forma escrita		
<i>Synformy</i> ²⁰		
Propiedades gramaticales		
Significado		
Forma y significado en relación con la L1	Traducción a la L1	
Registro		
Combinaciones de palabras: colocaciones		
	Longitud	Número de letras
		Número de fonemas
		Número de sílabas
	Concreta / imaginable	Grado de <i>imagineabilidad</i>
		Grado de concreción
		Grado de relevancia significativa.
		Categoría gramatical

Tabla 4. Comparación entre factores que influyen en la carga de aprendizaje

Además, otros dos factores suelen estar presentes en los estudios sobre la carga de aprendizaje léxico: la frecuencia y la congruencia. El primero es uno de los más estudiados, en algunos casos, incluso considerado factor determinante (Milton, 2013). Las razones que han llevado a priorizar la frecuencia frente a los otros aspectos no solo es el interés particular sino la objetividad y facilidad procedimental y metodológica con la que se analiza. La congruencia se define como la presencia o ausencia de una traducción literal o equivalente a la L1 (Nesselhauf, 2003). La influencia de la congruencia se ha demostrado, sobre todo, desde la lingüística de corpus.

La *carga de aprendizaje* de una palabra no es la misma para cada aprendiz. El grado de dificultad variará en función de sus conocimientos lingüísticos previos. Esta idea se concreta en lo que denomina *principio general* de la carga de aprendizaje: cuanto más se acerca una palabra a los patrones y conocimientos con los que está familiarizado el aprendiz, más ligera es la carga de aprendizaje de dicha palabra (Nation & Webb, 2013). Y es que diferentes palabras tienen diferentes cargas de aprendizaje para aprendices con

¹⁸ Únicamente incluye factores intrínsecos y objetivos.

¹⁹ Los presenta como regla de oro: una palabra de una lengua extranjera es probablemente más fácil de aprender si...

²⁰ Concepto que define como similitud de las formas léxicas (Laufer, 1997).

diferentes bagajes lingüísticos (Nation, 2013b, p. 44). Esta familiaridad se marca a partir de tres condicionantes: la L1, el conocimiento de otras lenguas y el conocimiento previo de la L2. Uno de los objetivos de la investigación sobre léxico es reducir esta carga buscando patrones y analogías de forma sistemática. En relación con este principio, está el concepto de *dificultad absoluta* (Dunn, 2013) de aprendizaje de una palabra que se resume en tres causas:

- Las circunstancias del aprendiz como las diferencias individuales, la motivación, la aptitud ante el aprendizaje, etc.
- El contexto de aprendizaje que viene determinado por la L1, el bagaje cultural previo, la situación en la que el aprendizaje tiene lugar, etc.
- Las características lingüísticas de las palabras que se pretenden aprender, entre las que están las características de las palabras en sí mismas, el papel que ésta juega en el sistema lingüístico del que es parte, etc.

Debería ser posible tomar todos estos factores separados y ponerlos en funcionamiento en solo modelo (Milton, 2013, p. 38). Dos estudios proponen un modelo (Daller et al., 2007; Willis & Ohashi, 2012) en el que se estudia el comportamiento de tres variables frente a la dificultad del aprendizaje: la frecuencia de coocurrencia, la longitud de la palabra y la influencia de la L1. El modelo propuesto por Milton & Daller (2007) tiene un doble objetivo: crear un modelo e identificar cuánto contribuye cada factor a la probabilidad de aprender una palabra. Las conclusiones revelan que la frecuencia prevalece sobre los otros factores en el resultado de **aprendibilidad*.

El segundo estudio se puede considerar continuación del primero, con la intención de mejorar la conceptualización tanto de la longitud de la palabra como de la influencia de la L1. Si en el primer estudio, la influencia de la L1 se medía en función de las letras coincidentes en cada palabra, y en el segundo, se comprobaba si coincidían con una lista específica de cognados. Aunque se mantuvo la forma binaria (cognado o no cognado) la forma de corrección también sufrió cambios. En cuanto a la longitud también se detectaron diferencias. En el segundo trabajo, además del número de letras se tuvieron en cuenta el número de sílabas y el número de fonemas. Este cambio supone que la influencia de la L1 pasa de no presentar una influencia significativa, a convertirse en el primer factor de la dificultad. La cognición, la frecuencia y la longitud fonética son

los factores que influirían en la dificultad de aprendizaje, en orden descendente, desde el que más influye al que menos. Entre ellos, hay propuestas que dan un paso más frente a la simple enumeración (Dunn, 2013).

Esta autora va más allá y propone un modo de organizar las hipótesis que han marcado el estudio sobre la carga de aprendizaje, es decir, ordena las variables que influyen en la mayor o menor dificultad de la palabra (Dunn, 2013, p. 38). Represente este modelo relacionar en una imagen (vid. Figura 6). En el centro del diagrama está la palabra, que implica el conocimiento de los rasgos formales de esta. Rodeándola se presentan cuatro factores divididos a su vez en dos secciones: una, relacionada con el sistema lingüístico de la L2, y otra, con la mente del aprendiz. El modelo consta de cuatro vectores causantes de la dificultad:

- Las cuestiones formales relacionadas con la palabra
- La frecuencia de coocurrencia de la palabra en la lengua, pero también a lo largo de diferentes géneros y registros.
- El papel y la regularidad de la palabra en el sistema de la L2.
- Los enlaces semánticos, que hacen referencia a la conceptualización del mundo que tiene. Por ejemplo, casos de ambigüedad.
- La interrelación de la palabra, formada con n palabras ya conocidas por el aprendiz.

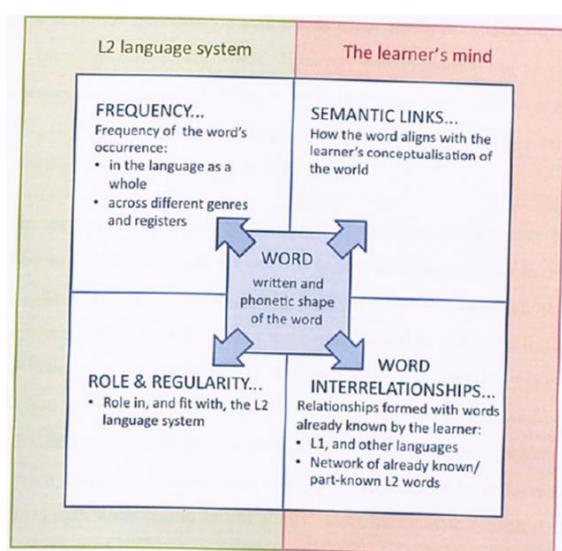


Figura 6. Modelo de dificultad (Dunn, 2013)

El modelo presenta varias novedades. Una es la diferencia que establece entre los rasgos de la L2 y de la L1. Deja fuera otras variables como la regularidad (rasgos lingüísticos propios de la L1) y la interrelación entre L1 y L2. Se trata de un modelo empírico cuyos factores estudian la contribución que cada uno de ellos aporta.

MILTON & DALLER	Willis & Ohashi	DUNN
Frecuencia	Frecuencia	Frecuencia
Congruencia	Congruencia	
Longitud	Longitud	Forma
		Ambigüedad (polisemia)
		Enlace semántico (concreción / imagineabilidad)

Tabla 5. Comparación entre varios modelos que explican la dificultad de la palabra

2.3. Las colocaciones

En este apartado, de forma paralela al dedicado al aprendizaje del léxico (vid. 2.2), abordaremos los mismos aspectos en relación con las colocaciones. Para esto haremos algunas consideraciones sobre qué entendemos por colocación, qué es el conocimiento y la competencia colocacional y cómo se aprenden y enseñan las colocaciones. También dedicaremos un espacio a la relación entre colocaciones y frecuencia, de gran relevancia en los estudios sobre colocaciones, y que nos llevará a desarrollar dos de los temas clave para la evaluación: la selección de la muestra y la asignación de nivel.

2.3.1. El concepto de colocación

La definición de la unidad de la lengua que se va a evaluar es uno de los pasos esenciales en el desarrollo del test. Pese a la cantidad de trabajos sobre las colocaciones que han aparecido en los últimos años (Almela & Cantos, 2018; Boers & Webb, 2018; Henriksen, 2013b; Tavares-Pinto, Rees, Frankeberg-Garcia, 2021), todavía no es un concepto carente de discusión. La dificultad reside en que se han usado diferentes términos para referirse a un mismo concepto o a conceptos con límites borrosos. Entre los términos más habituales está el de *colocación*, en la actualidad mayoritariamente aceptado, pero también se utiliza otros como *unidades multipalabra*²¹, *solidaridades*

²¹ Muchos de estos términos engloban lo que nosotros entendemos como diferentes unidades fraseológicas: colocaciones, locuciones, modismos...

léxicas, combinaciones de palabras, combinaciones semilibres, chunks, etc. Pese a esta disyuntiva, como apunta Alonso Ramos (2010) en el título de uno de sus artículos: “No importa si la llamas colocación o no, descríbela”.

Para describir la colocación es imprescindible precisar la definición y perfilar los criterios de identificación. Son varias las definiciones de colocación con las que contamos y una de las razones que se suele dar como causa de esta diversidad es la variedad de propósitos con los que estudia este fenómeno lingüístico (Bosque, 2011b, p. viii). Por lo tanto, la pluralidad de enfoques se atribuye a los distintos campos de investigación desde los que se ha tratado el concepto de colocación (lexicografía, lingüística de corpus, enseñanza de lenguas, etc.), que han ido adaptando la definiciones y descripciones a sus necesidades (Alonso Ramos, 2010; Gyllstad, 2007b, p. 7; Higuera García, 2006; Nesselhauf, 2005b). Los estudios que se dedican al análisis de corpus suelen tomar un *enfoque basado en frecuencias o estadístico* (Firth, 1957); los que se centran en la enseñanza de una L2 es común que se decanten por un *enfoque didáctico o pedagógico* (Higuera García, 2006, 2017); y los que trabajan en la elaboración de diccionarios (lexicografía) habitualmente optan por un *enfoque fraseológico* (F. J. Hausmann, 1991) y desde la evaluación de lenguas, por un *enfoque combinado* (Gyllstad, 2007a). Frente a la opinión de que el único común denominador entre la amplia variedad de interpretaciones de colocación es que el término se usa para referirse a algún tipo de relación sintagmática de palabras (Nesselhauf, 2005b, p. 11), pensamos que hay más puntos en común entre las diferentes propuestas, tal y como trataremos de analizar a continuación.

Una de las cuestiones más debatidas entorno al concepto de colocación es la relacionada con los límites frente a otras combinaciones de palabras o entidades léxicas (Koike, 2001, pp. 29-39). Por un lado, se detectan dificultades a la hora de distinguir entre las combinaciones libres de las colocaciones. Por otro lado, se encuentran problemas en la delimitación de la colocación frente a otras unidades fraseológicas. Varias propuestas tratan de establecer los límites entre combinaciones libres, colocaciones y otras combinaciones fraseológicas (Corpas-Pastor, 1996; Koike, 2001; Mel'čuk, 2017; Rufat, 2010). Como explicaremos a continuación, consideramos las colocaciones como un tipo de unidades fraseológicas. La naturaleza de la colocación, la naturaleza de los elementos

de la colocación y el número de elementos de la colocación son algunos de los rasgos que aparecen de forma constante en las distintas definiciones propuestas de colocación. Predominan los siguientes acuerdos:

- Las colocaciones son unidades del léxico de orden fraseológico.
- Los elementos de la colocación suelen ser dos.

Antes de centrarnos en la identificación, definición y descripción de las colocaciones desde los diferentes enfoques, presentaremos brevemente su clasificación, por tratarse de uno de los aspectos más debatidos y entre los que existe un mayor consenso en la actualidad. La tipología de colocaciones ha respondido a diferentes criterios, entre los más extendidos están el número de componentes y la categoría gramatical de los mismos. En primer lugar, Benson (Benson, 1990, 2010) categorizó las colocaciones en *colocaciones léxicas* y *colocaciones gramaticales*. Las *colocaciones léxicas* están compuestas por palabras léxicas (verbos, adjetivos, nombres y adverbios) que se combinan de diferentes formas dando lugar a subclases. Las *colocaciones gramaticales* son las que contienen una palabra léxica, pero la otra, necesariamente ha de ser gramatical como una preposición, un infinitivo o una cláusula. En su clasificación combina los esquemas sintácticos con una sucinta descripción semántica de los colocativos. En segundo lugar, atenderemos al criterio morfosintáctico, por el que se clasifican las colocaciones según la categoría gramatical de sus constituyentes, para lo que seguiremos la propuesta de Koike (2001, pp. 44-46). Así ordena las colocaciones en *simples* y *complejas*. Las *colocaciones simples* son aquellas compuestas por unidades léxicas (UL) simples, por ejemplo, *escalar la cima*; y las *colocaciones complejas* son aquellas que se componen de una UL y de otra unidad fraseológica, por ejemplo, *escalar las cimas del poder* (Koike, 2012). Las *colocaciones simples* se organizan en seis grupos según la categoría gramatical implicada. Así habla de colocaciones sustantivo + verbo (entre las que incluye *sustantivo + verbo*, *verbo + sustantivo*, *verbo + preposición + sustantivo*, *sustantivo + adjetivo*, *sustantivo + de + sustantivo*, *verbo + adverbio*, *adverbio + adjetivo / participio* y *verbo + adjetivo*). Nos centraremos únicamente en las *colocaciones léxicas simples* que definiremos para poder identificarlas y describirlas (vid.2.3.1.4).

Las descripciones y definiciones de la colocación suelen presentar una lista de rasgos o características. Al comparar las características que aportan los diferentes

enfoques (*estadístico, fraseológico, didáctico o combinado*) observamos que se repite la misma idea con diferente nombre o en una se integran varias. Aun así, alguno de los rasgos se define por su carácter controvertido (Bosque, 2011b). A continuación, enumeramos algunos de los rasgos o características más habituales.

- a. La naturaleza fraseológica o no de la colocación²²
- b. La categoría gramatical de los elementos de la colocación
- c. Número de elementos
- d. La estructura morfosintáctica de la colocación
- e. Composicionalidad
- f. Contigüidad de los elementos
- g. Transparencia u opacidad
- h. Deducibilidad
- i. Institucionalización
- j. Relación entre los elementos
- k. Dificultad de aprendizaje

Nos acercaremos a las propuestas de los diferentes enfoques a partir de las características o rasgos propios que han ido proponiendo. Estas características las ordenaremos según sean *identificativas (o definitorias)* o *descriptivas* para cada enfoque. Por *características identificativas* entendemos aquellas que permiten definir (codificar o descodificar) las combinaciones de unidades léxicas como colocaciones. Las *características descriptivas* son las que ofrecen los enfoques como propias de las colocaciones, pero que no sirven para identificarlas. Es decir, no resultan operativas para determinar si un sintagma se trata de colocación o no. Esta organización nos permitirá comparar de un modo más eficaz las descripciones propuestas por los distintos enfoques. La imprecisión en las diferentes definiciones fraseológicas lleva a que el número de rasgos descriptivos o identificativos, según la propuesta, sea mayor en número que los que se tienen en cuenta desde otros enfoques, especialmente el basado en frecuencias.

²² Con naturaleza de la combinación se alude a la caracterización o no de la colocación como unidad fraseológica. En la actualidad se acepta mayoritariamente esta cuestión (Corpas Pastor, 2001), aunque todavía hay trabajos recientes en los que se pone en duda o no se ve tan claro el límite entre las combinaciones libres y las colocaciones (Bosque, 2017).

2.3.1.1. Las colocaciones según el enfoque fraseológico

Incluimos dentro del enfoque fraseológico todas las propuestas que con mayor o menor claridad consideran la colocación una combinación fraseológica y que, sobre todo, priorizan como criterio de identificación la relación entre los elementos constituyentes. Este enfoque tiene su origen en los trabajos de los formalistas rusos que consideran las colocaciones como unidades fraseológicas y, tal y como hemos adelantado, este es el enfoque adoptado para definir las colocaciones en esta investigación. De forma general, se dice que se tienden a considerar las colocaciones en términos composicionales y a centrarse en las restricciones léxicas, sintácticas y semánticas que sus componentes ejercen uno sobre el otro (Barfield, 2013; Mel'čuk, 2017). A continuación, explicamos la característica definitoria que diferencia las colocaciones de las combinaciones libres y de otras combinaciones fraseológicas como las locuciones. Después, revisaremos el resto de características que describen las colocaciones pero que no sirven para identificarlas en este enfoque.

a) Característica definitoria

La característica definitoria que determina este enfoque es la relación entre los constituyentes de la colocación, ya sea de orden léxico (Alonso Ramos, 2016), morfosintáctico (Bosque, 2017) o semántico (De Miguel, 2006). Según la restricción priorizada, los componentes de la colocación toman diferentes nombres. Cuando la restricción priorizada es léxica, se le llamará *base* al elemento libre, elegido semánticamente, y *colocativo*, al elemento dependiente seleccionado léxicamente por el otro constituyente de la colocación (F. J. Hausmann, 1998; F. J. Hausmann, 1991). Cuando la restricción priorizada es la semántica, se habla de *predicados* y *argumentos*. También es habitual llamar *colocados* a los dos elementos de la colocación (Corpas Pastor, 2001), en la Tabla 6 vemos la comparativa terminológica.

	Elemento 1	Elemento 2
Tendencia léxica-semántica	BASE	COLOCATIVO
Tendencia gramatical	ARGUMENTO	PREDICADO

Tabla 6. Diferente terminología para los componentes de la colocación en el enfoque fraseológico

Las principales discrepancias entre los enfoques fraseológicos radican en la destreza desde la que se observa la lengua: mientras que en la tendencia semántica y

gramatical el análisis y descripción se realiza desde la comprensión, en la tendencia léxica, se hace desde la producción (Alonso Ramos, 2016). Por lo tanto, el punto de mira está en la selección y no en la interpretación de las combinaciones. Asimismo, mientras en otros enfoques, es solamente uno el rasgo definitorio, en este se suelen incorporar, además de la *restricción*, otros rasgos más o menos objetivos, muchos de ellos coincidentes en unas y otras propuestas (Nesselhauf, 2005a, p. 25). La *transparencia* (otros hablan de opacidad) y la *conmutabilidad* (Cowie, 1981, p. 994) son algunos de los más repetidos, pero otros muchos han ido incorporándose (Henriksen, 2013a).

b) Características descriptivas

El número y la naturaleza de los componentes, la *frecuencia* de aparición y la *transparencia* son las características descriptivas que se vienen enumerando en los trabajos sobre colocaciones desde un punto de vista fraseológico (Barrios Rodríguez, 2015; Bosque, 2006; Corpas-Pastor, 1996; Corpas Pastor, 2001; Koike, 2001). Es decir, cómo entendemos la frecuencia y transparencia de las colocaciones desde un punto de vista fraseológico, considerándose aspectos propios pero que no nos permiten identificarlas. Algunas de las características esenciales habituales son la estabilidad, institucionalización, tipicidad y construcción según las reglas del sistema de dicha lengua (Corpas Pastor, 2001).

En cuanto a la frecuencia de coaparición, que dos palabras coocuran con una determinada frecuencia no indica que formen una colocación, aunque la mayoría de los miembros de la colocación sí coocuran frecuentemente, lo que justifica que se trate de característica descriptiva y no definitoria. De hecho, en este enfoque, la frecuencia no tiene ninguna validez para considerar una determinada combinación como colocación (Alonso-Ramos, 1994). Corpas apunta que los elementos de una colocación suelen aparecer con frecuencia de forma conjunta, pero que también los elementos de la combinación se caracterizan por ser frecuentes también como unidades (Corpas-Pastor, 1996). Esta afirmación choca con algunos de los estudios actuales sobre frecuencia de coaparición o exclusividad (Gablasova et al., 2017): son frecuentes juntos, pero no tienen por qué ser frecuentes por separado. Sería el caso de *hacer la cama* y *cama mullida*. Ambas son colocaciones y coaparecen frecuentemente, pero mientras tanto *cama* como

hacer son frecuentes, el adjetivo *mullido*, no. Además, la frecuencia de coaparición se considera, según Bosque (2011a) o Corpas Pastor (2001) un aspecto extralingüístico que se sitúa fuera del sistema.

La *institucionalización* es, desde el punto de vista psicolingüístico o cognitivo, la otra cara de la característica de frecuencia. La podemos entender como la fijación en función de la reproductibilidad de estas unidades del discurso (Corpas-Pastor, 1996). La relación entre la frecuencia y la percepción que los hablantes o aprendices de una lengua tienen de estas construcciones es un tema discutido.

Desde el enfoque fraseológico, se entiende la *transparencia* como una propiedad de los *frasemas* (o unidades fraseológicas) de orden psicológico y orientada a la interpretación y no como una propiedad formal orientada a la producción (Alonso-Ramos, 2017). La *transparencia* sí es una propiedad subjetiva, gradual y cambiante, ya que puede depender del conocimiento previo del hablante ya que la cercanía de la L1 y de la L2 es uno de los factores que puede determinar el grado de transparencia. De forma opuesta, la *opacidad* es solo perceptible desde la decodificación, nunca desde la producción. Otro de los rasgos subjetivos, la *predictibilidad*, parece un criterio poco operativo para determinar si una combinación de palabras es una colocación o no. Desde el punto de vista de la producción cualquier combinación sería impredecible (Alonso Ramos, 2010). En cuanto al carácter *deducible* de las colocaciones, dos artículos marcan el desarrollo de este rasgo en la literatura española sobre colocaciones: *Deducing collocations* (Bosque, 2011a) o *On the conceptual bases of collocations* (Bosque, 2017) y su réplica, *Can collocations be deduced? A lexically driven analysis from the perspective of language production* (Alonso Ramos, 2016). En el primero se defiende el carácter *deducible* de las colocaciones y en el segundo se descarta este rasgo como propio de las colocaciones. En todo caso, concluimos que, desde el punto de vista de la comprensión, el sentido de las colocaciones puede ser deducible (o predecible), pero no siempre son predecibles desde el punto de vista de la producción. Este carácter deducible también tendría un punto de conexión con otro de los rasgos descriptivos de las colocaciones, la *tipicidad* (Koike, 2001, p. 28). Se entiende la tipicidad como rasgo semántico por el que se establece una relación típica entre los componentes de la colocación, esto es, la relación entre los dos componentes de la colocación suele estar presente en la definición

lexicográfica. El ejemplo característico para explicar la tipicidad es el de los instrumentos musicales y el verbo *tocar*: *toca la guitarra, el piano o la batería*.

2.3.1.2. *Las colocaciones según el enfoque estadístico*

Este enfoque nace unido a los estudios de *lingüística de corpus*, de ahí que algunos también lo llamen *enfoque textual* (Barfield, 2013) o *distribucional* (Gablasova et al., 2017). Sus principales representantes son Firth (Firth, 1958), Halliday (1966) y Sinclair (1966) que sientan su bases. Nesselhauf (2005a) establece una división dentro del enfoque distinguiendo el enfoque orientado a la estadística y el enfoque basado en frecuencias. O, dicho de otro modo: estudios de colocaciones con orientación textual (identificación de combinaciones atendiendo sobre todo a la distancia permitida entre ambas) y estudios de colocaciones basados en frecuencia (probabilidad de que una palabra aparezca en un texto al lado de la otra) (Men, 2018). Desde la *lingüística de corpus* se han trabajado las colocaciones con dos fines: la descripción de las unidades de la lengua (identificación) y, más recientemente, en relación con el proceso de aprendizaje.

a) Característica definitoria

Desde la primera definición propuesta en este enfoque hasta una de las últimas pasan cincuenta años y el rasgo que sirve para identificar las colocaciones es el mismo, la frecuencia (Firth, 1958; S. Gries & Durrant, 2019; Sinclair, 1991). Aun así, el criterio de frecuencia se ha ido precisando para paliar algunas de las críticas más habituales: qué medidas estadísticas son las más adecuadas y qué relación existe entre los miembros de la colocación (Gablasova et al., 2017) y, en consecuencia, cómo se refleja esa relación desde un punto de vista estadístico. Las soluciones propuestas plantean la utilización de medidas de asociación que tienen en cuenta la fuerza de o el grado de atracción entre los elementos de la combinación. La elección de una u otra medida de asociación dependerá del objetivo de la investigación (Gries, 2021a, 2021b).

Por lo general, los estudios que tratan las colocaciones desde una perspectiva estadística no se corresponden con lo que en este trabajo entendemos por colocaciones. La principal diferencia es que este enfoque se centra en la comprensión puesto que se trabaja con corpus, mientras que nuestro interés reside más en el conocimiento de los aprendices para producir colocaciones. Además, las medidas no suelen distinguir entre

diferentes unidades multpalabra e incluyen bajo el término de colocación diferentes unidades fraseológicas. Se plantearon diferentes medidas estadísticas de asociación para la identificación de las colocaciones. Estos métodos están asociados a unas características además de la frecuencia, como es la direccionalidad, la dispersión o la distribución, pero todavía se necesita un estudio más profundo sobre qué factores sirven para establecer qué palabras pueden considerarse colocaciones en un corpus (Gablasova et al., 2017).

b) Característica descriptiva

Las características descriptivas de este enfoque no dejan de estar relacionadas con el criterio identificativo, la frecuencia; cuestión en la que se avanza con nuevas medidas de asociación (Cantos-Gómez & Almela-Sánchez, 2018). Es el caso de la relación que establecen entre la frecuencia y tres dimensiones: la *dispersión*, la *exclusividad* y la *direccionalidad* (Gablasova et al., 2017). Dos de las más relevantes para nuestro estudio son la direccionalidad y la transparencia.

La *direccionalidad* atiende al diferente peso de un elemento de la combinación sobre el otro, frente a otras visiones en las que la relación entre los elementos de la combinación se considera bidireccionalidad o simétrica (Gries, 2013). Otra de las características que se suele incluir es la *transparencia* y esta sin relación con la frecuencia. En un intento por integrar la frecuencia y la transparencia como criterios para identificar colocaciones, se creó el FTF (*frequency-transparency framework*), que representa la relación que diferentes combinaciones pueden tener con las dos variables. En la *Figura 7* se puede ver esta propuesta de forma gráfica a partir de un plano de coordenadas en el que la frecuencia se representa en el eje vertical y la transparencia en el horizontal. Así, las colocaciones frecuentes y transparentes serían las más adecuadas para aprendices de los niveles más bajos o iniciales (por ejemplo, *take place*) e infrecuentes y opacas las más recomendadas para aprendices de niveles elevados (*take credit*). Los problemas empiezan con las categorías intermedias, de las que se ofrecen menos detalles (sobre todo de las frecuentes y opacas). Otra cuestión discutida es la relativa a la definición cuantitativa de la frecuencia, muy dependiente de las necesidades del aprendiz. Se define la transparencia más como una variable dependiente del aprendiz que como una característica no inherente a la combinación (Martinez & Murphy, 2011).

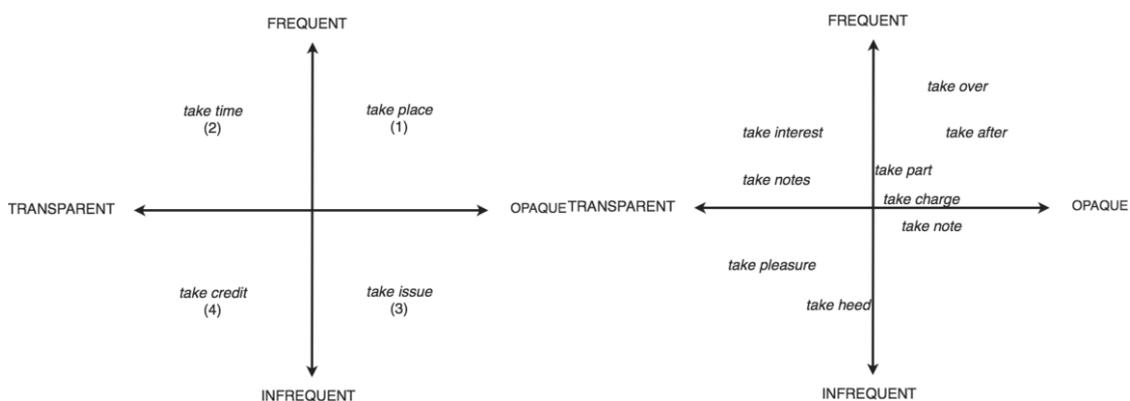


Figura 7. Representación del grado de transparencia y frecuencia de la colocación (Martinez & Murphy, 2011)

Las conclusiones extraídas de este enfoque tienen una importante repercusión en la selección de combinaciones, ya sea para fines didácticos o como en esta investigación, para convertirse en colocaciones meta de un test de lengua. Recordemos que el objetivo del enfoque basado en frecuencias no es describir las colocaciones sino identificarlas.

2.3.1.3. *Las colocaciones según el enfoque didáctico o combinado*

La perspectiva didáctica surge como consecuencia de la necesidad de enseñar colocaciones a los aprendices de segundas lenguas. De ahí que su objetivo principal no sea la identificación como en el enfoque basado en frecuencias, ni la definición como en el fraseológico, sino la descripción, de modo que los aprendices puedan reconocer este tipo de combinaciones y su dominio pueda facilitar su aprendizaje (Chamorro Guerrero, 2018; Serrano, 2015). La definición se simplifica y, por lo tanto, es más comprensible y rentable para los profesores de una lengua extranjera ya que permite la aplicación directa en el aula (Ferrando, 2018, p. 25). El *enfoque didáctico* no tendría espacio sin la aparición del *enfoque léxico* de Lewis (M. Lewis, 1993b, 1997, 2000). Desde este enfoque se define la colocación, entendida como *chunk*, como palabras que coocurren juntas más a menudo de lo esperado (Lewis, 2000, p. 127). En el ámbito del español como segunda lengua destacan los trabajos de Higuera (2006).

La relación de características descriptivas recoge muchas de las revisadas en los apartados anteriores en relación con el enfoque fraseológico y estadístico. Atendiendo al criterio de dificultad, Higuera (2006, p. 27) clasifica las colocaciones en prototípicas y no prototípicas, resaltando la importancia que para el aprendizaje presenta el primer tipo,

frente a las combinaciones libres. Las *colocaciones prototípicas*, como *reservar una mesa*, cumplirían con ocho rasgos: a) restricción léxica, b) direccionalidad, c) tipicidad, d) fijación arbitraria en la norma, e) regularidad sintáctica, f) transparencia semántica, g) institucionalización (pertenencia a una comunidad) y h) frecuencia de coaparición. Las *colocaciones no prototípicas* solo cumplen cinco, de manera que no existiría ni restricción léxica (a), ni direccionalidad (b), ni tipicidad (c), sería el caso de *hacer la compra*. Por su lado, las combinaciones libres se caracterizarían por la regularidad sintáctica (e) y la transparencia semántica (f), por ejemplo, *comprar un libro*.

En la misma línea, en el estudio del inglés como lengua extranjera, y teniendo en cuenta rasgos semánticos, en concreto la dualidad *transparencia* y *opacidad* como síntoma de la facilidad o dificultad ante el aprendizaje, Macis y Schmitt (2017, p.51) proponen otra clasificación de las colocaciones. Diferencian entre *colocaciones literales* (*literal collocations*), las *colocaciones figurativas* (*figurative collocations*) y las *colocaciones dobles* (*duplex collocations*). Las *colocaciones literales* son completamente composicionales, es decir, el significado de la colocación es la suma de los sentidos literales de sus componentes y que esquematiza como 'SIGNIFICADO A + SIGNIFICADO B'. En este caso el proceso de aprendizaje no consistiría en el significado de la colocación, sino en la elección del colocativo para cada base. En las *colocaciones figurativas* el significado no es derivable del de las palabras que lo componen porque la base mantiene el sentido literal y el colocativo adquiere un sentido específico para esa base. Las *colocaciones dobles* son polisémicas, tiene un significado literal y otro figurativo. En el caso del español podemos hablar de colocaciones como *dar un grito* como ejemplo de colocación literal y *soltar un grito* o *pegar un grito* como ejemplos de colocaciones figurativas, aunque esta distinción no sería completamente adecuada para los criterios que establecemos en este trabajo sobre qué es una colocación, pero sí puede servir como orientación.

Este enfoque también aparece de la necesidad de seleccionar una muestra de colocaciones adecuada para su evaluación, por lo tanto, está estrechamente relacionado con la identificación de la colocación, así que el criterio definitorio debe ser estrictamente objetivo, dada la insuficiencia del criterio de frecuencia y la complejidad del criterio fraseológico (restricción léxica, principalmente, aunque también semántica o gramatical).

Su singularidad reside en que se apropia de los dos rasgos definitorios exclusivos de los enfoques básicos: la frecuencia de coaparición y el tipo de relación entre constituyentes, es decir, el criterio identificativo de cada uno de los enfoques tradicionales (estadístico y fraseológico). Así, para identificar una colocación se tienen que encontrar los dos criterios en una misma combinación.

Los más claros representantes de este enfoque serían Nesselhauf (2005a), Gyllstad (2007a) o Nizonkiza (2012). Es importante señalar que no todos pertenecen a un mismo ámbito, sino que desde la lingüística de corpus y desde la evaluación de lenguas se trabaja con un concepto similar de colocación. Gyllstad (2007b), en concreto, propone una definición operativa de colocación, aunque limitada ya que se lo hace desde el punto de vista de la comprensión. También desde los estudios de corpus se busca la relación con las características de carácter más lingüístico, incluso semántico, con la frecuencia, más concretamente con la fuerza de asociación (Gablasova et al., 2017). Esta tendencia parece afianzarse en los estudios sobre vocabulario, sobre todo, en el ámbito anglosajón como se puede ver en la siguiente definición:

Pairs of words that are syntagmatically associated with each other because: they are frequently found together; they are restricted to each other, in the sense that they are rarely found outside of each other's company; or one word takes on a special meaning it does not carry in other contexts. Pairs may be flexible in the sense that their relative positions in an utterance can change, and, in some cases, they can be used in different parts of speech: e.g. *that was a **strong** argument; that **argument** was **strong**; she **argued strongly*** (Durrant et al., 2022, p. 17).

2.3.1.4. *El concepto de colocación en el TCE*

Tomo la definición de colocación para esta investigación de la *Lexicología explicativa y combinatoria* (Mel'čuk, 2017; Mel'čuk et al., 1995) que se enmarca en lo que hemos clasificado como enfoque fraseológico. En primer lugar, estableceremos los límites entre las colocaciones y las combinaciones libres y las colocaciones y otras combinaciones fraseológicas. Esto nos permite aclarar la definición con sus rasgos identificativos, a los que añadiremos los rasgos descriptivos que completan la caracterización de las colocaciones. Expondremos los tipos de colocaciones y haremos

una breve aproximación a una herramienta básica de este marco teórico, las *funciones léxicas*.

Las colocaciones son combinaciones no libres y se diferencian de las combinaciones libres porque la selección de uno de los componentes de la combinación está restringida léxicamente por el otro componente. Las colocaciones, además, son composicionales, es decir, su significado es la suma de los sentidos de sus dos componentes. Se diferencia de otras expresiones fraseológicas como las locuciones por este rasgo. Por lo tanto, las colocaciones están formadas por dos unidades léxicas. Tal y como avanzamos (vid. 2.3.1.1), los dos rasgos definitorios de este marco son la *composicionalidad* y la *restricción léxica* (Mel'čuk, 2012b), que explicaremos a continuación. Además, nos apoyaremos en dos conceptos clave que ayudan a delimitar el concepto de colocación: *control léxico* y *concordancia semántica* (Alonso Ramos, 2016). En el siguiente cuadro podemos ver la relación de las colocaciones con otras expresiones fraseológicas²³ (o *frasesmas*) atendiendo a los criterios que acabamos de mencionar: la composicionalidad y la naturaleza de la restricción. Así, las colocaciones se distinguen de los *locuciones* por la composicionalidad y de los *clichés*, por ejemplo, *en otras palabras* ;por su tipo de restricción (Mel'čuk & Mel'čuk, 2012).

Compositionality of phrasemes Nature of constraints	non-compositional	compositional
	lexical	IDIOMS
semantic-lexical	impossible	CLICHÉS

Figura 8. Clasificación de los frasesmas o unidades fraseológicas (Mel'čuk & Mel'čuk, 2012)

La TST cuenta con un sistema formal para describir las colocaciones. Se trata de un aparato que permite la representación formal y sistemática de las colocaciones (Mel'čuk, 1998). Su uso es especialmente útil en la elaboración de diccionarios (Alonso-Ramos, 2017) y en el procesamiento automático. Incluso se ha estudiado su aplicación en el ámbito de la enseñanza-aprendizaje de L2 (Barrios Rodríguez, 2012). La función

²³ No entraremos en la explicación del concepto de *locución* ni de *cliché*, según la TST, pero sí nos aproximaremos a su definición para dejar clara la diferencia con la noción de colocación que vamos a utilizar en este trabajo.

léxica (FL) se puede definir como una herramienta que modeliza la relación colocacional entre una base y un colocativo (Alonso-Ramos, 2006). En una colocación como *montar una fiesta*, en inglés *to make a party*, ambas se representan con la FL Oper, que podemos glosar como ‘hacer algo’. Entendemos por glosa una codificación en metalengua que parafrasee el sentido expresado (Alonso-Ramos, 2006, p. 72).

2.3.2. El conocimiento colocacional

Si definir el constructo de un test que mide el conocimiento léxico supone un reto, hacerlo de los test que miden colocaciones no es una tarea más sencilla. Esta mayor dificultad viene determinada por la complejidad de la propia unidad, dado su carácter fraseológico, la inconsistencia en la definición del conocimiento léxico, así como el menor número de estudios en comparación con los dedicados al conocimiento del vocabulario. Del mismo modo que mostramos la distinción entre el conocimiento y la competencia léxica, estableceremos una diferencia entre el conocimiento y la competencia colocacional. Entendiendo, de forma correlativa a la distinción que hemos hecho en relación con el léxico: la *competencia* la entendemos como la habilidad para usar las colocaciones de una lengua y *conocimiento* lo usamos para todo lo que hay que saber sobre una colocación en concreto, siguiendo en mayor o menor medida propuestas previas (Lee Revier, 2009). Así, relacionaremos la competencia con el test de nivel y el conocimiento con el test de diagnóstico.

Entre los estudios dedicados al conocimiento colocacional, según el tratamiento que se le ha dado a las colocaciones podemos diferenciar dos tendencias (Brown, 2014). Primero, la colocación como uno más de los aspectos que integran el conocimiento de las palabras, tendencia a la que nos referiremos como *de constructo dependiente*. Segundo, la colocación como elemento del léxico independiente y autónomo, por lo que los nueve aspectos del conocimiento (Nation, 2013) se aplican o se deben aplicar a las colocaciones, que entendemos como *constructo independiente*. Sin negar que el conocimiento fraseológico y, por supuesto, colocacional es uno más de los aspectos del conocimiento léxico, entendemos que el conocimiento colocacional también se puede medir como constructo independiente y, por lo tanto, es preciso describir qué aspectos de la colocación se deben conocer para comprender y producir una colocación, así como,

el grado en que un aprendiz puede conocer la colocación. También debemos tener en cuenta las dimensiones de la competencia colocacional, así como el alcance del dominio colocacional de cada aprendiz en el sistema, que estableceremos de acuerdo con los niveles establecidos por el MCER (inicial, intermedio y avanzado).

Entendiendo las colocaciones como un constructo dependiente, entre los componentes que integran el conocimiento de la palabra está el conocimiento de las posibles colocaciones a las que una palabra pueda dar lugar o pueda formar. Ya Richards (Richards, 1976) habla de las asociaciones como uno de los aspectos del conocimiento léxico entre los que están las colocaciones, aunque se use el nombre genérico de asociaciones y se evite mencionar el concepto directamente. Uno de los nueve componentes que Nation (2001, p. 27) menciona es la coocurrencia con otras palabras, es decir, las colocaciones (vid. Tabla 8). Es uno de los subcomponentes del uso y que, como el resto de componentes, presenta en su versión comprensiva y productiva aludiendo a la relación sintagmática entre la palabra a conocer y las que la acompañan.

Colocaciones	C	¿Qué palabras o tipo de palabras coocurren con esta?
	P	¿Qué palabras o tipos de palabras debemos usar con esta?

Tabla 7. Componente colocacional del conocimiento de las palabras. Extraído de Nation (2001, p. 27)

Como apuntábamos en el apartado dedicado al conocimiento léxico, una de las críticas a las propuestas componenciales centradas en la palabra (vid. 2.2.2) es el desconocimiento sobre la interrelación entre los componentes del conocimiento de las palabras. Esta interrelación se ha estudiado al comprobar la dificultad que conlleva cada componente para el aprendiz y el momento de aprendizaje en relación con los otros componentes (Dóczy & Kormos, 2016; González-Fernández & Schmitt, 2020). En uno de los trabajos el uso de colocaciones se limita al conocimiento productivo del vocabulario y se relaciona con el conocimiento que se requiere para usar las palabras en contexto (Dóczy & Kormos, 2016, p. 81). Además, con la intención de ordenar los componentes según el momento de adquisición en el proceso de aprendizaje, sitúa este tipo de conocimiento no demasiado tarde sino, en tercer lugar (vid. Figura 9), mostrando, así como las colocaciones ocupan un papel destacado e intermedio en la lista de componentes. Debemos apuntar que no son el punto más difícil de la escala, sino que ocupan un lugar intermedio. En otro estudio (González-Fernández & Schmitt, 2020), las

conclusiones son similares (vid. 3.3.2.2, para más información el test). De entre los cuatro componentes medidos en el TCE, cada uno en su versión de comprensión y producción, se presentan dos conclusiones: por un lado, la independencia del conocimiento colocacional con respecto al resto de componentes del conocimiento léxico y la diferencia de posición en la escala implicativa de dificultad en la que el reconocimiento de colocaciones se sitúa en el segundo y la recuperación en el quinto, desde el componente más fácil (uno) al más difícil (ocho).

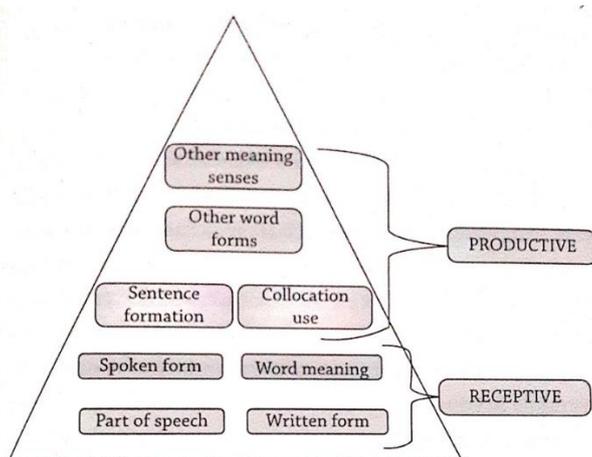


Figura 9. Modelo longitudinal de los aspectos del conocimiento de la palabra (Dóczi & Kormos, 2016)

Como señala Brown (Brown, 2014), tratando las colocaciones como constructo independiente, se consideran las colocaciones elementos léxicos por derecho propio y los nueve aspectos del conocimiento del esquema de Nation (Nation, 2013) podrían aplicarse al conocimiento colocacional. De hecho, proponen la adaptación de los nueve componentes al conocimiento de las combinaciones de palabras (Nation & Webb, 2011). Debemos apuntar, sin embargo, que la adaptación se realiza para las unidades multipalabra, entre las que incluyen las colocaciones (vid. Tabla 9). Los autores afirman que el conocimiento de estas unidades puede abordarse de la misma manera que el conocimiento de las palabras. Debemos destacar que las colocaciones se distancian de otras unidades fraseológicas, que constituyen una unidad léxica, pero este no es el caso de las colocaciones que están constituidas por dos unidades léxicas.

Form	spoken	R	What does the MWU sound like?
		P	How is the MWE pronounced?
	written	R	What does the MWU look like?
		P	How is the MWU written and spelled?
	word parts	R	What words are recognizable in this MWU?
		P	What MWU can be used to express this meaning?
Meaning	Form and meaning	R	What meaning does this MWU signal?
		P	What words are needed to express the meaning?
	Concept and referents	R	What is included in the concept?
		P	What items can the concept refer to?
	associations	R	What other words or MWUs does this make us think of?
		P	What other words or MWEs could we use instead of this one?
Use	Grammatical functions	R	In what patterns does the MWU occur?
		P	In what patterns must we use this MWU?
	collocations	R	What other words, MWUs, or types of MWEs occur with this one?
		P	What words, MWUs, or types of MWUs must we use with this one?
	Constraints on use	R	Where, when, and how often would we expect to meet this MWU?
		P	Where, when, and how often can we use this MWU?

Tabla 8. Componentes que se necesitan dominar para conocer una unidad multipalabra (Nation & Webb, 2011)

Para medir el conocimiento o la competencia colocacional es necesario definirlo como constructo independiente. Esta es la línea que siguen algunos de los trabajos sobre evaluación de colocaciones, por ejemplo el *Test of English Whole Collocations* (Lee Revier, 2009) o COLLEX Y COLLMACH (Gyllstad, 2007b, 2013). Pese a que ambos trabajos definen el constructo, su definición no es completamente aplicable a esta investigación por dos razones: a) no entienden la colocación desde una perspectiva fraseológica, sino combinada; b) describen el constructo de las colocaciones del inglés. En nuestro caso necesitamos una definición del constructo de las colocaciones del español según el enfoque fraseológico y más concretamente desde el marco de la TST (vid. 2.3.1.4).

A continuación, realizaremos una comparativa entre los componentes del conocimiento de una palabra (Nation, 2013) y los componentes del conocimiento de las unidades multipalabra (Nation & Webb, 2011). Aun tratándose de unidades diferentes, son más los componentes que comparten que los que los distinguen. Para el análisis nos centraremos únicamente en las diferencias. No atenderemos, por lo tanto, al significado ya que no hay variaciones, los componentes del conocimiento de la palabra presentes

son exactamente los mismos que los componentes propuestos para el conocimiento de las llamadas unidades multipalabra.

Con relación a la forma, hay una falta de correspondencia en el aspecto morfológico. Mientras para las palabras se habla de las partes de la palabra; en las unidades multipalabra se centra en las palabras ‘completas’ que componen la unidad multipalabra. Esto quiere decir que no tienen en cuenta la morfología de estas palabras. En las locuciones puede tener sentido, ya que son invariables, pero en las colocaciones no. Muchas mantienen flexión e incluso hay casos en los que se permite la derivación. En relación con la flexión tanto podemos decir *hacer la cama* como *cama hecha*; en el caso de la derivación, podemos decir que tenemos una colocación en casos como *deshacer la cama* o *cama deshecha*. En relación con el uso, sí percibimos diferencias. En el caso de las palabras se centra en las posibles combinaciones con esa palabra. En las unidades multipalabra, se limita a las palabras que se combinan con la combinación, punto que compartimos, ya que es posible encadenar colocaciones como *cometer un crimen* y *crimen atroz*, podemos decir *cometió un crimen atroz* o incluso locuciones que sean base de una colocación como en la locución *a las mil maravillas*, *salir a las mil maravillas*.

Según esta comparativa, los rasgos para conocer palabras y colocaciones son casi idénticos y no entrañan grandes diferencias cuando se trata de unidades diferentes. Las colocaciones tienen unas características propias que implican componentes del conocimiento también propios. En la siguiente tabla (vid. Tabla 10) podemos ver la comparación entre los componentes del conocimiento de las palabras y los componentes del conocimiento de las unidades multipalabra, por lo que también incluye a otro tipo de unidades como las locuciones.

			PALABRAS	MULTIWORDS	
Forma	oral	R	¿Cómo suena una palabra / UM?		
		P	¿Cómo se pronuncia una palabra / UM?		
	escrito	R	¿Cómo es se escribe?		
		P	¿Cómo es su ortografía?		
	partes de la palabra	R	¿Qué partes se reconocen en esta palabra?	¿Qué palabras se reconocen en esta UM?	
		P	¿Qué partes de la palabra son necesarias para expresar el significado?	¿Qué palabras se necesitan para expresar el significado?	
Significado	forma y significado	R	¿Qué significado expresa la UM?		
		P	¿Qué UM puede usarse para expresar este significado?		
	concepto y referentes	R	¿Qué se incluye en el concepto?		
		P	¿A qué elementos se refiere el concepto?		
	asociaciones	R	¿Qué palabras (o UM) nos hacen pensar en ésta?		
		P	¿Qué otras palabras (o UM) podrían usarse en lugar de ésta?		
Uso	funciones gramaticales	R	¿En qué patrones puede aparecer la UM?		
		P	¿En qué patrones debemos usar esta UM?		
	colocaciones	R		¿Qué palabras, UM o tipos de UM aparecen con ésta?	
		P		¿Qué palabras, UM o tipos de UM debemos utilizar con ésta?	
	restricciones de uso (registro, frecuencia, etc.)	R		¿Dónde, cuándo y cómo esperamos encontrar UM?	
		P		¿Dónde, cuándo y cómo de : menudo usamos UM?	

Tabla 9. Comparación entre el conocimiento componencial de la palabra (Nation, 2013) y de las unidades multipalabra (Nation & Webb, 2011)

En esta investigación definiremos el conocimiento colocacional siguiendo el enfoque independiente y componencial, teniendo en cuenta que el concepto que se toma de colocación es fraseológico y más concretamente, de la TST. De modo que en el apartado 4.1.2 concretaremos qué entendemos por conocimiento colocacional de una colocación según los criterios apuntados en el 2.3.1. Esto es, la definición del constructo del TCE con el que se describe qué implica conocer una colocación.

En la línea de lo que avanzábamos al inicio del apartado, el conocimiento colocacional se centra en la entidad léxica y la competencia en el sistema desde un punto de vista dimensional (Henriksen, 2013b; Schmitt & Schmitt, 2020). Entre las dimensiones propuestas para describir la competencia de las unidades fraseológicas se encuentra: la frecuencia, la precisión de uso y la rapidez en el procesamiento de uso (Schmitt &

Schmitt, 2020). En el TCE atenderemos a las dos primeras dimensiones. Ya de forma más concreta, se trata explícitamente la competencia colocacional y se presenta como un proceso complejo y acumulativo que demanda exposición y condiciones de consolidación, engloba los siguientes pasos desarrollados tanto desde la producción como desde la comprensión (Henriksen, 2013b): el reconocimiento de las colocaciones (señaladas en el *input*), el significado y función de la combinación, las restricciones de uso de las colocaciones, la elección entre diferentes opciones de colocaciones y la fluidez colocacional para acceder a las colocaciones con facilidad. En este caso la propuesta se engloba más en la tendencia de desarrollo, es decir, en la que gradúa el conocimiento.

2.3.3. Relación entre colocaciones y frecuencia

La relación entre las colocaciones y la frecuencia es más compleja que entre las colocaciones y el léxico. Sin embargo, en las últimas décadas se recogen multitud de trabajos que tratan este asunto gracias al uso de los corpus y el análisis cuantitativo (Cantos-Gómez & Almela-Sánchez, 2018; Cuberos Vicente et al., 2024). Una parte importante de los estudios que relacionan la frecuencia y las colocaciones proceden de la lingüística de corpus y no de la evaluación de lenguas, aunque los principios derivados son la base de la enseñanza-aprendizaje de colocaciones y, por lo tanto, también son punto de partida para la evaluación de lenguas.

Prevalecen los estudios con una perspectiva estática a los longitudinales, en los que el lenguaje meta suele ser el inglés, la competencia analizada se suele situar en el nivel B1 y, principalmente, en el medio escrito lo que limita las conclusiones extraídas (Granger, 2018). Aun así, de acuerdo con esta autora, se pueden extraer una serie de conclusiones limitadas por la falta de criterios uniformes. Las principales aportaciones se centran en los siguientes puntos: la identificación de las colocaciones, que aborda desde el punto de vista de la densidad y de la diversidad, con diferente grado de discrepancia; la precisión y adecuación, con relación al tipo y número de errores; la influencia de la L1; y la relación con la competencia, que parece influida por el tipo de colocación según la categoría gramatical de sus componentes. En cuanto a la densidad, unos estudios se inclinan porque los aprendices utilizan más colocaciones que los nativos y otros porque utilizan el mismo número; con relación a la diversidad, se ha detectado que predomina

la carencia, es decir, se repiten las mismas colocaciones, pero no se utiliza un número variado de ellas. Todos estos principios comparten una limitación en su investigación previa, la falta de unidad de los criterios, lo que hace que sea difícil comparar y establecer correlaciones entre los diferentes trabajos. La evaluación de colocaciones se nutre de estos principios y los complementa. Sin embargo, el trasvase de información de un ámbito a otro no es inmediato por la diferente conceptualización de las combinaciones de palabras. En apartados siguientes nos centraremos en la evaluación de las colocaciones (mediante test) y en las conclusiones a las que ya se ha llegado.

A continuación, abordaremos la relación de la frecuencia y la sección de la muestra que se va a evaluar y la relación de la frecuencia y la asignación de nivel. Nos son los únicos campos en los que la frecuencia tiene un efecto en la relación colocaciones-aprendizaje, también se está trabajando en la incidencia del número de encuentros del aprendiz con la combinación y como las colocaciones y otras combinaciones de palabras, se leen de forma más rápida que las combinaciones libres (Godfroid, 2020; Pellicer-Sánchez et al., 2022) u otras técnicas de origen psicolingüístico (Gyllstad, 2020, pp. 398-399) como la lectura autoguiada, *priming* o rastreo ocular, en las que no entraremos por exceder los límites de esta investigación.

2.3.3.1. *Selección de la muestra*

Tal y como apuntábamos en el apartado dedicado a la selección del léxico, esta puede tener dos propósitos diferentes. Desde un punto de vista pedagógico, existen la necesidad de elegir qué colocaciones enseñar en un curso o qué colocaciones incluir en un libro de texto. Desde el punto de vista de la evaluación de lenguas, hay que escoger una muestra representativa de la lengua que refleje la competencia colocacional. Para este propósito existen diferentes recursos, entre ellos, las listas de frecuencia, a las que aludiremos de forma genérica como listas de colocaciones frecuentes y a las que dedicaremos esta sección. Este recurso es ampliamente utilizado en la selección del léxico (Nation and Waring 1997; Leech et al., 2001; Nation, 2001; Ellis, 2002), especialmente de palabras, pero su aplicación al ámbito de las colocaciones presenta dudas (González Fernández & Schmitt, 2015). Aun así, hay otras herramientas para la

selección en las que no ahondaremos en este punto por tratarse de herramientas que no se basan en la frecuencia.

Schmitt y González (2015) explicitan las dudas con respecto a la relación, por extensión, que se presupone a las colocaciones y la frecuencia. Esta cuestión está también relacionada con la identificación de la colocación, que ya hemos abordado previamente (vid. 2.3.1.2). Podemos saber la frecuencia de una palabra, pero no es tan fácil establecer la frecuencia de una colocación. En caso de que hayamos aclarado qué entendemos por colocación frecuente, habría que confirmar si las colocaciones frecuentes son las más rentables o las fáciles de aprender. Uno de los métodos más recurrentes consta de tres pasos: la frecuencia de las unidades léxicas (base y colocativo) que componen la combinación, la frecuencia de la combinación y la identificación de la acepción en la que se emplea cada unidad léxica, siempre y cuando alguna de ellas sea polisémica (Koike, 2001, p. 26)

Al igual que ocurre en el léxico, las listas de colocaciones frecuentes son el recurso más útil para la elección de colocaciones, pero en este caso, se suman varias dificultades a nivel terminológico y lingüístico, pero sobre todo son dificultades derivadas de la identificación. El problema terminológico y conceptual de este tipo de combinaciones se extiende también a las listas de frecuencias y, por lo tanto, contamos con listas de unidades pluriverbales del inglés (que incluyen combinaciones de palabras entre las que encontramos todo tipo combinaciones fraseológicas) en las que se incorporan o no colocaciones, aunque todas de las que disponemos parten de un enfoque basado en frecuencias (Garnier & Schmitt, 2015; Martínez, 2011). Es decir, no contamos con una lista de colocaciones frecuentes en la que se conceptualice la colocación desde el enfoque fraseológico. En el siguiente esquema podemos ver la distribución de las listas de combinaciones frecuentes atendiendo a dos criterios: la lengua de las combinaciones, la naturaleza de las combinaciones y el propósito de la lista: pedagógico o destinado a la muestra para la evaluación.

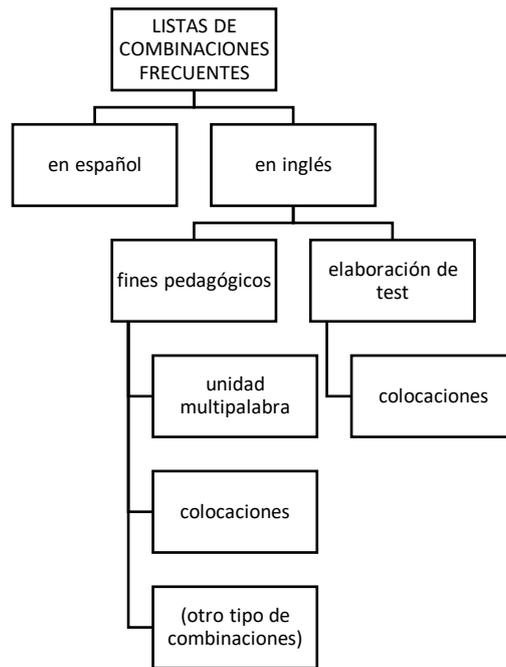


Figura 10. Tipos de listas de combinaciones frecuentes según su finalidad

En la siguiente tabla (vid. Figura 10) podemos ver una recopilación de algunas de las listas de combinaciones frecuentes con fines pedagógicos (Ackermann & Chen, s. f.; Durrant, 2008; Garnier & Schmitt, 2015; Martínez, 2011; R. Martínez & Schmitt, 2012; Shin & Nation, 2007; Simpson-Vlach & Ellis, 2010). Como podemos observar, no solo se incluyen colocaciones, sino que gran parte de las listas incluyen más de un tipo de unidades fraseológicas bajo los términos generales de unidades multipalabra, pluriverbales, etc. Aún, las que únicamente incluyen colocaciones no se podrían adaptar automáticamente ya que la conceptualización de la colocación y en consecuencia el método de identificación, pueden ocasionar problemas de validez y fiabilidad.

FECHA	AUTOR	TIPO DE UNIDAD	PROPÓSITO	OTROS
2008	Shin & Nation	Colocaciones	-	Inglés oral
2009	Durrant	colocaciones	-	Inglés académico
2010	Simpson-Vlach	formula	(nuevos métodos en la investigación fraseológica)	Inglés académico
2011	Ron Martínez	Expresiones formulaicas	Pedagógico	-
2012	Martínez & Schmitt	Expresiones fraseológicas		
2013	Ackerman	ACL colocaciones		Inglés académico
2016	Schmitt	PRHASE LIST Expresiones fraseológicas	-	-

Tabla 10. Recopilación de listas de colocaciones frecuentes

En español no contamos con listas específicas de colocaciones frecuentes. Sí podemos encontrar listas de combinaciones en distintos diccionarios: en formato en papel, tanto en *REDES* (Bosque, 2004) como en el *Práctico* (Bosque, 2006); en formato electrónico, también en el DiCE²⁴, en el que las colocaciones sí tienen asignada frecuencia e incluso están niveladas. Dentro de las listas específicas, encontramos HARTA²⁵ (*Herramienta de ayuda a la redacción de textos académicos*), un recurso que recoge expresiones académicas (colocaciones, fórmulas de varias palabras recurrentes en el discurso académico). No se presenta como lista de palabras frecuentes al uso, sino como una herramienta destinada a los usuarios del lenguaje académico que tratan de elegir la mejor opción en sus textos (Guzzi & Ramos, 2023).

2.3.3.2. *Métodos estadístico para la selección de la muestra*

Las medidas de asociación permiten identificar y medir la frecuencia de coocurrencia de colocaciones dadas, así como cuantificar la *fuerza y fiabilidad* (en inglés conocidos como *strength* y *reliability*) de la colocación; es decir, decidir “cuáles son más colocaciones que las otras”. Para explicar cada una de las medidas seguimos la revisión hecha por Bisht, Dhami, & Tiwari (2006).

La *medida de frecuencia (frequency measure)* busca dos palabras que coocurren de forma habitual juntas. Este es el método más simple e incluye cualquier tipo de combinación de palabras, ya sea colocación léxica, gramatical u otro tipo de combinación sin ningún tipo de relación entre ellas.

La *información mutua (mutual information)* busca una relación más estrecha entre las palabras de la combinación. Se puede definir como una medida de información que superpone la información de dos variables aleatorias (Bouma, 2009).

En la prueba de la hipótesis (*hypothesis testing*) a partir de un logaritmo se consiguen extraer las palabras que ocurren juntas más a menudo de lo que lo hacen por casualidad. Para este propósito se formula una hipótesis nula que consiste en que dos palabras aparecen independientemente en el texto y bajo esa hipótesis nula se calcula la

²⁴ <http://www.dicesp.com/paginas>

²⁵ <http://www.dicesp.com:8083/>

probabilidad de que aparezcan juntas. Se puede probar esta hipótesis nula utilizando varios test: *t-test*, *chi-square test*, *likelihood ratio test*, etc. La mayor sensibilidad de la medida junto con la facilidad tanto para calcular como para entender el cálculo son algunas de las ventajas que se presentan como argumento para su aceptación. Además, esta medida, parece acercarse al enfoque fraseológico, en tanto que podría ser útil en el ámbito de la lexicografía ya que podría ayudar a determinar cuál de las dos palabras que forman la combinación debería ser la entrada del diccionario. Y precisamente este, la perspectiva tomada desde la producción y la restricción, son dos de los rasgos que presenta el enfoque fraseológico. De aquí que podamos decir que incluso dos de las características definitorias que en principio separan los dos enfoques tradicionales estén íntimamente relacionadas.

2.3.3.3. *Asignación de nivel a las colocaciones*

La asignación de nivel consiste en establecer una relación entre los distintos niveles de aprendizaje del MCER (Consejo de Europa, 2002) y las colocaciones, lo que resulta esencial para la evaluación de lenguas, ya que es necesario partir de un constructo ordenado, ya sea mediante un *ranking* de frecuencias o la asignación de un nivel concreto. Todavía no hay un sistema estandarizado, pero sí diferentes propuestas con mayor o menor aceptación. La lexicografía es uno de los ámbitos desde el que más se ha investigado esta cuestión. Diversos diccionarios combinatorios han tratado de nivelar sus entradas siguiendo un criterio basado en frecuencias (Brrok Antle, 2014; García-Salido & Alonso-Ramos, 2019), pese a que su criterio para identificarlas o su conceptualización se encuentra más o menos alejado del enfoque basado en frecuencias. Es decir, pese a concebir las colocaciones como unidades fraseológicas, la frecuencia se presenta como el criterio elegido debido a su objetividad. Usar la frecuencia como único criterio para la asignación del nivel de competencia puede no ser la mejor opción. Uno de los argumentos base de la relación entre frecuencia y léxico es la rentabilidad del léxico frecuente, pero también se debe tener en cuenta que no hay total correspondencia entre frecuencia y rentabilidad (Izquierdo Gil, 2004) ya que hay palabras o colocaciones (y léxico en general) poco frecuente muy rentable y que, por lo tanto, podría modificar el nivel estimado si solo tenemos en cuenta el criterio cuantitativo.

En el ámbito de inglés, el *English Vocabulary Profile*, incluye en las distintas entradas las posibles combinaciones de palabras a las que también asignan nivel, siguiendo el mismo criterio que para las palabras (vid. 2.2.3.2). Combinan diferentes criterios para la nivelación, entre ellos el juicio del hablante nativo, la dispersión o la pertenencia a diferencias temáticas (Capel, 2010, 2012). En el ámbito del español, también contamos con herramientas lexicográficas que han incorporado índice de frecuencia a las combinaciones. Uno de ellos es el DICE, *Diccionario de colocaciones del español* (Alonso Ramos, 2004), en el que se asignó a cada colocación su frecuencia, de modo que las colocaciones se podían ordenar de la más a la menos frecuente (Vincze et al., 2013). En español contamos con una propuesta metodológica para asignar niveles de aprendizaje a las colocaciones (García-Salido & Alonso-Ramos, 2019) y aunque está diseñada para nivelar las colocaciones del DICE se trata de una propuesta generalizable. Utilizan como criterio la frecuencia y como punto de referencia el PCIC, que cuenta con colocaciones niveladas de acuerdo con el MCER. De cada colocación se obtiene la siguiente información: a) la frecuencia de uso, que ya aparece en el DiCE; b) el nivel de aprendizaje, extraído del PCIC. Las colocaciones comprenden desde el nivel B1 hasta el C2. El análisis estadístico muestra que existen una correlación negativa, es decir, a valores bajos en una correlación le corresponden valores altos en la otra. En este punto se decide a partir de qué frecuencia a una colocación dada se le asigna un nivel u otro. Se toma como punto de corte el valor coincidente con el primer cuartil de cada nivel que actúa como límite inferior. Esta propuesta metodológica implica una tendencia a la asignación de niveles altos a las colocaciones. Relativo a uno de los criterios del PCIC por el que el simple hecho de ser *frase hecha* implica la asignación a un nivel alto y choca con el criterio de rentabilidad, por el que combinaciones poco frecuentes son útiles, y deberían aprenderse en los primeros niveles.

En el ámbito del español, varias propuestas centradas en el aprendizaje de ELE se ocupan del nivel de colocaciones. Es el caso del trabajo de (Rojo-Mejuto, 2015) en el que propone una lista de colocaciones niveladas según el PCIC en el que tiene en cuenta los índices de frecuencia de los elementos constituyentes en el CORPES XXI, no de la colocación como unidad. Se trata, en definitiva, de una lista de colocaciones extraídas del *PCIC* (Instituto Cervantes, 2006) que puede ser útil como punto de partida. En otro estudio,

utilizando también el PCIC como fuente principal, se hace un análisis sobre la evaluación (Laya Gómez, 2014). A partir de una serie de colocaciones extraídas del PCIC un grupo de profesores de ELE asignan nivel a dichas colocaciones. Del análisis se extraen dos conclusiones: primero, la falta de coherencia entre profesores y nivel del *Plan*, de hecho, solo coinciden en tres de las cien colocaciones seleccionadas; segundo, la relativa coherencia entre los propios profesores. Por lo tanto, el criterio de autoridad (profesorado especializado) tampoco se podría utilizar como criterio único para la nivelación de colocaciones. Otro estudio (Cuberos Vicente et al., 2024) trata de establecer los criterios para nivelar colocaciones del español a partir del análisis de corpus. En este trabajo se hace evidente el problema de la identificación de colocaciones, pero también la ineficacia de los criterios ligados a las propiedades de este tipo de combinaciones (transparencia, estructura sintáctica o fuerza de asociación) relegándolos frente a los criterios más objetivos de densidad y diversidad.

2.3.4. Aprendizaje de colocaciones

Como componente del conocimiento y competencia léxica, las principales hipótesis sobre el aprendizaje de las palabras se han venido trasladando al de las colocaciones. Las tendencias más asentadas en el aprendizaje del vocabulario son las siguientes: se aprenden antes las palabras más frecuentes que las menos frecuentes y se mantiene la frecuencia alta como factor de utilidad del aprendizaje (Durrant et al., 2022, p. 211; Milton, 2013; Webb & Nation, 2017, p. 38). En principio, no habría razón para que con las colocaciones pasara lo contrario. Sin embargo, sí hay una diferencia: una palabra frecuente es aquella que ocurre más de un número determinado de veces en un corpus, pero no hay consenso sobre qué es una colocación frecuente. Una colocación frecuente podría ser la que tiene sus elementos integrantes frecuentes, las que usan diferentes medidas estadísticas o las que aportan mayor peso a uno de los elementos de la combinación (vid. 2.3.3).

El interés por el aprendizaje de colocaciones se debe en gran medida a las aportaciones llegadas desde el *enfoque léxico* (Lewis, 1993, 1997, 2000), así como desde el enfoque comunicativo. Diversos trabajos se han centrado en el proceso de enseñanza-aprendizaje de colocaciones, tratado tanto desde la lingüística aplicada en la enseñanza

del inglés como L2 (Boers & Webb, 2018; Gyllstad & Wolter, 2016; Pellicer-Sánchez, 2015; Peters, 2014), pero también desde la enseñanza del español como lengua extranjera (Ferrando, 2018; Higuera García, 2018; Higuera, 1997). En ninguno de los trabajos referenciados el concepto de colocación coincide con el que presentamos en este trabajo. De hecho, en ocasiones, partiremos de trabajos que apuntan a unidades fraseológicas en general a las que se alude usando una terminología variada.

Los documentos de referencia, MCER y PCIC, no solo incorporan las palabras como *unidades del léxico*, sino que incluyen una serie más amplia de unidades, que sin usar el término *colocación*, sí hacen referencia al concepto. El tratamiento que estos documentos hacen de las colocaciones no es ni sistemático ni riguroso (Pérez-Serrano, 2015), entre otras razones, por la confusión terminológica. Además, se prioriza el estudio de este tipo de combinaciones en los niveles más altos, obviando su función y papel en los niveles inferiores. En el MCER, se presentan una serie de escalas niveladas que recogen los estándares sobre la riqueza y el dominio del vocabulario (vid. Tabla 8) que recogen algunas directrices sobre el reconocimiento y utilización de los elementos del léxico. En el PCIC, el carácter abierto de la lista de nociones dota de cierta libertad a la incorporación de nuevas combinaciones, pero las incluidas no vienen avaladas por unos criterios de selección ni claros ni específicos.

RIQUEZA DEL VOCABULARIO		DOMINIO DEL VOCABULARIO	
B1	Tiene suficiente vocabulario para expresarse con algún circunloquio sobre la mayoría de los temas pertinentes para su vida diaria, como, por ejemplo, familia, aficiones e intereses, trabajo, viajes y hechos de actualidad.	B1	Manifiesta un buen dominio del vocabulario elemental, pero todavía comete errores importantes cuando expresa pensamientos más complejos, o cuando aborda temas y situaciones poco frecuentes
B2	Dispone de un amplio vocabulario sobre asuntos relativos a su especialidad y sobre temas más generales. Varía la formulación para evitar la frecuente repetición, pero las deficiencias léxicas todavía pueden provocar vacilación y circunloquios.	B2	Domina un limitado repertorio relativo a necesidades concretas y cotidianas.

Tabla 11. Ejemplo de escala del MCER (Consejo de Europa, 2002, p. 109)

La idea general de que las colocaciones son uno de los puntos débiles de los aprendices de L2 (Levitzky-Aviad & Laufer, 2013; Nesselhauf, 2003, 2005b) y que el aprendizaje de las colocaciones es el reto que supone para quienes aprenden una segunda lengua (Laufer, 1997) o que las construcciones fraseológicas son problemáticas para los aprendices va perdiendo peso. En el ámbito del español, se concluye que la diferencia entre nativos y aprendices no radica en el número de colocaciones utilizadas,

que es muy similar entre ambos grupos, sino en otros factores que se pueden resumir en las siguientes apreciaciones sobre las colocaciones utilizadas por los aprendices: a) repiten muchas colocaciones, b) predominan colocaciones con constituyentes muy frecuentes y c) algunas de las combinaciones utilizadas, donde debería usarse una colocación, son erróneas, es decir, en la densidad y diversidad (Cuberos Vicente et al., 2024; Vincze et al., 2016). La percepción negativa del uso de las colocaciones por parte de los aprendices y la idea de que suponen un reto suele provenir de los estudios de corpus que, obviamente, miden la producción. Los resultados obtenidos, sin embargo, a partir de test suelen ser más optimistas, entre otras cosas porque muchos de ellos miden la comprensión o en algún caso la producción, pero desde el reconocimiento. Esto ya nos adelanta una de las posibles claves del aprendizaje de colocaciones: supone un reto su dominio en un nivel de producción, pero en ningún punto su comprensión.

La elección de un enfoque para la definición de la colocación condiciona tanto el procesamiento y adquisición ya que están fuertemente afectados por su conceptualización (Gyllstad & Wolter, 2016). El conocimiento sobre los factores que condicionan la carga de aprendizaje de las colocaciones no es muy abundante y deriva, principalmente, de los estudios de corpus (Laufer & Waldman, 2011; Nesselhauf, 2003; Wolter & Gyllstad, 2011, 2013). La mayor parte de los trabajos que encontramos hasta el momento atienden algunos factores determinantes como la *frecuencia* (Wolter & Gyllstad, 2013) y la *congruencia* (Gyllstad & Wolter, 2016). La congruencia se puede definir como el grado de coincidencia entre la L1 y la L2. De hecho, han demostrado que los errores más comunes responden a colocaciones no congruentes (Laufer & Waldman, 2011) pero también al tipo de colocación según su clase de palabra, sobre todo las que siguen el esquema verbo + nombre, o la longitud de las palabras (Peters, 2016). Sin embargo, no tendremos en cuenta en esta investigación ni la frecuencia ni la congruencia con factores de la dificultad o facilidad de los aprendices ya que la muestra seleccionada no tiene en cuenta ni la frecuencia ni los cognados (vid. 4.5.3). Otros se centran en los factores que facilitan la superación de la barrera del aprendizaje de las colocaciones y que incluyen las reglas gramaticales, el carácter composicional y arbitrario, el conocimiento del mundo, la *memorabilidad* o el significado general de la combinación (Webb & Nation, 2017, pp. 32-33). Los últimos estudios también apuntan a que ninguno

de los factores anteriores condicionarían el reto de su dominio sino la falta de exposición explícita durante el proceso de aprendizaje (Pellicer-Sánchez, 2020).

En cuanto al modelo de procesamiento de colocaciones²⁶, los trabajos son limitados (Gyllstad & Wolter, 2016; Macis & Schmitt, 2017; Martinez, 2013; Sonbul & Schmitt, 2013a; Sonbul & Siyanova-Chanturia, 2019; Wolter & Gyllstad, 2011) e íntimamente relacionados con aquellos que pretenden describir la carga de aprendizaje. Sin embargo, estos están encaminados hacia el análisis de los factores que determinan o condicionan el procesamiento de aprendizaje de colocaciones, lo que los acerca lógicamente a la mayor o menor dificultad de aprendizaje, y los separa de la descripción del proceso de aprendizaje en sí, que se relacionaría desde la psicología lingüística o la lingüística cognitiva. No hay un modelo de adquisición de las colocaciones que describa de un modo jerárquico los factores que afectan a la mayor o menor dificultad de las colocaciones. Este vacío únicamente se llena con trabajos que confirmen las hipótesis que ya han sido confirmadas en el ámbito del léxico y otras propias del enfoque adoptado para definir este tipo de unidades.

2.4. Conclusión

En este capítulo hemos aclarado alguno de los conceptos clave necesarios para elaborar desde el inicio un test de diagnóstico de vocabulario. Principalmente conceptos que serán imprescindibles tanto para entender la estructura del TCE como para la interpretación de resultados: el de sofisticación como alternativa a los de extensión y profundidad, así como el de reconocimiento y recuperación para complementar los de producción y comprensión. Sobre todo, hemos establecido una diferencia clara entre el conocimiento y la competencia léxica, por un lado, y entre el conocimiento colocacional y la competencia colocacional, por el otro.

De forma paralela, hemos tratado diferentes cuestiones relativas al léxico en general y a las colocaciones en particular, comprobando que muchas de las cuestiones

²⁶ Con *modelo de procesamiento* nos situamos en la línea de las *estrategias cognitivas* que podemos definir como las actividades y procesos mentales que los aprendientes realizan de manera consciente o inconsciente; con ellas mejoran la comprensión del lenguaje, su asimilación, su almacenamiento en la memoria, su recuperación y su posterior utilización (Martín Peris, 2008).

son aplicables. Nos hemos detenido en dos aspectos: la definición del conocimiento, que nos permitirá atender a los requisitos de los test de diagnóstico, y el papel de la frecuencia en relación con la evaluación y aprendizaje, lo que nos permitirá atender a los requisitos del test de diagnóstico. La comparativa entre los avances en el aprendizaje del léxico y su adaptación al aprendizaje de colocaciones nos permite detectar como, en lo relacionado con el conocimiento desde una perspectiva componencial, no hay tantas diferencias (forma, significado, uso), pero sí se hacen más evidentes en lo relacionado con la frecuencia (como criterios para identificar), cuestión que ha obstaculizado en gran medida los avances en torno al aprendizaje de las colocaciones.

3. LA EVALUACIÓN DE LENGUAS: MEDIR EL CONOCIMIENTO LÉXICO Y EL CONOCIMIENTO COLOCACIONAL

La elaboración del TCE debe ser entendido desde distintas ramas de la lingüística aplicada, especialmente desde la evaluación de lenguas. Trataremos algunos aspectos teóricos generales imprescindibles, nos acercaremos a la explicación de algunos de los métodos estadísticos que posteriormente utilizaremos para el análisis de los datos y nos centraremos, sobre todo, en la evaluación del léxico y en la evaluación de colocaciones. Tanto en la evaluación del léxico como, especialmente, en la evaluación de colocaciones, los fundamentos teóricos están basados en estudios de inglés como lengua extranjera, ya que es sobre la lengua que más se ha estudiado. Dedicaremos especial atención a la revisión de una selección de test de léxico y, de forma más exhaustiva, a una selección de test que miden la competencia colocacional.

3.1. La evaluación de lenguas

La *evaluación de lenguas (Language Testing)* se reconoce en el momento actual como un campo de la *Lingüística Aplicada*. Sin embargo, no siempre ha sido así y, como apunta McNamara (2008), hasta hace no muchos años no se incluía en los manuales de *Lingüística Aplicada* o aparecía como un apartado más de la enseñanza de lenguas. Desde la aparición de asociaciones específicas, como *International Language Testing Association*²⁷ (ILTA) o *European Association for Language Testing*²⁸ (EALTA), y la publicación de revistas especializadas como *Language Testing*²⁹, *Assessing Writing*³⁰ o *Language Assessment Quarterly*³¹, ocupa un lugar destacado en el ámbito de la *Lingüística Aplicada*.

Explicar qué es la *evaluación* y en qué consiste la *evaluación de lenguas* es un aspecto demasiado amplio para nuestros objetivos e intereses, pero sí necesitamos establecer algunos conceptos e ideas básicos que nos permitan construir de forma eficaz

²⁷ <http://www.iltaonline.com/index.php/enUS/>

²⁸ <http://www.ealta.eu.org/>

²⁹ <http://journals.sagepub.com/home/ltj>

³⁰ <https://www.journals.elsevier.com/assessing-writing>

³¹ <http://www.tandfonline.com/toc/hlaq20/current>

el TCE. La *evaluación* se puede definir como la acción de recoger información para tomar decisiones (Bachman & Palmer, 2010b). Esta afirmación se ha aplicado a la *evaluación de lenguas* de forma que se concreta el tipo de información recogida y de decisiones tomadas. Así, se entiende *la evaluación de lenguas* como la acción de recoger información sobre la competencia lingüística de un individuo o, en general o en algún aspecto concreto, para tomar decisiones sobre su aprendizaje (Bachman, 1990). Pese a que existe cierta confusión sobre el uso de los términos ingleses *test*, *measurement* y *evaluation*³² y se propusieron definiciones estrictas para cada uno de ellos, en la actualidad, parece que la diferencia se diluye al compartir un mismo objetivo, recoger información.

La herramienta clave de la evaluación de lenguas es el test en sus diferentes formas. Su propósito responde a dos acciones: recoger información y tomar decisiones. La mayor parte de la bibliografía se centró durante años en el estudio de la primera de estas acciones, la recogida de información respondiendo a preguntas como qué información se recoge o cómo se recoge la información requerida. En los últimos años, se ha reflexionado más sobre la toma de decisiones, hasta convertirse en uno de los aspectos clave de la evaluación (Kane, 2017; Schmitt, Nation, et al., 2020). Así que la pregunta a la que se trata de dar respuesta en la actualidad es cómo recoger información de forma sistemática para tomar decisiones. En el TCE se recoge información sobre la competencia colocacional de los aprendices de español como lengua extranjera para detectar las debilidades y proponer soluciones que faciliten el proceso de aprendizaje.

3.1.1. Clasificación de los test

Los test se pueden clasificar atendiendo a varios criterios como el propósito de la evaluación, el marco de referencia tomado o tomando los agentes de la evaluación como referencia (Martín Peris, 2008). El propósito está vinculado con el ámbito de uso, ya sea a nivel de aula, de una institución, etc.; el marco de referencia, a los fundamentos teóricos que lo definen; y los agentes de evaluación, al tipo de participantes. Pese a la cantidad de

³² Se pretende, con esta propuesta, diferenciar de forma sistemática tres términos similares: se define el término *assessment* como *evaluación*; el término *measurement* como *evaluación* o *medición*; y el concepto de *test* (Bachman, 1990). Otros autores diferencian únicamente entre *assessment* y *testing* (Rodríguez-Sánchez, 2022)

tipos de test de lengua que nos podemos encontrar, no existe una taxonomía estandarizada que recoja todas las variantes. En el esquema recogemos algunos de los tipos de test más habituales atendiendo a los diferentes criterios (vid. Figura 11).

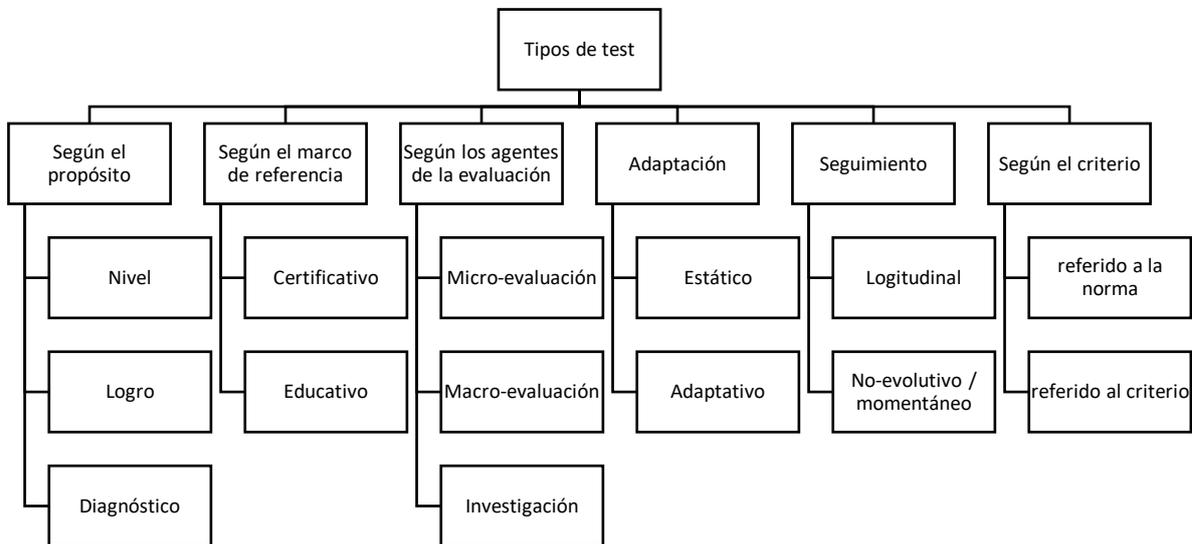


Figura 11. Clasificación de los test atendiendo a diferentes criterios

Según el propósito para el que se utiliza el test, los podemos clasificar en test de nivel, test de logro y test de diagnóstico. Se entiende por *test de nivel* o de *competencia* el que busca asignar un lugar en una escala competencial, muy a menudo un nivel de MCER (A1, A2, B1, B2 C1 o C2). Los *test de logro* se enmarcan en el entorno académico y buscan evidenciar lo que los aprendices han aprendido previamente en curso. Finalmente, los *test de diagnóstico* buscan detectar las carencias de los participantes. Nos centraremos en los últimos, por tratarse del tipo de test que se va a diseñar en esta investigación según los propósitos que avanzamos en la introducción. Aunque este es el principal propósito, el TCE también nos va a permitir asignar un nivel de competencia a los participantes.

Según la función del tipo de decisiones que se vayan a tomar podemos hablar también de diferentes tipos de test. En primer lugar, si las decisiones repercuten directamente sobre (la vida de) de las personas evaluadas estamos ante *test certficativos (proficiency test)*; la evaluación se hace para comprobar si la habilidad lingüística es adecuada para unas necesidades estándar, por ejemplo, alcanzar un nivel de lengua. En segundo lugar, si las decisiones repercuten sobre el proceso de enseñanza-aprendizaje,

hablamos de *test educativo* (*educational test*); el objeto evaluado suele haber sido enseñado / aprendido previamente y suele llevarse a cabo por profesores. Mientras que el primero se centra en los resultados del aprendizaje, el segundo se centra en el proceso de aprendizaje, en el que se situamos el TCE.

Según los agentes de la evaluación, individuos, los programas metodológicos o la investigación científica, las decisiones afectan a un ámbito o destinatario concreto (Bachman & Palmer, 2010b). Este trabajo se inscribe en el último de los grupos: la evaluación para la investigación lingüística. Y, en el ámbito de la investigación lingüística, es posible decidir sobre unas nuevas preguntas de investigación o metodología para futuras investigaciones, cambiar nuestra visión de un particular fenómeno lingüístico o modificar nuestra forma de entender o explicación de un particular fenómeno lingüístico (Bachman & Palmer, 2010b, p. 25). En nuestro caso, el test permitirá dar una explicación del conocimiento y competencia colocacional de aprendices de español para mejorar el aprendizaje de estas unidades del léxico.

Frente a los *test estáticos* (o no adaptativos) están los *test adaptativos* informatizados. Un test estático se caracteriza por tener una estructura fija en la que todos los participantes contestan al mismo número de preguntas, resultando algunas demasiado sencillas y otras demasiado complejas. Los test adaptativos, por su parte, permiten flexibilizar el número y tipo de preguntas adaptando la estructura del test a la habilidad que el aprendiz demuestra a medida que avanza su participación (Chapelle & Douglas, 2006; Kremmel, 2022). Este nuevo formato permite la mejora de la validez pero para ponerlo en marcha se necesita un algoritmo y un banco de ítems, lo que en muchos casos puede condicionar su aplicación (Tseng, 2022). Aun así, la mayor parte de los test adaptativos se centran en la extensión del vocabulario, salvo excepciones (Mizumoto et al., 2019), que se centra en uno de los componentes del conocimiento de las palabras, las partes de la palabra, que se ocupa de cuestiones morfológicas, sobre todo, la derivación. El TCE es estático y en el apartado 4.4.2 explicamos las razones que nos llevaron a no utilizar un modelo adaptativo.

El penúltimo criterio es el carácter *longitudinal* o *no longitudinal* de los test en relación con la evolución del aprendizaje en un período más o menos largo y no solo en un momento puntual. En el ámbito de la enseñanza-aprendizaje de lenguas es habitual,

pero, sin embargo, en la del vocabulario no tanto. Todavía hay pocos trabajos que aborden la evaluación del vocabulario desde un enfoque longitudinal, aunque ofrecerían una importante contribución a este campo (Pellicer-Sánchez & Schmitt, 2012). Estos test suelen medir el progreso del aprendizaje de vocabulario analizando el efecto del vocabulario incidental o intencional, alguna medida metodológica, etc. Se caracterizan por realizar un pre-test y un post-test para comparar los resultados previos con los finales y así poder extraer conclusiones (Kremmel, 2020)

Desde una perspectiva estadística, una de las clasificaciones más relevantes por su implicación en el desarrollo del test, ya que condiciona todo el proceso de la corrección y posterior calificación, es la aplicación o no de un criterio o de unos estándares. Así, los test pueden ser *referidos a la norma* (TRN) o *referidos al criterio* (TRC). En la siguiente tabla (vid. Figura 12) se pueden ver cómo los test varían según el objeto evaluado, el objetivo de evaluación y el ámbito de aplicación, lo que condiciona el análisis de la fiabilidad del test (Barrios & Coscolluela, 2013). El TCE es un test referido a criterio en el que, por lo tanto, el proceso de estandarización juega un papel crucial en el desarrollo del instrumento de evaluación.

	TRN	TRC
Objeto que se evalúa	Mide una variable psicológica o rasgo	Mide un conjunto de conocimientos o competencias (criterio)
Análisis de la fiabilidad	A partir de las diferencias entre los sujetos	A partir del dominio que presenta el individuo sobre el tema
Objetivo de evaluación	Posición relativa de un individuo respecto al resto del grupo	Grado de conocimiento que presentan los individuos en el dominio
Ámbito de aplicación	Personalidad, actitudes, etc.	Educación, ámbito laboral, evaluación de programas, etc.

Figura 12. Diferencia entre test referidos a la norma (TRN) y test referidos al criterio (TRC) (Barrios & Coscolluela, 2013)

Tener claro el tipo de test que vamos a crear nos permitirá tomar decisiones adecuadas al aplicar los procedimientos necesarios en la elaboración del test de colocaciones del español. Por lo tanto, el TCE es un test de diagnóstico, perteneciente al

ámbito educativo, diseñado para la investigación, no adaptativo ni evolutivo y referido a criterio. Sin duda, el tipo de test que más condicionará el proceso de construcción de este será la diagnosis, por lo que dedicaremos un breve espacio a explicar en qué consiste y cómo condicionará el nuevo instrumento de evaluación.

El *test de diagnóstico* sirve para hacer evidente qué es lo que va mal en un proceso de aprendizaje o aquello que podemos considerar defectuoso, insuficiente o precario en una competencia particular, es decir, inferir los puntos fuertes y débiles de la habilidad de interés y utilizarlos para mejorar el proceso de aprendizaje (Jang, 2017). En el área de la evaluación de lenguas se conoce como *evaluación lingüística de diagnóstico* (*diagnostic language assessment*) desde la que Lee (2015) define la evaluación de diagnóstico del modo siguiente: “[...] the processes of identifying test-takers (or learners’) weakness, as well as their strengths, in a targeted domain of linguistic and communicative competence and providing specific diagnostic dieback and (guidance for) remedial learning”. Es así como se incluyen tanto las fortalezas como las debilidades del aprendizaje bajo el paraguas del test de diagnóstico, pese a que normalmente se identificaba este tipo de test únicamente con las debilidades, a la vez que se incluye la generalización como propósito a alcanzar.

Para describir el test de diagnóstico debemos hacerlo a partir de sus características y componentes. Por un lado, en este tipo de test se necesita conceptualizar no solo los objetivos sino también el concepto de *normalidad*³³, esto es, el estado del conocimiento de la entidad evaluada, en el caso del TCE, la colocación, asociado a las expectativas (conocimiento de un hablante nativo). Asimismo, la *estandarización* del constructo que se suele dividir en subcomponentes que, a su vez, se convierten en indicadores de la mayor o menor competencia de la habilidad evaluada, se trata de una concepción fragmentada. Esta representación multicomponencial del constructo permite en el diseño prever las fortalezas y debilidades y en el análisis de los resultados diferenciarlos. El proceso cuenta con tres fases: diagnosis, retroalimentación y reparación del proceso de aprendizaje (Jang, 2017) que se definen por sus objetivos y

³³ Asociado al concepto de *normalidad* está el de *discrepancia vertical* (Little, 2006) que alude a distancia existente entre la actuación, entendida aquí como muestra de lengua, y lo esperado o, dicho de otro modo, entre la realidad y las expectativas, enfocadas en este caso a la normalidad descrita. Esta idea requiere el desarrollo de teorías y escalas orientadas a la competencia, como las propuestas en el MCER.

el instrumento utilizado, que explicamos en la *Tabla 13* y que nos servirán como guía para el TCE.

	DIAGNOSIS	RETROALIMENTACIÓN	REPARACIÓN DEL PROCESO DE APRENDIZAJE
OBJETIVO	Identificar las debilidades y fortalezas de los participantes. Localizando la raíz de las debilidades	Mejorar el aprendizaje de los participantes en el test.	Facilitar la superación de las debilidades y alcanzado los objetivos de aprendizaje.
INSTRUMENTO	Diseño, desarrollo y validación del TCE.	Publicación de respuestas correctas.	Diseño y desarrollo de actividades de aprendizaje que abordan las debilidades detectadas y sus causas.
	Identificación y definición de los atributos y dimensiones del objeto de medida.		

Tabla 12. Componentes del test de diagnóstico. Adaptación (Jang, 2017).

3.1.2. Principios de la evaluación de lenguas

Los principios de la evaluación de lenguas son los criterios que definen la calidad de un instrumento de evaluación, en este caso el test. Se trata de un aspecto relevante en el ámbito de la evaluación de lenguas, de ahí que también se hayan presentado diferentes propuestas para describirlos. No obstante, en todas ellas prevalece la *validez* como principio central de toda actividad de diseño e investigación en evaluación de lenguas (Fulcher & Davidson, 2017). Pese a esta consideración, se trata de un concepto en continuo cambio y evolución³⁴, del que se han planteado diferentes enfoques definiciones y descripciones (Chapelle, 2012a; Cook et al., 2015; Cumming, 2012a; Green, 2014; Kane, 2013). Esta evolución ha condicionado la visión y relevancia del resto de principios que han ido variando en número y relevancia. Nos centraremos en la *validez basada en argumentos* por tratarse de la versión más actual que supera algunas de las dificultades previas como la multiplicidad tipológica (Cook et al., 2015). Posteriormente, atenderemos a la *fiabilidad* y, para finalizar, repasaremos el resto de principios que se tienen en cuenta desde la evaluación de lenguas sin detenernos en aquellos modelos propios que definen y describen los principios de la evaluación.

3.1.2.1. La *validez basada en argumentos*

Con la intención de crear un TCE válido y apropiado para las interpretaciones y usos propuestos, en esta investigación tomaremos el enfoque de la *validez basada en*

³⁴ Validez de contenido, validez de constructo, validez de criterio hasta validez basada en argumentos.

argumentos (Kane, 2012a) como referente, igual que en otros test actuales de vocabulario (Knoch & Chapelle, 2018; Schmidgall, 2017; González-Fernández & Schmitt, 2020), incluido el modelo que seguimos en este trabajo, CATSS (Aviad-Levitzky et al., 2019). En estudios recientes se defiende la necesidad de adecuar los test de vocabulario a los nuevos criterios de validez que dotan de mayor rigor su desarrollo, porque una parte sustancial de las investigaciones que tienen como objetivo la evaluación del léxico no aplican como norma el *argumento de validez* (Schmitt et al., 2020). Entre las causas, señalan la tendencia natural al continuismo, la no incorporación a la evaluación del léxico de los avances propuestos desde la evaluación de lenguas y, también, la densidad de las propuestas. De hecho, el título de un artículo publicado por algunos de los principales especialistas (Schmitt, Nation y Kremmel) en léxico es “Moving the field of vocabulary assessment forward: The need for more rigorous test development and validation” (Schmitt, Nation, et al., 2020). En él analizan de forma detallada las razones por la que consideran que los test de léxico no son ni rigurosos ni están suficientemente validados.

Desde esta perspectiva, la *validez basada en argumentos*, se deja de priorizar el instrumento de evaluación o la habilidad evaluada para centrarse en la validación de la interpretación de los resultados y, sobre todo, en la validación de las decisiones que se toman a partir de las interpretaciones realizadas (Chapelle et al., 2010). Es decir, con este enfoque lo que se valida son los resultados a partir de una serie de argumentos teóricos justificados estadísticamente, mientras que, en enfoques anteriores sobre la validez, el objeto de la validación era el constructo, los resultados, etc. Así, la validez basada en argumentos se puede describir como el grado en el que la evidencia y la teoría sustentan la interpretación de los resultados del test para el propósito del mismo (ALTE, 2011). De este modo, la validez basada en argumentos deja de entenderse como una propiedad del test para concebirse como una propiedad de los resultados y usos del test, lo que supone un cambio con respecto a los enfoques anteriores ya que amplía el espacio para la mejora de la calidad de la evaluación.

El nuevo marco de la validez tiene como referentes teóricos, por un lado, el *Enfoque basado en argumentos (Argument Based Approach)*, de Kane (2010, 2012a), que seguiremos en esta investigación; y, por el otro, el *Argumento de uso de la evaluación (Assessment Use Argument)*, de Bachman & Palmer (2010b). Se trata de dos propuestas

muy similares que comparten más puntos de los que las separan. El *enfoque basado en argumentos* presenta dos requisitos: con el primer argumento, se exponen las explicaciones de orden teórico que justifican las decisiones que se van tomando; con el segundo, se evalúa la credibilidad de estas explicaciones (Kane, 2017). Ambas propuestas presentan dos argumentos de los que uno sustenta teóricamente al otro; el *argumento interpretativo* ofrece el apoyo teórico necesario al *argumento de validez*. Entre las diferencias más evidentes está la terminológica, que podemos ver en la *Tabla 14* con la propuesta de traducción, por la que cada uno de los argumentos recibe un nombre diferente.

	EBA / ABA (Kane, 2012b)		AUE/ AUA (Bachman, 2005)	
	Traducción	<i>Terminología original</i>	Traducción	<i>Terminología original</i>
Nombre del enfoque	ENFOQUE BASADO EN ARGUMENTOS	<i>Argument Based Approach</i>	ARGUMENTO DE USO DE LA EVALUACIÓN	<i>Assessment Use Argument</i>
Primer argumento	Argumento interpretativo	<i>Interpretative argument</i>	Argumento de utilización de la evaluación	<i>Assessment utilization argument</i>
Segundo argumento	Argumento de validez	<i>Validity argument</i>	Argumento de validez de la evaluación	<i>Assessment validity argument</i>

Tabla 13. Comparación terminológica entre el ABA y el AUA.

Por un lado, el *argumento interpretativo* consiste en la especificación de las interpretaciones propuestas y los usos de los resultados y, al hacerlo, especifica las aclaraciones o explicaciones detalladamente. Por el otro, el *argumento de validez* proporciona una evaluación general de las interpretaciones y usos propuestos en el argumento interpretativo. De modo que, uno no es independiente del otro, sino que se complementan y se desenvuelve en dos pasos que se suelen solapar: el *paso de desarrollo* y el *paso de valoración*. De forma sencilla, el *argumento de validez* equivale a un *argumento interpretativo* en el que se ha proporcionado el respaldo teórico para constatar las *suposiciones* (Chapelle et al., 2010).

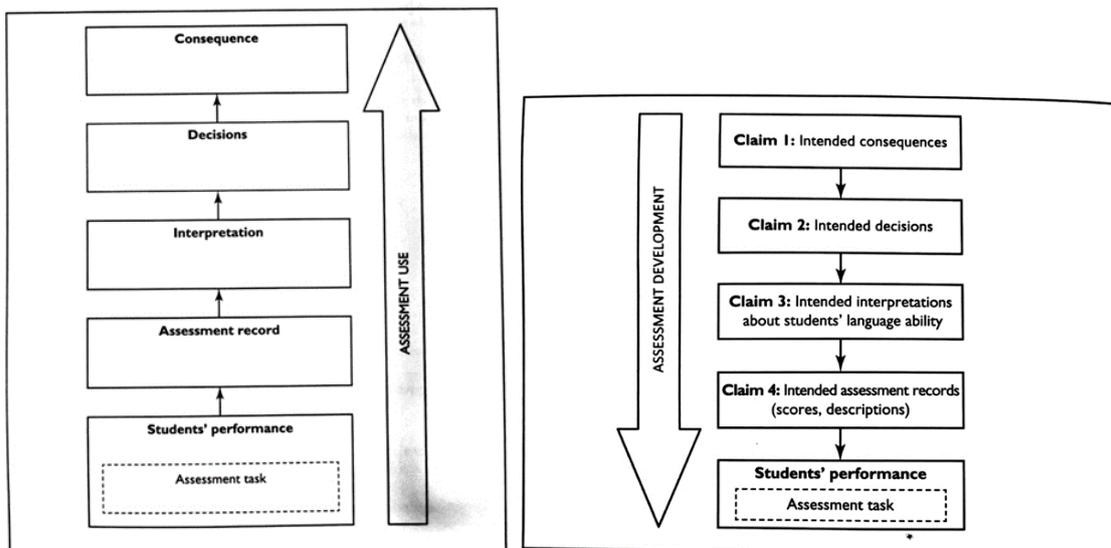


Figura 13. Argumento de uso de la evaluación (Schmidgall, 2017)

El proceso de construcción y definición de estos argumentos es bastante complejo y consta de varios pasos que pasamos a explicar. El argumento de validez se construye a partir de una serie de *inferencias* que relacionan la *observación* con la *explicación*. Estas inferencias cuentan con una *justificación* (teórica) previa y dan lugar a una *suposición o argumento*. A las que es posible añadir una *refutación*. El número de inferencias suele ser cuatro (*evaluación, generalización, explicación y utilización*) (Bachman & Palmer, 2010a; Green, 2014; Kane, 2013) aunque algunas propuestas aplicadas, como es la de TOELF, las amplían a seis (*descripción del dominio, evaluación, generalización, explicación, extrapolación y utilización*) (Chapelle et al., 2010) y que permiten alcanzar el objetivo propuesto, la validación de los resultados del test. En la validación del TCE atenderemos a las cuatro inferencias estándar propuestas por Kane, que explicamos a continuación.

- La *evaluación* une las tareas de los test con la habilidad evaluada a través de los resultados obtenidos. En esta inferencia, el resultado debe ser un reflejo justo de la actuación mostrada por el participante en las respuestas dadas en la tarea de evaluación. Las refutaciones a esta proposición oscilan entre la sobrestimación y el error en la calificación.
- La *generalización* conecta los test con las especificaciones. Se infiere si los resultados son un claro reflejo de las verdaderas habilidades de la persona participante en el área lingüística evaluada. Una posible refutación sería un resultado muy diferente en una prueba paralela que mida la misma competencia o habilidad lingüística.

- La *explicación* vincula las especificaciones con el diseño. Con la inferencia explicativa se busca reflejar la coherencia teórica que puede explicar la actuación de la persona participante. Así, la desacreditación de esa teoría podría ser la refutación de esta proposición.
- La *utilización* conecta el diseño de la evaluación con las decisiones que se pretenden tomar. Pasar la prueba sin tener los conocimientos demandados sería la refutación de esta proposición.

ENFOQUE BASADO EN ARGUMENTOS		ARGUMENTO DE USO DE LA EVALUACIÓN	
Evaluación		Registro	
Tareas del test	Habilidad evaluada	Actuación	Resultados
Generalización		Interpretaciones	
Test	Especificaciones	Decisiones	Habilidad evaluada
Explicación		Decisiones previstas	
Especificaciones	Diseño	Decisiones	Habilidad evaluada
Utilización		Consecuencias previstas	
Diseño	Decisiones	Diseño	Consecuencias

Tabla 14. Comparación entre inferencias en el Enfoque basado en argumentos y el Argumento de uso de la evaluación

Una de las cuestiones más estudiadas desde la evaluación de lenguas es la definición del *constructo*, incluso se consideró la *validez de constructo* y se presentaron estadios para su definición: teórica, operativa y procesual (Bachman, 1990), que se sigue en el diseño de COLLECH y COLLMACH (Gyllstad, 2007a). En este enfoque no se requiere un constructo *per se* sino una interpretación explícitamente denominada *argumento interpretativo*. Por lo tanto, el marco que previamente sustenta la interpretación de los resultados, en este enfoque se relaciona con el argumento interpretativo. Acogiéndonos a la propuesta que sigue la validación de TOELF, definiremos también el constructo sin dejar de aplicar la validez basada en argumentos (Chapelle et al., 2010) ya que es necesaria para guiar (la fase de) desarrollo del test y para ofrecer una justificación a las interpretaciones previstas de la evaluación (Bachman & Palmer, 2010a, p. 212).

No es suficiente con que el test sea válido, también necesita ser fiable, útil, eficaz, etc. A continuación, dedicaremos un apartado tan solo a la fiabilidad. Pese a la relevancia de la validez como principio, la fiabilidad también ha estado presente en todas y cada de las propuestas que explican los principios de la evaluación. Posteriormente, dedicaremos el último apartado a resumir el resto de propiedades que, a otro nivel, son también esenciales en la evaluación de lenguas.

3.1.2.2. La fiabilidad

La *fiabilidad*³⁵, al igual que la *validez*, no se considera una cualidad del test, sino una propiedad de los resultados del test. Además, es clave en cualquier instrumento de medida y contribuye a la validez en general (ALTE, 2020). Su relevancia en la evaluación no es discutible. De hecho, en la actualidad se trata como una parte importante de la mayoría de los argumentos de la validez (Chapelle, 2012c). El concepto se suele relacionar con dos términos: la *consistencia* y la *estabilidad*. Tradicionalmente la *fiabilidad* se explica acudiendo a la obtención de unos resultados semejantes en la repetición del mismo test a participantes diferentes. Queda un margen para el error que se da por motivos ajenos al test: el día, las circunstancias personales de la persona que realiza el test, los evaluadores u otros factores que están fuera del control de los administradores (ALTE, 2011). Por lo tanto, el concepto clave en la descripción de la fiabilidad es el de *error*. Todo test debe ser fiable, o, de forma más precisa, los resultados del test deben ser fiables. Sin embargo, que un test sea fiable no es sinónimo de que sea válido. Pero sí, al contrario: si el test es válido, ha de ser fiable.

La fiabilidad se valora a partir de coeficientes de fiabilidad que vienen dados por procesos estadísticos. A la hora de calcular la fiabilidad de un test, es preciso conocer el tipo de test con el que estamos trabajando (vid. 3.1.1): en un test basado en la norma los mecanismos para comprobar su fiabilidad son unos, y para un test basado en el criterio los mecanismos son otros (Barrios & Cosculluela, 2013; Shrock & Coscarelli, 2008). El Alpha de Cronbach es la medida estándar para informar sobre la fiabilidad en la evaluación de lenguas en general. Aun así, podría no ser la mejor medida de fiabilidad en los estudios sobre el vocabulario, ya que cada palabra es su propio constructo y conocer una no implica conocer otra. Lo mismo ocurre con las colocaciones. La aplicación de medidas de fiabilidad más sofisticadas podrían ser más adecuadas para los instrumentos de evaluación del vocabulario (González-Fernández, 2021, p. 130): la escala implicación o el modelo ecuativo estructural.

³⁵ Próximo al concepto inglés de *reliability*, traducido aquí como *fiabilidad*, está el de *fairness*, comúnmente utilizado como sinónimo del anterior.

3.1.2.3. Otros principios: utilidad y eficacia

La *validez* y la *fiabilidad* no son los únicos principios o propiedades de la evaluación. Además de estos, se han aplicado otros buscando la *utilidad* (Bachman & Palmer, 2010b), la *eficacia* (Green, 2014) o las *buenas prácticas* de la evaluación (ALTE, 2020). Tanto la utilidad como la eficacia y las buenas prácticas son propuestas que tratan de sistematizar los principios de la evaluación de lenguas. En la *Tabla 15* podemos ver de forma comparada los principios de la evaluación que enumeran las diferentes propuestas. En todas las propuestas aparece la *fiabilidad* y la *validez*, incluso en algún caso, se tienen en cuenta diferentes tipos de *validez* (*de contenido*, *de constructo* y *basada en criterios*). También todas incluyen la *practicabilidad* (o *funcionalidad*) y las *consecuencias* o *impacto* que la evaluación produce en los participantes. Además de estas cuatro, la *utilidad* tiene en cuenta la *autenticidad* y la *interactividad*; mientras que las *buenas prácticas* incorporan la *imparcialidad* y la *calidad del servicio*. En primer lugar, revisaremos cada una de las propuestas y, después, explicaremos brevemente cada una de las propiedades.

UTILIDAD / usefulness (Bachman et al., 2010)	Efectivo y útil (Green, 2014)	Utilidad, buenas prácticas (ALTE, 2020)
Fiabilidad (<i>reliability</i>)	Fiabilidad (<i>reliability</i>)	Validez de contenido (<i>content validity</i>)
Validez de constructo (<i>construct validity</i>)	Validez (<i>validity</i>)	Validez de constructo (<i>construct validity</i>)
Autenticidad (<i>authenticity</i>)	-	Fiabilidad (<i>reliability</i>)
Interactividad (<i>interactiveness</i>)	-	Evidencia basada en criterios (<i>criterion-related evidence</i>)
Impacto (<i>impact</i>)	Consecuencias beneficiosas (<i>beneficial consequences</i>)	Imparcialidad (<i>Fairness</i>)
Practicabilidad (<i>practicality</i>)	Practicabilidad (<i>practicality</i>)	Calidad de servicio (<i>quality of service</i>)
-	-	Funcionalidad (<i>practicality</i>)
-	-	Impacto (<i>impact</i>)

Tabla 15. Comparación entre diferentes propuestas de los principios de la evaluación

La *utilidad* (Bachman et al., 2010), frente al resto de propuestas que recogen de forma global los principios de la evaluación, se centra en el propio instrumento de medida más que los resultados y su interpretación. Muestra las propiedades en una ecuación que presenta la suma de seis propiedades, equiparándolas así en relevancia. La *eficacia* (Green, 2014) restringe el número de propiedades o principios a cuatro, que presenta gráficamente en forma de pirámide, estableciendo una relación desigual entre las propiedades. Por último, las *buenas prácticas* (ALTE, 2020) ponen en el centro la validez basada en argumentos, dejando clara su prevalencia frente al resto de propiedades. En

las siguientes imágenes se puede ver la representación gráfica de cada una de las propuestas.

$$usefulness = reliability + construct\ validity + authenticity + interactiveness + impact + practicality$$

Ecuación 1. Modelo de Utilidad del test (L. Bachman et al., 2010)



Figura 14. Modelo efectivo y útil (Green, 2014)



Figura 15. Modelo de buenas prácticas (ALTE, 2020)

Además de la validez y la fiabilidad, una de las propiedades o principios que comparten todas las propuestas es el *impacto de la evaluación*, sobre todo en lo referente a la interpretación de los resultados. Se dice que el impacto puede ser a nivel *macro*, entonces recae sobre la sociedad y el sistema educativo, o a nivel *micro*, sobre los individuos (ALTE, 2020). Es lo que se denomina *consecuencias beneficiosas* (A. Green, 2014, p. 86), ya que considera que una verdadera evaluación útil es aquella que tiene más beneficios que perjuicios para todos los implicados. Esta propiedad es mucho más relevante en test de certificación o académicos ya que tienen consecuencias directas sobre los participantes. Mucho menos en relación con los test de diagnóstico, como el TCE, cuya repercusión directa sobre los participantes dependen del grado de interés de los participantes. Las decisiones que se vayan a tomar a partir de los test de diagnóstico son a largo plazo (metodologías, normalmente). En este principio se integran, más que en ningún otro, tres componentes: el currículo, el aprendizaje y la evaluación para que el examen sea justo, no sesgado, y ejerzan un efecto positivo y no negativo.

Aun así, estudios recientes tratan este tema en los test de diagnóstico (Kunnan & Jang, 2009). En este caso, la *retroalimentación* o *feedback* es el punto de interés. La

retroalimentación puede ir desde el resultado numérico a la facilitación de las respuestas del test. Además, alguna propuesta más elaborada organiza la retroalimentación en diversos componentes, uno de ellos, el perfil en el que se explicita el grado de competencia alcanzado, así como las debilidades y fortalezas de cada participante. Tanto el impacto como la repercusión de la evaluación son consecuencias de uso de la evaluación, que en muchas ocasiones se ven condicionadas por otra de las propiedades del test, como la *practicabilidad*, que limita las intenciones. Además, este principio también se relaciona con el *efecto Washback*, que se refiere a la importante influencia que los test tienen en la prácticas realizadas en el aula, en particular, los efectos que tienen en la enseñanza y el aprendizaje (Wall, 2012).

La *funcionalidad* o *practicabilidad* es otro de los principios para tener en cuenta (L. Bachman et al., 2010). Diferente en naturaleza a las otras cinco cualidades de la utilidad del test (fiabilidad, validez d constructo, autenticidad, interactividad e impacto) porque mientras las anteriores se centran en los resultados de los test, esta, en el test en sí mismo, en cómo ha sido desarrollado. Se define como la relación entre los recursos necesarios para el diseño, desarrollo, uso del test y los recursos de los que se dispone (ALTE, 2020). Se suele representar a partir de una fórmula en la que dividimos los recursos necesarios entre los disponibles. Si el resultado es igual o superior a uno, podemos hablar de un test funcional y si es inferior no será funcional. Estos recursos pueden ser humanos, materiales o temporales. Esta propiedad plantea la necesidad del número y experiencia necesaria de creadores, administradores, correctores y testantes; las facilidades, equipamiento y materiales que cada agente implicado en el proceso necesita; el calendario y la asignación de tiempo que cada uno de los pasos que se dan en el ciclo; y, por último, el presupuesto (Green, 2014).

La *autenticidad* y la *interactividad* son dos principios que incorpora la propuesta de la *utilidad* (Bachman & Palmer, 2010a). El primero de los principios implica la correspondencia entre la actuación (*performance*) expuesta en el test y el dominio de lengua medido. El segundo, la *interactividad* se puede definir como el alcance y el tipo de implicación que suponen las características individuales del examinador en relación con la prueba. La *autenticidad* y la *interactividad* comparten algunas características: ambas son relativas, es decir, graduales y no exclusivas / excluyentes; para evaluarlas debemos

tener en cuenta las características de los participantes, de las tareas y de los evaluadores; dependen de la finalidad (para una pueden ser altas y para otras bajas); los mínimos dependen de la situación de la evaluación.

Conjuntamente a estas, en el manual de buenas prácticas de ALTE, incluyen otros principios: la *imparcialidad (fairness)*, la *calidad del servicio* o la *evidencia de criterio*. La *imparcialidad* atiende a los antecedentes culturales diversos de los participantes y la búsqueda del trato justo y equitativo en todas las pruebas y a todos los participantes, tomando incluso medidas de flexibilización. Por la naturaleza de este test y las condiciones de administración no será uno de los principios más relevantes en el TCE. Está íntimamente relacionada con el impacto, que en el caso del TCE tampoco recae directamente sobre los participantes, y la fiabilidad; así como con el sesgo (Elder, 2012), otro de los principios que se ha tenido en cuenta en la elaboración del test y que no hemos desarrollado ya que ninguna de las propuestas la tenía directamente en cuenta, pero sobre todo, porque dado el propósito y circunstancias del TCE no es pertinente.

3.1.3. Ciclo y especificaciones

La elaboración de un test no empieza por la adquisición de una serie de propiedades o principios ni tampoco por la construcción de un *argumento de validez*, sino que se trata de un proceso complejo más o menos estandarizado al que se denomina *ciclo* en el que se siguen unos pasos concretos. Para ilustrar este proceso se han dado una serie de explicaciones de carácter metafórico que ayudan a entender el proceso, pero también reflejan la evolución de este en la teoría de la evaluación de lenguas. Se compara la construcción de un test con la de una vivienda (Fulcher, 2010). Cuando un arquitecto diseña edificios y elige los materiales que van a usarse en la construcción sabe para qué se va a usar el edificio, y entonces planifica el diseño conociendo las necesidades específicas de sus supuestos ocupantes. También se compara con la defensa en un juicio ante los tribunales (Bachman & Domböck, 2017). En este caso, el abogado de la defensa presenta una serie de argumentos que justifican la decisión que toma el juez.

El hecho de que se presente el proceso de creación del test a modo de *ciclo* implica que dar un paso no es sinónimo de olvidarlo, sino que se ha de volver a él en

cualquier momento de la fase de construcción, en un continuo proceso de mejora. Desde la evaluación de lenguas existen propuestas más o menos estandarizadas y genéricas aplicables a cualquier evaluación (Fulcher, 2010), sea cual sea su propósito. Sin embargo, también hay propuestas específicas de ciclos para la elaboración de test de léxico en inglés³⁶ (Read & Chapelle, 2001) y en español (López-Mezquita Molina, 2005). Tras una revisión de alguno de los ciclos propuestos, comprobamos que las diferencias entre ellos no son sustanciales y que se asemejan bastante. Las discrepancias residen en el número de pasos, pero no por exclusión o incorporación de pasos novedosos, sino por un ejercicio de síntesis terminológica frente a un mayor desglose o especificación en otros casos. El número de pasos de cada ciclo varía en función de la propuesta, en la *Imagen 8* mostramos uno de ellos. Sin embargo, todos los ciclos recogen las siguientes acciones: definición lo que se va a medir, técnicas para recoger la información relevante, modo de administración y evidencia de que la evaluación es válida (Green & Fulcher, 2021).

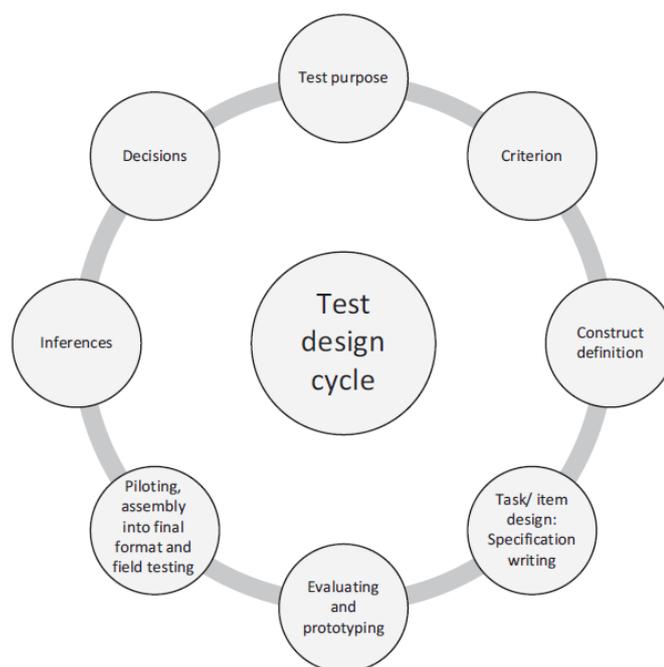


Figura 16. Adaptación (Green & Fulcher, 2021) del ciclo propuesto por Fulcher (2010)

Alguno de los ciclos propuestos no tiene en cuenta la *validez basada en argumentos*, de modo que su adaptación, aunque posible, no puede ser directa. Es el caso, por ejemplo, de la propuesta de ciclo para la evaluación del léxico mencionada

³⁶ Esta propuesta tiene en cuenta el propósito del test, consideraciones de la validez, los factores de mediación, el diseño de test y la validación en relación con lo inferencia, el uso y el impacto del propio test.

previamente (Green & Fulcher, 2021). Otras propuestas ya lo tienen en cuenta (ALTE, 2011; Bachman et al., 2010; Green, 2014). En el caso del TCE, como se trata de un test de diagnóstico, haremos una adaptación de las propuestas estándar, haciendo convivir estas con los componentes y características de los test de diagnóstico.

Las *especificaciones* se corresponden con el documento clave en el que se explican los estándares, y cómo se lleva a cabo la evaluación (Green, 2014). Su grado de concreción puede variar desde el grado más general hasta un alto grado de detalle. Antes de llegar al documento definitivo, las especificaciones van variando y transformándose durante el proceso de diseño, construcción y uso del test. Además, mantienen una importante relación con varios de los principios de la evaluación, sobre todo con la validez.

La estructura del documento de especificaciones es flexible y depende en gran medida del propósito y diseño del test (ALTE, 2011). Unas especificaciones simples podrían contener el propósito del test, el constructo del test, el número y tipo o tipos de ítem que aparecen (Fulcher, 2010) Las *especificaciones* suelen contar con tres componentes: *la descripción del diseño*, *el prototipo o plano de la evaluación* y *la explicación de las tareas* (Green, 2014). La *descripción del diseño* contiene una definición del constructo evaluado, una explicación de la evaluación y una explicación de por qué el enfoque adoptado por el diseñador es el adecuado. En el *prototipo de la evaluación* se debe incluir la extensión del test, el número de ítems, la metodología usada y cuál es el contenido cubierto, el tiempo de la prueba, los materiales usados, etc. La *tabla de especificación* (Miller & Biber, 2015, p. 147) es un concepto práctico que alude a otra forma de presentar la información relativa al contenido de la evaluación y el informe de los resultados. Se refiere a una tabla en la que en el eje horizontal se pongan los conocimientos o habilidades evaluadas y en el eje vertical las tareas usadas para evaluar cada uno de ellos. De modo que en el interior de la tabla se especifican los objetivos que se persiguen formulados a modo de *poder hacer*. Por último, *la explicación de los ítems* o de las tareas en el que se explica qué hay que hacer y cómo hacerlo, indicando cuál es el *input* y cuál es la información esperada. En el apartado 4.3.3 se presentan las especificaciones para el TCE, que se concretan también en el ANEXO III.

3.1.4. Estandarización

La *estandarización* es un proceso estrechamente vinculado a los test de clasificación y de diagnóstico y, sobre todo, en un test referido al criterio (TRC) como el TCE. Es necesario *establecer unos estándares* que nos permitan relacionar los resultados del TCE con diferentes niveles de competencia y diferentes grados del conocimiento colocacional. Aunque uno de los objetivos sea observar cómo es el conocimiento que los aprendices de ELE tienen de las colocaciones y diagnosticar cuáles son los mayores obstáculos en su aprendizaje, también está el propósito de crear un instrumento adecuado que mida la competencia colocacional y el de comprobar si es índice de la competencia léxica y de la competencia lingüística. Así, resulta imprescindible asignar un nivel de competencia colocacional a cada una de las pruebas y a cada uno de los participantes para lo que debemos marcar un punto de corte. Para ello, es imprescindible disponer de los recursos necesarios y aplicar la metodología apropiada de forma correcta, pero sin dejar de lado que los puntos de corte se construyen, no se encuentran (Cizek, 2012).

Así, la *estandarización* es el proceso según el cual se establece uno o más puntos de corte en un test para crear categorías de competencia y clasificar los resultados de los participantes (Cizek, 2012; Kenyon, 2012a). Es indispensable en el desarrollo del argumento de validez (Papageorgiou & Tannenbaum, 2016), por el que todas las decisiones que se tomen a partir de los resultados deben partir de una base teórica que la sostenga. Este proceso se suele dividir en dos fases: a) el *estándar de contenido*, comparación a nivel horizontal de contenidos; y b) el *estándar de actuación*, que consiste en determinar el resultado (o constelación de resultados) en un test dado que mejor se corresponde con el nivel o habilidad requerido (Green, 2017). Uno de los conceptos básicos asociados al establecimiento de estándares es la *discrepancia vertical* que marca la diferencia entre la actuación real de los participantes y la actuación esperada (Jang, 2017), es decir, las expectativas creadas una vez definida la *normalidad*, competencia esperada. Esta discrepancia vertical es la base de la categorización. La *Figura 17* muestra de forma gráfica una escala tipo a partir de la cual unas categorías preestablecidas deben relacionarse con las actuaciones recogidas en el instrumento de evaluación y con los niveles del MCER (Kenyon, 2012). En una dimensión horizontal, el contenido debe

estandarizarse, para lo cual se utilizan diferentes métodos, aunque uno de los más habituales es la reutilización de documentos de referencia. En nuestro caso utilizaremos como documentos de referencia el PCIC y, sobre todo, el MCER, en el que se recoge este como uno de sus usos, tal y como se muestra en la *Tabla 16*.

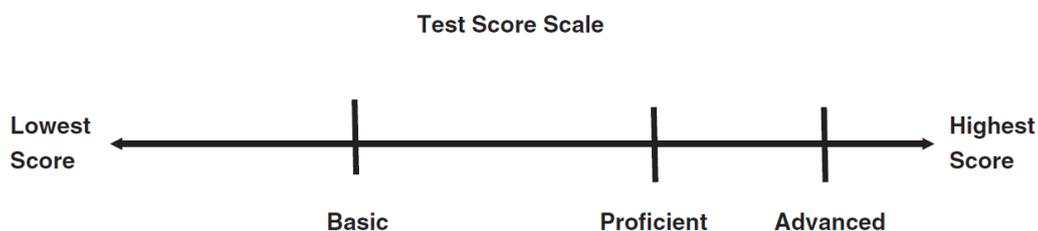


Figura 17. Escala tipo de tres niveles

1	Para especificar el contenido de las pruebas y de los exámenes.	<i>lo que se evalúa</i>
2	Para establecer los criterios con los que se determina la consecución cómo se interpreta la actuación de un objetivo de aprendizaje	<i>cómo se interpreta la actuación</i>
3	Para describir los niveles de dominio lingüístico en pruebas y exámenes cómo se pueden realizar las existentes, permitiendo así realizar comparaciones entre distintos sistemas de certificados:	<i>cómo se pueden realizar las comparaciones</i>

Tabla 16. Usos del MCER

La *estandarización*, como todos los procesos, requiere un progreso a través de una serie de fases. Parte de las propuestas de estandarización están enfocadas a test de gran escala, esto no es el caso del TCE, por lo tanto, es necesario hacer una adaptación. Las fases más habituales son tres: 1) selección del método, 2) diseño para el estudio, 3) preparación de las descripciones de las categorías de desempeño o niveles de competencia. Descartamos otros, presentes en los modelos estandarizados (Cizek, 2012), como la selección del panel, ya que en este caso no hay un equipo encargado de realizar los estándares y el entrenamiento de los panelistas para el uso del método.

Existen varios métodos de estandarización (*Angoff, Ebel, Jaeger, etc.*) y según las características del test en cuestión, se escoge uno u otro método. Teniendo en cuenta las particularidades del TCE adaptaremos el método *Bookmark* (Lewis et al., 2012) que requiere asignar un nivel a todos los ítems del test en orden ascendente de dificultad. Algunas de las condiciones que nos permiten elegirlo son el número de niveles, entre tres y cuatro, y el tipo de los ítems, tanto de selección como de respuesta construida. La solidez de la relación entre los estándares, el significado de los resultados y sus interpretaciones son parte esencial de la validez de argumento. En el método

seleccionado se hacen juicios basados en la revisión de materiales de evaluación y rúbricas de calificación basándose en cuatro factores: a) el tipo de ítems, b) el tiempo disponible para establecer los estándares, c) la experiencia previa del método y d) las percepciones y evidencias sobre la validez del método. La elección del método de estandarización se realiza según diversos criterios. Aunque todos reciben críticas porque implica siempre un juicio de valor, su presencia se hace imprescindible en aquellos test en los que la clasificación de las actuaciones de los participantes es necesaria.

Uno de los criterios habituales a la hora de marcar un punto de corte en test basados en criterio estadístico, seleccionando la muestra a partir de la frecuencia de las palabras seleccionadas en un corpus ordenado en bandas de frecuencia. La respuesta aproximada a preguntas como cuántas palabras hay en una lengua, cuántas palabras conoce un hablante o cuántas palabras necesita un hablante nativo para comunicarse y, sobre todo, cuántas palabras conocen los aprendices de una L2 (Schmitt, Schmitt, et al., 2020) son algunos de los puntos de partida. A estos interrogantes se les dio respuesta en relación con el inglés como L2, pero en otras lenguas como el español no contamos ni con la misma cantidad de estudios ni de datos recopilados. El número de palabras conocido se relaciona con un nivel del *Marco* (Milton, 2013, pp. 180–192), esto es lo que se hace, por ejemplo, con el test de léxico X-LEX (vid. 3.2.2.2), creando escalas del conocimiento léxico y asociándolas en cada uno de los niveles del *Marco* en algún grado con la extensión del vocabulario. Además, mientras en los niveles inferiores, los descriptores se centran en el número de entidades léxicas conocidas, en los niveles superiores, es más habitual que los descriptores se enfoquen a cuestiones relacionadas con el grado de conocimiento de cada entidad léxica. Esta distinción hace referencia a una dualidad imprescindible tanto en el ámbito de la evaluación de lenguas como del aprendizaje de segundas lenguas, la oposición o relación entre extensión y profundidad, (vid.2.1.2).

Tomamos como estándar un modelo de actuación lingüística (de recepción o producción), descrito en términos cualitativos y cuantitativos, que se requiere en un contexto determinado —por ejemplo, el mínimo grado de dominio que se precisa para cursar un programa de estudios en una universidad— y con el que se comparan las actuaciones de los candidatos de un examen determinado. Los estándares suelen

presentarse en escalas de descriptores del rendimiento, que muestran grados del desarrollo de una capacidad o de aspectos parciales de esta (Llorián-González, 2015). En apartados sucesivos concretamos los estándares cualitativos (vid. 4.3.3.2 A) y el estándar cuantitativo (vid. 4.3.3.2 B),

3.1.5. Estadística para la evaluación de lenguas

El análisis estadístico nos permite examinar los resultados obtenidos en la administración del test. La estadística para la evaluación de lenguas, basándose en la psicometría, cuenta con su propio sistema de análisis. Para poder hacer uso de los diferentes modelos estadísticos es necesaria la introducción de los datos en un programa estadístico. Estos programas permiten llevar un control de los resultados, así como todas las operaciones estadísticas que requiere el análisis de los datos. Los más usuales son R (libre), SPSS, JASP o JAMOVI. Es cierto también que, de un modo más simple, se puede usar Excel (la herramienta de Microsoft) para realizar algunos de los análisis estadísticos sencillos.

El primer paso, una vez hemos introducido los datos en un programa estadístico, será la de extraer la *estadística descriptiva*. Esto nos permite descubrir de qué forma se agrupan los resultados, se utilizan las medidas de agrupación o de tendencia central: *moda*, *media* y *mediana* (Bachman & Palmer, 2010), así como su dispersión, tanto la amplitud como la *desviación estándar* y la *varianza*. También utilizaremos *histogramas*, herramienta útil para ilustrar la curva de la distribución basada en la media y la desviación estándar. Una distribución perfectamente *normal* es extremadamente rara en un test de lengua ya que implica simetría y la coincidencia de las medidas centrales (media, moda y mediana). Si el test es demasiado difícil para la población, la curva es positiva, si el test es demasiado fácil la curva es negativa. Otras distribuciones pueden ser las que muestran una *distribución homogénea (peaked distribution)* o la población no sigue una *distribución estándar (flat distribution)*.

La distribución de la muestra se lleva a cabo a partir del análisis de la normal. Este análisis se puede realizar mediante tres métodos: gráfico (con histogramas, gráficos de densidad o gráficos Q-Q), analítico o mediante el contraste de hipótesis. El método

analítico implica el cálculo de la asimetría y de la Curtosis (que representa el grado de concentración de los valores con respecto a la media). Si el valor de uno y otro se encuentra entre el siguiente intervalo³⁷, -2 y +2, la distribución será normal o moderadamente normal (Bachman, 2004; Green, 2013). El método de contraste de hipótesis más habitual es Shapiro-Wik para muestras de menos de cincuenta participantes y Kolmogorov-Simonov para más. La distribución no se considera normal si el valor p es inferior a 0,5 ($p < 0,5$).

El segundo paso será pasar de la *estadística* descriptiva a la *inferencial*, ya que el análisis únicamente descriptivo es insuficiente y necesitamos comprobar si las primeras conclusiones son significativas y se pueden generalizar. En este punto, atenderemos a varios aspectos clave, entre otros, el análisis del ítem y la comparación entre variables, que nos permitirán ahondar en los principios de la evaluación, sobre todo en lo que respecta a la fiabilidad y validez, así como responder a las preguntas de investigación.

Para elegir qué método estadístico es más apropiado para el análisis hay que tener en cuenta si el test es *referencial* (NRT por *Norm-referenced test*) o *basado en criterio* (CRT por *Criterion-referenced test*) (Kunnan & Carr, 2012). En el primero, el TRN, se observa la actuación de uno de los testantes comparado con la de los otros que realizan el mismo test. En el segundo, el TRC, la actuación del testante se compara con unos criterios previos. Así, los TRN (test referidos a la norma) buscan saber si el resultado obtenido por el participante A es superior al obtenido por el participante B, los TRC (test referidos al criterio) tratan de averiguar si A y B alcanzan el criterio previamente establecido (Barrios et al., 2013). El TCE es un test basado en criterio y, por lo tanto, los métodos estadísticos elegidos deben adecuarse a sus características.

3.1.5.1. *Análisis del ítem*

El análisis del ítem es uno de los procedimientos estadísticas más importantes. Este análisis ha de pasar por tres fases: a) la primera en la que se analizan los *valores de facilidad* y los *niveles de dificultad*, b) la segunda en la que se observan los *niveles de*

³⁷ Otros trabajos más estrictos indican que para que la distribución sea normal el intervalo debe estar entre -1 y +1 o, incluso entre -0,5 y +0,5.

discriminación y c) la última, en la que se comprueba la *consistencia interna* (Green, 2019). Tal y como veremos a continuación, estos pasos varían en función del tipo de test con el que estemos trabajando. Así, la TCT (Teoría Clásica del Test) y la TRI (Teoría de Respuesta al Ítem), que explicamos brevemente a continuación, son las teorías apropiadas para los test referidos a la normal. Ambas tienen ventajas y desventajas, por lo tanto, la combinación de todos los modelos es la mejor de las opciones (Green, 2013). Tanto la TCT como la TRI está destinada a los test basados en la norma y no en los test basados en criterio. Por lo tanto, en este tipo de test como el TCE, aunque pueda resultar útil si tenemos en cuenta su naturaleza, se requieren procesos propios para examinar la consistencia de estos test (Barrios & Cosculluela, 2013; Muñiz, 2010; Sawaki & Xi, 2019). De modo que en los test de criterio se debe medir la fiabilidad a partir de otros mecanismos que veremos en el siguiente subapartado.

La TCT y la TRI se diferencian en distintos aspectos, entre ellos, el tamaño de la muestra necesario, ya que para la TRI se recomiendan más de 500 participantes y para aplicar la TCE sería suficiente contar con entre 200 y 500. La principal diferencia está en el centro de su análisis, mientras la TCT se centra en la fiabilidad del test, la TRI lo hace en el ítem (Muñiz, 2010). Siendo la TCT la más abundante en los estudios de evaluación, no son pocas sus limitaciones, como comentaremos a continuación, por lo que se ha propuesto una flexibilización de este enfoque conocida como Teoría G (Sawaki & Xi, 2019).

A. La TCT

La base de la *Teoría Clásica del Test* (TCT) de este modelo es el reconocimiento de la existencia del error. En la actualidad este modelo no siempre se utiliza, sino que a veces se relega, lo que supone un perjuicio ya que sigue siendo adecuado (Dean-Brown, 2012, p. 362) porque continúa contribuyendo al desarrollo del test y es la base para entender la evaluación de lenguas. Debemos tener en cuenta varios conceptos: los resultados observados (*observed scores*), error, variable (*variance*), error verdadero (*true scores*) y variable de error verdadera (*true score variance*). Los principales métodos de este modelo son los análisis del ítem (*facilidad del ítem*, *discriminación del ítem* y *análisis de la eficiencia del distractor*), la fiabilidad estimada, la medida del error estándar y varios

análisis de validez (que incluyen el criterio relacionado, contenido y estrategias de la validez de constructo).

Dean-Brown (2012) enumera algunas de las ventajas y desventajas del uso de la *Teoría Clásica del test*. Entre las ventajas podemos hablar, primero, de la difusión de este modelo y, por lo tanto, de lo conocido de sus métodos. Segundo, de la sencillez del cálculo de estos métodos y del cumplimiento, en gran medida, con las necesidades de la evaluación sobre todo en las pruebas de nivel. Entre las desventajas, primero, para utilizarlo es necesario que los resultados sean homogéneos, de modo que su uso está restringido a la dispersión de los resultados. Más concretamente, dependen de la distribución normal de los resultados y sólo son últimos para los test no referenciales. Segundo, suele ser muy sensible a los extremos de la escala. Suele penalizarse la sobreestimación de los resultados.

La fiabilidad también se mide a partir de la consistencia interna de los resultados del test. Para esto se vienen utilizando dos métricas de forma habitual: α de Cronbach y ω de McDonald's (Trizano-Hermosilla & Alvarado, 2016). En cualquiera de los dos coeficientes es imprescindible que el constructo sea independiente. Entre las ventajas del Alpha está su objetividad y, por lo tanto, su fácil interpretación. Sin embargo, el coeficiente Alfa no está recomendado para test con ítems ordenados en términos de dificultad y, en este caso, se recomienda utilizar el *alfa estratificada*³⁸ (McKay & Plonsky, 2021), como ocurre con el TCE. Omega presenta el valor multidimensional, ya que parte de un análisis factorial, pero repercute en la dificultad para calcularlo. La adscripción de cada ítem a una u otra valoración suele implicar la eliminación de dicho ítem.

B. TRI y MTRI

La *Teoría de respuesta al ítem*, TRI, por su parte, se centra en los resultados del participante, es decir, relaciona la actuación observada de cada participante con su posición en el continuo del constructo observado (Ockey, 2012). Esto se hace posible al relacionar la probabilidad de las respuestas correctas de los ítems y la habilidad de los

³⁸ También es posible calcular el alfa de cada grupo o dimensión.

participantes en el test. Es una poderosa metodología estadística empleada comúnmente para test estandarizados a larga escala y muy útil, entre otras cuestiones, para controlar la dificultad del ítem y del test así como para estimar los niveles de la habilidad de los testantes (Carr, 2011, p. 15). Pese a su utilidad la TRI tiene varias limitaciones (Green, 2013), que nos obliga a no utilizarla o a combinarla con la TCT. Primero, se necesita una población muy amplia. Segundo, hasta hace poco, los *software* que podían realizar este análisis tenían poca usabilidad, lo que aumentaba su dificultad.

Dos requisitos son necesarios para poder aplicar estas medidas: la *unidimensionalidad* y la *independencia local* (Fan & Bond, 2019). La unidimensionalidad, se entiende como sinónimo de constructo. ¿Qué obtenemos con este modelo? El TCE del español mide un constructo, el conocimiento de las colocaciones, y cada una de las modalidades representa un modo de acceso a dicho conocimiento. Con la independencia local se refiere a la nula relación de dependencia entre un ítem y otro, por lo que el conocimiento de uno en ningún caso depende del otro o es necesario. Así, el TCE se adapta a los requisitos, igual que el VST (Beglar, 2010), que además demuestra su coherencia con la validez basada en argumentos que también aplicamos y seguimos en el TCE.

Podemos concluir, por lo tanto, siguiendo a Green (2013) que lo mejor es realizar un análisis que combine ambos modelos, de modo que obtendremos un análisis mucho más rico que nos va a permitir profundizar en las conclusiones y decisiones que vamos a tomar a partir de los resultados obtenidos. Ambas teorías, TCT y TRI miden la dificultad del test, pero existe una diferencia en la forma en la que lo miden. En la TCT la dificultad del ítem se mide en función del número de testantes que responden correctamente a dicho ítem. La TRI es un modelo probabilístico que se centra en la posibilidad de que un ítem sea respondido según la habilidad de cada testante. Por último, los datos mínimos que se deberían conocer para un test de lengua: el valor de facilidad de cada ítem, si el ítem discrimina, cómo contribuye cada ítem a la fiabilidad del test, como la población del test se distribuye y cómo perciben el test los testantes.

Como hemos avanzado, este tipo de enfoques no son los más adecuados para los test basados en criterio, aunque sí nos puedan resultar de utilidad y, por lo tanto, debemos valorar el uso de otros enfoques y medidas, aunque menos habituales. Existen

alternativas que permiten calibrar la fiabilidad de la clasificación de los aprendices según el o los puntos de corte establecido como el coeficiente de Hambleton y Novick, el de Kappa de Cohen o el de Livinston; para los que es necesario aplicar el test varias veces o contar con una forma paralela dentro del mismo, como es el caso de TCE. Así como la Teoría de Respuesta del Ítem multidimensional (Bonify, 2022). Con este tipo de coeficientes podemos valorar si los criterios de clasificación son fiables.

3.1.5.2. *Comparación entre variables*

Si en la investigación hay más de una variable, lo cual suele ser habitual en evaluaciones destinadas a la investigación, es necesario contrastar las variables, además de con la estadística descriptiva, con la estadística inferencial: comparando la media de las variables, comúnmente a partir de técnicas estadísticas como *t-test* o ANOVA, y teniendo en cuenta las *correlaciones*, a partir de índices como Pearson o Spearman. Además, existe el análisis factorial, que ahonda en las causas de las relaciones entre variables, y que están menos extendidas que los anteriores recursos estadísticos.

Podemos definir la *correlación* como el grado de relación entre dos variables (Carr, 2011). Las correlaciones son una de las posibles dimensiones para comparar variables y nos aportan diferente información: a) confirman la dirección de la relación (positiva o negativa), b) determinan el grado de confianza de los análisis, es decir, si pueden generalizarse o no a una población amplia (Green, 2013). También se obtendrá la fuerza de relación oscilando entre débil, moderada o fuerte.

Antes de avanzar hacia los siguientes pasos de análisis es necesario tener claro cuáles son los tipos de variables con los que vamos a trabajar, lo que nos va a permitir elegir tanto el método de varianza como el *coeficiente de correlación* que debemos utilizar. Podemos encontrarnos con cuatro tipos de variables correlacionadas (Carr, 2011, p. 262). Primero, las *variables nominales* como edad, sexo, etc. Segundo, las *variables ordinales* que se refieren esencialmente a las categorías que podemos ordenar (lista, rango). Tercero, la *variable de intervalo*, igual que la anterior, se refieren a categorías que podemos ordenar (escala); pero en este caso, y a diferencia del otro, la distancia entre cada punto de la escala es el mismo, de ahí que se denominan *variables de intervalo*.

Cuatro, las *variables de ratio / proporción*. Igual que en las variables de intervalo, hay la misma distancia de puntos entre cada punto de la escala, pero en este caso también se incluye un punto cero (*true zero point*).

El *coeficiente de correlación* nos permite conocer la fuerza y la dirección de la relación entre dos o más variables. Las dos más comunes son Pearson y Spearman y aunque cada una de ellas cuenta con varios tipos, no profundizaremos más en este asunto. No se pueden usar indistintamente, sino que es necesario atender a unos requisitos para cada una de ellas. Para poder utilizar Pearson es imprescindible que la relación entre las dos variables sea lineal, que las dos variables constituyan escalas de intervalo y que las dos variables estén en distribución normal (Bachman & Palmer, 1996). Además, una de las variables suele ser dicotómica y la otra variable dependiente continua (Green, 2013). Con dos variables ordinales, es habitual usar el *coeficiente de correlación de Spearman* que muestra un rango, por ejemplo, los niveles de competencia lingüística del MCER, y este coeficiente no tiene en cuenta la distribución de las variables. Es importante señalar que el hecho de que dos variables estén relacionadas no significa que una sea la causa de la otra. Las tablas de contingencia, herramienta con la que se representan datos categóricos en términos de conteos de frecuencias, también nos ayudará a comparar las dos opciones de cada modalidad. Este recurso también ha sido utilizado en otros estudios que analizan la validez de test referidos a criterio (Gyllstad et al., 2015).

El análisis de la varianza nos permite generalizar sobre la relación entre dos o más variables y comprobar si esta relación es o no significativa. Antes de elegir el análisis estadístico que vamos a llevar a cabo para comprobar que la relación entre las variables es significativa, debemos comprobar la distribución de la variable. El tipo de distribución de cada variable determina el estadístico que vamos a elegir. La elección de un u otro estadístico depende de los requisitos de cada uno según las características de la muestra. Así, para utilizar t-test no puede haber más de dos variables. Para utilizar ANOVA se pueden usar más de dos variables. Hay muchas más que no enumeramos porque no son útiles para el análisis de esta investigación. T-test se usa con pruebas paramétricas en las que se comparan las medias de dos grupos que realizan el mismo test (dos muestras independientes) o un mismo grupo que realiza dos test (muestras pareadas). Además,

debe cumplir los siguientes requisitos: independencia de las observaciones, distribución normal. Para usar ANOVA se requiere una distribución normal, una varianza equitativa entre variables y la independencia de las variables que se van a tener en cuenta.

El análisis factorial no está todavía tan asentado en la evaluación de lenguas, aunque ya hay trabajos, incluso sobre evaluación de vocabulario, sobre todo los llamados estudios componenciales en los que sí se utiliza (González-Fernández & Schmitt, 2020). Este tipo de modelo ya no solo busca comprobar si la relación entre dos variables es significativo o en qué grado existe esta relación, sino que busca establecer una relación de causalidad entre las variables (Ferrando et al., 2022), entre los que está, por ejemplo, el *Structural Equation Modeling* (SEM) (Yang & Green, 2011).

3.2. La evaluación del léxico

La *evaluación del léxico* ha asumido dos funciones principales: por un lado, determinar el nivel de la competencia léxica, y por extensión, el nivel de la competencia lingüística, al funcionar como indicio; por el otro, analizar el comportamiento del léxico en relación con otros aspectos de la competencia lingüística como la comprensión oral y escrita, en mayor medida. Pese a su arraigo en los estudios de enseñanza-aprendizaje, todavía acarrea una serie de limitaciones o dificultades en su desarrollo y, sobre todo, en su validación (Masrai & Milton, 2018; Schmitt, 2019a). Entre las dificultades se menciona la desconexión de la *evaluación del léxico* con respecto a la *evaluación de lenguas*. Esta se ve reflejada en la reticencia de los especialistas a la aplicación de los avances y nuevas teorías como el modelo *de validez basada en argumentos*. La necesaria actualización lleva a la colaboración entre ambas líneas de estudio, la del léxico y la de la evaluación de lenguas, en el mismo sentido que complementamos la teoría de la evaluación de las colocaciones con la teoría de la evaluación del léxico.

Los test de léxico, por lo tanto, deben seguir las directrices que aplicamos a los test de lengua en general: principios, ciclo, selección de la muestra, etc. Sin embargo, la principal particularidad es la complejidad de la unidad evaluada, el léxico. En esta sección haremos un repaso sobre algunos de los test de vocabulario que han sido y son referente en este campo. Para hacerlo los clarificaremos atendiendo a distintos criterios.

3.2.1. Tipos de test de vocabulario

Con la intención de ordenar los procesos de evaluación del léxico, Read (Read, 2011; Read & Chapelle, 2001) propone tres dimensiones que nos permiten clasificar los test que miden el conocimiento del vocabulario.

3.2.1.1. *Test discretos frente a test incorporados*

Un test *discreto* toma el conocimiento del vocabulario como un constructo diferenciado del resto de componentes de la competencia lingüística. Un test *incorporado* es aquel que contribuye a la evaluación de un constructo más amplio. De modo que, en un *test discreto* únicamente se mediría el conocimiento léxico, pero no otra destreza. En este punto debemos plantearnos si esta premisa es viable en tanto que no sería posible diseñar un test en el que no esté presente ninguna destreza, pero sí que otra competencia, por ejemplo, la comprensión lectora, esté presente y sea la destreza objeto de la evaluación. Los *test incorporados* son los más habituales, especialmente, aquellos que combinan la evaluación del léxico con la comprensión escrita frente al resto de posibles destrezas. Esta distinción no se centra en cómo el vocabulario se presenta a los participantes sino en la intencionalidad de quien lo diseña. Para saber si un test de vocabulario es *discreto* o *incorporado* es necesario atender al propósito del test y a la interpretación de los resultados y no solo a su forma.

3.2.1.2. *Test selectivos frente a test ampliados*

La segunda dimensión propuesta concierne al rango de vocabulario que se va a incluir en la evaluación. Un *test selectivo* elige las palabras concretas que se van a evaluar y se debe demostrar el conocimiento de su significado o su uso. Un *test ampliado* no selecciona las palabras concretas previamente, sino que evalúa el léxico de un texto dado.

3.2.1.3. *Test contextualizados frente a test no contextualizados*

El papel del contexto es la base de la tercera dimensión. Así, se contraponen los *test independientes contextualmente* a los *dependientes contextualmente*. Por contextualización no solo se entiende que las palabras clave se presenten en contexto o aisladas sino a si la respuesta depende o no del contexto. Un ejemplo claro es el de las palabras polisémicas en las que, si no se parte de un contexto, no se puede presentar el ítem.

A pesar del peso de los estudios de Read (2011) sobre el léxico y su evaluación, la caracterización de los diferentes test no se suele guiar por esta clasificación sino por una serie de criterios relacionados con la definición del constructo, es decir, con la descripción tomada del conocimiento léxico, que se centran más en el propósito y contenido medido que en la forma. Los estudios dedicados a la recopilación, análisis o revisión de los test que miden el vocabulario suelen utilizar este tipo de criterios para ordenar los que van a revisar y describir. Uno de los más habituales es la dicotomía *extensión-profundidad* (vid.2.1.2); otro, la dicotomía *producción-comprensión* (vid. 2.1.1); de ahí que se suele hablar de test que miden la extensión y la producción, los que miden la extensión y la comprensión, los que miden la profundidad y la producción o los que miden la profundidad y la comprensión. En esta línea, se han propuesto otras clasificaciones de los test de vocabulario que incluyen cuatro dimensiones (Leńko-Szymańska, 2020, p. 46): el conocimiento, la habilidad, los ítems específicos y el léxico (de forma completa).

Desde la preocupación por la naturaleza del conocimiento léxico y, en concreto, desde el enfoque componencial, se ha propuesto otro tipo de test con un propósito más amplio y global, los *test en batería*. La búsqueda de un conocimiento general de una palabra hace que los estudios se puedan organizar según se centren en un solo componente (*single components studies*) o si se centran en varios (*multiple components studies*) (González-Fernández & Schmitt, 2020; Zhang & Pei, 2022). Estos test en batería son pruebas que miden al menos tres componentes del conocimiento léxico, tanto en comprensión como en producción para poder trabajar estadísticamente con las diferentes variables. Obviamente, predominan los estudios simples de componentes, en los que solo se analiza un aspecto del conocimiento. Esto es debido a la complejidad que supone la conceptualización del conjunto y su difícil aplicación, es decir, la dificultad de evaluar de forma global.

Haremos un breve recorrido por alguno de los test de vocabulario más usados y que han servido como modelo para posteriores versiones o adaptaciones. Nos centraremos en los test que miden la competencia léxica del inglés como L2 porque son la referencia en el campo y también la lengua más estudiada. En español como L2 la mayor parte de los test de vocabulario se centran en el léxico disponible de la lengua (M. Ainciburu, 2012; Santos Díaz et al., 2020), el trabajo sobre la extensión y profundidad es

todavía limitado, aunque dedicaremos el subapartado siguiente a hacer un breve recorrido por estos trabajos. En la siguiente tabla avanzamos los test que vamos a describir organizándolos atendiendo a los cuatro criterios que definen la competencia léxica: profundidad, extensión, producción y comprensión, dando lugar a partir de su combinación en un clasificación en cuatro áreas que ya siguen otros trabajos (Cervatiuc, 2007; Durrant et al., 2022).

Analizaremos cada una de las cuatro áreas atendiendo a sus posteriores versiones. Por un lado, analizaremos el test X-LEX (Meara & Milton, 2003; James Milton & Alexiou, 2010), que mide la extensión en un bajo grado de profundidad, únicamente desde el reconocimiento de la forma, pero que es relevante en nuestra investigación ya que lo usamos como fuente para conocer la competencia léxica de los participantes en el estudio. También, el test VLT (Nation, 1990; Schmitt et al., 2001a) que mide la extensión en un grado más alto de profundidad que el anterior. Por otro lado, VKS (Wesche & Paribakht, 1994; Wesche & Paribakht, 1996) mide el grado de profundidad del conocimiento de cada ítem, principalmente, según el criterio del aprendiz. Asimismo, analizaremos Lex-30 (Meara & Miralpeix, 2017b; Zareva & Wolter, 2012) que mide la profundidad centrándose en uno de los aspectos del conocimiento de las palabras. Por último, nos centraremos en otro que no se adapta a esta descripción, CATSS (Aviad-Levitzky et al., 2019; Laufer & Goldstein, 2004), pero que es básico para esta investigación, ya que el TCE se basa en él.

3.2.2. Test de vocabulario en inglés

Como hemos avanzado, describiremos un test representativo de cada una de las áreas que deja la combinación de los criterios más habituales en la descripción de test de vocabulario y que, a su vez, en conjunto, mostrarían una imagen global de los test sobre conocimiento léxico.

3.2.2.1. Vocabulary Level Test³⁹ (VLT)

El *Vocabulary Level Test* (VLT) (Nation, 1983, 1990; Schmitt et al., 2001) o test de nivel de vocabulario es una herramienta de evaluación que mide el léxico desde la comprensión; en posteriores versiones también se creó una adaptación que medía la producción, PVLT (Laufer & Nation, 1999). Pertenece a la categoría de test basado en frecuencia en los que los ítems se seleccionan de acuerdo a *ranking* extraídos de listas de frecuencias (Peters et al., 2019) y las palabras meta se organizan en secciones que identificamos con bandas de frecuencia. Parte de la idea asumida de que las palabras más frecuentes se aprenden antes que las menos frecuentes. En este caso, los ítems se ordenan en cuatro bandas de las diez mil palabras más frecuentes. Puede considerarse el test de vocabulario estandarizado, por ser el más utilizado, citado y adaptado, tanto por el número de estudios en los que se ha utilizado como en el número de adaptaciones (Nation, 1983, 1990; Schmitt et al., 2001). Se puede acceder a él de forma libre a través de LEXTUTOR⁴⁰, y aunque en la actualidad su finalidad es la de conocer el nivel del conocimiento léxico, en su versión original se creó para ayudar a los profesores a diseñar sus cursos de idiomas, es decir, como herramienta de diagnóstico (Read, 2011) y se ha usado para medir la extensión a falta de otra herramienta equiparable.

You must choose the right word to go with each meaning. Write the number of that word next to its meaning.

1 concrete	
2 era	_____ circular shape
3 fiber	_____ top of a mountain
4 hip	_____ a long period of time
5 loop	
6 summit	

Figura 18. Ejemplo de ítem de VLT (Schmitt et al., 2001, p. 58)

En su primera versión, la versión original, *Vocabulary Level Test* (VLT), la prueba consiste en relacionar palabras con su definición, construida de la forma más breve posible. De modo que en cada ítem se presentan tres sustantivos, dos verbos y un adjetivo para tres definiciones a partir de las que el aprendiz debe relacionar adecuadamente cada palabra con su definición correspondiente (vid. Figura 19). Los

³⁹ No debemos confundir con VST (Vocabulary Size Test), otro instrumento estandarizado que mide la extensión, pero su intención no suele ser diagnóstica.

⁴⁰ <https://www.lexutor.ca/> [Último acceso 18/08/23]

resultados muestran que el conocimiento de las palabras menos frecuentes implica el de las de más alta frecuencia y que el número de aciertos disminuye a medida que desciende el índice de frecuencia de la palabra meta. En una versión posterior (Versión C), se propuso hacer una adaptación que midiera la producción controlada del conocimiento léxico (Laufer & Nation, 1999). El formato se convirtió para esta adaptación en un *fill-in-the-gap*, en el que los aprendices deben recuperar la palabra meta a partir del contexto con la ayuda de las letras iniciales de la palabra meta. En la *Figura 19* se puede ver un ejemplo. Se trata de un test de diagnóstico, pero también de competencia, a partir del cual se puede asignar un nivel a los aprendices.

One of two equivalent versions of the A LEVELS TEST OF PRODUCTIVE VOCABULARY: Parallel Version 1 (Version C)

Complete the underlined words. The example has been done for you.

He was riding a bicycle.

The 2000-word level

1. I'm glad we had this opp_____ to talk.
2. There are a doz_____ eggs in the basket.
3. Every working person must pay income t_____.
4. The pirates buried the trea_____ on a desert island.
5. Her beauty and cha_____ had a powerful effect on men.

Figura 19. Ejemplo de ítem del PVL (Laufer y Nation, 1999)

Son varias las versiones del test propuestas, y muchas propuestas de mejora de la herramienta (Schmitt et al., 2001; Stoeckel et al., 2021). Recientemente se han desarrollado nuevas versiones que intentan solventar las limitaciones de tanto de VLT (Vocabulary Level Test) ⁴¹ como del VST (Vocabulary Size Test) (Nation & Beglar, 2007b). Por ejemplo, el UPVLT (Webb et al., 2017), *Uptodate Vocabulary Level Test*, o el *VocabLab* (Peters et al., 2019). Este último se trata de un test de diagnóstico del conocimiento léxico del francés y del inglés. Esta nueva versión elimina algunas de las cuestiones más problemáticas de las versiones previas: a) reduce el factor del azar al incluir una nueva opción en cada ítem; b) actualiza las listas de palabras frecuentes, al usar las más recientes; y c) se adapta a la L1 de los aprendices, evitando cognados. El objetivo es comprobar si de él se deduce un perfil escalable (gradual) y si el test discrimina entre diferentes niveles de competencia. Plantea dos objetivos con agentes diferentes: el

⁴¹ El VLT se diferencia del VST en su propósito: el primero es un test de diagnóstico y el segundo un test de extensión. Mientras en el primero se incluyen palabras específicas (por ejemplo, de la *Academic Word List*), en el segundo no. Además, usan diferentes formatos: en el VLT hay que relacionar palabras aisladas con definiciones y en el VST se usa el formato de elección múltiple (para una palabra contextualizada el aprendiz debe escoger de entre cuatro la definición más adecuada).

primero es relativo al conocimiento de los aprendices, pero el segundo atiende a las posibilidades de uso del propio test. El análisis de los resultados muestra que el conocimiento se puede representar en un perfil escalar de acuerdo con la frecuencia. Es decir, pese a la sobreestimación de los anteriores y las otras limitaciones, la relevancia de la frecuencia en el aprendizaje del léxico sigue demostrándose.

English example	French example
<i>talk:</i>	<i>mort:</i>
o to speak	o qui ne vit plus
o to remove something	o qui dit ce qu'il pense
o to give in return for money	o qui ne perd pas son calme
o to start an important activity	o qui ne prend pas de risques
o I don't know the answer	o je ne connais pas la réponse

Figura 20. Ejemplo de ítem de VocabLab (Peters et al., 2019)

3.2.2.2. X-LEX

X-LEX es el test de léxico que usamos en esta investigación y que nos permitirá comparar los resultados de la competencia léxica con los de la competencia colocacional extraída del TCE. Este test mide la extensión del vocabulario en relación a la frecuencia (Meara & Milton, 2003). Utiliza un formato de sí / no (o *checklist*) en el que los aprendices se encuentran con una lista de palabras y deben señalar cuáles conocen. Las *palabras meta* se extraen de una lista de palabras frecuentes empleando el siguiente criterio: un número de palabras de cada una de las bandas de frecuencia; así, se extraen veinte palabras lematizadas de la lista de las mil palabras más frecuentes, veinte palabras de las dos mil más frecuentes, veinte de las tres mil más frecuentes, veinte de las cuatro mil y otras veinte de las cinco mil. Uno de los puntos más destacados es la vinculación entre el *perfil basado en frecuencias* (Milton, 2013) y la competencia léxica. El perfil representa que el número de palabras conocidas disminuye a medida que aumenta la frecuencia de las palabras en cuestión. Además, a estas cien palabras se añaden veinte *pseudopalabras* para evitar la sobreestimación a partir de falsas alarmas, aquellas no palabras que los aprendices marcan como conocidas. Estas siguen el modelo de formación de palabras y la combinación de sonidos propias cada lengua. También este será útil porque será el que utilizemos para comparar la competencia léxica con la colocacional.

El test original mide la extensión del vocabulario del inglés (Meara & Milton, 2003). Hay otras versiones paralelas y otras están como son la del sueco, italiano y portugués. Una de las versiones publicadas es la que mide la extensión de vocabulario del griego (Batista & Horst, 2016; Milton & Alexiou, 2010). En el caso del test realizado a aprendices de griego, al igual que ocurre en inglés y en francés, las hipótesis se confirman y la fiabilidad y validez del test también. Los resultados reflejan que el efecto de la frecuencia está presente y que la media de los resultados incrementa con cada nivel. Esto no hace más que corroborar el fuerte impacto de la frecuencia en el conocimiento léxico al observar diferencias entre aprendices de diferentes niveles y coherencia con las predicciones del perfil. Debemos señalar, sin embargo, que uno de los puntos diferenciales entre las distintas versiones del X-LEX es el número de palabras necesarias en cada banda de frecuencia para llegar al nivel. En griego, el número requerido para cada nivel podrá ser más que el requerido para alcanzar el nivel en inglés o en francés. Este será uno de los aspectos que debemos tener en cuenta a la hora de analizar los resultados en la versión en español de X-LEX.

Los formatos de lista de verificación o sí / no son unos de los habituales para medir la extensión de vocabulario. Una de sus fortalezas es su simplicidad (Anderson & Freebody, 1981); pero una de las debilidades reside en su validez cuestionable. Esta cuestionada validez está condicionada por la autoevaluación. Es el aprendiz el que decide si conoce o no conoce la palabra en cuestión, lo que implica sobreestimación o, en algunos casos subestimación. Este punto se trató de corregir a través de las pseudopalabras. En los últimos años se han comparado los resultados de test de extensión de diferentes formatos, entre ellos los de verificación sí/no con otros formatos y con posteriores entrevistas en las que se trata de confirmar el conocimiento de una palabra (Stubbe & Stewart, 2012; Raymond Stubbe, 2012a). Los resultados muestran que, aunque la sobreestimación es una realidad, existe una correlación importante entre los resultados que ofrece este formato con otros formatos e incluso con las entrevistas. Pese a que el tipo de formato elegido debe depender del tipo de información que el investigador necesita (obtener) (Kremmel & Schmitt, 2017), la elección de los formatos suele venir determinada por la tradición (Read, 2011). Los test de extensión de vocabulario suelen basarse en la propuesta que hoy por hoy se considera la estándar, su

relación con la frecuencia. Los formatos de lista de verificación o sí / no son unos de los habituales para medir la extensión de vocabulario.

Otro de los puntos críticos de este formato es el relacionado con el conocimiento de la palabra. Únicamente se está evaluando el reconocimiento de la forma, ningún aspecto más del conocimiento de las palabras se tiene en cuenta. Desde un punto de vista cualitativo, la diferencia entre qué entiende el investigador y qué entiende el aprendiz por conocer una palabra puede hacer variar el sentido de las respuestas. En este caso no se habla de empleabilidad como el conocimiento real de una palabra sino solo de su identificación o aprendizaje o conocimiento previo. Para utilizar este formato se deben interpretar adecuadamente los resultados, no se evalúa el conocimiento de la palabra de una forma global / completa, sino simplemente el reconocimiento de una forma, lo que en este caso resulta suficiente ya que no es el objeto de estudio sino un punto de referencia.

Además de este, Meara cuenta con una serie de herramientas electrónicas que permiten centrarse en otras cuestiones y utilizar herramientas ya creadas para medir el vocabulario (Meara & Miralpeix, 2017b). Además de otras herramientas⁴², incluye varios test: V-Unique v1.0, que mide la originalidad del léxico a partir de una imagen estímulo; V-YesNo, que mide la extensión del vocabulario desde el reconocimiento.

3.2.2.3. *Vocabulary knowledge Scale (VKS)*

Uno de los estudios pioneros en la evaluación de la profundidad es el test basado en escalas *Vocabulary Knowledge Scale* (Wesche & Paribakht, 1994; Wesche & Paribakht, 1996) como intento de capturar los diferentes niveles del conocimiento propio percibido a partir de palabras específicas. En este caso se entiende la profundidad como la calidad del conocimiento de cada ítem en concreto, por lo tanto, se considera ejemplos de *test de desarrollo*. Pese a que este test se encuadra como herramienta para medir la profundidad, otros indican que no es un instrumento válido para medirla, sino que mide el significado de la palabra de forma indirecta (Laufer & Goldstein, 2004).

⁴²Estas herramientas incluyen V-Words, V-List, D-Tools, P-Lex, V-Size, en su mayoría programas que miden, en alguna medida, el léxico a partir de corpus.

Se trata de un formato escalar, con las palabras clave descontextualizadas, que permite reconocer cambios en el conocimiento de una palabra. Los niveles de la escala oscilan desde la ausencia de familiaridad con la palabra (desconocimiento) pasando por el reconocimiento de la palabra escrita o la muestra de alguna idea del significado hasta la habilidad para usar la palabra en una oración. Además, mide con un único formato tanto la comprensión como la producción. Los dos primeros niveles dependen de la sinceridad del testante (*auto-información*), pero la respuesta en los siguientes pasos depende de su demostración lingüística. En la imagen podemos ver la escala (vid. Figura 21) tal y como se presenta a los participantes.

Instructions: Please read each item on this activity very carefully. If you are not sure of the at each of the levels (labeled I, II, III, IV, and V) mean, please ask your researcher for help. Then indicate your knowledge about the FIVE words below.

resmo

- I. I have never seen this word. _____
- II. I saw this word in the passage, but I don't know what it means. _____
- III. I saw this word in the passage, and I think it means _____ (provide an English translation.)
- IV. I know this word. It means _____ (provide an English translation.)
- V. I can use this word in a sentence (if you answer this item, you MUST also answer item IV)
-

Note. The VKS used in the present study is adapted from Paribakht & Wesche (1993, p. 15)

Figura 21. Ejemplo de ítem de VKS (Wesche & Paribakht, 1996)

En la escala de los aprendices se pretende demostrar qué categoría representa mejor el conocimiento que se tiene de la palabra. Los pasos en la escala son los siguientes: I) no reconocimiento, II) reconocimiento, pero no enlace entre la forma y el conocimiento, III) poca seguridad en la traducción, IV) más seguridad en la traducción y V) producción, que consiste en la creación de una frase con la palabra. Al contrario de lo que ocurre con otros test que hemos mencionado, en este se necesita además del juicio del aprendiz evidencia de la respuesta. De modo que el aprendiz que cubre la escala debe demostrarlo, excepto en los primeros escalones por razones evidentes.

Pese a la amplia difusión del formato, son muchas las críticas que ha recibido, sobre todo por su falta de objetividad y ausencia de validez. A continuación, recogemos algunos de los puntos críticos que desde diferentes trabajos se han ido haciendo a este formato (Bruton, 2009; Laufer & Goldstein, 2004; Yanagisawa & Webb, 2020):

- a. el test solo puede ser administrado a una pequeña muestra ya que debe ser corregido de forma manual;
- b. la dificultad que entraña la valoración de una escala: número de niveles, equidistancia entre ellos, etc.;
- c. falta de correspondencia de los resultados con el MCER, de modo que no es evidente la correlación de un nivel de competencia léxica ni lingüística con resultados del test;
- d. falta de fiabilidad, sobre todo, en los dos primeros pasos de la escala, en los que las respuestas dependen de la buena voluntad, no son verificables;
- e. carencia de validez;

En cuanto a los resultados, según la escala de corrección, estos no nos permiten describir un modelo gradual de conocimiento de las palabras, sino que únicamente permite discernir entre reconocer una palabra descontextualizada y producirla de forma contextualizada. Lo que es un salto importante en la gradación del conocimiento en relación con su dificultad. Si nos guiamos por el número de componentes que se evalúan en un paso y otro de la escala, el número se incrementa considerable. En el reconocimiento se tiene en cuenta ese únicamente, en la producción contextualizada entrarían además los rasgos gramaticales, combinatorios, etc.

En el ámbito del español también se usó una versión adaptada del VKT (Pulido, 2004) para distinguir como variable en su estudio sobre el efecto de la familiaridad cultural en la adquisición del vocabulario de forma incidental a través de la lectura de aprendices de español. La adaptación del VKT también mantiene las tareas autoevaluación adaptando los niveles de la escala.

3.2.2.4. *Lex30*⁴³

Este test se enmarca en el grupo de test que analizan el léxico a partir de la relación entre las palabras (Zareva & Wolter, 2012) y que se conocen como WAF (*Word Associates Form*), un tipo de test que miden la profundidad desde uno de los aspectos del conocimiento léxico: las asociaciones. Hay un número importante de investigaciones

⁴³ <https://www.lognostics.co.uk/tools/Lex30/index.htm>

en relación a este componente del conocimiento léxico (Fitzpatrick & Munby, 2014) en los que se trabaja en la búsqueda de la correlación entre las asociaciones y la competencia lingüística, aunque los resultados dispares han denostado este tipo de test. Este problema se achaca a la falta de sistematicidad en el método de corrección. El análisis suele centrarse en el tipo de asociaciones realizadas (sintagmáticas o paradigmáticas) más que en los elementos específicos que se dan como respuesta. Aunque ha habido varias propuestas sobre test asociativos, nos centraremos en Lex30⁴⁴ (Fitzpatrick & Meara, 2004a; Meara & Fitzpatrick, 2000).

Lex30 es un test que mide el conocimiento productivo, en concreto, el conocimiento productivo a partir de uno de los aspectos del conocimiento léxico: las asociaciones. Los primeros resultados presentan la correlación entre las asociaciones y la competencia lingüística (Meara & Fitzpatrick, 2000). Es fundamental como herramienta de elicitación ya que representa una de las características del lexicón mental. Se trata de una herramienta fiable y con cierta carga de validez que, como apuntábamos, tiene como hándicap el diseño de los criterios de la evaluación (Yanagisawa & Webb, 2020). Entre las ventajas de esta prueba está la facilidad para administrarlo y la adaptación a los diferentes niveles de competencia (Zhang & Koda, 2017). El formato consiste en una lista de treinta *palabras estímulo* a partir de las cuales los aprendices deben escribir las tres primeras palabras que les vengan a la mente. Este formato representa un nivel mínimo de conocimiento productivo, ya que las palabras meta no se usan, sino que simplemente se asocian con otras.

Posteriormente, se realizó otra versión (Nissen & Henriksen, 2006) en la que el análisis se centró en la influencia de la clase de palabra al asociar otra. Los datos sugieren que se producen (*elicit*) nombres en una proporción más alta que verbos y adjetivos. La influencia de la clase de palabras se discute en términos adquisición y organización sintagmática y que, también condiciona la tendencia a elegir asociaciones sintagmáticas o paradigmáticas. Wolter y Gyllstad (2011), en una análisis de este tipo de formatos, parten de tres hipótesis sobre las asociaciones de palabras: a) en la L1 se producen más

⁴⁴ Con un formato similar, años después se construyó *LexCombi* (Barfield, 2009) que, posteriormente, sería revisado. Este se centra en la evaluación de las colocaciones, un tipo de asociación de palabras, pero que incluso es otro de los componentes del significado de las palabras y que veremos en apartados posteriores (Vid. 3.3.2.1, sección D).

respuestas paradigmáticas, b) en la L2 hay más respuestas sintagmáticas y c) los nombres suscitan más respuestas que el resto de clases de palabras. Estas hipótesis se basan en investigaciones previas en las que se considera que el lexicón mental de la L1 está estructurado paradigmáticamente y el lexicón mental de la L2 lo está sintagmáticamente (Jiménez-Catalá & Moreno-Espinosa, 2004).

Pesa las limitaciones que hemos apuntado, sigue siendo un avance importante en cuento a los test de profundidad, y se siguen haciendo versiones de este test, como el *Evaluation of Academic Vocabulary* (EAV) (Crosson et al., 2019), que adapta el estilo de los test de asociaciones a adolescentes y al léxico del ámbito académico o una herramienta clásica, que también se considera referente en este tipo de test, WAT (*Word associative test*) (Read, 2011).

3.2.2.5. CATSS

El *Test adaptativo y computacional de extensión y sofisticación* (Aviad-Levitzky et al., 2019; Laufer et al., 2004; Laufer & Goldstein, 2004), de ahora en adelante CATSS (*Computer Adaptive Test of Size and Strength*) tiene tres versiones. La última, al igual que el TCE atiende a criterio argumentativo de la validez. El test se centra en medir la solidez⁴⁵ (*strength*) y la extensión del conocimiento léxico (Laufer et al., 2004). El fundamento del test reside en la idea de superar la confusión latente de qué significa conocer una palabra. La solución pasaría por distinguir cuatro grados en el conocimiento del significado basándose en la combinación de dos dicotomías: producción y comprensión (vid. 2.1.1), por una parte, y recuperación y reconocimiento (vid. 2.1.1), por la otra, teniendo como centro del conocimiento la unión entre la forma y el significado.

El diseño del test parte de una serie de supuestos que guían su construcción. Primero, el componente del conocimiento más importante es la unión entre forma y significado. Segundo, el conocimiento del significado no es un todo o nada, sino que se puede explicar en cuatro niveles de solidez. Tercero, conocer un número alto de palabras superficialmente es más importante que conocer un número bajo en profundidad. Además, CATSS se rige por tres principios: extensión, solidez y adaptación informatizada.

⁴⁵ En el apartado 2.1.3 se explica de forma pormenorizada en qué consiste el concepto de solidez.

Con *extensión* se refiere al número de palabras conocidas y con *solidez*, al grado en que conoce cada palabra. En cuanto a la *adaptación informatizada*, es un test adaptativo con respecto a las modalidades, no en relación con los ítems, lo que determina los criterios de corrección. Si se contesta correctamente a un ítem de la modalidad más difícil ya no contestará a ese ítem en las modalidades hipotéticamente más sencillas. Si, por el contrario, no se acierta un ítem en la modalidad más difícil, tendrá que volver a responder a dicho ítem en la modalidad siguiente según la escala de dificultad y así sucesivamente (Aviad-Levitzky et al., 2019). La puntuación de cada ítem también se relaciona con la modalidad en la que se acierta la palabra meta: 1 punto por acertar en la modalidad recuperación de la forma, 0,75 por acertar en la modalidad de recuperación del significado; 0,5, por reconocimiento de la forma; y 0,25, por acertar en reconocimiento del significado. Es decir, el acierto vale diferente en cada una de las modalidades; por ejemplo, no es lo mismo traducir de la L2 a la L1 que de la L1 a la L2.

En cuanto a la muestra, nos resulta más útil para esta investigación centrarnos en la metodología de selección que en las herramientas utilizadas para hacerlo, listas de frecuencias del inglés y del hebreo. Mientras en las primeras versiones se elegían palabras de bandas alternas, en la nueva versión se eligieron palabras de todas las bandas de forma consecutiva. Se utilizan catorce bandas de frecuencia, atendiendo únicamente al número de palabras que un aprendiz debe conocer y no a niveles de competencia (lo que lo diferencia de nuestro propósito). Se basan en los estudios que afirman que entre las 14000 palabras frecuentes reside el 99% del léxico de un hablante (Nation, 2006; Schmitt et al., 2015, 2017; Schmitt & Schmitt, 2014).

En la prueba original se eligieron cuatro formatos, uno para cada uno de los grados de dificultad del conocimiento de las palabras. Para medir la producción y la comprensión desde la recuperación, utilizaron un ítem de traducción. El aprendiz parte de la equivalencia en su L1 y debe buscar la palabra meta en la L2. En comprensión desde la recuperación, también se usa la traducción, pero en este caso en sentido inverso. El aprendiz parte de la palabra meta (en la L2) que debe reconocer y escribir su equivalente en la L1. En ambos casos se ofrece la inicial de la palabra que deben recuperar para limitar el número de opciones y acota el número de respuestas. Para evaluar la producción desde el reconocimiento se acude a un formato de opción múltiple. Se parte de la

equivalencia a la palabra meta en la L1 y se presentan cuatro opciones: la palabra meta en la L2 y tres distractores, elegidos por sus relación semántica y misma banda de frecuencia, manteniendo así la misma dificultad entre todas las opciones. En el formato que mide la comprensión desde el reconocimiento, de nuevo se usa un formato de opción múltiple y la traducción, pero en el sentido opuesto. Se parte de la palabra meta, en la L2, y el aprendiz debe elegir entre cuatro opciones: la equivalencia en la L1 más adecuada y tres distractores. Estos tres distractores son la traducción a la L1 de los usados en el otro formato. En la *Tabla 17* se ve de forma esquematizada. En los cuatro casos la palabra meta está descontextualizada y únicamente se pregunta por el enlace entre forma y significado.

	Producción	Comprensión
Recuperación	Traducción desde la L1 a la L2 <i>Leche → milk</i>	Traducción desde la L2 a la L1 <i>Milk → leche</i>
Reconocimiento	Opción múltiple desde la L1 a la L2 <i>Leche → milk, beer, wine</i>	Opción múltiple desde la L2 a la L1 <i>Milk, beer, wine → leche</i>

Tabla 17. Muestra de ítems en CATSS, versión bilingüe (Laufer & Goldstein, 2004).

En la última versión se produjeron algunos cambios en el diseño de los formatos en las tres versiones disponibles: monolingüe inglés-inglés, bilingüe inglés-hebreo e inglés-árabe. Nos centraremos únicamente en la primera. Al ser de carácter monolingüe se dejó de lado la traducción y los ítems se contextualizaron en su mínima expresión, uso del contexto mínimo, de modo que se juega con la sinonimia. En la modalidad *recuperación de la forma* se presenta la palabra meta contextualizada y dos pistas: la inicial de la palabra meta y un sinónimo entre paréntesis. En la modalidad *recuperación del significado de la palabra* se ofrece la palabra meta contextualizada y hay que dar una palabra sinónima. En *reconocimiento de la forma* escrita hay que seleccionar de entre cuatro opciones la palabra meta, como pista se da un sinónimo de la palabra meta. En *reconocimiento del significado* también se escoge entre cuatro opciones el sinónimo más adecuado para la palabra meta. En la *Tabla 18* se puede ver un ejemplo de cada uno de los ítems. Se buscó que la frase del contexto fuera igual en las cuatro modalidades buscando la consistencia interna.

	Producción	Comprensión
Recuperación	She is a _____ girl. (small)	'She is a little girl' means that she is _____
Reconocimiento	She is a _____ girl. (small) a. little b. great c. nice d. single	'She is a little girl' means that she is _____ a. small b. great c. nice d. single

Tabla 18. Muestra de ítems de CATSS, versión monolingüe (Aviad-Levitzky et al., 2019)

En cuanto a la validez, en la última versión se dio un paso más y se adaptó el test de vocabulario a la nueva tendencia. Así, la validez se guía por el enfoque basado en argumentos. Se utilizaron los siguientes métodos estadísticos: alpha de Crombach, modelo Rasch y ANOVA para comprobar la validez y fiabilidad de la herramienta. Sobre las preguntas de investigación, al igual que había pasado con la primera versión, la diferencia entre las dos modalidades de reconocimiento no es significativa, de modo que la escala de dificultad no es totalmente escalar. En la segunda versión, en cada modalidad se obtuvo un resultado diferente; por lo tanto, la forma de preguntar, el tipo de ítem, la relación entre proceso cognitivo y destreza influyen en el conocimiento. Pese a esto, se concluye que se obtienen mejores resultados en comprensión que en producción y en reconocimiento que en recuperación. Así, quedan perfectamente definidos los polos: producción recuperación, como el más difícil, y comprensión reconocimiento como el más sencillo. La jerarquía en el centro de la escala es más difusa o menos clara, pero aun así parece que es el procesamiento cognitivo el que inclina la jerarquía dejando la comprensión-recuperación como más difícil que la producción-reconocimiento. Además, los resultados de cada modalidad no varían en cada nivel, de modo que la jerarquía no es dependiente de la frecuencia o nivel de la palabra. Estos resultados se asemejan a los obtenidos en nuestro TCE, que es la adaptación de este modelo a las colocaciones. El argumento central de la propuesta reside en que el componente principal del conocimiento de las palabras, que es el enlace forma-significado, sólo puede considerarse una jerarquía en cuatro grados de solidez, que se mantiene en los cuatro niveles de frecuencia y mantiene la correlación en los cuatro grados o modalidades.

En resumen, en todas las versiones de CATSS se mantiene como centro del estudio el enlace entre la forma y el significado, el conocimiento como la sucesión de cuatro modalidades escalares y una puntuación diferente para cada acierto en cada modalidad. Las diferencias con el original se resumen en los siguientes puntos: uso de catorce bandas secuenciales, uso de listas de inglés británico y americano, uso del contexto mínimo en

los ítems, el número de letras de la pista varía en función de las necesidades y el *feedback* que se le ofrece al informante: tanto el número de palabras conocidas como el grado de conocimiento de cada palabra. Algunos estudios defienden que la diferencia entre los cuatro grados viene dada por la sobreestimación de los formatos de opción múltiple, sin tener en cuenta el efecto del azar (Stewart, 2014).

Entre los aciertos de CATSS está el paso que supone la propuesta de una categorización del conocimiento del vocabulario y la demostración de que el conocimiento se puede jerarquizar. Pese a que algunas de las limitaciones que los propios autores identifican en las primeras versiones, como por ejemplo la sustitución de las tareas de traducción por los contextos reducidos, de modo que los aprendices a los que se les administra el test no tienen por qué tener una misma lengua materna, sino que puede variar, otras todavía no parecen estar resueltas. Por ejemplo, la relacionada con la jerarquización y la categorización. Dejan claro los límites, pero las diferencias entre los dos formatos o niveles de reconocimiento no son significativas. Esta falta de diferencia significativa entre el reconocimiento desde la producción o desde la comprensión puede que se deba a otros factores además del enlace entre la forma y el significado (Schmitt, 2010). Otra de las limitaciones por resolver de CATSS sería la confusión terminológica (Schmitt, 2010) a la que se propone una alternativa que se considera más descriptiva y transparente: modificación la dualidad comprensión-producción por significado y forma. Es decir, dos habilidades por dos componentes del conocimiento léxico. En definitiva, parece que no solo cambia el término, sino que cambia el foco, aunque se mantienen la correspondencia.

3.2.3. Test de vocabulario en español

En el contexto de ELE, la evaluación del léxico no tiene el recorrido que tiene esta materia en el estudio del inglés como L2 (Barcroft et al., 2022). La mayor parte del material encontrado pertenece a test de disponibilidad léxica (Bartol Hernández, 2010; Jiménez-Catalán, 2017; Kiriakí Palapanidi, 2012; Santos Díaz et al., 2020) y test

certificativos, como el DELE⁴⁶, SIELE⁴⁷ o CERTACLES⁴⁸, que no cuentan con una prueba específica del léxico sino que incluyen el léxico como una competencia integrada en otras como la comprensión auditiva o la expresión e interacción escrita. Con todo, existen algunas propuestas de test que miden la competencia léxica con fines específicos (Bowles, 2022)⁴⁹; test adaptativos que miden el conocimiento léxico del español de México en niños (Kachergis et al., 2022); una propuesta de cómo hacer un test y propone un ejemplo de test de Sí / No (Rodríguez-Sánchez, 2022), pero que únicamente se construyó pero que no se administró. También contamos con la versión de X-LEX (Meara & Milton, 2003) en español, que hasta este momento no ha sido publicada. Otros test se usan como complemento de otras investigaciones (Pulido, 2007). En este caso, se presentan dos modelos de traducción, uno de producción (traducción directa) y otro de comprensión (elección múltiple) para corroborar el análisis previo realizado a partir de corpus y que tiene por objeto comprobar el valor de la interferencia en el aprendizaje del léxico, en relación con la comprensión lectora.

Un ejemplo claro de adaptación de los modelos del inglés al español es el SVLT (*Spanish Vocabulary Levels Test*) (Chandler, 2022) que sigue el propósito y estructura del VLT, un test que mide la extensión a partir de la comprensión. Basado en listas de frecuencia, extrae las palabras meta del corpus *A Frequency Dictionary of Spanish* (Davies, 2006). Se mantiene el mismo formato: seis palabras y tres definiciones, estas definiciones deben relacionarse con las palabras correspondientes de entre las ofrecidas (vid. Figuras 22 y 23), pero con algunas modificaciones como el número de ítems testados en cada banda (de 18 se pasa a 24). El criterio de corrección consiste en la multiplicación del número de aciertos por 42. También se presenta una versión que mide la comprensión, PVLT (*Productive Spanish Vocabulary Level Test*), conocimiento productivo controlado,

⁴⁶ Pruebas certificativas organizadas por el Instituto Cervantes, que cuentan con el sello de calidad ALTE y siguen las directrices del MCER y del Manual para relacionar exámenes al MCER. Además, del PCIC. Cubren todos los niveles del MCER, desde el A1 hasta C2, incluidas dos versiones para escolares (A1 y A2/B1). Con exámenes des. <https://exámenes.cervantes.es/es/dele/que-es>

⁴⁷ SIELE <https://siele.org/examen>

⁴⁸ Pruebas certificativas organizadas por ACLES. Cubren niveles desde el A2 hasta el C2. Las destrezas incluidas son las siguientes: comprensión auditiva, lectora, interacción oral, expresión oral y escrita

⁴⁹ Se trata de un test con fines muy concretos y en un contexto muy determinado, en EEUU en donde la inmigración de hispanohablantes es muy elevada. Las universidades necesitan realizar test de clasificación a alumnos que eligen español como L2 o L3 y que algunos son aprendices nuevos y otros aprendices de lengua heredada

siguiendo la propuesta del del PVLТ. El principal problema de esta propuesta radica en la presentación de los resultados ya que no se presentan datos concretos como el número de participantes que permitirían generalizar o no los resultados obtenidos. Entre las conclusiones presentadas está la de que, a mayor nivel, mejores resultados en el plano léxico.

1.	Destino	
2.	Estudio	___ cambio
3.	Influencia	___ fin
4.	Paso	___ poder
5.	Reforma	
6.	Situación	

Figura 22 Ejemplo de ítem de SPVLT

1.	Antes de acostarme me gusta leer un buen li ___
----	---

Figura 23. Ejemplo de ítem de PVLТ

Por lo tanto, los trabajos sobre evaluación del léxico en español todavía se encuentran en desarrollo y sin duda, la mayor aportación al estudio del estudio del español como L2 viene desde los estudios de corpus (Berton, 2020; Díez-Ortega & Kyle, 2023; Gouverneur, 2008) desde los que se plantea diferentes cuestiones sobre si las conclusiones extraídas de lenguas como el inglés como L2 son aplicables al español como L2.

3.3. La evaluación de las colocaciones

Las conclusiones extraídas sobre diversos factores que condicionan la evaluación del léxico se puede extrapolar al aprendizaje de colocaciones (González Fernández & Schmitt, 2015). En nuestro caso también cabría preguntarse si es posible extrapolar las conclusiones extraídas de la evaluación del léxico del inglés como L2 al español como L2. En líneas generales podremos aplicar los procedimientos de la evaluación del léxico del inglés a la evaluación de las colocaciones del español en gran medida. Además, contamos con importantes avances en la evaluación de colocaciones, aunque en muchos casos los estudios no se centren en las unidades tal y como las entendemos en este trabajo sino en unidades fraseológicas de carácter más amplio.

Incluimos en este repaso los test que comparan la competencia colocacional, desde la comprensión o la producción, incluso ambas, con la competencia léxica o con la competencia lingüística. Debemos tener en cuenta también test que tratan las colocaciones de forma indirecta. Por un lado, los test longitudinales que incluyen entre sus baterías de test uno o dos test de colocaciones (González-Fernández & Schmitt, 2020). Por otro lado, test de asociaciones de palabras que incluyen las colocaciones como uno de los criterios para detectar relaciones sintagmáticas (Eguchi et al., 2022; Fitzpatrick, 2013; Fitzpatrick & Munby, 2014; Fitzpatrick & Thwaites, 2020; Schmitt & Meara, 1997; Wolter, 2002; Zareva & Wolter, 2012). Dejamos de lado también algunos test que, aún con fines pedagógicos, se centran más en cuestiones relativas al procesamiento (Sonbul & Siyanova-Chanturia, 2019; Wolter & Gyllstad, 2013; Wolter & Yamashita, 2014, 2018), desde una perspectiva psicolingüística, y que, por otro lado, hemos ido referenciándolos al hablar de conocimiento y procesamiento de las colocaciones. Por su parte, en el ámbito del español, hasta el momento predominan los trabajos en los que se aborda la evaluación de competencia colocacional a partir del análisis de corpus (Almela et al., 2013; Almela & Cantos, 2018; Alonso-Ramos, Wanner, Vázquez Veiga, et al., 2010; García-Salido, 2016; Vincze et al., 2016).

No hay duda de que para conseguir un uso fluido de la lengua se necesita dominar la competencia colocacional ni que para llegar al nivel de uso de lengua de un nativo es imprescindible contar con este dominio. Sin embargo, hay estudios en los que no se aprecia una diferencia significativa en cuanto a la *densidad* de colocaciones entre el hablante nativo y el aprendiz (Orol-González et al., 2013; Vincze et al., 2016). Otro de los principios es que las colocaciones son un desafío para los aprendices, más debemos precisar que no siempre: las colocaciones no suelen suponer un problema a nivel de comprensión, pero si pueden suponerlo en la producción. Pese al número de estudios, todavía no hay test totalmente estandarizado que mida las colocaciones en inglés ni tampoco en español como L2. Aún sin haberse convertido en estándar, COLLEX y COLLMACH (Gyllstad, 2009a), se pueden considerar referentes en la evaluación de las colocaciones del inglés. A partir de la aportación del autor, otros test de colocaciones han seguido avanzado y perfeccionando los métodos de creación.

En primer lugar, al igual que hemos hecho con los test del léxico, presentaremos una clasificación de los test de colocaciones, en este caso también tendremos en cuenta el aspecto cronológico. Después nos detendremos en la descripción de los test dedicados a la evaluación de colocaciones. No conocemos ningún test que se centre en la evaluación de colocaciones del español como L2, aunque sí hay test que evalúan colocaciones con aprendices con español como L1, de ahí la necesidad del TCE que cumple con estos requisitos.

3.3.1. Clasificación de los test de colocaciones

En este punto vamos a hacer un breve análisis de algunos de los test que se han diseñado para medir la competencia colocacional en inglés y que nos servirán como base y referencia para el diseño y construcción del TCE. Son varios los trabajos que han dedicado un espacio a la revisión de los test previos de colocaciones (Gyllstad, 2009a; Moreno-Jaén, 2007b), aun así, debemos tener en cuenta que ambos son de los años dos mil, hace más de una década, en la que se ha trabajado ampliamente en este campo, de ahí que también sean relevantes los análisis de los últimos trabajos (Gyllstad & Schmitt, 2019).

Para el análisis de cada uno de los test atenderemos a varios aspectos que enumeramos a continuación y, además, en los ANEXOS I y II contamos con dos tablas que recogen de forma más esquematizada la información de cada uno de los test (nombre, año, autor, comprensión-producción, proceso cognitivo, formato o valores comparados).

- 1) La definición de colocación que se utiliza en cada test. Aunque sea un aspecto poco tratado en los análisis, es de vital importancia, ya que diferentes definiciones implican diferentes formas de identificación y, por lo tanto, unidades analizadas diferentes.
- 2) Los objetivos de cada test, centrándonos en si la evaluación es de la comprensión o la producción y, sobre todo, con qué habilidad se compara la competencia colocacional (con la competencia léxica o con la competencia lingüística) o qué factor de aprendizaje influye más (frecuencia, clase de palabra o dificultad).

- 3) La selección de colocaciones, íntimamente relacionado con la definición y la identificación.
- 4) La población, apartado en el que veremos pocas diferencias entre los distintos test, tanto en cuanto al nivel de competencia testada como a otras características de los participantes.
- 5) El formato, aspecto en el que encontramos bastante heterogeneidad, pero lo que también dificulta la comparación entre los diferentes test (Schmitt, 2019a), sobre todo a la hora de extrapolar o generalizar las conclusiones extraídas.
 - a. Si miden la comprensión o la producción.
 - b. Si miden desde el reconocimiento o desde la recuperación.
 - c. Si se centran en la identificación de la colocación o en alguno de los aspectos de su conocimiento.
- 6) Las conclusiones a las que han llegado los análisis de resultados de cada test y que serán los puntos de partida, algunos contradictorios, en los que se sientan las bases del desarrollo de nuevos test y referentes en la enseñanza-aprendizaje de colocaciones.
- 7) También, las limitaciones de cada instrumento de evaluación.

3.3.2. Descripción de una selección de test de colocaciones

En este apartado revisaremos una selección de test que miden la competencia y el conocimiento colocacional. Debemos aclarar una cuestión terminológica: el concepto de colocación que utilizan estos test no siempre coincide con el que se utiliza en este estudio, normalmente es un concepto más amplio. Además, también hay casos que ni siquiera se utiliza el concepto de colocación, sino uno más amplio como el de *relaciones sintagmáticas*, *construcciones formulaicas* o *unidades fraseológicas*, etc. En todo caso, todos los test que vamos a describir se corresponden con uno de los nueve componentes del conocimiento léxico: las colocaciones (Nation, 2013b). Además de los test que miden el conocimiento de colocaciones en exclusividad, también nos encontramos con test, que siguiendo la perspectiva multicomponencial de la evaluación de lenguas, crean baterías de test entre los cuales se incluyen las colocaciones, o una categoría en la que se incluyen

las colocaciones, como un componente más (Chui, 2006; González-Fernández & Schmitt, 2020; Nontasee & Sukying, 2021).

Realizamos el análisis de los test de colocaciones existentes centrándonos, sobre todo, en dos puntos. Primero, la forma de selección de la colocación para la muestra, lo que nos permitía conocer el enfoque para la definición de colocación, que, en la gran parte de las ocasiones, se alejaba del que adoptamos en este trabajo. Segundo, los formatos elegidos para la evaluación. Esto nos permitió ilustrar la información que se demandaba en cada uno de ellos, así como la forma de acceso a esta información. Para la descripción de los formatos tuvimos en cuenta principalmente tres aspectos: la destreza (la comprensión y la producción), el proceso cognitivo (el reconocimiento y la recuperación), y, por último, la habilidad testada (la identificación de la combinación o el reconocimiento del enlace entre la forma y el significado). Entendemos por *identificación* la habilidad de discernir si una combinación de palabras es habitual en la L2, es decir, si existe y se considera colocación según los criterios preestablecidos, pero no se demuestra su capacidad para entender su sentido, tal y como ocurre en los test que sí miden el conocimiento.

Distribuimos los test de forma cronológica ordenándolos en tres grupos o generaciones. Los límites vienen dispuestos a partir de una serie de publicaciones que comparten tiempo de publicación con COLLECH Y COLLMACH, apareciendo varios publicados en el manual, *Researching collocations in another language* (Barfield & Gyllstad, 2009), o contemporáneos de estos por el avance que supusieron con respecto a los trabajos anteriores y la impronta que dejaron en los posteriores, marcando una clara tendencia. El grupo de test previos a ese momento pertenecerán a la *generación cero* (Bonk, 1996; Gitsaki, 1996). Los contemporáneos a *Researching Collocation in another language* (Gyllstad, 2011) conformarán lo que hemos denominado *primera generación*. Los trabajos que se publicaron con posterioridad y que tienen como referentes los previos, formarán parte de la *segunda generación* (Brown, 2018). En el ámbito de ELE contamos con un número muy reducido de test que miden la competencia colocacional (De Alba Quiñones, 2016) en el que se atiende a la variable de género con un número reducido de participantes.

3.3.2.1. *Primera generación*

Researching Collocations in Another Language (Barfield & Gyllstad, 2009) es el manual que marca la tendencia de esta primera generación de test que tiene por objetivo diseñar una herramienta de evaluación válida y fiable. Incluimos cinco test, tres de ellos se incluyen en este manual (COLLEX y COLLMACH⁵⁰, CONTRIX y DISCO) y los otros dos se publicaron al margen de este manual (LexCombi y ADELEX).

A. COLLEX y COLLMACH

COLLEX y COLLMACH son el nombre de los test que Gyllstad desarrolla en sus tesis (2007a) y posteriormente en un artículo (Gyllstad, 2009a) del libro sobre colocaciones que él mismo editó junto a Barfield. El objetivo no es pedagógico, por lo menos no directamente, sino que se enmarca en el ámbito de la evaluación del léxico, de forma general, y, más concretamente, dentro de la evaluación de diagnóstico. Se trata de los primeros test que definen claramente el constructo, tienen un número de participantes que permite la generalización y las medidas estadísticas ratifican las dos premisas básicas de una buena evaluación: la fiabilidad y la validez. Podríamos decir que es el test que marca tendencia, aquel que los posteriores siguen o intentan mejorar.

La definición del constructo ocupa una parte importante del estudio y sigue la propuesta tripartita que define el constructo en tres fases: *teórica, operativa y procesual* (Bachman & Palmer, 1996). Se define la colocación como dos palabras ortográficas relacionadas sintácticamente y que coocurren con frecuencia (Gyllstad, 2009a). Esta definición es válida, pero no suficiente para la evaluación; de modo que, a esta definición añade la idea de comprensión que va a utilizar para elaborar una *definición operativa*. Esta atiende al constructo colocacional comprensivo: conocimiento necesario para reconocer dos o más palabras que coocurren juntas frecuentemente como combinaciones de palabras convencionalizadas en la lengua y accediendo a su significado en algún sentido (Gyllstad, 2009a).

⁵⁰ Incluyo COLLEX y COLLMACH como un único test porque son modelos de una misma investigación como explico en el apartado correspondiente.

El propósito se centra más en el diseño del instrumento de evaluación que en la interpretación de los resultados que pueda ofrecer desde un punto de vista pedagógico. El primer objetivo consiste en saber si los test son fiables y consistentes. El segundo, en conocer si la herramienta de medida tiene capacidad para discriminar entre hablantes nativos y no nativos de la lengua y entre aprendices de diferentes niveles en relación con su conocimiento colocacional. El tercero busca relación entre la competencia léxica y la comprensión de colocaciones.

Para la selección de la muestra toma un enfoque centrado en las palabras, es decir, que atiende a la frecuencia de los elementos constituyentes de la combinación, pero utiliza una metodología diferente para cada una de las muestras de los test o formatos que utiliza. En COLLMACH la elección se realiza en dos fases: en la primera se seleccionan los verbos de una lista de palabras frecuentes; en la segunda, se seleccionan los colocativos. La muestra se compone de setenta colocaciones y treinta *pseudocolocaciones*⁵¹. En COLLEX se cambia el criterio y mediante un método estadístico se busca la frecuencia de la combinación, en este caso, a partir de z-score en el BNC (*British National Corpus*). Los distintos criterios para la selección de la muestra hacen que se reduzca la fiabilidad de la prueba, ya que la falta de homogeneidad en los criterios es uno de los hándicaps de la investigación del léxico (Schmitt, 2019b).

Además de los dos test de colocaciones, previamente se administró un test de vocabulario, VLT (Nation, 1983, 1990; Schmitt et al., 2001), compuesto por ciento cincuenta ítems (vid. 3.2.2.1). Trescientos siete informantes de diferentes niveles, todos aprendices de inglés y con el sueco como L1, realizaron los cincuenta ítems de COLLEX y los cien ítems de COLLMATCH. No solo se hizo una versión piloto previamente, sino que en la tesis se muestra la evolución y evaluación de los formatos hasta la versión definitiva, quinta de COLLEX (con modificación del formato incluida) y tercera versión de COLLMACH. El test fue administrado a más de trescientos estudiantes suecos de inglés de nivel intermedio.

⁵¹ Combinaciones de palabras en la lengua meta que podrían haberse dado en la lengua pero que no dan y que funcionan como distractores.

En ambos test se mide la comprensión desde el reconocimiento, la colocación correcta se presenta como *input*. COLLEX, que viene de la frase inglesa *collocating lexis*, consiste en la unión de los colocativos, es decir, un formato de elección forzada. Esto quiere decir que se elige entre combinaciones ya construidas. Se presentan tres secuencias de *verbo + nombre*, las cuales comparten la base, entre las que hay dos distractores (dos *pseudocolocaciones*) y una colocación (combinación convencionalizada), entonces, el objetivo es elegir la buena colocación. Las instrucciones se ofrecen en la L1 de los aprendices para que no haya ningún tipo de confusión a la hora de realizar la tarea. En la *Figura 24* se puede ver un ejemplo.

1	a. receive a cold	b. achieve a cold	c. catch a cold
---	-------------------	-------------------	-----------------

Figura 24. Ejemplo de ítem de COLLEX 5 (Gyllstad, 2009)

COLLMATCH evolucionó como formato en las sucesivas versiones. En una primera fase se presentaba una parrilla en la que el aprendiz debía escoger el colocativo (o los colocativos) más adecuados para una base dada. En una segunda fase se presentó un formato de sí o no en la que el aprendiz debía decidir si la secuencia es o no colocación, aunque se evita usar el término *colocación* y se pregunta si la combinación de palabras es natural, es decir, si está convencionalizada. Ambos formatos presentan las colocaciones descontextualizadas porque así es posible un mayor número de ítems y también es válido para informantes con diferentes L1 (vid. Figura 25).

1	raise objections	2	bear witness
<input type="checkbox"/>	yes	<input type="checkbox"/>	yes
<input type="checkbox"/>	no	<input type="checkbox"/>	no

Figura 25. Ejemplo de ítem de COLLMATCH 3 (Gyllstad, 2009)

Con relación a los propósitos relacionados con la herramienta en sí, se comprueba tanto la validez como la fiabilidad. El test es fiable en términos de consistencia interna, pero no tanto según la TCT (teoría clásica del test), de modo que una de las limitaciones son los ítems y, por lo tanto, este es uno de los propósitos de mejora. El análisis estadístico también confirma la validez del test, aunque con unos valores límite que se atribuyen a la selección de las colocaciones meta y el componente de azar o adivinación de los ítems, en los que, además, no se puede observar de forma explícita el

conocimiento, sino que la información analizada es la auto-información del participante. En cuanto a los propósitos relacionados con el conocimiento colocacional, las correlaciones son significativas. Tanto la variable del conocimiento léxico como el componente discriminatorio muestran los resultados esperados. Es decir, el conocimiento léxico está correlacionado con el conocimiento colocacional y es posible discriminar entre nativos y aprendices por su conocimiento colocacional.

B. CONTRIX

La prueba CONTRIX (*New Test of Whole English Collocations*) rompe con la tendencia general de la época y no evalúa la comprensión del conocimiento colocacional. Se trata de un test de producción desde el reconocimiento (Lee Revier, 2009).

Se parte de la colocación como un constructo independiente, en la línea seguida por otros trabajos (Nation & Webb, 2011). De modo que considera las colocaciones como ítems léxicos de pleno derecho con propiedades formales, semánticas y de uso (igual que las palabras). Así, define la *colocación* como una combinación recurrente de palabras que forman una unidad. Posteriormente, amplió esta definición añadiendo propiedades semánticas. Pese a que la inclinación es al enfoque fraseológico, en la definición no se hace explícito y se alude al término recurrente que se asocia más al de frecuente que a los rasgos definitorios propios del enfoque fraseológico. Sin embargo, establece tres subcomponentes del conocimiento colocacional que describe según el grado de *transparencia* (u *opacidad*) diferenciando entre colocaciones *transparentes*, *semi-transparentes* y *no transparentes*. También especifica el sistema de conocimiento productivo que requiere dominio de los constituyentes léxicos centrales (conocimiento de cada una de las palabras que integran la combinación), el significado combinado y los elementos gramaticales. Pese a que la definición no se adapta al marco teórico que seguimos en este trabajo, es próxima y útil en la medida en la que propone una clasificación de los componentes necesarios para dominar la producción de una colocación.

El propósito del test es atender la necesidad de evaluar las colocaciones desde la producción y como elementos del léxico independiente. Entre otros objetivos está el de comprobar la fiabilidad y validez de la herramienta, comprobar si discriminaba los distintos niveles de competencia de los aprendices y si el criterio semántico de

transparencia influye significativamente en los resultados. Además, se entregó un cuestionario previo y un test de vocabulario. La elección de la muestra de cuarenta y cinco ítems se rigió por tres ideas: la representatividad, la inclusión de las tres categorías semánticas (transparente, semitransparente y no transparente) y la frecuencia, ya que las combinaciones debían encontrarse dentro de los cien millones del BNC.

Para la creación de los ítems, parte de la idea de que la evaluación de las colocaciones ha de centrarse en el reconocimiento y la producción de toda la combinación, no en el reconocimiento y producción de sus componentes de forma separada. Por lo tanto, en el formato propuesto el aprendiz no debe elegir un colocativo para una base dada, sino que debe elegir una base apropiada y el colocativo adecuado para ella en el contexto ofrecido. En este punto parece que la definición se acerca más a la de otras combinaciones fraseológicas que no son composicionales. El formato busca la identificación de la colocación descontextualizada para hacer el test apto para aprendices de cualquier L1 y consiste en una oración con un hueco (*sentence cloze*) al que incorporar la colocación. En la parte derecha de cada oración se encuentra una parrilla con los elementos de la colocación y varios distractores que el aprendiz debe identificar y montar. En el diseño se han tenido en cuenta cuatro puntos: selección de los ítems, escritura rápida de oraciones, selección de los distractores, norma de los hablantes nativos.

The quickest way to win a friend's trust is to show that you are able to <input type="text"/> .	tell	a/an	joke
	take	the	secret
	keep	—	truth

Figura 26. Muestra de ítem (Lee Revier, 2009, p. 129)

Se asume la fiabilidad y validez del test a partir de análisis estadísticos, con ciertas carencias en la consistencia interna debido a que el test está en un estado inicial de desarrollo. Las conclusiones a las que se llega son las siguientes: el test es capaz de distinguir entre los aprendices de diferentes niveles y discrimina entre los niveles de conocimiento: transparente, semi-transparente y colocaciones no transparentes, aunque las correlaciones no son significativas. Entre las limitaciones destacan la pequeña población y también el reducido número de ítems. En CONTRIX encontramos dos problemas: pese a que consideramos la colocación como un constructo independiente que se puede medir de forma aislada con respecto al léxico, en ningún caso las

colocaciones se consideran unidades léxicas, como sí lo son otras combinaciones fraseológicas, sino que son de carácter composicional y, por lo tanto, están formadas por al menos dos unidades léxicas. Además, el formato elegido está muy alejado de la realidad y no resulta natural, los aprendices no suelen encontrarse con un conjunto de piezas a “montar” como si se tratara de un puzle

C. DISCO

DISCO, *Discrimination Collocation Test*, (Eyckmans, 2009) implica un estudio longitudinal que mide la comprensión de colocaciones. La noción de colocación se basa en el concepto de Nesselhauf (2005a) de restricción arbitraria en la capacidad para ser sustituido.

Frequently co-occurring V+N combinations that are different from free V+N combinations in that there is a restriction on the substitutability of their parts (Nesselhauf, 2005).

Se parte de un criterio de frecuencia para la selección de colocaciones, pero en lugar de tener en cuenta la frecuencia de los componentes de la colocación por separado, como hacen otros test, atiende a la frecuencia de ocurrencia de la combinación a partir de modelos estadísticos, tratando la colocación como un *chunk*. Además, como veremos en la selección de las combinaciones, parten del nodo, que identifican con verbo, para elegir la combinación, por lo que la restricción que presenta es sintáctica y no léxica. Se establece la relación del conocimiento comprensivo con el uso comprensivo del lenguaje fraseológico y de su competencia lingüística. Se plantean dos cuestiones: primero, si el test de colocaciones discrimina lo suficiente para reflejar el progreso. Segundo, si el método usado permite conocer el grado de conocimiento productivo.

Veinticinco estudiantes belgas de inglés realizaron el test. Al tratarse de una prueba de carácter longitudinal, hicieron el test en dos ocasiones. La primera, antes de un curso de inglés de sesenta horas; la segunda, tras finalizar el curso. Además, realizaron un test de extensión de vocabulario que mide la comprensión (Eyckmans, 2009) que combina palabras y pseudopalabras para descubrir cuáles son las que el aprendiz conoce. El test tiene cincuenta ítems de colocaciones *verbo + nombre*. La muestra final, cincuenta ítems, contiene quince colocaciones de frecuencia alta, quince de frecuencia media y

veinte de frecuencia baja. Se eligió un formato de elección múltiple con tres opciones: dos colocaciones, elegidas de la misma banda de frecuencia, y como distractor una combinación libre. La habilidad medida es la de distinguir colocaciones de combinaciones libres.

Las medidas estadísticas avalan la fiabilidad del test. Pese a que el eje central de la prueba es la de validar la herramienta de evaluación, también atienden a otras cuestiones que se ven limitadas por el bajo número de participantes y que, por lo tanto, le impide generalizar los resultados. Posteriormente se realizó un test que comparó la competencia colocacional con la producción oral con aprendices daneses de inglés y español (Stengers et al., 2011).

D. LexCombi

LexCombi (Barfield, 2009) mide la producción desde la recuperación de colocaciones a partir de una adaptación de un test de asociaciones, adaptado de Lex30 (vid. 3.2.2.4). No se da una definición de colocación como tal, pero tanto por los objetivos como por el método de selección podemos inscribir este test dentro del enfoque basado en frecuencias.

Todos los objetivos están directamente relacionados con la frecuencia. En primer lugar, se plantean si los perfiles léxicos de aprendizaje se parecen a los de la producción de colocaciones. Segundo, si existe diferencia entre los diferentes niveles de competencia lingüística. También, qué clase de palabras usan los aprendices con más frecuencia para formar colocaciones. En síntesis, investiga la relación entre perfiles léxicos de frecuencia, nivel de la competencia lingüística y la clases de palabras. El objetivo final es la creación de un mapa que recoja cien posibles respuestas para cada palabra meta, ítem.

La elección de la muestra se desarrolló en dos fases. En la primera se seleccionaron cincuenta sustantivos extraídos de la lista lematizada extraída del BNC. Los treinta elegidos fueron los que mejor discriminaron entre aprendices y nativos. En el segundo paso, se buscaron posibles respuestas en dos diccionarios de colocaciones. Una vez recogidas estas posibles respuestas se subdividieron entre colocativos frecuentes (siempre entre las dos mil más frecuentes) y colocativos no frecuentes. El test se repartió

en papel y el tiempo utilizado en hacer lo fue de veinte minutos aproximadamente. Los aprendices comprenden tres niveles de competencia y todos tienen como L1 el japonés. Previamente, se realizó un test piloto. Ochenta y nueve participantes con un tiempo delimitado para cada prueba.

Como ya avanzamos, se trata una adaptación del Lex30 (Fitzpatrick & Meara, 2004b; Meara & Fitzpatrick, 2000). El test se construyó con treinta ítems según un instrumento basado en la respuesta-estímulo. Es decir, se ofrece una palabra base y se solicita a los aprendices que escriban tres palabras que relacionen con ella. Un máximo de tres colocativos por palabra estímulo, como se puede ver en la imagen del ejemplo de un ítem, para eliminar el porcentaje de error en el análisis. Entre las ventajas, también está la facilidad para recoger la información en menos tiempo.

decision	1. _____	2. _____	3. _____
----------	----------	----------	----------

Figura 27. Ejemplo de ítem de LexCombi (Barfield, 2009)

En la corrección se comprobó si los colocativos propuestos por los aprendices coincidían con los previamente seleccionados como colocativos. Así se clasifican las palabras en apropiadas o no apropiadas y en frecuente o infrecuente. Para analizar la dificultad de cada ítem se utilizan diferentes medidas estadísticas: *t-scores* y *Rasch* análisis. En la corrección tampoco se tuvieron en cuenta los errores tipográficos. La primera conclusión es que existe correlación entre la competencia lingüística de los participantes y los buenos resultados en el test, es decir, en su conocimiento o competencia colocacional. Los aprendices de nivel avanzado presentan un mayor porcentaje de colocaciones apropiadas, entendiendo por apropiadas una serie de colocaciones esperables.

En la elaboración del test se propone como reto la superación de una serie de críticas anteriores a trabajos sobre evaluación de lenguas: la pequeña cantidad de datos analizados, el foco en los errores más que en los aciertos y en los aprendices iniciales más que en los avanzados. Con la herramienta de evaluación de las colocaciones se permite recolectar un amplio número de colocaciones de forma eficiente, se centra solo en lo que los aprendices saben y puede usarse para aprendices de diferentes niveles. La principal limitación de este estudio es que el concepto de colocación no se define. Además, las

respuestas de los aprendices pueden ser colocaciones o no. Pero no solo eso, sino que la relación que establece con la palabra base puede ser sintagmática o paradigmática.

E. ADELEX

El test ADELEX (Moreno-Jaén, 2007a, 2007b), aunque ideado para medir competencia comprensiva de colocaciones, si atendemos a diferentes variables: el proceso cognitivo (reconocimiento o recuperación) y el contexto (presente, ausente o dado en la L1) algunos de los formatos puede considerarse de producción. La confusión viene por la identificación entre reconocimiento y comprensión, que como vimos anteriormente (vid.2.1.1) da lugar a equívocos. Resulta también interesante en la medida en la que la L1 de los participantes en el estudio es el español.

Posteriormente se creó TAI ADELEX⁵², un test adaptativo. La transformación de la herramienta lineal, frente a la adaptativa, supuso ventajas que superaron algunas de las limitaciones de su primera versión. En cuanto a la definición de colocación, no se entra en debates terminológicos, pero sí define el constructo atendiendo a rasgos formales y funcionales. Con relación a los rasgos formales, considera la colocación como una combinación de dos elementos (base y colocativo) que no comparte estatus lingüísticos, ya que la base mantiene autonomía semántica mientras que el colocativo no. Se alude en este punto a la restricción léxica, propia del enfoque fraseológico. Con relación a los rasgos funcionales, atiende a la frecuencia de la combinación. Esta dualidad sitúa al estudio en el marco del *enfoque combinado*, que se hará explícito en el diseño del método de identificación de la muestra seleccionada para testar.

Los objetivos del estudio son el conocimiento de la competencia colocacional de aprendices españoles de inglés y la construcción de un test válido y fiable para medirla. Además, fue necesaria la construcción de una lista de colocaciones frecuente que permitiera una selección de la muestra basada en frecuencias, lo que ocupa un gran espacio en la investigación. Se creó una base datos de 2688 colocaciones (Moreno-Jaén, 2009), combinaciones a partir de los 412 sustantivos más frecuentes. Presentan dos

⁵² El texto en el que se explica carece de autoría. Se encuentra en el repositorio de la Universidad de Granada: <https://wpd.ugr.es/~adelex/documentos/extractoMT.pdf> [Último acceso 20/08/23].

versiones de ADELEX previas a la versión adaptativa: en la primera versión del test ADELEX (Moreno-Jaén, 2007a) se administró a sesenta y tres alumnos y cuenta con dos formatos; en la segunda, se administró a trescientos once alumnos de cuatro centros universitarios españoles (Moreno-Jaén, 2009) y se utilizaron cuatro formatos. Explicaremos esto últimos, que fueron parcialmente modificados con respecto a la anterior versión.

En el primero de ellos, C-test, que mide la producción controlada desde la recuperación, se presenta la colocación contextualizada y falta el colocativo, pero del que se da la primera letra.

TASK 1

Fill in the blanks in each sentence by adding only one word. The word you need to add can only be an adjective, a noun or a verb. The first letter of each word is provided to help you.

Example:

0. N _____ light is preferable to artificial light.

Answer:

0. Natural light is preferable to artificial light.

Figura 28. Ejemplo de ítem del primer formato de ADELEX (Moreno-Jaén, 2007b, p. 301)

El segundo es un formato de traducción que mide la producción desde la recuperación. Se da la colocación en la L1 y deben escribir la traducción en la L2. De nuevo se mide la producción a partir de la recuperación, pero en este caso el procesamiento no se deduce a partir del contexto, sino que se presenta la combinación en la L1. No se traduce la colocación entera, la base ya aparece traducida, únicamente se debe recuperar la traducción del colocativo. Se trata de un formato útil para ver la influencia de la L1, pese a que en este test todos comparten la misma, pero no tan útil como podría. A la hora de seleccionar la muestra hay que tener en cuenta que no todas las colocaciones entre una L1 y una L2 tiene una traducción literal, de hecho, hay lenguas en las que una idea se formula de manera analítica, a partir de una colocación, o de manera sintética, a partir de una única palabra. En el mismo caso que se presenta en el ejemplo en español contamos con la versión analítica, *atender*, y con la versión analítica, *prestar atención*. En otros casos podría no darse la opción.

TASK 2

Translate the following collocations into English. Add either one single word or one hyphenated word in each case.

Example:

0. Prestar atención: To pay attention

Figura 29. Ejemplo de ítem del segundo formato de ADELEX (Moreno-Jaén, 2007, p. 302)

El tercer formato invierte el objetivo, en lugar de recuperar el colocativo de la base dada, el aprendiz debe reconocer la palabra que no forma colocación con la base dada. Este formato es uno de los que se eliminan en la versión final.

TASK 3

Choose the word which does not collocate with the noun given (there is only one wrong collocation in each case). If all of them are correct, tick the option “no wrong collocation”.

Example:

0. To a list

compile compose make no wrong collocation

Figura 30. Ejemplo de ítem del tercer formato de ADELEX (Moreno-Jaén, 2007b, p. 303)

El cuarto formato, contextualizado, también mide la producción, pero en este caso a partir del reconocimiento. La base de la colocación se presenta en una oración con un hueco para el que se debe elegir el colocativo de entre cuatro opciones (dos distractores y un “ninguna de estas”).

TASK 4

Choose the correct collocation and tick the appropriate box. There is only one correct collocation in each case. If none of the 3 first options is correct, tick the option “none of these”.

Example:

0. A place devoted to entertainment is called a/an area.

break game play none of these

Figura 31. Ejemplo de ítem del cuarto formato de ADELEX (Moreno-Jaén, 2007, p. 303)

El análisis de los resultados revela que el conocimiento colocacional de los aprendices universitarios de inglés es insuficiente (por debajo del 50%) y que, por lo tanto, es necesaria una intervención pedagógica eficiente que pueda atender las carencias de la competencia colocacional. El instrumento de medida permite discriminar entre los niveles que muestran una progresión escalar en los resultados, aunque entre dos de los niveles no hay una diferencia significativa. En cuanto a la estructura gramatical de las colocaciones, solo hay posibilidad de comparar dos tipos de combinaciones (V+N y N+A) entre las que tampoco se puede hablar de diferencia significativa sino una tendencia. Entre las limitaciones, la autora menciona la ausencia de variantes externas relevantes, como las estancias en los países de la L2, y la necesidad de un estudio longitudinal para comprobar el impacto de los procesos de enseñanza-aprendizaje. Pudiendo así realizar un estudio más profundo sobre las dificultades que entraña el dominio de la competencia colocacional.

El test muestra registros aceptables de validez y fiabilidad en tres de los cuatro formatos. Esto, junto con el hecho de que algunos de los ítems y distractores tuvieran que ser sustituidos según las diferentes medidas estadísticas, obligó a proponer una segunda versión del ADELEX Test de colocaciones. Esta versión se construyó en formato electrónico⁵³ (a través de la plataforma Moodle) y aunque se reserva para futuras investigaciones su administración y análisis de resultados, nunca se llegó a publicar nada en el que pudiéramos leer estos resultados.

3.3.2.2. *Segunda generación*

A esta generación pertenecen aquellos test que se publicaron a partir de 2010 y que buscan superar las limitaciones de los anteriores, sobre todo en el área de la validez y fiabilidad. Predominan, también, los formatos que buscan medir la producción, una de las carencias de los test previos sobre colocaciones. Aunque la frecuencia sigue estando presente en la definición de la colocación, muchos toman como guía para la identificación de las colocaciones meta el enfoque combinado. Con lo que las conclusiones extraídas

⁵³ Se accede a esta información a partir de un artículo publicado en la web de la Universidad Granada <https://wpd.ugr.es/~adelex/documentos/extractoMT.pdf> [último acceso: 02/02/23)].

sobre la unidad evaluada, las colocaciones, serán más próximas a lo que en esta investigación entendemos por colocación.

A. LexCombi2

El test de conocimiento productivo de colocaciones del inglés (Brown, 2014, 2018) no es completamente original, sino que se trata de una adaptación de *LexCombi* (Barfield, 2009). Toma la definición de colocación establecida por Durrant (Durrant, 2014) que combina tanto el enfoque fraseológico como el basado en frecuencias, es decir, se enmarca en el enfoque combinado. Esta definición se centra en dos criterios: la frecuencia y la dualidad opacidad-restricción (Brown, 2018). Considera este último punto esencial ya que el objetivo del test no es la identificación (reconocimiento de la forma) en la competencia colocacional sino un objetivo de carácter diagnóstico, desde una perspectiva pedagógica.

El propósito del test es principalmente salvar algunas de las limitaciones de los sistemas de evaluación precedentes, de modo que, mide la producción, centrándose más en los aciertos que en los errores, limitándose el número de ítems sin preocupación por los habituales problemas de administración (normalmente, el tiempo) y adaptándose a cualquier nivel de competencia. Además, atiende al análisis de las respuestas para detectar las diferencias entre colocaciones más que en las diferencias entre los aprendices. De modo que las preguntas de investigación versan sobre qué clases de palabras son las más recurrentes, cuáles de las repuestas más recurrentes son aceptables según el criterio de hablantes y según diversos métodos estadísticos: información mutua (MI) o la frecuencia.

El test consta de treinta palabras clave, todos nombres, que se corresponde con la información presentada a los participantes para obtener la respuesta. Las treinta palabras clave se seleccionaron nuevamente ya que las elegidas en el test original tenían varios problemas: sustantivos como verbos, *pistas malinterpretadas*, relación entre las palabras y rentabilidad pobre. De modo que se modificaron los criterios para hacer una elección más eficaz: la alta frecuencia, ocurrencia principalmente como sustantivos, tendencia a provocar respuestas paradigmáticas o ausencia de relación entre las palabras meta. Se detiene específicamente en recoger una lista de posibles respuestas, ya que la

coincidencia de respuestas, recurrencia, será la llave para considerar una combinación como error o acierto. El criterio para aceptar una respuesta como válida es la repetición recurrente de dicha respuesta, más concretamente, que haya un 10% de respuestas coincidentes entre las dadas por los participantes (Brown, 2013). Además, la corrección de estas posibles repuestas se comprueba con dos herramientas: un diccionario y la frecuencia extraída de corpus (t-score). Así las colocaciones pueden ser recurrentes (cuando varias respuestas coinciden) y a su vez pueden ser canónicas (cumplen con la frecuencia) o no canónicas (no se ajustan al corte de frecuencia estipulado).

Participaron en la investigación ciento veinticinco estudiantes japoneses de inglés con un nivel intermedio bajo y además de la adaptación de *LexCombi*, los participantes realizaron un test de extensión de respuesta cerrada sí / no. En *LexCombi2*, como se puede ver en el ejemplo del ítem (vid. Figura 32) se presenta una palabra y los aprendices deben recuperar todas las palabras que combinen con la palabra dada, es decir, todos los colocativos posibles para las bases ofrecidas. El formato original se adaptó con la idea de limitar el número de asociaciones (de tipo paradigmático) y priorizar el de colocaciones, el análisis en ambos formatos muestra que el cambio supone una mejoría al aumentar el número de colocaciones y disminuir el número de errores y respuesta perdidas.

_____ house _____	_____ house _____	_____ house _____
_____ issue _____	_____ issue _____	_____ issue _____
_____ research _____	_____ research _____	_____ research _____
_____ power _____	_____ power _____	_____ power _____

Figura 32. Ejemplo de formato de *LexCombi 2* (Brown, 2018, p. 201)

El principal hallazgo de esta investigación es que no se encuentra una relación significativa entre las dos variables: extensión del léxico y competencia colocacional. En cuanto a la categoría gramatical de las palabras que combinan con la base, el 70% de las palabras propuestas son sustantivos o adjetivos y el 30% son repuestas que pertenece a otra categoría gramatical. La aceptabilidad (el criterio de los hablantes nativos) visibiliza que más de la mitad (un 65% de las combinaciones) son aceptables según el criterio de

los hablantes nativos. Las razones que llevaron a la no aceptación van desde las limitaciones del diccionario, el entorno lingüístico o la aparición de palabras compuestas, escritas de forma conjunta. Este nuevo test, LexCombi2, se sitúa en el lado de los test que apuestan por no considerar la frecuencia como factor de la dificultad de las colocaciones, por lo menos no considerarlo determinante.

B. Test de conocimiento colocacional en relación con la competencia lingüística

En este apartado recogemos una serie de trabajos realizados por el mismo autor⁵⁴ (Nizonkiza, 2011, 2012, 2015; Nizonkiza et al., 2014) cuyo objetivo es la relación entre el conocimiento colocacional y la competencia lingüística. Además, propone dos versiones: una, para medir la comprensión; otra, para medir la producción. Pese a que en ningún trabajo presenta los dos test de forma conjunta o paralela, el hecho de que compartan un objetivo y mantengan la misma definición de colocación, hace que los abordemos de forma conjunta en este apartado. Aun así, como iremos viendo, existen diferencias entre ambas herramientas. La investigación comienza con la publicación de la tesis (Nizonkiza, 2011) y va evolucionando en diferentes publicaciones a lo largo de la década (Nizonkiza, 2012, 2015, 2017).

Toma el enfoque combinado como marco para definir el concepto de colocación. Esta decisión se refleja en el modo de selección de la muestra representativa de las colocaciones meta. Se tiene en cuenta la frecuencia como criterio base, pero no único, la composicionalidad y la restricción (incluso teniendo en cuenta la direccionalidad de esta) se presentan bajo criterio del hablante nativo. Se busca establecer la relación existente entre la competencia colocacional productiva o comprensiva, la competencia lingüística y la frecuencia. La hipótesis de partida es que tanto la frecuencia como la competencia lingüística mantienen una correlación positiva con el conocimiento comprensivo y productivo de las colocaciones. Los objetivos de este test también se inclinan hacia el ámbito pedagógico. Se pretende, por un lado, relacionar el conocimiento comprensivo y productivo de las colocaciones con la competencia lingüística de los aprendices. Por otro

⁵⁴ Desde la perspectiva de la comprensión (Nizonkiza, 2015) y desde la perspectiva de la producción (Nizonkiza, 2011, 2017).

lado, se busca relacionar la variación de la competencia comprensiva y productiva colocacional en función de la banda de frecuencia en la que se sitúan las palabras que componen la colocación.

La selección de colocaciones se realiza en dos fases. En la primera, se eligen las palabras de dos listas de palabras frecuentes y selecciona diez palabras de cada una de las bandas de frecuencia. En la segunda, se seleccionan los colocativos en un diccionario, el *Oxford Collocation Dictionary*. Además, busca tanto la frecuencia como el grado de coocurrencia de cada combinación resultante en corpus. La población a la que se administró el test está integrada por aprendices de inglés de Burundi y de Bélgica. Además, se incluyen diferentes niveles de competencia. Primero se hizo un test piloto y, además del test de colocaciones, se administró un test de nivel lingüístico, el TOELF, que permitió correlacionar las dos competencias y así cumplir con uno de los objetivos propuestos.

Para la comprensión utiliza un formato de opción múltiple en el que la se omite el colocativo y el aprendiz debe elegir entre varios a partir de una oración en la que aparece la base. Para la producción se utiliza un test que mide la producción de forma controlada, inspirado en el VLT, *Vocabulary Level Test* (Coxhead et al., 2015). En él se presenta una colocación contextualizada de la que se elimina el colocativo y el aprendiz debe recuperar. Al tratarse de un formato controlado se ofrecen las dos primeras letras de la palabra meta, en este caso, siempre del colocativo. Para la elección de los ítems se apostó por una adaptación de ejemplos recogidos del BNC. El formato elegido para medir la comprensión es de elección múltiple⁵⁵. La base de la colocación está contextualizada y falta el colocativo que los aprendices deben elegir entre varias opciones. Los distractores son *pseudocolocaciones*.

Instruction: Complete the underlined words in the sentences below.

Example: She is conducting campaigns to at..... new clients.
She is conducting campaigns to attract new clients.

Figura 33. Ejemplo de ítem de producción (Nizonkiza, 2012)

⁵⁵ Al contrario que en los artículos que miden la producción, en el artículo dedicado a la evaluación de la comprensión no se facilita un ejemplo del formato.

Una de las limitaciones es que no establece relación con la competencia léxica, como sí hacen otros trabajos (Brown, 2018; Gyllstad, 2009a) y como haremos también en el TCE. Los resultados muestran que existe una correlación positiva / significativa entre la competencia colocacional comprensiva y la competencia lingüística. Aunque la correlación se muestra más fuerte en los niveles más bajos. También se muestra que el conocimiento comprensivo de las colocaciones varía en función de la banda de frecuencia de las palabras constituyentes. Otra de las limitaciones del test es que en ambos formatos se mide la producción, pero en uno a partir del reconocimiento (lo que el autor clasifica como de comprensión) y en otro a partir de la recuperación (lo que el autor clasifica como producción) (vid. 2.1.1 para aclarar este punto).

C. Test productivo de colocaciones

El test productivo de colocaciones (González Fernández & Schmitt, 2015) es un test enfocado también hacia el ámbito de enseñanza-aprendizaje, es decir, con fines pedagógicos y cuyos participantes tienen como L1 el español. Se define la colocación como dos o más palabras que tienden a coocurrir juntas, pero se presenta el problema de la identificación de la colocación, punto en el que se hacen evidentes las dificultades que entraña. Siguiendo la definición basada en frecuencias, se plantea qué medida estadística es la que mejor se adapta a la definición de la colocación, ya que cada una permite extraer diferentes unidades, diferentes colocaciones. De ahí que utilicen para la identificación tres de las medidas más habituales: *frecuencia de corpus*, *mutual information* y *t-score*. El test se administró a ciento ocho aprendices españoles de inglés de diferentes niveles.

Buscan establecer la relación entre los factores que pueden afectar al aprendizaje de las colocaciones como son la *frecuencia* y la *exposición a la L2*. Se seleccionaron noventa y seis colocaciones a partir de los tres métodos estadísticos anteriormente indicados en COCA, *Corpus of Contemporary American English* (Davies, 2008), y según cinco criterios: a) las dos palabras que componen la colocación debían tener una distancia de 2-gram, b) todas las combinaciones deben considerarse colocaciones naturales a juicio de un nativo, c) las palabras constituyentes debían considerarse frecuentes, d) además,

deben considerarse *dispersas* según las tres medidas estadísticas y e) no equivalentes a correlato español (no deben ser cognados).

Además del test para medir el conocimiento colocacional con cincuenta colocaciones meta se pasó un cuestionario para ver la relación del aprendiz con la L2 y así poder correlacionar los resultados del test con el grado de exposición a la L2. El test se realizó en formato papel. Primero se llevó a cabo un test piloto en el que un nativo de español y tres nativos de inglés realizaron el test comprobando la claridad, simplicidad y naturalidad de cada ítem. También seis aprendices de inglés que confirmaron las apreciaciones de los expertos.

El formato elegido mide la producción desde la recuperación para evitar el efecto de adivinación⁵⁶. Se eligió la tarea de rellenar huecos en la que se ofrece un contexto en la L1 de los participantes y una frase en inglés con dos huecos, para completar con los dos elementos de la colocación. Con la intención de controlar la respuesta, se dan las iniciales de las dos palabras y las instrucciones aparecen en español, facilitando así la comprensión por parte de los participantes. En la *Figura 34* se puede ver un ejemplo de ítem.

*28. Mi tía está siguiendo una dieta muy estricta porque el vestido que se compró para la boda de mi hermana le queda pequeño, y quiere entrar en él.
She wants to l_____ some w_____ by next month.*

Figura 34. Ejemplo de ítem (González Fernández & Schmitt, 2015)

Los resultados muestran que el porcentaje de aciertos es superior al cincuenta por ciento, con lo que confirman que el dominio de las colocaciones no es tan deficiente como se podría esperar, en contra de los resultados de otros test de colocaciones (Moreno-Jaén, 2009; Nguyen & Webb, 2017). Con relación a uno de los factores que intervienen en el aprendizaje, la frecuencia, se llega a la conclusión de que tanto la frecuencia como el t-score están bastante igualados. Sin embargo, la MI no muestra correlación con el conocimiento colocacional, al igual que con otros test (Nguyen & Webb, 2017). Por lo tanto, se concluye que el conocimiento colocacional no mantiene

⁵⁶ El efecto de adivinación consiste en la habilidad de los participantes en un test en adivinar la respuesta por razones ajenas a su conocimiento. En el caso de que haya dos modelos, uno de producción y otro de comprensión, se debe empezar por aquel en el que no aparece la respuesta ya que así se evita que el aprendiz acierte por el recuerdo de la prueba anterior y no por sus propios conocimientos.

una fuerte dependencia con la frecuencia, de forma contraria a lo que ocurre con el conocimiento léxico. Así, la frecuencia no es factor determinante en el aprendizaje de colocaciones, ya que por ser más frecuente una colocación, no se conocerá más o menos. Por otro lado, el grado de exposición a la L2 sí se correlaciona: cuanto mayor es el *input*, mayor es el conocimiento colocacional. Plantean, por lo tanto, que las colocaciones deberían dejar de considerarse un desafío para los aprendices. Entre las limitaciones del estudio está el método de selección; la limitación de los participantes, todos comparten la misma L1; y el tipo de formato, producción controlada, por lo que faltaría comprobar si con producción libre los resultados son similares.

D. Test de comprensión de colocaciones

Este test de colocaciones mide la comprensión a partir de un único formato (Nguyen & Webb, 2017). Los autores siguen el enfoque basado en frecuencia, es decir, consideran la colocación como una combinación de palabras frecuentes. El argumento para decantarse por este enfoque y no por el fraseológico es el criterio de objetividad. Aun así, se reconoce una limitación en cuanto a la frecuencia de las colocaciones y es que no se tiene en cuenta la frecuencia de la colocación sino la frecuencia de las palabras que componen la colocación.

El principal objetivo del test es comparar la competencia colocacional con la competencia léxica y analizar la posible correlación entre ambas. Asimismo, relacionar la frecuencia de las palabras que componen las colocaciones, la competencia léxica, con la competencia colocacional. También buscan los factores que predicen el conocimiento comprensivo colocacional con una clara perspectiva pedagógica; estos factores son la frecuencia de los elementos de la colocación, la frecuencia de la colocación, la relación estadística entre los elementos de la colocación (MI), la congruencia y el esquema gramatical (V+N y A+N). La muestra elegida consta de ciento ochenta colocaciones meta entre las mil, dos mil y tres mil palabras más frecuentes. Para la selección de la muestra han seguido los siguientes pasos: a) selección del nodo, que en nuestro marco teórico consideraríamos base, en listas de palabras frecuentes, b) selección de los colocativos a partir de corpus y c) análisis final de hablantes nativos, pero no para identificar sino para

marcar si son congruentes o no lo son. Incluso se seleccionan los distractores a partir de los mismos criterios de frecuencia.

Además del test de colocaciones, se administró un test de vocabulario con las palabras que formaban parte de las colocaciones previamente. El test de colocaciones se administró a cien estudiantes vietnamitas de inglés con las instrucciones en vietnamita para facilitar el entendimiento. El formato elegido es el de opción múltiple descontextualizado en el que se parte de un nodo, que entendemos como base, y se busca reconocer la combinación más frecuente entre varios colocativos, es decir, la colocación. Siguiendo la intención de objetividad que predomina todo el diseño del test, los distractores también se seleccionan bajo criterio de frecuencia. En la *Figura 35* podemos ver un ejemplo.

1. money	a. check	b. drop	c. make	d. miss
2. figure	a. central	b. funny	c. happy	d. usual

Figura 35. Ejemplo de ítem (Nguyen & Webb, 2017)

La primera conclusión que se saca del análisis de los resultados de este test es el desafío que supone adquirir la competencia colocacional en L2 dado que el porcentaje de errores fue superior al de aciertos. Por lo tanto, se mantiene en la creencia de que el aprendizaje de las colocaciones supone un desafío. Debemos tener en cuenta que con este formato solo se tiene en cuenta la identificación de la forma, sin intervenir en ningún caso la relación entre la forma y el significado y que, además, está descontextualizado. Sin embargo, sí es preciso establecer el vínculo entre la base y el colocativo, componente que como explicaremos en nuestro modelo de competencia colocacional es indispensable. Así, consideramos que, aunque el test se plantea como de comprensión, en realidad, mide la producción, de ahí que las colocaciones sí supongan un reto. Entre las limitaciones, está la ausencia de una versión del test en la que se mida la comprensión, la variedad de niveles y también la repercusión de otros factores en el aprendizaje como la congruencia o la transparencia que también pudieran afectar al aprendizaje. Otra de las carencias del estudio reside en la selección de las combinaciones; dejan claro que el criterio de frecuencia afecta las partes de la colocación, pero no al conjunto, es decir, a la combinación.

E. CPCT de inglés académico

El CPCT (*Test de colocaciones productivo controlado*) se usó en dos investigaciones con diferente objetivo. En la primera, el test productivo controlado de colocaciones restringido al lenguaje académico (Frankenberg-García, 2018), el segundo (Sonbul et al., 2022), que recuperó el formato anterior ampliando los objetivos y que trataremos en el siguiente apartado (vid. 3.3.2.2 (F)). El CPCT se crea por la idea de que los estudios de corpus únicamente pueden captar una visión limitada del conocimiento que los aprendices tienen de las colocaciones.

La definición de colocación elegida sigue el enfoque basado en frecuencia, que observamos por el método de identificación. La selección de la muestra parte de la base, sustantivos, elegidos de entre las palabras más frecuentes de *Academic Vocabulary List*. En lugar de inventar ejemplos, como suele ser habitual, se buscaron ejemplos reales de inglés académico en un. Se eligió un formato de rellenar hueco en el que se ofrecen tres opciones, entre las que está el colocativo meta; por lo tanto, se trata de un formato de producción desde el reconocimiento. Los criterios de corrección estaban basados en criterios de frecuencia. En la *Figura 36* se pueden ver los diez ítems de los que consta el test.

1. The objective is to _____ a system that... <i>design, implement, develop</i>	6. Another _____ change observed was... <i>significant, major, dramatic</i>
2. Current research has _____ that... <i>indicated, shown, found</i>	7. These decisions play a/an _____ role in... <i>important, key, crucial</i>
3. The data _____ during the process... <i>collected, obtained, gathered</i>	8. The analysis was _____ in two stages... <i>performed, conducted, carried out</i>
4. The information _____ in table 3... <i>provided, summarized, contained</i>	9. An additional factor that _____ these results was... <i>affected, influenced, contributed to</i>
5. They attempted to _____ the effect of.... <i>measure, examine, analyze</i>	10. Johns (1991) _____ a different approach to.... <i>adopted, used, took</i>

Figura 36. Ejemplo de ítem del CPCT académico (Frankenberg-García, 2018)

El test se administró a casi cien estudiantes de los cuales más de la mitad tenía como L1 el inglés y menos de la mitad otra L2. Los resultados muestran que la diferencia entre los resultados de los participantes que tienen el inglés como L1 y los que tienen el inglés como L2 no es significativa aun siendo ligeramente mejor los resultados para el

primer grupo. Sin embargo, sí hay una diferencia significativa entre aquellos participantes con una mayor experiencia en el mundo académico, quienes obtienen mejores resultados en la producción de colocaciones.

F. CPCT de inglés no académico

El *Controlled Productive Collocation Test* (Sonbul et al., 2022), CPCT, tiene por objetivo analizar la relación entre el conocimiento productivo de las palabras y el conocimiento productivo de las colocaciones, rellenando así un hueco, ya que son varios los estudios encargados de comparar el conocimiento léxico comprensivo con el conocimiento colocacional productivo. Para esto, en la investigación se ponen en marcha dos test: uno para medir el conocimiento productivo de las palabras y otro para medir el conocimiento productivo de las colocaciones. Para el primer objetivo utilizan una versión propia del CPWT (Laufer & Nation, 1999), *Controlled Productive Word Test*, y para alcanzar el segundo objetivo, utilizan el CPCT, una versión del CPCT académico (Frankenberg-Garcia, 2018). Además, de utilizar dos test se incluyen en el grupo de participantes hablantes nativos y no nativos, que también sirven como medida de control de la fiabilidad de la herramienta de evaluación.

Aquí se define la colocación atendiendo a tres criterios: el carácter apropiado de la combinación, la frecuencia y la fuerza de asociación. Utilizan lemas como base de entre las tres mil palabras más frecuentes del inglés y usan MI como medida estadística para extraer las combinaciones de COCA. La muestra seleccionada cuenta con veinte lemas por banda de frecuencia. Así el test consta de sesenta ítems, que tras las revisiones en la fase de pilotaje quedaron en cuarenta y siete. La muestra incluye colocaciones verbo + nombre y modificador + nombre. Se entiende por modificador adjetivo, principalmente, pero también cualquier otra palabra que quepa en el contexto y que modifique al nombre. Para cada colocación meta se ofrece una frase que sirve como contexto y el nodo, que en nuestro marco teórico es la base, aparece subrayado. En la *Figura 37* aparece un ejemplo de ítem para cada tipo de colocación con una serie de posibles respuestas. Una de las ventajas de este formato frente a otros es que acepta diferentes respuestas, lo que se acerca más a una situación natural de comunicación.

MN: *This is a/an _____ chance. (possible responses: good, real, great, fair, excellent)*

VN: *They _____ a chance to win. (possible responses: have, get, take, stand)*

Figura 37. Ejemplo de ítem del CPCT no académico (Sonbul et al., 2022, p. 12)

El test se administró a un grupo de veintisiete nativos de inglés, que funcionó como grupo de control, y a un grupo de cincuenta y cinco estudiantes de Arabia Saudí. Los resultados muestran que el conocimiento productivo de las palabras y el conocimiento productivo de las colocaciones es superior en los hablantes nativos que en los aprendices. También muestra que un alto conocimiento productivo de las palabras precede a un alto conocimiento productivo de las colocaciones. Sin embargo, los resultados con las variables de frecuencia no son tan esclarecedores, sobre todo en comparación con otros estudios similares. La frecuencia, en ninguno de los tres factores valorados, presenta una incidencia significativa en los resultados.

G. Test de conocimiento léxico multicomponencial que incluye el conocimiento colocacional

Esta sección contiene dos test que incluyen el conocimiento colocacional en estudios multicompetenciales del conocimiento léxico (González-Fernández & Schmitt, 2020; Nontasee & Sukying, 2021). Este tipo de trabajos exploran la interrelación entre los componentes del conocimiento léxico (Nation, 2013a) a través de una batería de test, cada uno de los cuales se centra en uno de los componentes del conocimiento de las palabras. Estos son interesantes para esta investigación en la medida en que sitúan el conocimiento colocacional dentro del conocimiento léxico, lo que nos servirá de ayuda para la descripción y diseño del constructo del TCE, así como su relación con los otros componentes.

a) Primera propuesta multicomponencial

En el primero de los estudios multicompetenciales (González-Fernández & Schmitt, 2020) se introducen cuatro componentes: enlace entre forma y significado, derivados, múltiples significados y colocaciones. Incluso mide cada uno de los componentes desde el reconocimiento y desde la recuperación. Además, la población del

test está formada por ciento cuarenta y cuatro estudiantes de inglés con español como L1. El objetivo de la investigación es comprobar cómo es la relación entre los componentes del conocimiento de las palabras y si hay un orden en la adquisición de estos. Las veinte palabras meta son las mismas en para los cuatro componentes y los ocho test y para la selección de la muestra se tienen en cuenta los siguientes criterios: frecuencia, categoría gramatical, palabras polisémicas, que den lugar a al menos tres palabras derivadas y que varias sean cognados (35% del total).

En el formato de recuperación de colocaciones optaron por la opción de rellenar huecos y en el formato de reconocimiento de las colocaciones por uno de opción múltiple, ambos miden, por lo tanto, la producción. En el caso de la recuperación se ofrece un breve contexto en español y tiene que rellenar una frase en inglés con el colocativo que falta; además, la base se señala gráficamente, como podemos ver en el ejemplo (*Figura 38*). En el formato en el que se mide el reconocimiento, los aprendices deben elegir entre varias opciones el colocativo más adecuado. Las opciones cuentan con tres distractores y la opción de señalar que desconocen qué palabra falta, en un intento por evitar la influencia del azar en la elección de las respuestas (vid. *Figura 39*). Ambos formatos muy similares a los utilizados en el TCE.

En temporada alta es cuando más gente se va de vacaciones.

When you plan to go on holidays you should bear in mind that hotels are always more expensive in p_____season.

Figura 38. Ejemplo de ítem de recuperación (González-Fernández & Schmitt, 2020).

There are always more tourists in _____season.

- (a) Main
- (b) Peak
- (c) Big
- (d) Top
- (e) I don't know

Figura 39. Ejemplo de ítem de reconocimiento (González-Fernández & Schmitt, 2020).

La batería de test se administró siguiendo un orden que minimizara los efectos de las secciones previas en las posteriores de modo que el test de colocaciones fue el

penúltimo en administrarse: el formato de recuperación se administró en anteúltimo lugar y el reconocimiento de colocaciones en último lugar, intercalando el test que medía la polisemia de la palabra meta, base de las colocaciones meta. En el análisis de datos se utilizó la escala implicativa (*Implicational Scaling*) para detectar una jerarquía entre los componentes del conocimiento léxico y *Modelo ecuativo estructural (Structural Equation Modelling)* para inferir la relación entre los componentes. Repasaremos únicamente los resultados en relación con los dos test que miden el conocimiento de las colocaciones.

Igual que en el resto de componentes del conocimiento léxico, en ambos test (comprensión de colocaciones y producción de colocaciones) los aprendices conocían más de la mitad de los ítems con mejores resultados en la versión que mide el reconocimiento que en la versión que mide la recuperación. Aunque en general la correlación entre los componentes del conocimiento de la palabra es positiva, la correlación entre otros componentes del conocimiento léxico y los que miden el conocimiento de las colocaciones es reducido. En la escala implicativa de dificultad⁵⁷ (que va desde el componente más sencillo, en el que los aprendices obtienen mejores resultados, al más complejo, en el que los aprendices obtienen peores resultados) los cuatro test que miden los diferentes aspectos del reconocimiento son más sencillos que los cuatro test que miden la recuperación. El componente colocacional medido desde el reconocimiento está en segundo lugar en la escala, es decir, es el segundo componente más sencillo del conocimiento colocacional; lo que desterraría la creencia generalizada de que las colocaciones son difíciles. Mientras, el componente colocacional medido desde la recuperación ocupa el quinto lugar de la escala, aún hay tres componentes por debajo, lo que quiere decir que tampoco es el aspecto más complejo del conocimiento léxico.

El conocimiento de las colocaciones, tanto desde el reconocimiento como desde la recuperación, es el aspecto más independiente del conocimiento del vocabulario con respecto al resto. En principio pudiera parecer sorprendente que resulte más fácil

⁵⁷ La escala completa que empieza por el componente más sencillo al más difícil es la siguiente: reconocimiento del enlace entre la forma y el significado > **reconocimiento de la forma del colocativo** > reconocimiento de múltiples sentidos > reconocimiento de la forma de palabras derivadas > **recuperación de la forma del colocativo** > **recuperación del enlace entre la forma** y el significado > recuperación de la forma de palabras derivadas > recuperación de la forma de múltiples sentidos de la palabra.

recuperar el colocativo que la base, sin embargo, los colativos suelen ser palabras más comunes o frecuentes que las bases. En cuanto a la aparente no dificultad de las colocaciones ofrecen dos explicaciones: su posición no indica que sean fáciles, sino que son más fáciles que otros componentes (polisemia y derivación), situación que pudiera estar condicionada por la acumulación de cognados, uno de los factores que más se más condicionan el aprendizaje. Sin embargo, los datos muestran que el conocimiento de los colocativos es el que tiene una menor relación con el conocimiento de las palabras en general, lo que también se deduce por la baja correlación entre los dos test que miden el conocimiento colocacional y los que miden el resto de componentes.

b) Segunda propuesta multicomponencial

En el segundo de los estudios multicomponenciales (Nontasee & Sukying, 2021) no se define la colocación ni el conocimiento colocacional explícitamente, simplemente se menciona como aspecto de los componentes del conocimiento léxico. Ahondar en el conocimiento de la adquisición del vocabulario a partir de una batería de test permite el estudio de varios componentes del conocimiento léxico de forma simultánea. Además de las colocaciones, los componentes que se tienen en cuenta son el enlace entre forma y significado⁵⁸ y la categoría gramatical de las palabras⁵⁹; todos desde la producción y comprensión.

Fue aplicado a doscientos sesenta y un aprendices tailandeses de inglés. Utilizan dos formatos, uno para medir la comprensión y otro para medir la producción. El *Collocation Recognition Test* (CRT) se basa en el formato de *Vocabulary Level Test* y las adaptaciones posteriores. Mide el conocimiento comprensivo de la colocación⁶⁰ con cuarenta ítems y los aprendices deben elegir dos colocativos para dos bases entre cuatro opciones posibles. El *Productive Collocation Recall Test* (PCRT) mide el conocimiento productivo desde la recuperación y consiste en rellenar el colocativo de una colocación meta contextualizada. Para evitar el carácter aleatorio se presentan las dos primeras letras de cada uno de los veinte colocativos.

⁵⁸ *The Word Segmentation Test* (WST) y *The Affix Elicitation Test* (AET)

⁵⁹ *The Affix Elicitation Test* (AET) y *The L1 Translation Test* (L1TT)

⁶⁰ Se corresponde parcialmente con PRECO, test que mide la producción desde el reconocimiento en el TCE.

Instructions: Choose the best word choice in the first column to describe the noun in the second			
Adjective	Noun	Correct Answer	Point
big			
external	1. <u>flip</u> chart	1. flip	1
large	2. <u>large</u> success	2. big	0
flip			

Figura 40. Ejemplo de ítem de reconocimiento (Nontasee & Sukying, 2021)

Instructions: Complete the sentence below with an appropriate word	Correct Answer	Point
1. The Beatles is a fa <u>mous</u> band in Thailand	famous	1
2. Do you want to hear the good or ba <u>d</u> news first?	bad	1

Figura 41. Ejemplo de ítem de recuperación (Nontasee & Sukying, 2021)

Desde esta propuesta los dos test que miden la competencia colocacional lo hacen en ambos casos desde la producción, uno desde la recuperación y otro desde el reconocimiento, aunque el objetivo es usar un test para la comprensión y otro para la producción. Se trata de una prueba fiable, atendiendo a parámetros estadísticos, en la que se obtienen mejores resultados en comprensión que en producción, también en el de colocaciones. A partir de la correlación entre los componentes, y partiendo de la idea de que existen varios grados en el conocimiento léxico, la competencia colocacional tendría el menor peso con respecto al resto de componentes, igual que ocurría en el test multicomponencial anterior.

c) Tercera propuesta multicomponencial

Por último, este test multicomponencial (Read & Dang, 2022) mide la extensión y la profundidad del conocimiento léxico. Los objetivos de esta investigación son medir el conocimiento del vocabulario académico de aprendices universitarios de inglés y probar nuevos formatos para medir la profundidad. Para este fin, se utilizan tres test: uno para medir la extensión, el UVLT⁶¹ (Webb et al., 2017), una batería para medir la

⁶¹ Updated Vocabulary Level Test, una nueva version del VLT.

profundidad y otro test de traducción. El test de profundidad está compuesto por tres partes: la parte A es un test de sinónimos, la parte B es un test de colocaciones y la parte C es un test de derivación. Nos centraremos únicamente en la parte B que mide las colocaciones. Se utiliza un formato original en el que se presentan cinco combinaciones de palabras descontextualizadas y los participantes debe elegir en Yes/No, es decir, si la combinación existe o no. Con la intención de reducir las posibilidades de acierto por azar, se añadió una tercera opción (*Not Sure*).

	Yes	No	Not Sure
a. a low level	✓		
b. to write an article	✓		
c. a closed difference		✓	
d. the process looks loud		✓	
e. the population grows each year	✓		

Figura 42. Opción C del test que mide las colocaciones.

La fuente elegida para la muestra difiere en cada caso: listas de frecuencia, corpus. Los veinte ítems que componen este test fueron extraídos de un corpus. En cada ítem hay al menos dos combinaciones que entendemos como colocaciones y dos distractores. En ningún punto explica cómo se eligen las combinaciones, es decir, desconocemos cómo se define y se identifica la colocación más que por ejemplo del que disponemos. Los resultados de los test que miden la profundidad fueron normalizados. Al comparar las medias, se observó cómo la parte A era la más sencilla para los participantes, la parte C se encuentra en un punto intermedio y la parte B, el test de colocaciones, el que resultaba más complejo para los aprendices. La correlación entre los resultados del UVLT y la profundidad son moderados, más moderada es la relación entre UVLT y la parte B, el test que mide las colocaciones. Dentro de la dificultad que implicó esta parte, se percibe que fue más fácil para los aprendices identificar las colocaciones como tal que los distractores como no colocaciones, ahí es donde residían la mayor parte de los errores. No rechazaban las no colocaciones, usando en muchos casos la opción de “No estoy seguro”. A medida que aumenta el nivel de competencia léxica, los participantes se sienten más seguros en sus juicios sobre las colocaciones.

H. Test productivo de colocaciones en francés

El test productivo de colocaciones en francés (Forsberg Lundell et al., 2018) es uno de los pocos test que miden el conocimiento colocacional en una L2 diferente del inglés. La versión que recogemos aquí es una adaptación de versiones previas (Forsberg Lundell & Lindqvist, 2014). Se definen las colocaciones desde un punto de vista estadístico sin tener en cuenta ningún criterio fraseológico ni para su identificación ni descripción. Su principal objetivo es establecer la relación entre la competencia colocacional y los niveles del MCER. También es conocer si el conocimiento productivo de las colocaciones es progresivo, es decir, si aumenta a medida que se incrementa el nivel de competencia lingüística de los aprendices. La selección de la muestra es representativa, por lo tanto, se hace siguiendo criterios estadísticos. Se reproduce un método habitual que consiste en la selección de bases en una lista de frecuencias, búsqueda de los colocativos en un corpus de frecuencia y selección de las combinaciones a partir de medidas estadísticas. Además, se eliminaron los cognados de sueco porque una parte importante de los participantes tenía esa lengua como L1.

Para la construcción de los ítems se siguió un formato contextualizado de producción controlada a partir de la recuperación. La base se presenta en una oración a la que le falta el colocativo, del que se da la primera letra, y el aprendiz debe escribir el resto de la palabra, como se puede ver en la imagen (vid. Figura 43). Siguiendo con el enfoque estadístico, incluso los ejemplos utilizados para la construcción de los ítems fueron extraídos de corpus. El test consta de treinta y un ítems administrados a casi sesenta aprendices de diferentes L1, todos con nivel entre B2 y C2 asignado a partir de un test de competencia lingüística en francés.

- (1) a. *L'ONU est fermement résolue à r_____ justice aux victimes du génocide. (The United Nations are determined to d_____ justice to the victims of the genocide.)*
- b. *Si les petites entreprises é_____ des difficultés à embaucher, c'est aussi parce qu'elles disposent de moyens qui sont moindres.*

Figura 43. Ejemplo de ítem de test productivo de colocaciones en francés (Forsberg Lundell et al., 2018)

Los resultados demuestran que existe una diferencia significativa entre los niveles de referencia en cuanto a su competencia colocacional y que, por lo tanto, es progresivo; a medida que avanza el nivel de competencia lingüística, avanza el nivel de competencia colocacional. Finalmente, entre las limitaciones recogen la dificultad para identificar colocaciones a partir de métodos estadísticos.

3.3.3. Recapitulación y comparación

En el primer apartado de este capítulo hemos esbozado algunas nociones básicas sobre la Evaluación de Lenguas que desarrollaremos en los capítulos siguientes de forma práctica y que nos servirán a su vez como guía para la construcción, administración y análisis de los resultados del TCE. En los dos apartados siguientes nos hemos centrado en la evaluación del léxico y de las colocaciones respectivamente. Hemos dedicado un espacio a la descripción de algunos de los test de léxico más representativos en inglés y otro a los test que miden el léxico del español. Estos sirven como referente para la elaboración del test de colocaciones. Algo que no hemos podido realizar en el caso de las colocaciones es la descripción de algún test que mida la competencia colocacional del español. Por lo tanto, nos hemos centrado en los test que miden las colocaciones, principalmente en inglés, pero también en otras lenguas como el francés.

El análisis nos permitió establecer una serie de parámetros que siguen los test de colocaciones e intuir el origen de algunas confusiones con respecto a las colocaciones en relación con el procesamiento y proceso de enseñanza-aprendizaje. Atendiendo a subcomponentes del conocimiento léxico (conocimiento de una palabra), en los test que se centran en la identificación solo miden el conocimiento de la forma y en los test que miden algún otro aspecto del conocimiento suelen centrarse en la conexión entre forma y significado. Algunos test se centran en la identificación de la forma de la colocación desde el reconocimiento (COLLEX) y otro desde la producción (LexCombi). Este tipo de test únicamente resuelve o atiende a una cuestión metalingüística, que es la conciencia por parte del aprendiz de que existe esa combinación. Entre los test que miden la conexión entre forma y significado, encontramos los que lo hacen desde la comprensión y la producción. Sin embargo, hay cierta confusión por la asociación inherente de la

comprensión al reconocimiento y la producción a la recuperación. Esto hace que algunos de los test que se describen como de producción realmente midan la comprensión desde la recuperación.

La mayor parte de las consideraciones comúnmente aceptadas sobre las colocaciones en el contexto de la enseñanza-aprendizaje del léxico derivan de las conclusiones extraídas de estos test. Entre ellas, el reto que suponen las colocaciones por los aprendices y la correspondencia entre la competencia colocacional y la competencia léxica y lingüística. Estudios posteriores, discordes con los parámetros anteriormente expuestos, ya han discutido este tipo de afirmaciones (González Fernández & Schmitt, 2015). Al igual que comprobamos esta investigación, en otras, se ha llegado a la conclusión de que las colocaciones suponen un reto en la producción (desde la recuperación) pero no en la comprensión (desde el reconocimiento). La correlación entre la competencia colocacional y la léxica viene dada, comúnmente, por la correspondencia entre la muestra del test que mide el léxico y las bases de la colocación o por el establecimiento de la correlación teniendo en cuenta solo un aspecto del conocimiento colocacional, normalmente la comprensión desde el reconocimiento. Esta visión es en cierto modo limitante ya que no parece totalmente generalizable si para que se da correlación se debe usar la misma muestra.

Entre los test de la primera generación, solo dos miden la comprensión y los dos lo hacen desde el reconocimiento, teniendo en cuenta únicamente la identificación de la forma. El resto miden la producción: por un lado, ADELEX y CONTRIX centrándose en el enlace entre la forma y el significado, pero LexCombi atendiendo únicamente al reconocimiento de la forma; por el otro, ADELEX (en su última versión) y LexCombi lo hacen desde la recuperación y CONTRIX desde el reconocimiento. En la *Tabla 19* podemos ver algunas de las características de los test de esta primera generación de forma genérica.

	CONOCIMIENTO (Enlace forma significado)		IDENTIFICACIÓN (Forma)	
	Comprensión	Producción	Comprensión	Producción
Reconocimiento		CONTRIX	COLLEX y COLLMACH DISCO	
Recuperación		ADELEX		LexCombi

Tabla 19. Clasificación de los test de la primera generación según sus características

Una de las limitaciones de estos test es que ninguno de ellos ofrece una visión global del conocimiento colocacional, en todos los casos miden el conocimiento de forma parcial. Solo COLLEX y COLLMACH y ADELEX utilizan una variedad de medidas que permiten observar el conocimiento desde diferentes perspectivas al testar una misma colocación meta⁶² de diferentes formas. El problema es que las diferentes formas de evaluar no consiguieron obtener diferente tipo de información al compartir propiedades de activación e incluso los mismos modelos de acceso a la información requerida. Si en la pre-generación se dio el primer paso para el desarrollo de los test, esta segunda generación de test dio el paso definitivo en el desarrollo de la evaluación de los de colocaciones. Los test de la segunda generación tienen en común el hecho de poner en primer plano la validez y la fiabilidad de la herramienta de evaluación desde un punto de vista estadístico, situando así la evaluación de colocaciones al nivel del léxico y de otras destrezas como la comprensión lectora.

En la *Tabla 20* se puede ver la clasificación de los test de colocaciones a partir de sus características de la que podemos extraer las siguientes conclusiones. Al contrario de lo que pasaba con los test de la anterior generación, en esta predominan los test que llamamos de conocimiento (evalúan la relación entre la forma y el significado) frente a los de identificación (evalúan el reconocimiento de la forma). Aun así, sigue habiendo test que evalúan la correcta identificación de colocaciones, en concreto, la producción desde el reconocimiento y que, además, suelen confundir con test que miden la comprensión por hacerlo desde el reconocimiento. Entre los test de conocimiento, la mayoría miden la producción desde la recuperación o desde el conocimiento, pero solo uno mide la comprensión desde el reconocimiento y ninguno la comprensión desde la recuperación. Pensamos que la identificación de colocación ya se mida desde la comprensión o desde la producción sitúa el grado del conocimiento colocacional en un estadio inicial. Sin embargo, el conocimiento, entendido como el enlace entre la forma y el significado, en cualquiera de sus modalidades, supone un mayor esfuerzo cognitivo y, por lo tanto, mayor grado de conocimiento de la colocación.

⁶² Sin embargo, en COLLEX y COLLMACH no se buscan las mismas colocaciones meta, sino que se usan diferentes colocaciones meta para cada formato. Esto limita la fiabilidad de la comparación entre los dos formatos del test.

	CONOCIMIENTO (Enlace forma significado)		IDENTIFICACIÓN (Reconocimiento de la forma)	
	Comprensión	Producción	Comprensión	Producción
Reconocimiento	Nizonkiza	CPCT académico CPCT no académico Glez-Fdz y Schmitt 2		LexCombi2 Nguyen y Webb Notasee y Sukying
Recuperación		Nizonkiza Glez-Fdz y Schmitt 1 Glez-Fdz y Schmitt 2 Nontase Forsberg et al		

Tabla 20. Clasificación de los test de la primera generación según sus características

De la comparación de las conclusiones extraídas a partir del análisis de resultados de cada uno de los test, concluimos los resultados tienden a ser negativos cuando se evalúan las colocaciones desde la producción y la recuperación, mientras que tienden a ser positivos cuando se miden desde la comprensión y el reconocimiento. De ahí que exista cierta contradicción entre unos estudios y otros. Por lo tanto, se necesita una imagen global que atienda a todas las perspectivas de la evaluación de las colocaciones. Además, esto debe realizarse utilizando una misma definición de colocación y unos mismos criterios de identificación y selección y, por supuesto, una misma lista de colocaciones meta, es decir, unas mismas condiciones que permitan comparar los resultados de diferentes formatos. Con el TCE buscamos un test que cumpla con estas condiciones.

4. DISEÑO Y CONSTRUCCIÓN DEL TCE

Existen trabajos que evalúan la competencia léxica de aprendices de inglés con español como L1 (Gitsaki, 1996; González-Fernández & Schmitt, 2020) y también estudios que miden la competencia colocacional de aprendices de español con inglés como L1 (Gitsaki, 1996), aunque no en todos los casos sea el eje central del estudio. No se conocen test que midan el conocimiento ni la competencia colocacional del español de forma global, tal y como ya explicamos y dejamos claro en el capítulo anterior al hacer un repaso por los test que miden la competencia o el conocimiento colocacional. No obstante, sí hay trabajos que evalúan el uso de colocaciones en español; estos trabajos se enmarcan, de forma general, en el ámbito de la lingüística de corpus. Entre otros, el análisis de errores en corpus de aprendices (Alonso Ramos, et al., 2010; Alonso Ramos, et al., 2010; Pérez Serrano, 2014), la relación entre frecuencia (en corpus), asociación de niveles del MCER y las colocaciones (García-Salido & Alonso-Ramos &, 2019). De los test que miden la competencia o el conocimiento de las colocaciones del inglés (y en español) como L2 ya se ha hecho un análisis pormenorizado en el capítulo anterior (vid. 3.3).

Como hemos comprobado en el capítulo anterior, no hay ningún test que se adapte a los propósitos de esta investigación y, por lo tanto, es necesario diseñar y construir uno. Este test ha de cumplir una serie de condiciones: ceñirse a la definición de colocación, contar con una muestra representativa de colocaciones del español y medir el conocimiento de forma global. Estas razones determinan las características que definirán el TCE: un test en español, con una concepción fraseológica de la colocación y que mida tanto la extensión como la profundidad, es decir, el conocimiento de una forma global y la competencia colocacional. Para evaluar el conocimiento global de las colocaciones no será suficiente con un único formato, sino con varios que tengan en cuenta más de un aspecto del conocimiento de las colocaciones. Asimismo, es necesaria una batería de test para poder obtener una imagen amplia de la competencia colocacional de los aprendices de español. De este modo, no solo se busca conocer cuántas colocaciones se conocen sino cómo se conocen para atender a las dificultades de aprendizaje buscando un diagnóstico del conocimiento y calibrando el peso de su aprendizaje. Este objetivo condiciona todo el proceso: desde la planificación hasta el análisis, pasando por el diseño y administración.

Este capítulo se centrará en la planificación, diseño y construcción del test para medir la competencia y el conocimiento que los aprendices de español tienen de las colocaciones. En primer lugar, explicaremos la metodología de la investigación, centrándonos en la validez del test (vid. 4.1.1). En segundo lugar, definiremos de forma operativa lo que significa el conocimiento colocacional, es decir, el constructo (vid. 4.1.2). En tercer lugar, expondremos de forma pormenorizada los pasos seguidos para el diseño del test, desde el propósito, hasta el formato elegido (vid. 4.2, 4.3 y 4.4). En cuarto lugar, recogeremos los criterios que han servido para elegir las colocaciones que se van a testar (vid. 4.5). En quinto lugar, describiremos el test de extensión de vocabulario que va a permitir establecer las correlaciones necesarias tanto con la competencia colocacional como con la competencia lingüística (vid. 4.6). Por último, explicaremos brevemente el test de extensión de vocabulario X-LEX (vid. 4.7).

4.1. Planificación del TCE

En este apartado nos centraremos en el cumplimiento de dos condiciones teóricas que determinarán todos los puntos del diseño del test: su carácter como test de diagnóstico y el argumento de validez como enfoque para dotar al instrumento de evaluación de esta propiedad. Además, en este punto definiremos el constructo de forma operativa. Finalmente, atenderemos a las fases de construcción por las que ha pasado el test hasta llegar a su forma definitiva.

4.1.1. Condiciones: test de diagnóstico y argumento de validez

Al construir un test, establecer el propósito es un requisito clave porque sin él no es posible validarlo (Schmitt et al., 2019, p. 6). El objetivo del TCE es identificar las fortalezas y debilidades detectadas en el conocimiento de las colocaciones y proporcionar pistas sobre cómo mejorar el aprendizaje de las colocaciones por parte de estudiantes de español como lengua extranjera. Además, debe permitir clasificar a los aprendices según su competencia colocacional de acuerdo con los parámetros del MCER. El tipo de test necesario para este propósito y el proceso de validación elegido

condicionarán el diseño del test y nos permite planificar su diseño y construcción de forma ordenada.

En el capítulo dos explicábamos que los test se podían clasificar atendiendo a varios criterios, según la parte afectada por las decisiones o en función del tipo de decisiones que se van a tomar. Los propósitos de este test y las preguntas de investigación (vid. 1.2) se centran en la descripción del conocimiento global de las colocaciones y la gradación del conocimiento. Esto nos sitúa en el ámbito de los *test de diagnóstico* que se centra en la detección de las fortalezas y debilidades de una habilidad lingüística con un fin didáctico (Jang & Wagner, 2013, p. 114). Esto implica dos cuestiones previas al diseño del test. Por un lado, definir de forma operativa la unidad de medida, en nuestro caso, la colocación, y describir el modelo de conocimiento de esta (vid. 4.1.2.2.). Por otro lado, la definición de un nivel estándar y de los grados del conocimiento colocacional, que nos permitirán diagnosticar las dificultades que presenta el aprendizaje de colocaciones. La clasificación de los aprendices según su competencia colocacional conforme a una escala estandarizada nos sitúa en el ámbito de los *test de competencia (proficiency test)*, lo que implica la nivelación de la competencia colocacional en base a unos criterios de clasificación operativos (vid. 4.3.3.2).

Otra de las condiciones imprescindibles para el diseño y construcción de un test es el proceso de validación, como ya hemos apuntado (vid. 3.1.2). Validaremos el TCE adoptando el enfoque de la *validez basada en argumentos* (Bachman & Palmer, 2010a; Chapelle, 2012a; Kane, 2017). En un intento de que la práctica de la evaluación sea amplia y consistente, presentaremos una serie de *afirmaciones* sobre la calidad de los resultados y cómo deberían interpretarse para tomar decisiones. Esas afirmaciones están respaldadas tanto por la investigación como por el análisis estadístico de los resultados. Para articular el *argumento* partimos de la concreción de una serie de elementos, de los cuales las *inferencias* son los principales: *consecuencias de los resultados, utilización, extrapolación, explicación, generalización, evaluación y definición del dominio* (ALTE, 2011; Bachman & Domböck, 2017; Chapelle, 2021). La fase de *justificación de evaluación* del *argumento de validez* ha quedado ya ampliamente explicada en los capítulos anteriores 2 y 3 (evidencia teórica), y que concretaremos a lo largo de este capítulo y del siguiente (evidencia a partir de la aplicación teórica y del análisis estadístico de los

resultados). Dado que la validez es un proceso continuo y que los test nunca están completamente validados, es preciso iniciar la validación desde la planificación. A continuación, presentamos el argumento de validez del TCE:

El investigador pretende usar la evaluación para analizar el conocimiento que los aprendices de español tienen de las colocaciones, así como su competencia en este ámbito. Recordamos que cuando hablamos del conocimiento de la colocación lo hacemos del grado en que se conoce y que cuando hablamos de competencia lo hacemos del nivel adquirido. Se procura obtener resultados que se puedan interpretar como indicadores de la competencia colocacional y del conocimiento de forma global. El resultado mostrará la clasificación en una escala de la competencia, el grado del conocimiento, además las dificultades que cada uno de los formatos plantea para los aprendices. Todo ello permitirá esbozar unas directrices para un <<aprendizaje más eficaz.

4.1.2. Propuesta de un modelo de conocimiento colocacional

En el capítulo dos (vid. 2) hemos definido qué entendemos por colocación como unidad lingüística, siguiendo el enfoque fraseológico y, más concretamente, la TST. En este capítulo veremos qué entendemos por colocación como unidad evaluable y de aprendizaje, es decir, un concepto operativo de colocación para unos fines concretos: guiar el desarrollo de las pruebas de evaluación del conocimiento colocacional, aportar una justificación teórica a las interpretaciones previstas y organizar el aprendizaje de colocaciones de L2. No se trata de dos definiciones opuestas, sino complementarias, en la que la segunda (*definición operativa*) es dependiente de la primera (*definición teórica*) ya que la definición operativa para ser práctica y útil debe sostenerse en una definición teórica precisa. Describir la habilidad evaluada de forma detallada permite observarla para medirla y comprobar el grado en el que esa habilidad se domina.

La necesidad de una definición operativa responde a lo que en evaluación de lenguas se denomina definición de *constructo*, como elemento esencial de la *validez de constructo*⁶³. Este concepto se define como la habilidad medida definida en los componentes particulares que la componen de forma apropiada para la situación de la evaluación (Bachman & Palmer, 2010, p. 212). Como test de diagnóstico, la descripción pormenorizada de la competencia que se va a medir es uno de los rasgos esenciales del segundo componente de la estructura diagnóstica⁶⁴ (Lee, 2015); por lo tanto, también en este sentido es importante establecer una definición operativa clara para el TCE. En esta descripción se debe caracterizar el concepto de *normalidad* de la competencia, es decir, el conocimiento y dominio estándar, así como los subcomponentes en los que dicha competencia se puede organizar para encontrar las fortalezas y debilidades de los aprendices de ELE. Describir qué significa conocer una colocación y cuáles son los conocimientos previos y adquiridos que supone es lo que vamos a describir en este apartado.

4.1.2.1. *Dificultades y antecedentes*

La tarea de proponer un modelo de conocimiento colocacional, definir el constructo, establecer los estándares y, en definitiva, delimitar una definición operativa de la colocación no es una tarea exenta de dificultades, desde la naturaleza de la unidad del léxico a las vacíos y limitaciones de la literatura precedente. En este apartado revisaremos esas dificultades y acudiremos a algunos de los antecedentes que nos permitan analizar de forma adecuado el dominio de interés de esta evaluación y así realizar observaciones apropiadas sobre la actuación obtenida a partir de los test de los aprendices de ELE.

En los test de evaluación de léxico se ha definido el *constructo* habitualmente a partir de la identificación y definición de los atributos de la competencia y, para este fin,

⁶³ Aunque en esta investigación tomamos un enfoque más actual del concepto de validez, en concreto la validez basada en argumentos, el constructo como concepto sigue estando vigente y siendo imprescindible. Como evaluación basada en el argumento de validez, sigamos el enfoque que sigamos, la justificación teórica del argumento es esencial.

⁶⁴ Recordamos que seguimos las directrices de la *evaluación lingüística de diagnóstico* que se define por unas características y unos componentes. Ver 2.4.2.

se toma como referencia la descripción del conocimiento del vocabulario de (Nation, 2013), la propuesta estándar en SLA, que divide el conocimiento léxico en diferentes componentes: forma, significado y uso. Para la definición del constructo de unidades fraseológicas también se tomó como punto de partida esta propuesta (Nation & Webb, 2011), pero es necesario hacer una adaptación de modo que los subcomponentes se adecuen a las características de la unidad lingüística en concreto, las colocaciones. En los últimos estudios la idea de conocer determinada unidad lingüística se ha ido alejando de la simple identificación de componentes para centrarse en el uso; por lo tanto, uno de los retos en la evaluación del vocabulario es el desarrollo del proceso de aprendizaje del léxico que permita avanzar desde la identificación de componentes hasta el orden o la relación de adquisición y uso de dichos componentes (Barclay & Schmitt, 2019). No es sencillo aplicar este cambio ya que existe un vacío en la investigación en términos de procesamiento y representación del léxico y, sobre todo, en el procesamiento y representación de colocaciones⁶⁵.

Gyllstad (2007) dedica un capítulo a este aspecto, que titula como “Conocimiento colocacional – definiendo el constructo” (*Collocational knowledge – defining the construct*), y toma la *validez de constructo* como punto de partida. Pese a no ser la definición que se tomará en este trabajo, su aportación es indispensable para avanzar y proponer una definición operativa del conocimiento colocacional. La propuesta limita el constructo al reconocimiento de las colocaciones, haciéndose la siguiente pregunta como guía para definir su constructo: ¿el aprendiz puede reconocer las colocaciones de forma apropiada? Y dejando de lado la recuperación de las colocaciones. En la definición operativa de la definición del constructo, analiza las formas de medir la habilidad atendiendo a diferentes dualidades: *reconocimiento y recuperación* más *profundidad y extensión*. La definición que aporta es parcial, ya que no solo es imprescindible reconocer una colocación cuando la leemos o la escuchamos, sino que también es necesario recuperarla cuando la queremos escribir o pronunciar al hablar. Además de esta insuficiencia, la definición teórica parte casi totalmente del enfoque basado en frecuencias, concepción que dista de la tomada en este trabajo, en que la frecuencia es

⁶⁵ A este respecto se han presentado varios trabajos que ahondan en cuestiones relativas al procesamiento de colocaciones y de otras unidades fraseológicas (Hoey, 2012).

un rasgo caracterizador más pero no definitorio. La definición operativa utilizada en COLLEX y COLLAMACH (Gyllstad, 2007) no nos sirve, además de por las razones anteriores, porque no tiene en cuenta el procesamiento y, tal y como se defiende en los últimos trabajos sobre la materia, es imprescindible definir el conocimiento sin tenerlo en cuenta. Por lo tanto, debemos proporcionar una definición operativa propia que abarca el conocimiento colocacional de manera panorámica, es decir, atendiendo a todas las dimensiones del conocimiento (Wolter & Yamashita, 2014).

Si entendemos la colocación como un constructo dependiente, es decir, la colocación como uno más de los rasgos que implican el conocimiento de las palabras, la *colocabilidad* de cada base es uno de los factores determinantes. Esto es, la capacidad de una palabra para no combinarse léxicamente de forma libre o restrictiva con otras palabras. La capacidad combinatoria de cada palabra determina (puede determinar) en mayor o menor medida su dificultad. A mayor *colocabilidad* (capacidad para formar colocaciones) de una palabra, más compleja será a la hora de aprenderla. Sin embargo, y de forma paralela a lo que ocurría en relación con el conocimiento léxico, creemos que cada colocación tiene su propia carga de aprendizaje. Una palabra como *dinero* tendría un alto grado de *colocabilidad* debido al alto número de combinaciones posibles, consideradas colocación, de las que puede formar parte: *dinero negro*, *dinero fácil*, *correr dinero*, *gastar dinero*, etc. Sin embargo, no tiene la misma dificultad para un aprendiz de L2 la combinación *gastar dinero* que *dinero fácil*, ya sea por la frecuencia de la combinación, de cada uno de los elementos que la componen o del grado de transparencia que perciba cada aprendiz. Si entendemos la colocación como un constructo independiente, podemos afirmar que no todas las colocaciones son igual de difíciles. Por ejemplo, serán más fáciles las colocaciones generalizables, como es el caso de *tocar un instrumento* y sus derivados *tocar la guitarra*, *el acordeón* o *la flauta*; que colocaciones no derivables (no hiperónimos). Por ejemplo, *dar un abrazo* y *dar un beso* frente a *hacer una caricia* o *un cariño*. Todos son actos para demostrar amor y sin embargo utilizan colocativos diferentes. Sin descartar la funcionalidad de la colocación como constructo dependiente, para su evaluación, entendemos la colocación como un constructo independiente.

4.1.2.2. Definición operativa de colocación

En la propuesta que presentamos en este trabajo seguimos la tendencia que define el conocimiento colocacional entendiendo la colocación como una unidad autónoma del léxico (vid. 2.3.2) y tenemos en cuenta varios aspectos: a) la definición teórica de colocación, b) la propuesta del conocimiento basada en componentes, c) el concepto de constructo aplicado a la validez basada en argumentos, y d) la identificación y subdivisión de los componentes, propia de los test de diagnóstico. Además de esto, debemos considerar el procesamiento de las colocaciones (Gyllstad & Wolter, 2016; Vilkaite & Schmitt, 2019; Wolter & Gyllstad, 2011; Wolter & Yamashita, 2014) para lo que nos acercaremos a alguna de las ideas de los trabajos que se acercan a la representación y procesamiento del léxico, próximos a la psicolingüística, área que sobrepasa los límites de esta investigación.

De este modo, proponemos un modelo del conocimiento colocacional que parte de la definición teórica (TST) y de la idea componencial del conocimiento, que nos servirá para definir las subcompetencias, y alguno de los conceptos tomados del procesamiento para establecer la relación entre los componentes y avanzar hacia la aplicación tanto de la interpretación como de la aplicación a la enseñanza-aprendizaje. Entre los conceptos o ideas que tomaremos de esta perspectiva está la definición del concepto de *vaguedad* (Bordag et al., 2021) como una propiedad del léxico de las segundas lenguas según la que el desarrollo de aprendizaje no siempre llega a su grado óptimo y la *normalidad* como el conocimiento óptimo de esa unidad del léxico, según el nivel esperable de un hablante nativo de la lengua⁶⁶.

A. Componentes del conocimiento colocacional

El modelo propuesto cuenta con cuatro subcomponentes. Debemos recordar que la colocación es una combinación de palabras, por lo tanto, hay que conocer la *base*, palabra “seleccionadora”, y el *colocativo*, palabra “seleccionada”. Además, debemos añadir dos componentes más que dan carácter de colocación a la combinación de

⁶⁶ El concepto de *fuzziness* se tradujo como *vaguedad*, propiedad fundamental del léxico de L2. Concepto que se enmarcan en el *Ontogenesis Model of the L2 Lexical Representation* (OM). Gyllstad pone a prueba este modelo para las unidades multipalabra al encontrar falta de claridad en su tratamiento.

palabras: el vínculo fraseológico y las particularidades propias de la colocación como unidad del léxico.

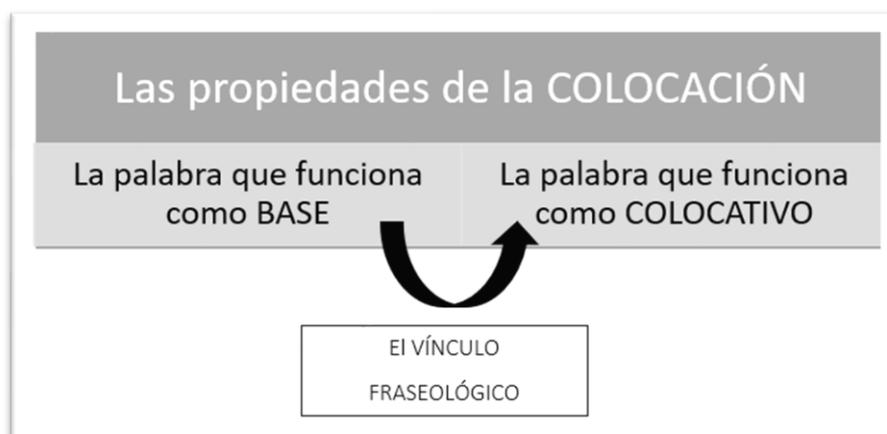


Figura 44. Componentes básicos del modelo de conocimiento de la colocación

A continuación, explicamos cada uno de estos componentes de forma pormenorizada, atendiendo a los aspectos establecidos en la propuesta componencial del conocimiento léxico: forma, significado y uso (Nation, 2013). Con la idea de obtener una imagen global del conocimiento colocacional, hemos definido el conocimiento colocacional desde cuatro perspectivas en relación con las destrezas y procesos cognitivos aplicados. Seguimos así la propuesta global del conocimiento de las palabras, ya que el conocimiento óptimo puede darse desde cuatro ángulos que surgen de la combinación de los siguientes aspectos: comprensión, producción, reconocimiento y recuperación.

a) Conocimiento de la base de la colocación

La *base* es el elemento de la colocación que el emisor elige por su sentido, mientras que el colocativo es elegido en función de esa base; desde el punto de vista de la comprensión, es el elemento que contiene la mayor carga semántica. En la siguiente tabla se recogen los aspectos necesarios para conocer la base de la colocación. Son los mismos que en cualquier palabra o unidad fraseológica. A continuación, explicamos cada uno de los componentes a través de ejemplos y en la siguiente tabla se ve la distribución de los componentes de forma más visual (vid. Tabla 21).

Base	Forma	Forma oral y escrita
		Estructura morfológica
	Significado	Sentido
		Relaciones paradigmáticas
	Uso	Rasgos gramaticales
		Registro

Tabla 21. Conocimiento de la base

De la base debemos conocer aspectos relacionados con la *forma*, el *significado* y el *uso*. En cuanto a la *forma*, es necesario saber identificar tanto la forma oral como la escrita, es decir, cómo se dice y cómo se escribe la palabra que va a funcionar como base sin cometer faltas de pronunciación (corrección fonética) y de ortografía y que son imprescindibles a nivel de producción. En relación con la *estructura morfológica*, debemos puntualizar que en la mayoría de las colocaciones es posible la flexión de la base, pero no su derivación. Por ejemplo, podemos decir *tomar una decisión* (singular) o *tomar decisiones* (plural), pero no **tomar indecisiones* sino *tener indecisiones*. Es decir, la colocación se mantiene con la flexión de la palabra, pero no siempre con la derivación; la palabra derivada *indecisiones* selecciona léxicamente otro colocativo como *tener* o *reflejar*. En el caso de la derivación debemos aclarar que ya estaríamos ante otra unidad léxica, por lo que nos encontramos ante otra colocación con otra base. Al contrario que ocurre con las locuciones, que no aceptan ninguna variación, la colocación permite ciertas transformaciones sintácticas (Koike, 2001, pp. 31-33) como la pronominalización, transformación en pasivación, etc. Por ejemplo, *trabrar amistad / amistades trabradas*. Al comparar los componentes del conocimiento colocacional en cuanto a la forma, con los componentes del conocimiento de las palabras (Nation, 2013), vemos como se corresponden totalmente con los subcomponentes que son necesarios para conocer una palabra.

Con relación al *significado* de la *base*, es imprescindible conocer el sentido de la unidad léxica que va a funcionar como base. Se trata de uno de los componentes más importantes ya que este se mantiene inalterable en la colocación, debido a la *composicionalidad*, rasgo definitorio de las colocaciones desde la corriente fraseológica. En este punto sí hay diferencias con respecto a la propuesta de referencia (vid. Nation). Mientras en la definición de los componentes de la palabra se incluye a) la forma y significado, b) los conceptos y c) los referentes y las asociaciones; en el conocimiento de la base de la colocación el componente se sintetiza en dos subcomponentes: a) el sentido

y b) las relaciones paradigmáticas. No es necesario conocer todos los posibles sentidos que posee esa forma de palabra sino solo el que toma en la colocación en concreto. Si se quiere aprender la colocación *carta abierta*, por ejemplo, en la que *carta* es la base, solo necesitas saber el sentido que tiene en esa colocación: ‘escrito enviado a alguien para algo, pero no el sentido de ‘naipe’. Lo mismo ocurre en una colocación como *carta alta*⁶⁷ no es necesario conocer el sentido de ‘escrito enviado’. Se da en casos en el que la base presenta polisemia u homonimia.

Con respecto a las asociaciones paradigmáticas, en algunos casos, la colocación se puede generalizar a los hipónimos (o cohipónimos). Es el caso de *beber una bebida*, *beber vino*, *agua* o *leche*; en donde *bebida* es el hiperónimo y *vino*, *agua* o *leche* son sus hipónimos a los que se generaliza la colocación y que, por lo tanto, también son seleccionados léxicamente por el mismo colocativo. Lo mismo ocurre con un *sentimiento profundo*, ‘intenso’, y *pena* o *amor profundos*. En cambio, pocas veces podemos hablar de sinonimia en colocaciones. No significa lo mismo *amor ciego* que *amor incondicional*, aunque en ambos casos se intensifique la emoción. No obstante, en ciertas ocasiones sí es posible, pero parece no ser lo más recurrente: *amor grande* y *amor intenso*; solo en aquellos contextos limitados en los que la colocación sea intercambiable. En colocaciones prototípicas, es decir, donde la coocurrencia léxica está poco basada semánticamente, es común que dos bases sinónimas elijan cada una su colocativo, pero no es frecuente encontrar colocaciones puras en este sentido (Alonso Ramos, 2016).

Con relación al *uso* de la base, atenderemos a los rasgos gramaticales y al registro. Primero, es útil conocer los rasgos gramaticales de la base ya que en muchas ocasiones se mantienen en la colocación, aunque no siempre. Por rasgos gramaticales, entendemos principalmente el régimen sintáctico de la misma. Como ejemplo podemos ofrecer *pensar concienzudamente (en algo o alguien)*. La colocación *pensar concienzudamente*, en la que *pensar* es la base, mantiene sus exigencias gramaticales, la preposición *en*. Los rasgos gramaticales también se incluirían la clase de palabra porque se necesita conocer los patrones gramaticales en que puede caber (Nation, 2013), es decir, la clase de palabra con la que se combina para expresar el sentido que se busca. Por ejemplo, colocaciones

⁶⁷ Colocación que se refiere a la que se usa en los juegos de azar y que está representada por uno de los números de mayor valor: as, rey, caballo...

que solo admiten el uso de la base en plural como *dar instrucciones* o la necesidad de un determinante y no otro como en *dar una opinión, dar tu opinión*, pero no **dar opinión*. Segundo, el registro de la base, que se suele mantener en la colocación (Vázquez Veiga, 2014). Por ejemplo, tenemos dos colocaciones como *perder dinero* y *o perder pasta* en las que como *dinero* tiene un registro neutro y *pasta* un registro coloquial.

El último de los componentes del conocimiento de una palabra, según el enfoque componencial (Webb & Nation, 2017), es el conocimiento de las posibles colocaciones de las que forma parte la palabra en cuestión. En el caso del conocimiento de una colocación, descartamos este aspecto porque para conocer una colocación no es necesario conocer qué otras colocaciones se pueden construir con esa misma base. Es, por lo tanto, la única diferencia con respecto al conocimiento de una unidad léxica cualquiera. Para conocer la colocación *cometer un crimen* no necesitamos saber que también existe *pagar por un crimen* o *crimen atroz*. Sin embargo, para conocer la palabra *crimen* en profundidad sí se debería conocer de qué colocaciones puede formar parte, es decir, con qué palabras se suele combinar restringidamente, además de otro tipo de conocimiento; recordemos que las colocaciones, al igual que las palabras, no se adquieren de forma aislada, sino que forman una red en el lexicon de cada aprendiz.

b) Conocimiento del vínculo fraseológico

Para explicar el vínculo fraseológico debemos acudir a la lexicografía y recuperar conceptos clave en la definición de colocación según la TST: *composicionalidad, control léxico* o *restricción léxica* y *concordancia semántica*. ¿Qué palabra acompaña a la base para expresar un sentido dado? ¿Qué sentido aporta el colocativo a la base? Conocer una colocación implica sobre todo conocer el vínculo fraseológico existente entre sus componentes. La característica definitoria de la colocación es su coocurrencia léxica restringida por la que la base controla y determina la elección del colocativo (Alonso-Ramos, 2017). El vínculo fraseológico consiste en conocer qué palabra selecciona la base para expresar un sentido dado, en producción, y qué sentido aporta el colocativo a la base, en comprensión. Es el componente imprescindible, determina si se conoce o no la colocación, tanto al producirla como al comprenderla.

Es el único de los cuatro componentes de la colocación (base, colocativo, vínculo y colocación) que no vemos ni oímos y forma parte del conocimiento de los hablantes de una lengua. Podemos decir que se trata del componente abstracto del conocimiento colocacional, frente a los otros tres que son concretos. Desde la teoría se ha tratado de explicar o de concretar de diferentes formas, desde la TST se han usado las funciones léxicas. Aunque es un instrumento útil, no resulta práctico ni operativo, requiere de un conocimiento profundo que no está al alcance de todos los usuarios, estaríamos hablando de metalenguaje. En un intento de facilitar este conocimiento que explica la relación entre la base y el colocativo se han propuesto glosas, sobre todo desde la lexicografía. Obviamente no hace falta conocer las funciones léxicas o glosas para demostrar el conocimiento del vínculo fraseológico, pero son la herramienta lingüística utilizada para poder explicarlo. Aunque podemos decir que las glosas son bastantes antinaturales, son mucho más cercanas a la realidad y cualquier hablante con un nivel aceptable podría glosar el sentido del colocativo en una colocación que domine.

Desde los estudios de enseñanza-aprendizaje de segundas lenguas, y más en concreto desde los estudios sobre el procesamiento de las colocaciones, se ha presentado esta restricción léxica como un condicionante ventajoso para el aprendizaje de las colocaciones. Se formula como una asociación que, dependiendo de su fuerza, se atrae más o menos. Esta ventaja se asocia a varias teorías⁶⁸ y está condicionada por distintas variables ajenas a la lexicografía como la adyacencia. Tanto la frecuencia de los elementos constituyentes como la congruencia se consideran factores determinantes que condicionan la velocidad del procesamiento, que como ya avanzamos, es el rasgo que, en comparación con otras unidades del léxico, más se ha tenido en cuenta en el procesamiento colocacional dejando otros aspectos como la delimitación de sus componentes, el orden o la relación entre ellos...

Otra de las particularidades del vínculo fraseológico es su carácter generalizador, que se relaciona con dos de los aspectos que hemos mencionado anteriormente, la *composicionalidad* y la *concordancia semántica*. Desde el punto de vista de la comprensión es posible, sin tener el conocimiento previo del vínculo, deducirlo a partir

⁶⁸ Por ejemplo, la *lexical priming theory* (Hoey 2005, 2012) y los *spreading activation models* (Collins and Loftus 1975).⁶

del conocimiento de otras combinaciones ya conocidas en la L2, pero también a partir de la similitud con colocaciones de la L1. Esta inferencia automática será, sin embargo, fuente de muchos de los errores colocacionales más comunes entre los aprendices. Podemos generalizar en casos claros de hiperonimia, si conocemos la palabra *batería* y la palabra *guitarra*, sabemos que con *guitarra* va a el verbo *tocar*, por ambos son instrumentos y la base *instrumento* selecciona el colocativo *tocar*, igual sus hipónimos. Esto no siempre funciona así, por ejemplo, muchos sentimientos seleccionan el verbo *sentir*, como *sentir amor*, *sentir tristeza*, pero no es válido para todos los sentimientos, no podemos decir **sentir manía* sino *tener manía*.

c) Conocimiento del colocativo de la colocación

El colocativo es el elemento restringido de la colocación que se selecciona léxicamente a partir de la elección de la base. Para conocer el colocativo también se trasladan los componentes de los conocimientos de las palabras y son muy similares al de las bases. En este caso no se tienen en cuenta las relaciones paradigmáticas en el apartado del significado. En la *Tabla 23* podemos ver los componentes y a continuación se explica uno a uno con ejemplos.

Colocativo	Forma	Forma oral y escrita
		Estructura morfológica
	Significado	Sentido
	Uso	Rasgos gramaticales
		Registro

Tabla 22. Conocimiento del colocativo

En cuanto a la *forma*, resulta imprescindible tanto reconocer como recuperar la palabra escrita y oral. También es importante conocer la estructura morfológica de los colocativos. Por ejemplo, en el caso de los verbos, se flexionan para adecuarse al contexto discursivo: *tocará el piano* a *habría tocado el piano*. Al hablar de la base decíamos que la colocación se mantenía normalmente con la flexión, y no siempre la derivación en la mayoría de los casos. Por ejemplo, *fruta madura* y **fruta inmadura*, en este caso no funciona la derivación, sino que utilizaríamos otro colocativo, *fruta verde*. Es posible, sin embargo, ver derivación en algunas clases de palabras que funcionan como colocativo, es el caso de las nominalizaciones de los verbos (Koike, 2001, p. 33) como en la colocación *perder el tiempo* que podemos transformar en *pérdida de tiempo*. Otros tan frecuentes

como el *sol se pone* (nombre + verbo) y *puesta de sol* (nombre de nombre) o *tomar una decisión* (verbo + nombre) y *decisión tomada* (nombre + participio) o *toma de decisión* (nombre de nombre).

Con relación al componente del *significado*, el sentido del colocativo es el aspecto más comprometido del conocimiento de la colocación, precisamente por su diferente grado de autonomía o dependencia semántica (Alonso Ramos, 2002). Se pueden clasificar las colocaciones según el grado de dependencia o autonomía semántica del colocativo como sigue:

- Palabras que solo usamos como colocativo al formar parte de una colocación concreta, son las unidades léxicas de *coocurrencia única*, como por ejemplo [*año*] *bisiesto* (Alonso Ramos, 2002; Mel 'čuk, 2012).
- Palabras que funcionan tanto en colocaciones como en combinaciones libres, de las que podemos diferenciar dos subtipos:
 - Las formas que adquieren un nuevo sentido al formar parte de la colocación; es el caso de *solo* que en combinación con *café* significa 'con un poco de leche', pero en ningún otro contexto tiene este sentido.
 - Las palabras que mantienen el mismo sentido cuando forman parte de una colocación y cuando lo hacen en una combinación libre; es el caso de *decir la verdad*. En estos casos el colocativo suele tener un sentido muy vago como el verbo *hacer, tener...* A este tipo de palabras se les suele llamar colocativos deslexicalizados; en el caso de los verbos hablaríamos de verbos de apoyo (Alonso-Ramos, 2004).

Por lo tanto, un colocativo es semánticamente autónomo cuando posee un sentido dado independiente de las UL con las que se combina. Por el contrario, es semánticamente dependiente cuando uno de sus sentidos está condicionado por las UL con las que se combina (Alonso Ramos, 2002, p. 84). Cuando el colocativo es semánticamente autónomo es necesario conocer el sentido de la UL del colocativo. Cuando el colocativo es semánticamente dependiente de la base, no es necesario

conocer todos los sentidos posibles de la colocación, pero sí el sentido que aporta a la colocación.

Al igual que ocurre con la base, la colocación, como unidad, puede heredar los rasgos gramaticales del colocativo, así que es necesario conocerlos. No obstante, es usual que el colocativo pierda las propiedades que tiene su contrapartida libre y que adquiera otras que no tiene cuando forma parte de la colocación (Alonso Ramos, 2002). Es el caso de *dar pena*, por ejemplo.

Finalmente, el registro o las restricciones de uso de las palabras que actúan como colocativo condiciona o puede alterar su registro. Ante una base neutra, un colocativo vulgar puede añadir un matiz coloquial, de grosería o de cortesía. Por ejemplo, podemos hablar de *cruzar la meta* o de *pasar la meta*. Aunque en algunas ocasiones la colocación permite cambiar el matiz, es el caso de *abrigar la confianza* frente a *tener confianza*. En un caso tenemos una colocación de un registro formal y en el otro una colocación de un registro más coloquial (Vázquez Veiga, 2014).

d) Conocimiento de la colocación

Los componentes del conocimiento de la colocación propios, como unidad, se reducen considerablemente en número con respecto a los de sus partes. Esto se debe a que la mayoría de ellos se derivan de las partes constituyentes, sobre todo, los relacionados con la forma. Entre la base y la colocativo puede aparecer un elemento, preposición, que une ambos componentes es el caso de *morir de alegría*, en donde se exige la presencia de la preposición *de*.

Colocación (como unidad)	Forma	Forma oral y escrita
		Estructura morfológica
	Significado	Definición
		Relaciones paradigmáticas
	Uso	Rasgos gramaticales
		OTRAS colocaciones
Registro		

Tabla 23. Conocimiento (propio) de la colocación

En relación con el significado, el sentido de la colocación es derivable por su carácter composicional. En cuanto al uso, tenemos en cuenta los aspectos sintácticos y las restricciones de uso, que pueden venir determinadas por la combinación o por alguno

de sus elementos. Finalmente, el conocimiento de otras colocaciones, en el que incorporamos el concepto de *colocaciones encadenadas* (Alonso Ramos & Wanner, 2007). Un ejemplo de este tipo de construcciones es el de *llamar la atención poderosamente* en el que partimos de la colocación *llamar la atención* que a su vez funciona como base de *poderosamente*. Esto es así porque no podemos decir solo **atención poderosa* o **llamar poderosamente*.

B. Estándar y grados del conocimiento colocacional

El conocimiento de la colocación y, por lo tanto, el estándar conlleva el conocimiento de todos los componentes de la colocación. En la *Tabla 24* se pueden ver todos de forma sistematizada. El pleno dominio de cada uno de estos componentes y subcomponentes del conocimiento de la colocación implica el grado *óptimo* de conocimiento, es decir, describe lo que se conocería como *normalidad* o *estándar* del conocimiento de las colocaciones. El conocimiento de una colocación es todavía más complejo que el de la palabra: hay dos palabras y tiene unos rasgos propios concretos; además, la suma de los significados de las palabras que la compone no siempre nos permite inferir el significado de la colocación. Dado que el conocimiento necesario para comprender difiere del necesario para producir, abordaremos el conocimiento colocacional desde estas dos perspectivas. Producir una colocación es una tarea más compleja, implica un proceso cognitivo más elevado que comprender una colocación. El propósito ya no es describir qué significa conocer una colocación, sino qué se tiene que conocer de la colocación para comprenderla y qué se tiene que conocer de la colocación para producirla.

Base	Forma	Forma oral y escrita
		Estructura morfológica
	Significado	Sentido
		Relaciones paradigmáticas
Uso	Rasgos gramaticales	
	Registro	
Vínculo fraseológico		
Colocativo	Forma	Forma oral y escrita
		Estructura morfológica
	Significado	Sentido
		Relaciones paradigmáticas
Uso	Rasgos gramaticales	
	Registro	
Colocación	Significado	Sentido
		Relaciones paradigmáticas
	Uso	Rasgos gramaticales
		Colocaciones
	Registro	

Tabla 24. Estándar del conocimiento colocacional en su grado óptimo

Sobre el conocimiento y adquisición del léxico se aceptan tres premisas (Nontasee & Sukying, 2021) que podemos trasladar al conocimiento y adquisición de las colocaciones: el conocimiento se organiza en diferentes aspectos, se adquieren en diferentes etapas y a diferente ritmo. Lo que se relaciona con la *carga de aprendizaje*, que podemos definir como grado de dificultad en el aprendizaje de una palabra (Webb & Nation, 2017, p. 25). Podemos aceptar también estas premisas para el conocimiento colocacional. Por lo tanto, la *facilidad colocacional* o la *carga de aprendizaje*⁶⁹ de una colocación en concreto se caracteriza por el dominio de todos sus componentes, adquiridos en diferentes etapas y a diferente ritmo.

Así, distinguimos diferentes grados de dominio o conocimiento de una unidad del léxico, la colocación, que de forma similar ya se aplicaron al conocimiento de las palabras: conocimiento parcial o preciso, conocimiento profundo o producción-comprensión (Henriksen, 1999). El pleno conocimiento de cada uno de los componentes se identificará con este conocimiento *óptimo*, que condensamos en el ANEXO V y VI. Sin embargo, el dominio de todos los componentes que integran el conocimiento óptimo de una colocación no es necesario para entenderla o usarla en ciertos contextos. En el caso de las colocaciones, al igual que en el caso de las palabras, podemos aplicar el concepto de *vaguedad*, ya que la colocación puede no conocerse, conocer parcialmente o, incluso, desconocerse. La cantidad de aspectos que hay que conocer de una palabra y los grados del conocimiento es el punto de partida del enfoque de los componentes del conocimiento que propone Nation y que hemos seguido también en este apartado. Si lo trasladamos al conocimiento de las colocaciones, podemos afirmar que hay muchos aspectos que conocer sobre una colocación en particular y que hay muchos grados de conocimiento. El criterio para determinar el estado del conocimiento son las expectativas específicas de aprendizaje (Jang, 2017).

En el proceso de comprensión, la forma de la combinación viene dada y, por lo tanto, el objetivo será recuperar su sentido o, en algún caso, menos habitual, reconocer su sentido. La recuperación vendrá condicionada por el contexto. La comprensión desde el reconocimiento no es habitual en situaciones comunicativas reales, pero sí en las

⁶⁹ En el apartado 2.4.5.2. sobre el aprendizaje de colocaciones se dedica un apartado a las causas lingüísticas y no lingüísticas que influyen en la carga de aprendizaje de una colocación.

tareas habituales de enseñanza-aprendizaje, en las que el aprendiz debe, por ejemplo, seleccionar el sentido de una colocación entre varias opciones. El primer aspecto imprescindible es el reconocimiento de la forma de la base, ya sea de forma escrita u oral y su conexión con su sentido, es decir, la recuperación del sentido de la base. Recordemos que las colocaciones son unidades fraseológicas composicionales. También, los aprendices deben inferir⁷⁰ el vínculo fraseológico, es decir, la necesidad de combinar esa palabra con otra. Esto se traduce en la búsqueda de otra palabra que complete el sentido de la base y que no se lleva a cabo de forma arbitraria. Además, es necesario reconocer la forma del colocativo y conectarlo con el sentido que aporta a la combinación. Esta conexión no tiene por qué conocerse de antemano. Según el tipo del colocativo ante el que nos encontremos, se podrá establecer la conexión directamente, forma-significado, o se inferirá a partir del contexto o los conocimientos previos del aprendiz, por ejemplo, la L1. De este modo, completan el último de los aspectos, la conexión entre la forma de las palabras constituyentes y el sentido que transmiten combinadas. Cuando la combinación es transparente el proceso es automático, como si de una combinación libre se tratara, lo que no es posible hacer en la producción. Aun tratándose de una colocación transparente el aprendiz necesitará saber qué palabra acompaña a la otra, así, podrá decir *sorpresa grande*, pero también *sorpresa mayúscula*.

Una de las cuestiones problemáticas es la incorporación o no del aspecto que alude a las relaciones paradigmáticas de la base como imprescindible para el dominio de la colocación en cuestión. No podemos afirmar que, para conocer una colocación, al menos parcialmente, haya que conocer la base en su grado óptimo. Por lo tanto, no será imprescindible conocer las relaciones paradigmáticas de la base. Desde la comprensión, sabemos que el vínculo y el sentido del colocativo en la combinación se puede inferir a partir de diferentes fuentes, principalmente del contexto. Desde la producción tampoco sería imprescindible ya que tanto el vínculo como el colocativo son deducibles, aunque en menor medida, a partir de diferentes fuentes: contexto, influencia de la L1, conocimiento previo en la L2... Esta información debe ser valorada antes de ponerla en uso, ya que también suele ser la causa de los errores colocacionales más típicos, es el

⁷⁰ Oponemos *inferir* a *deducir*. *Inferir* como el proceso por el que se sacan conclusiones desde el reconocimiento y *deducir* como la obtención de información a partir de los conocimientos previos que no se ven ni se oyen.

caso de **golpear los huevos* en lugar de *batir los huevos* por una traducción literal de la colocación *beat eggs*.

BASE	FORMA	Forma	Es imprescindible reconocer la forma de la base, ya que seas escrita u oral
	SIGNIFICADO	Sentido	Es imprescindible recuperar el sentido de la base.
VÍNCULO			Es imprescindible al menos inferir la conexión con el colocativo.
COLOCATIVO	FORMA	Forma	Es imprescindible reconocer la forma del colocativo, ya sea escrita u oral.
	SIGNIFICADO	Sentido	Es imprescindible inferir el sentido del colocativo.
COLOCACIÓN	SIGNIFICADO	Sentido	Es imprescindible al menos inferir el sentido de la colocación

Tabla 25. Lista de aspectos del conocimiento de la colocación mínimos para comprender una colocación

El proceso de producción de la colocación es inverso, partimos del sentido, y debemos recuperar la forma, en condiciones normales, o reconocerla en situaciones concretas de aprendizaje que permitan usar la combinación. El número de aspectos del conocimiento implicados aumenta, ya que se deben recuperar aspectos tanto de las palabras integrantes como de la colocación. Así, a partir del sentido de la base, es imprescindible recuperar la forma, ya sea escrita u oral. Además, es necesario deducir la conexión con el colocativo, esto es, el vínculo fraseológico. Igualmente, se podrá establecer la relación entre el sentido y la forma del colocativo. De la palabra que funciona como colocativo, es también necesario conocer tanto su estructura morfológica como sus rasgos gramaticales. Finalmente, conocer los rasgos gramaticales que le son propios a la combinación.

BASE	FORMA	Forma	Es imprescindible recuperar la forma de la base, ya sea escrita u oral.
		Estructura morfológica	Es imprescindible, al menos conocer los aspectos relativos a la flexión de la palabra que funciona como base.
	SIGNIFICADO	Sentido	Es imprescindible recuperar el sentido de la base.
	USO	Rasgos gramaticales	Es imprescindible, al menos, identificar la categoría gramatical de la palabra que funciona como base.
VÍNCULO			Es imprescindible al menos deducir la conexión con el colocativo.
COLOCATIVO	FORMA	Forma	Es imprescindible recuperar la forma del colocativo, ya sea escrita u oral.
		Estructura morfológica	Es imprescindible.
	SIGNIFICADO	Sentido	Es imprescindible conocer el sentido del colocativo.
	USO	Rasgos gramaticales	Es imprescindible.
COLOCACIÓN	SIGNIFICADO	Sentido	Es imprescindible conocer el sentido de la colocación
	USO	Rasgos gramaticales	Es imprescindible.

Tabla 26. Aspectos imprescindibles del conocimiento de la colocación para su producción

Estos puntos del aprendizaje deben encontrarse en *escalas de dominio* en forma de rúbrica orientadas a la competencia colocacional. Presentamos así dos rúbricas, una

para la comprensión de una colocación (ANEXO V); otra para la producción de la colocación (ANEXO VI). Con esto, tratamos de concretar la representación léxico-semántica gradual para desarrollar una intervención pedagógica efectiva (Sunderman & Forcelini, 2022), fin último del test de diagnóstico. Una colocación, igual que cualquier otro elemento de la lengua, se puede usar mal o bien, es decir, con un diferente grado de corrección. Por ejemplo, si a la hora de usar la colocación se elige bien el colocativo, es decir, se conoce el vínculo fraseológico, pero se comete una falta de ortografía, sería posible afirmar que no se conoce la colocación de forma óptima sino solo de una forma parcial. La ortografía de la palabra no parece ser uno de los componentes que primero se aprende de una palabra. La interrelación entre los componentes del conocimiento (González-Fernández & Schmitt, 2020) es una de las claves para el desarrollo de las escalas de dominio.

Veamos el proceso a partir de la colocación *café cortado* en una conversación como la que sigue:

_ ¿Quieres café con leche?

_ No, prefiero un café cortado, me gusta con poca leche.

A la hora de entender esta colocación, caben dos opciones. En la primera opción se conoce la base, la palabra *café*, y se conoce el colocativo *cortado*; de modo que no habría ningún problema. En la segunda opción, se conoce la base, pero no el sentido del colocativo. Sin embargo, es posible inferir su sentido ya que el contexto lo permite. A la hora de producir esta colocación es imprescindible partir del conocimiento de la base, tal y como pasa en la comprensión. También es imprescindible conocer el vínculo fraseológico, el colocativo y las particularidades de la combinación. No es posible usarla si no sabes que con *café* se usa el adjetivo *cortado* para indicar ‘con un poco de leche’.

4.2. Ciclo del TCE

La tarea de construir un test de lengua que busque la calidad de la evaluación requiere seguir los pasos de un plan claro para el diseño, desarrollo y administración. Como ya explicamos, desde la Evaluación de Lenguas se propone seguir un ciclo que sirva como guía en este proceso (vid. 3.1.3). Debido a las particularidades del TCE presentamos

un ciclo a medida de sus necesidades respetando los pasos básicos y basándonos en los antecedentes más próximos.

Proponemos un ciclo en el que distinguimos seis pasos: 1) desarrollo del test (definir el constructo y producir especificaciones), 2) montaje (producir los materiales reales), 3) administración (recoger la información sobre la habilidad de cada participante), 4) corrección (recoger un resultado preciso y fiable de cada participante), 5) calificación (clasificar la actuación de cada participante en una categoría significativa, de modo que los resultados puedan interpretarse) y 6) información de los resultados (analizar los resultados para su interpretación). Cada uno de los estadios se presenta con un objetivo asociado. Aunque la decisión de diseñar y crear el test está presente en el esquema, no se considera parte del ciclo, sino un paso previo. Además, se distinguen dos fases: la primera, únicamente ocupada por el desarrollo del test, cuyo objetivo es el de producir las especificaciones (vid.4.3.3); y la segunda, que corresponde al uso del test, más concretamente con la evidencia, que engloba cuatro pasos: el montaje, la administración, la corrección y la calificación. Cada uno de los pasos de la construcción del test cuenta con una serie de microtarefas necesarias para completar el proceso. En la *Tabla 27* se especifica el ciclo que utilizaremos para el TCE.

FASES			OBJETIVO
Proyección			Comprobar que no existe ningún otro test ya creado con el mismo propósito
Desarrollo del test	Planteamiento	Participantes. Propósito. Contexto educativo. Definición del constructo.	Producir las especificaciones del test.
	Diseño	Estandarización	
	Pilotaje	Estructura. Tareas.	
Uso del test	Montaje	Puesta en marcha del test.	Producir los materiales reales del test
	Administración	Producción de materiales. Control de calidad. Construcción del test.	
	Corrección	Recopilación de la información sobre la habilidad de cada participante.	
	Calificación e interpretación	Adjudicar una nota rigurosa y fiable de cada aprendiz.	
	Retroalimentación	Clasificar y valorar la actuación de cada participante de acuerdo con los parámetros establecidos.	
Reparación del proceso de aprendizaje	Informe de los resultados de las pruebas		Proponer una propuesta metodológica
	Identificación de debilidades y fortalezas		
Elaboración de una propuesta metodológica			

Tabla 27. Ciclo del TCE

En la fase inicial, o fase de *proyección* hay que asegurarse de que no existe un test previo que reúna los requisitos necesarios para satisfacer los objetivos elegidos. La fase

de desarrollo cuenta con tres fases: planteamiento, diseño y pilotaje. En el *planteamiento* se describen las características de los potenciales participantes en el test, se define el propósito y se pone en relación con el contexto educativo, por ejemplo, el MCER, y se explica para qué se va a utilizar el test. También es el punto en el que se define la habilidad que va a ser medida, es decir, el constructo. Además, se debe hacer hincapié tanto en los efectos de la evaluación como en cuestiones prácticas. Ya entran en juego algunos de los principios expuestos en el apartado anterior como son la *validez de constructo*, el *impacto* y la *practicabilidad* (o *funcionalidad*).

En el *diseño* el objetivo es la estructura general del test (forma) y todos los aspectos relacionados con el contenido de este. Algunos de los rasgos técnicos que se muestran en esta fase son la duración, el número de ítems, el tipo de ítem, el formato, la forma de calificación, etc. Se busca el equilibrio entre las necesidades y las limitaciones. El propósito de la fase de *pilotaje* es la de rodar el test, ponerlo en marcha para detectar los posibles errores y así poder rectificar el borrador de las especificaciones. Esta fase demuestra que los pasos del ciclo no son secuenciales, sino que se va pasando de uno a otro: *planteamiento* > *diseño* > *montaje* > *administración* / *pilotaje* para hacer una revisión y volver a empezar a partir de los datos obtenidos. De forma que el proceso completo sería *planteamiento* > *diseño* > *montaje* > *administración* / *pilotaje* > *planteamiento* (*revisión*) > *diseño* (*revisión*) > *montaje* > *administración*, etc. Es la razón de que el proceso de construcción se presente como un ciclo, de modo que el dar un paso no significa olvidar el anterior sino mantenerlo en un continuo proceso de mejora. Una vez que las especificaciones están preparadas se pasa al proceso de *montaje* del test, que a su vez consta de tres etapas o *microtarear*s: la producción de materiales, el control de calidad y la construcción del test propiamente dicha.

En la fase de *producción*, de acuerdo con una serie de requisitos exigidos por las conclusiones a las que se llegaron en la fase anterior, se determina cuántos ítems son necesarios y de qué tipo. Asimismo, se ciñe a una serie de parámetros que tienen por objetivo evitar la confusión y los malentendidos. A este respecto hay que aplicar guías para su elaboración según el tipo de formato seleccionado (Hoshino, 2013), por ejemplo. En cuanto a la fase de *control*, es esencial realizar un examen de calidad, una fase determinante en la que se revisan los ítems construidos y, además, se hace un simulacro

con un grupo similar de participantes, con un mismo contexto, para rectificar aquellas cuestiones mejorables. Por último, la *construcción*, una vez que todos los materiales están creados y validados se forman los materiales.

La *administración* tiene como objetivo recoger la información de forma rigurosa y fiable sobre la competencia evaluada. Este proceso plantea algunos retos: la colaboración tanto de centros como de los participantes, las circunstancias externas al test como puede ser el lugar, el ruido, etc. o las condiciones desfavorables que se puedan dar. Después, se inicia la fase de corrección que se basa en los parámetros establecidos en una guía de corrección previa. Hay una diferencia importante entre los test de respuesta cerrada y los de respuesta abierta, aunque los dos pueden ser corregidos mediante una guía, en los de respuesta abierta es común que aparezcan respuestas ambiguas, incluso alguna que no esté prevista en la guía de corrección por lo que es necesario regresar a ella y modificarla. Finalmente, la *calificación* en la que se toman las decisiones expuestas en el argumento de validez.

4.3. Diseño del TCE

Para abordar el reto de diseñar el TCE es importante partir de sus características como instrumento de evaluación y así definir la forma en la que se va a medir el conocimiento colocacional concretando la estructura y forma del test. Tanto el propósito como la información que se quiere obtener se han tenido en cuenta para la planificación del tipo de evaluación, que hemos definido como un test de diagnóstico validado a partir de un enfoque basado en argumentos. El TCE se define, además, por su carácter *discreto*, *selectivo* y *dependiente* (del contexto) atendiendo a las *dimensiones de la evaluación*⁷¹ (Read, 2011). Es *discreto* en tanto que mide un constructo independiente, es decir, mide el conocimiento colocacional y no otro componente de la competencia lingüística. También es *selectivo*, ya que se seleccionan las colocaciones concretas que se van a utilizar y no se extraen de un texto dado. Por último, es un test *contextualizado*, es decir,

⁷¹ Las dimensiones propuestas responden a tres dualidades: el carácter *discreto* o *incrustado* (*embedded*); el carácter *selectivo* o ampliado (*comprehensive*); y *dependiente* o *independiente* del contexto.

es un test en el que no se muestran las colocaciones de forma aislada, sino que las unidades de la lengua evaluada simulan una situación posible del uso de la lengua.

Cumplir con los requisitos que imponen la definición del TCE como instrumento de evaluación y alcanzar los propósitos del test nos va a permitir diseñar una estructura y elegir unos formatos concretos. El estudio de los antecedentes en la evaluación del léxico y de las colocaciones hace que podamos adaptar alguna de las propuestas previas que se adecúan a estos requisitos y propósitos. El concepto del *conocimiento gradual y escalar implicativo* nos lleva a tomar como modelo la estructura propuesta por CATSS (Aviad-Levitzky et al., 2019). La elección de los formatos elegidos varía con respecto al modelo: se corresponde con el propósito y depende del tipo de información que se quiere extraer (Kremmel & Schmitt, 2016, p. 387), tomando como modelo algunos de los formatos elegidos en los test estandarizados del léxico y algunos test de colocaciones (Nizonkiza, 2015, 2017). Además de tener en cuenta la competencia, en el diseño se deben de tener en cuenta las *estrategias de aprendizaje de idiomas y estrategias de gestión de pruebas / realización de test* que también van a determinar el éxito de la evaluación (Cohen, 2017).

En primer lugar, explicaremos en qué consiste la adaptación del test de léxico adaptativo. Posteriormente, definiremos la estructura y forma del TCE, centrándonos en el diseño de cada uno de los formatos que se usarán. Por último, detallamos las especificaciones que guiarán la administración, análisis e interpretación de los resultados.

4.3.1. Adaptación de CATSS

El diseño del TCE sigue el modelo propuesto en CATSS⁷², *The New Computer Adaptive Test of Size and Strength* (Aviad-Levitzky et al., 2019; Laufer & Goldstein, 2004): una batería de cuatro test que proporcionan una visión global y jerarquizada del conocimiento léxico. Esta estructura nos permite cumplir con los propósitos de nuestro TCE, comprobar cómo es el conocimiento colocacional de los aprendices para hacer un diagnóstico que discrimine entre fortalezas y debilidades. El test original parte de una

⁷² Test adaptativo que mide el conocimiento léxico de forma global y gradual atendiendo a la producción, la comprensión, el reconocimiento y la recuperación. En el apartado 3.3.2.1 sección E) se hace una descripción más pormenorizada del test.

hipótesis que jerarquiza el sistema de conocimiento en base a cuatro aspectos del conocimiento: la *comprensión* y la *producción*, por un lado; el *reconocimiento* y la *recuperación*, por el otro. De esta manera el conocimiento se presenta como una realidad gradual y no como un concepto dual, que opone el conocimiento al desconocimiento como única oportunidad. El test consta, por lo tanto, de una batería de cuatro formatos diferentes, resultantes de la combinación de los cuatro aspectos del conocimiento anteriormente citados, que pretende ofrecer una visión panorámico de la competencia y, sobre todo, del conocimiento colocacional.

Existen diferencias notables entre las características del TCE y CATSS. La primera es la unidad evaluada, mientras que CATSS es un test de léxico, centrándose en el conocimiento de las palabras, el TCE es un test que mide el conocimiento que los aprendices de ELE tienen de las colocaciones. La segunda diferencia radica en la forma, mientras CATSS es un test informatizado y adaptativo, el TCE, aunque cuenta con una versión informatizada⁷³, no es adaptativo. Esto hace que los criterios de corrección no puedan tomarse del modelo. Por último, la población a la que el test está destinada también es diferente. CATSS está pensado para participantes con inglés como L1 o, por lo menos, con el inglés como lengua de intercambio en su primera versión, en las sucesivas versiones ya presenta formatos utilizando únicamente la L2, el inglés. Sin embargo, el TCE se administra en una situación de inmersión por lo que la L1 de los participantes no es siempre el inglés. Esto hace que en el TCE nos veamos obligados a no utilizar los formatos de traducción (que sí se usan en el formato original de CATSS). Por lo tanto, el test se ha diseñado únicamente en la L2, y no de forma combinada (hebreo e inglés), como lo es CATSS.

4.3.2. Estructura y forma del TCE

El TCE está formado por una batería de cuatro versiones o modalidades para obtener una imagen global del conocimiento que los aprendices de español tienen de las colocaciones. Pese a que las cuatro modalidades representan el conocimiento de una forma panorámica, cada aprendiz solo realiza dos modalidades del test, que organizamos en opciones (A y B). En la construcción del test hemos tenido en cuenta tanto las

⁷³ Por cuestiones relacionadas con la administración de test, fue necesario crear una versión informatizada paralela a la original en papel (vid. 5.1).

directrices generales para la creación de ítems como las particularidades formales de cada uno de los modelos elegidos. Estas cuatro modalidades se corresponden con las etapas y los grados del conocimiento de la colocación, desde el que hipotéticamente implica más dificultad al que menos. Son el resultado de la combinación de *producción* y *comprensión*, *recuperación* y *reconocimiento*.

	Producción	Comprensión	
Recuperación	TEST 1: PRECU	TEST 2: CRECO	Modalidad A
Reconocimiento	TEST 3: PRECO	TEST 4: CRECU	Modalidad B

Tabla 28. Estructura del TCE

Así, en la primera modalidad, PRECU, se medirá la *producción* desde la *recuperación* (ANEXO XXV); en la segunda, CRECO (ANEXO XXVI), la *comprensión* desde el *reconocimiento*; en la tercera, PRECO, *producción* desde el *reconocimiento* (ANEXO XXVII); y en la última, CRECU, *comprensión* y desde la *recuperación* (ANEXO XXVIII), como se puede ver en la *Tabla 28*. Para testar cada una de estas modalidades se utiliza un formato adecuado a la información que se pretende obtener. Además, no todos los componentes del conocimiento de la colocación están presentes en todos los modelos, lo que va a permitir observar e interpretar las debilidades y fortalezas del conocimiento / de la competencia. Dos de las modalidades son de *respuesta construida*⁷⁴ (PRECU y CRECU) y dos de respuesta seleccionada (PRECO y CRECO).

4.3.2.1. Modalidad 1: PRECU

La modalidad que mide la *producción* desde la *recuperación*, y que denominamos PRECU (ANEXO XXV), sigue uno de los formatos tradicionales para medir la producción del conocimiento léxico, el C-Test (Chapelle, 1994; Laufer & Nation, 1999) y que ya se ha utilizado para medir la *extensión* del conocimiento colocacional (Nizonkiza, 2012). El C-test o *cloze test* se inscribe dentro del enfoque integrativo que plantea la evaluación de la lengua en contexto. Consiste en la presentación de una frase (*stem*) en la que aparece la base de la colocación contextualizada como estímulo, pero no el colocativo, en su lugar aparece un hueco (que representamos con puntos suspensivos), y el participante debe

⁷⁴ Los formatos de respuesta construida son aquellos en los que el aprendiz debe recuperar su respuesta y se opone a los formatos de respuesta seleccionada en los que los participantes deben elegir una de entre varias propuestas (Haladyna et al., 2002).

recuperar el colocativo de la colocación. En este caso no se ofrecen opciones de respuesta a los participantes. Para controlar la respuesta abierta y evitar las distintas variantes se ha decidido dar como pista la primera letra del colocativo (Aviad-Levitzky et al., 2019; Moreno-Jaén, 2007b).

Al contrario que en CATSS, en donde se opta por un contexto mínimo, que se concreta en la aportación de un sinónimo (que queda descontextualizado), en el TCE, el contexto ofrece la información clave para que el aprendiz pueda elegir el colocativo. En el caso de las colocaciones, el contexto al aprendiz le da la opción de conocer el vínculo fraseológico, componente imprescindible a la hora de recuperar el colocativo. En algunas ocasiones el contexto dado se presenta en forma de *glosa* y otras se limita a brindar un uso habitual de la colocación, como en el *Ejemplo 1*.

La niña tiene el pelo r...

Ejemplo 1. Muestra (descartada) de ítem de PRECU

Las opciones podrían ser *rizo* o *rubio*, pero si nuestra colocación meta es *pelo rizo* debemos ofrecerle al aprendiz información que le permita inferirlo. Read (2011) afirma que el uso o no de contexto depende de la situación o del propósito del test. En un test de certificación, por ejemplo, necesitas saber lo que se conoce o lo que estudió cada participante, pero en un test de diagnóstico el hecho de que se parezca lo máximo posible a la situación real, en la que la adivinación es un componente más de la realidad, no es un hecho disuasorio. El contexto podría verse afectado por la congruencia con la L1 o el propio conocimiento de la L2, pero estos dos aspectos forman parte de la realidad comunicativa, no los podemos omitir porque forman parte del aprendizaje y uso de la lengua. Sin embargo, en un contexto como el que sigue, si se conoce la colocación (vínculo fraseológico y colocativo, es decir, la palabra que la base exige para el sentido “buscado”), no habrá duda de que la palabra que se busca es *rizo* y no *rubio*. De modo que, evitamos ofrecer un contexto *extralingual* y el *interlingual* (Carton, 1971), aunque no evitamos el *intra lingual*.

La niña tiene el pelo r..., forma ondas.

Ejemplo 2. Muestra de ítem de PRECU

Este formato, además del grado de profundidad del conocimiento (en este caso, la producción desde la recuperación), se utiliza en la modalidad que entraña *a priori*

mayor dificultad para los aprendices. PRECU evalúa, en concreto, los siguientes componentes del conocimiento colocacional: a) el reconocimiento de la forma escrita de la base, b) el conocimiento del sentido de la base, c) la buena elección del colocativo, d) la producción de la forma escrita del colocativo y e) el conocimiento del sentido del colocativo. Es cierto que en la corrección se estableció como criterio la aceptación como correcta de colocativos en los que no se demostrara conocer la forma. En el siguiente ejemplo se puede ver un ítem.

INSTRUCCIONES

Completa la frase con la palabra que no está. Como pista tienes la primera letra que la palabra que buscamos.

Tiene el pelo r..., está muy guapa con esas ondas.

Ejemplo 3. Muestra de ítem de PRECU con instrucciones

4.3.2.2. Modalidad 2: CRECO

Se mide la comprensión desde el reconocimiento y se utiliza el formato típico de opción múltiple (ANEXO XXVI). El ítem (*stem*) es una frase incompleta que contiene la colocación completa, con sus dos elementos (base y colocativo), a la que le falta una continuación natural en la que refleje o demuestre que conoce el significado de la colocación. La información que se busca es una glosa que sintetice el sentido de la colocación. Esta continuación natural o glosa se debe escoger de entre cuatro opciones, tres de las cuales son incorrectas, los distractores, y solo es una correcta. Los distractores pretenden representar errores comunes como la asociación fonética con otras palabras, diferentes relaciones semánticas como la confusión con el contrario / antónimos u otros sentidos de la base (incluso la homonimia).

El formato de opción múltiple es una de las opciones más populares para medir el vocabulario. Entre sus ventajas están las de a) ofrecer resultados objetivos, b) ser prácticos tanto para la administración como para puntuación (recuento de resultados), c) permitir testar un grupo amplio de palabras meta en un tiempo limitado (Kremmel & Schmitt, 2016, p. 378). Sin embargo, no se trata de un formato perfecto, sino que suscita críticas, entre las más habituales, la alta influencia de los distractores, el factor del azar y

la sobreestimación (Stewart, 2014). Una de las posibles soluciones es la de optar en lugar de un distractor al uso por una opción “no sé”, lo que incrementaría la fiabilidad. En nuestro caso no parece útil, ya que la información que nos proporcionan los distractores es tan relevante como la de la opción correcta. Recordemos que se trata de un test de diagnóstico en el que se busca detectar las debilidades del conocimiento colocacional, para lo que el análisis de los distractores es de gran utilidad. Otra crítica es la falta de naturalidad de este tipo de tareas (Hoshino, 2013), que en este caso tratamos de suplir a partir del contexto.

En la construcción de los test de opción múltiple, la creación de los distractores y la relación entre estos influye directamente sobre la dificultad del test. De ahí que sea una de las tareas que plantea mayor esfuerzo en la fase de construcción, ya que deben mantener la verosimilitud⁷⁵ (Hoshino, 2013). Otras cuestiones que se suelen tener en cuenta es la frecuencia de las palabras que aparecen como distractores, incluso perteneciendo a la misma banda de frecuencia, siempre para mantener la misma dificultad tanto en los distractores como en la opción correcta. En el caso del TCE, como ni las colocaciones meta ni las palabras integrantes se extrajeron de listas de palabras frecuentes (ni de corpus directamente), tampoco se utilizó como criterio para la elaboración de los distractores. Sí se tuvo en cuenta la dificultad de las palabras seleccionadas para la creación de los distractores atendiendo a la dificultad de estas y a su presencia en el PCIC. En este caso, la elaboración de los distractores y, sobre todo, de la opción correcta nos lleva a darle un sentido a la colocación, a buscar su definición o descripción semántica, que se hacemos o mediante glosas (Barrios Rodríguez, 2012), recurso más adecuado para este contexto. Para la elaboración de las glosas hemos seguido una serie de pautas (Alonso-Ramos, 2006) . En ese artículo se busca la mejor formulación desde un punto de vista lexicográfico, pero que también se podría aplicar al ámbito que nos ocupa. Pese a que en los últimos años se ha trabajado en la generación de

⁷⁵ Este autor presenta en su estudio tres tipos de distractores: relacionados paradigmáticamente (PARA), sintagmáticamente (SYN) o los que no tienen relación sintagmática ni paradigmática. Se demostró que tanto los distractores que mantienen una relación paradigmática como sintagmática funcionan mejor que los que no tienen ninguna relación, más específicamente los sintagmáticos, aunque esto añade una mayor dificultad al test.

paráfrasis de forma automática (Zhang et al., 2014), en la construcción del TCE se hacen de forma tradicional.

Con este formato medimos principalmente el sentido de la colocación, sin embargo, también entran en juego los siguientes aspectos del conocimiento colocacional: a) reconocimiento la forma escrita de la base, b) reconocimiento la forma escrita del colocativo e c) identificación del sentido de la colocación. Dando por hecho que el vínculo fraseológico es el aspecto clave que nos indicaría si se conoce o no la colocación, el formato aporta información relevante, pero parece el menos determinante a la hora de dibujar la imagen global del conocimiento colocacional. Y, por lo tanto, aquel modelo en el que creemos que los aprendices obtendrán mejores resultados. Como se puede ver en el ejemplo, los distractores muestran diferentes tipos de relación con la colocación meta: algún juego semántico (doble sentido de la base o sobre todo del colocativo), sentidos cuasi-sinónimos e hiperónimos, etc...

INSTRUCCIONES

Elige la continuación de la frase que mejor indique el sentido de la colocación.

Tiene el pelo rizo, está muy guapa...

- a) ... con esas ondas.
- b) ... con el rubio.
- c) ... con la coleta.
- d) ... con la curva.

Ejemplo 4. Muestra de ítem de CRECO

4.3.2.3. Modalidad 3: PRECO

Se mide la producción desde el reconocimiento también a partir de un formato de elección múltiple (ANEXO XXVII). El *ítem* consiste en una frase en la que falta el colocativo, pero se presenta la base de la colocación y el aprendiz debe escoger entre las cuatro opciones, cuatro posibles colocativos. Tres funcionan como distractores y el otro como opción correcta. Se presenta la colocación en dos partes, es decir, no como *chunks* sino la base en el ítem y el colocativo como uno de los distractores. Lo que se mide de forma clara es el vínculo fraseológico, la elección del colocativo. Los distractores, además, tienen como función representar posibles errores comunes que se cometen y que pueden dar lugar a confusión.

Se muestra la colocación fragmentada, por un lado, la base contextualizada; por el otro, el colocativo, que deben elegir entre las cuatro opciones. Recientemente, desde el *procesamiento analítico*⁷⁶ (como *chunk*) (Boers et al., 2014), se favorecen las actividades contextualizadas, pero también aquellas en las que se presenta la colocación “intacta”, esto es, sin fragmentarse, los dos componentes como un *chunk*. De ahí que en formatos como estos se presente un hueco para toda la colocación y no para el colocativo. Encontramos varias opciones: que todas las colocaciones compartan la base (*cometer un crimen, combatir un crimen o investigar un crimen*) o que se presenten diferentes bases con diferentes colocativos (*cometer un delito, dictar sentencia o aprobarse una ley*). En el segundo caso pensamos que la actividad carece de sentido al confrontar con su propia definición; es la base la que selecciona léxicamente el colocativo y por lo tanto sería suficiente con que el aprendiz identificara la base de la colocación. Así que debemos dejar claro qué se pregunta, en este caso el foco está en el vínculo fraseológico y, por lo tanto, lo más lógico es dar la base, contextualizada, y que el aprendiz tenga que elegir entre los colocativos. Debemos aclarar también que los estudios a los que nos referimos no tratan los formatos de los test, sino las actividades de los manuales de aprendizaje. De esta manera se evitarían errores comunes en las colocaciones como las asociaciones cruzadas. Un argumento válido para una actividad de aprendizaje no lo es para una actividad de diagnóstico en la que precisamente se trata de corroborar cuáles son los problemas en el aprendizaje de las colocaciones.

En nuestro caso, otro punto para tener en cuenta en la creación de ítems son los errores que suelen cometer los aprendices con las colocaciones. Como orientación nos servimos de una de las tipologías de errores basadas en corpus (Alonso Ramos, Wanner, Vincze, et al., 2010). Partimos de esta por compartir con ella la base teórica sobre qué se entiende por colocación. La clasificación que se presenta atiende a tres dimensiones: la *localización*, la *descripción* y la *explicación*. Dadas las características del formato, solo nos centraremos en los errores propios del colocativo a nivel léxico. Atendiendo a la dimensión explicativa: los errores pueden ser *interlinguales* (procedentes de la L1) o

⁷⁶ Desde el procesamiento analítico se entienden las colocaciones como combinaciones procesadas cognitivamente de forma analítica, es decir, como un todo (o *chunk*), sin distinciones. Esta idea es compartida por el enfoque léxico (Lewis, 2000), pero no es el que se sigue en este estudio. Desde el enfoque tomado en este estudio, solo las locuciones se procesarían de esta manera.

intralinguales (propios de la L2). Priorizamos los errores intralinguales sobre los interlinguales porque, como ya hemos avanzado, el test se administra en contexto de inmersión y existe una amplia variedad de lenguas maternas entre los participantes. Aun así, en algún caso resulta útil e inevitable crear distractores en relación con el inglés como lengua franca. Así, utilizamos como distractores falsas elecciones, derivaciones erróneas, generalizaciones, etc. Sin embargo, el principal requisito para crear distractores es el de no ofrecer ninguna palabra o forma que no exista en español y, por lo tanto, incurrir en un error en los distractores.

Con este formato, al igual que con el anterior, medimos además los siguientes componentes de conocimiento colocacional: a) conocer el sentido de la colocación, b) reconocer la forma de la base y c) reconocer la forma del colocativo.

INSTRUCCIONES

Elige la opción correcta para completar la secuencia.

Dicen que las mujeres de pelo oscuro son más guapas que las de pelo ...

- a. rubio
- b. amarillo
- c. encrespado
- d. rizo

Ejemplo 5. Muestra de ítem de PRECU

4.3.2.4. Modalidad 4: CRECU

Para medir la comprensión desde la recuperación se ha elegido también un formato de completar (ANEXO XXVIII). En el *ítem* se presenta la colocación contextualizada y el aprendiz debe continuar la frase de la forma más natural posible, de modo que muestre que conoce el sentido de esta. Se le pide al aprendiz, por lo tanto, que muestre que conoce el significado sin acudir al metalenguaje directamente, es decir, sin pedirle que defina la colocación. En las instrucciones está la clave para entender el formato: continuación natural de la frase. Se busca que el aprendiz identifique el sentido de la colocación y lo demuestre en el uso de la lengua. Se trata de un ítem abierto de difícil corrección, pero que nos muestra de una forma natural si se conoce el sentido de la colocación. Los aspectos del conocimiento evaluados con este formato son los

siguientes: a) conocimiento de la base, b) reconocimiento de la forma del colocativo y c) sentido de la colocación.

Esa niña tiene el pelo rizo, hace ... (ONDAS, ONDULACIONES, OLAS...)

Ejemplo 6. Muestra de ítem de CRECU

Este tipo de formato choca con las asociaciones más estables: producción con recuperación y comprensión con reconocimiento. A este respecto, debemos apuntar que ya se demostró que esta asociación no es correcta, que tanto la comprensión como la producción pueden ser medida a partir de cualquiera de estos dos procesos cognitivos (Nation, 1990). Es cierto que, hasta el momento, no hay ningún otro formato que evalúe el conocimiento de las colocaciones desde este punto de vista. Irremediamente en este test también influirán las características personales aludiendo a la creatividad del aprendiz. Pedir que el aprendiz escriba el sentido de la colocación puede ocasionar discrepancias entre la competencia y la habilidad. Debemos tener en cuenta el perfil de la población que va a realizar el test ya que no todos los aprendices de una L2, ni de ELE, en particular, son estudiantes universitarios que pueden jugar con el metalenguaje. Este suele ser el perfil de la población de las investigaciones que se llevan a cabo en la universidad. Sin embargo, no representan la realidad de todos los aprendices de ELE, que no siempre tienen o no les supone ningún problema trabajar con el metalenguaje, sobre todo, en contextos de inmersión.

4.3.2.5. *Los cuatro formatos del TCE: recapitulación*

Adaptar CATSS a las necesidades del TCE supone un cambio más formal que conceptual. Mantuvimos las combinaciones entre las destrezas y los procesos cognitivos para perfilar la sofisticación del conocimiento colocacional y centrada en el enlace entre la forma y el significado de la combinación. Además de adaptar los formatos a otra unidad léxica, tuvimos que adaptarlo a una población con diferentes L1, de ahí que la traducción fuera eliminada como posible formato. La *Tabla 28* recoge un ejemplo de cada uno de los formatos a partir de la misma colocación meta, en este caso, *café cortado*. En los cuatro formatos la colocación meta aparece contextualizada y

- El formato de PRECU, de rellenar huecos, presenta la base contextualizada y el aprendiz debe recuperar el colocativo.
- El formato de PRECO es de elección múltiple y el participante debe elegir entre varias opciones para selección el colocativo adecuado.
- En CRECO, también de opción múltiple, el objetivo es reconocer el significado de la colocación, contextualizada, a partir de una serie de paráfrasis.
- CRECU tiene el formato más original y menos habitual. Se presenta la colocación y contextualizar y es el aprendiz el que tiene que recuperar una paráfrasis que explique el sentido de la colocación.

CAFÉ CORTADO		
	PRODUCCIÓN	COMPRENSIÓN
RECUPERACIÓN	<p>PRECU Pepe siempre quiere café c..., le gusta con poca leche.</p>	<p>CRECO Un café cortado es un café con ...</p> <ul style="list-style-type: none"> a. un poco de azúcar. b. un poco de leche. c. mucho café. d. mucha leche.
RECONOCIMIENTO	<p>PRECO A Pepe le gusta el café ... con un poco de leche.</p> <ul style="list-style-type: none"> e. cortado f. con leche g. largo h. blanco 	<p>CRECU Un café cortado es un café con un poco de...</p>

Tabla 29. Ejemplo de ítem en todas las modalidades en TCE

4.3.3. Especificaciones

La elaboración de las especificaciones es el paso previo a la construcción del test. Una vez que el diseño está claro, es necesario sintetizar los pasos que se deben dar en la construcción, administración y corrección del test, aspectos esenciales todos tanto para la generalización como explicación e interpretación de los resultados. Las especificaciones se pueden organizar en tres grandes bloques (Green, 2014): a) declaración del diseño, b) guía del test y c) especificaciones de las tareas y de los ítems. La *declaración del diseño* se ha especificado a lo largo de todo el trabajo, más concretamente en la introducción, al justificar el objeto de esta investigación, en la que se exponen las razones que justifican el diseño de este test. En este punto nos

centraremos en la *guía del test* y en las *especificaciones de las tareas y de los ítems*, que nos permitirán construir y analizar los resultados del test.

Para la correcta elaboración de las especificaciones debemos definir los criterios de corrección, clasificación y administración del test. Con los criterios de corrección especificamos de qué forma se van a corregir las cuatro modalidades del test. Los criterios de gradación nos permitirán establecer el grado del conocimiento colocacional y, sobre todo, la detección de las dificultades. Los criterios de clasificación nos van a permitir relacionar la competencia colocacional con los niveles del MCER. Por último, establecemos los criterios de administración, es decir, las instrucciones que permiten recoger la información proporcionada por los aprendices de ELE. En las siguientes tablas adelantamos el resultado de los criterios que desarrollaremos a continuación.

En la *Tabla 31* vemos de forma esquemática la guía del test en la que se especifica su estructura. No todos los participantes realizan las cuatro modalidades, sino que cada uno realiza solo dos: la Modalidad A, con PRECU y CRECO; o la Modalidad B, con PRECO y CRECU. En cada una de las opciones cada participante realiza una modalidad que mida la producción y otra que mida la comprensión de su conocimiento colocacional. En la *Tabla 32* se resumen las características más formales de cada una de las modalidades indicando el número de ítems, el tiempo de realización, el formato elegido y el soporte utilizado.

Modalidad	Modalidad	Capacidad que se evalúa	Tarea que se debe realizar	Material que se ofrece para realizar la prueba
A	PRECU	Producción de la colocación a partir de la recuperación del colocativo.	El candidato debe seleccionar el colocativo adecuado.	La base de la colocación contextualizada y la primera letra de la colocación.
	CRECO	Comprensión de la colocación reconociendo su sentido.	El candidato debe seleccionar una de las cuatro glosas que mejor expliquen el sentido del colocativo o de la colocación.	La colocación contextualizada y cuatro glosas, de las cuales solo una describe el sentido de la colocación.
B	PRECO	Producción de la colocación a partir del reconocimiento de uno de sus elementos.	El candidato debe seleccionar el colocativo adecuado de entre cuatro opciones.	La base de la colocación contextualizada y cuatro posibles colocativos, de los que solo uno es válido en ese contexto dado.
	CRECU	Comprensión de la colocación recuperando su sentido.	El candidato debe continuar una oración en la que aparece la colocación meta, de forma que demuestre que recupera su sentido.	La colocación semi contextualizada.

Tabla 30. Guía del TCE

Por último, en las especificaciones también debemos incluir otros dos aspectos: a) la asignación de nivel de LA competencia colocacional y b) el grado de conocimiento colocacional. Para evaluar la competencia colocacional creamos una escala en la que se especifica el número de colocaciones que cada participante debe conocer para asignarle un nivel, que explicamos a continuación en el apartado dedicado a los criterios de clasificación (vid. 4.3.3.2). En el ANEXO IV se puede ver de forma visual la correlación entre el número de colocaciones acertadas y el nivel del MCER asignado. En cuanto al grado de conocimiento colocacional, ya se ha explicado en un apartado previo, se puede ver de forma esquemática en la *Tabla 34* y con más detalle en las rúbricas).

	NULO	PARCIAL	ÓPTIMO
CONOCIMIENTO PRODUCTIVO	Desconocimiento de la base	Conocimiento de la base, pero desconocimiento del colocativo o del sentido de la colocación	Conocimiento de la combinación
CONOCIMIENTO COMPRENSIVO	Desconocimiento de la base	Conocimiento de la base, pero desconocimiento del colocativo o del sentido de la colocación	Conocimiento de la combinación

Tabla 31: Rúbrica básica para graduar el conocimiento colocacional

4.3.3.1. Criterios de corrección

El primer paso en el análisis cuantitativo de los resultados es convertir la actuación de los participantes, es decir, sus respuestas, en datos observables, que recogimos en un documento de Excel. En los test de recuperación se le asigna un 1 si la respuesta es correcta y un 0 si la respuesta es incorrecta. En los test de reconocimiento, se recoge la opción elegida por los participantes, lo que nos permitirá llevar a cabo un análisis cualitativo de los distractores, y también se asigna un 1 a la respuesta correcta y un 0 a la incorrecta. En los test de recuperación también se copia en una hoja de cálculo la respuesta dada y se resuelve la corrección de la misma forma. Para calcular el resultado de cada test se han sumado las respuestas correctas. La competencia colocacional de cada aprendiz se calcula a partir la media entre el número de aciertos de cada una de las secciones de la modalidad.

Cada una de las modalidades exigió la especificación de unos criterios de corrección particulares. Como decíamos en el apartado anterior, el diferente grado de complejidad entre los formatos de *reconocimiento* y los de *recuperación* hace que la

corrección de las respuestas abiertas (recuperación) también conlleva un mayor coste de corrección.

A. Modalidades de reconocimiento

Las dos modalidades de *reconocimiento*, tanto la que se combina con la de *producción*, PRECO, como la que se combina como la de *comprensión*, CRECO, utilizan un formato de opción múltiple. En este tipo de test de respuesta los efectos de su dificultad residen en su construcción, pero simplifica la corrección, de modo que, en cada ítem hay una respuesta correcta y tres distractores. Si la opción elegida es la esperada, el ítem se da como correcto; si se elige uno de los distractores, la opción se da por incorrecta. En este tipo de test será de mucho interés observar si algún distractor no funciona o si se elige más de lo esperado, algo que puede revelar un error común a la hora de reconocer colocaciones. En este caso no hay problema para determinar si la respuesta correcta es una colocación o no, ya que se ha partido de una lista de combinaciones que ya han sido juzgadas como colocaciones y que, además, están niveladas (vid. 4.5). La regla de corrección se puede formular de la siguiente manera: serán correctas solo las repuestas con la opción elegida y no con ninguno de los distractores.

B. Modalidades de recuperación

La corrección de las modalidades de recuperación implica mayor dificultad ya que la demanda cognitiva aumenta. En ellas el participante debe crear la respuesta, lo que se conoce como *ítem de respuesta construida*⁷⁷. Además, para cada una de ellas se requieren unos criterios diferentes, ya que para cada una se diseñó un formato diferente, pero la mayor discrepancia radica en la diferente naturaleza de la respuesta.

a) Criterios para corregir PRECU

En este formato falta el colocativo de la colocación-meta, por lo tanto, la respuesta esperable se corresponde con la palabra restringida léxicamente por la base. Consecuentemente, el corrector cuenta con una lista en la que están los colocativos-

⁷⁷ Traducción del inglés *constructed-response* (CR), que se opone a *selected-response* (SR) y que se corresponde con los ítems de opción múltiple (Haladyna & Rodríguez, 2013).

meta. Sin embargo, se han dado casos de colocativos no-esperados que se han dado por correctos. En las instrucciones no se indica que busquen el colocativo más adecuado, sino la palabra más adecuada para ese contexto. Es decir, no se tiene en cuenta la conciencia colocacional⁷⁸ de los aprendices. Así, en algún caso se eligió una palabra que no forma colocación con la base, es decir, no se trataba del colocativo meta, pero que aporta un sentido coherente a la oración en la que se inserta. Es el caso de *letra bonita* en lugar de *buena letra*, con un sentido similar, e intercambiable en algunos casos; o el caso de *éxito absurdo* en lugar de *éxito apoteósico*, en el que el sentido difiere bastante, pero es válido en el contexto del ítem. También encontramos casos en los que utilizan una colocación sinónima. Por ejemplo, en lugar de elegir el verbo *dar* se usó *dejar* para acompañar a *propina*; así consideramos como respuesta correcta tanto *dar propina* como para *dejar propina*⁷⁹ o en lugar de *apoteósico* como colocativo de *éxito*, se admitió como correcto desde *absoluto* a *alucinante* pasado por *arrollador*.

Asimismo, cabe la posibilidad de que, aun identificando el colocativo, el aprendiz no sea capaz de escribirlo con la corrección que exigen las reglas ortográficas y gramaticales. Aquí es donde entra en juego el grado de dominio de cada uno de los componentes. Nos podemos encontrar con las siguientes situaciones en las que se acepte el colocativo aun cuando la forma escrita no es correcta como un error de concordancia o un error ortográfico. En ambos casos damos la respuesta como correcta⁸⁰. En este caso los errores admitidos más comunes son la mala elección de formas verbales, ya sea por la confusión de persona, tiempo o modo (presente por pasado, indicativo por subjuntivo) o algún error tipográfico como **duble* por *doble*, en el caso de *habitación doble*.

Por lo tanto, la *regla de corrección* se podría formular como sigue: se darán por buenas en PRECU todas las respuestas que coincidan con el colocativo-meta, así como los que presenten algún error de índole gramatical (elección de forma verbal), ortográfico

⁷⁸ Con *conciencia colocacional* nos referimos al conocimiento que los aprendices tienen de que este tipo de combinaciones de palabras se denominan colocaciones. Estamos hablando de conciencia metalingüística.

⁷⁹ *Dar* y *dejar propina* son dos colocaciones muy cercanas que podríamos tratar con la misma función léxica, aunque aporten matices diferentes.

⁸⁰ Es cierto que este grado sería aceptable en un nivel bajo y no en un nivel alto. Sin embargo, hacer esta dirección en la corrección no resultó viable. Además, si en un nivel alto nos encontramos con una falta de ortografía o gramatical no estaríamos detectando un problema en la competencia colocacional en concreto sino en la competencia lingüística en general.

(ausencia de acento) o tipográfico (confusión de una letra por otra); incluso se admitirán como correctas las palabras, que, incluso no formando colocación, den sentido al ítem propuesto.

b) Criterios para corregir CRECU

En este formato la respuesta no se selecciona, sino que se construye, es decir, pertenece al tipo de *test de respuesta-construida*, concretamente, se identifica con los test de respuesta corta en los que la respuesta es objetiva (responde al valor de correcto o incorrecto). Igual que en CRECO, el objetivo es una glosa u explicación del sentido de la colocación contextualizada, pero a diferencia de este, la respuesta no se puede elegir entre varias opciones, sino que se debe crear. En términos de corrección esto supone la necesidad de una valoración o juicio por parte del corrector (Haladyna & Rodríguez, 2013) y esa decisión debe basarse en una guía o *regla de corrección*. Pese a que se mantiene el criterio de corrección objetivo, es decir, todas las respuestas serán valoradas como correctas o incorrectas y contamos con una lista de respuestas esperables, la variedad de respuestas que se pueden dar como correctas según las condiciones de la actuación son variadas, lo que dificulta el proceso de corrección.

El objetivo es que los aprendices completen la oración, que contiene la colocación que se está testando, indicando que han comprendido su sentido y de la forma más natural posible. Las respuestas esperables se recogen en una lista de glosas esperadas, o respuesta ideal, para cada uno de los ítems, que, en la mayoría de los casos, se corresponden con la respuesta correcta esperable de PRECO. Por ejemplo, en el caso de la colocación meta *pelo rubio*, la glosa esperada era *pelo amarillo*, pero se dieron otras que juzgamos como correctas como es el caso de *pelo dorado*. El objetivo, la información que pretendemos obtener es si los participantes conocen el significado de la colocación y si son capaces de explicarlo. De esta forma, evaluamos de una forma holística, al valorar la glosa en su conjunto. No es significativo, por lo tanto, para esta respuesta algún aspecto relacionado con cuestiones gramaticales u ortográficas relativas a la glosa que escribe.

4.3.3.2. Criterios de clasificación

En el TCE vamos a medir el conocimiento y la competencia colocacional. Para obtener el grado de conocimiento colocacional y nivel de competencia necesitamos unos criterios que nos permitan graduar la actuación de los participantes según unos estándares preestablecidos, que explicamos a continuación. La configuración de estos estándares tiene una estrecha relación con la validez ya que dotan de significado las interpretaciones que vamos a hacer de los resultados para llegar a unas conclusiones.

A. Nivelación de la competencia colocacional

En este apartado la conceptualizaremos la competencia colocacional de forma cuantitativa concretando los *estándares* en *puntos de corte* que nos permitirán clasificar la competencia colocacional de los aprendices participantes en un test de nivel de competencia. El establecimiento del estándar nos permite crear un enlace entre los resultados del test y un nivel de competencia. En el primer paso, tomamos como *sistema de referencia* el MCER y su concreción en español, el PCIC. Los descriptores verbales del MCER no están exentos de problemas para su adaptación, ya que necesitamos una serie de criterios cuantitativos que nos permitan establecer la correlación entre los resultados y los descriptores. En la *Tabla 35* podemos ver los descriptores, en este caso del nivel intermedio sobre la riqueza y dominio del vocabulario. En ellos se utilizan palabras vagas para aludir a la extensión del vocabulario como *suficiente*, *amplio* o *limitado* y se aceptan errores que no concreta.

RIQUEZA DEL VOCABULARIO		DOMINIO DEL VOCABULARIO	
B1	Tiene suficiente vocabulario para expresarse con algún circunloquio sobre la mayoría de los temas pertinentes para su vida diaria, como, por ejemplo, familia, aficiones e intereses, trabajo, viajes y hechos de actualidad.	B1	Manifiesta un buen dominio del vocabulario elemental, pero todavía comete errores importantes cuando expresa pensamientos más complejos, o cuando aborda temas y situaciones poco frecuentes
B2	Dispone de un amplio vocabulario sobre asuntos relativos a su especialidad y sobre temas más generales. Varía la formulación para evitar la frecuente repetición, pero las deficiencias léxicas todavía pueden provocar vacilación y circunloquios.	B2	Domina un limitado repertorio relativo a necesidades concretas y cotidianas.

Tabla 32. Escala del MCER

En el caso del TCE no es suficiente con determinar un *punto de corte* simple que convierta el test en un aprobado o un suspenso en función del número de aciertos, sino que necesitamos *punto de corte múltiple* que nos permita asignar un puesto en una escala

de competencia, ya que se corresponderá con cada uno de los niveles del MCER. La decisión de definir el punto de corte está condicionada por un criterio numérico que, a su vez, ha de estar justificado y apoyado por una base teórica. El criterio cuantitativo, que enlace los descriptores con el nivel, puede fundamentarse en la asignación de una proporción de aciertos por cada nivel, de tal modo que cada porcentaje se relacionará con un nivel o puede basarse en la relación entre el nivel y el número de aciertos. Nos hemos decantado por esta opción ya que la muestra de colocaciones meta, nivelada según el PCIC (vid. 4.5), nos permite asegurar la correlación.

El TCE, como test basado en criterios, necesita de una metodología que nos lleve a marcar los puntos de corte. Existen diferentes métodos según la valoración que los jueces hacen de los ítems del test, de la actuación del grupo de participantes y según la posición de un sujeto con respecto al resto del grupo (Barrios & Cosculluela, 2013). Por las características de este test, debemos atenernos a los métodos propuestos centrados en el test en los que la valoración se hace sobre el contenido, es decir, sobre los ítems del test ya que esto permite establecer los criterios y los puntos de corte antes de administrar el test⁸¹. El nivel de competencia colocacional de cada informante está basado en su actuación en las dos secciones de cada modalidad (por ejemplo, PRECU y CRECO en la Modalidad A), sin ningún requisito de llegar a un mínimo de resultados en cada una de las secciones de la modalidad. Las siguientes tablas son la guía para convertir los resultados en indicador del nivel de la competencia colocacional. Al analizar por separado los resultados de una y otra modalidad, pueden ser útiles como herramienta de diagnóstico para indicar las áreas de relativa fortaleza y también las de mayor debilidad.

Debemos mostrar unos criterios de clasificación diferentes para el TCEp (Test de colocaciones del español piloto) y para el TCE (Test de colocaciones del español) debido al diferente número de colocaciones meta, aunque el criterio es el mismo en ambos casos; por lo tanto, las tablas y los índices están relacionados. En la tabla de escala se muestra el número de aciertos que cada aprendiz debe obtener para alcanzar un nivel.

⁸¹ Tomamos como referencia el método de Angoff (1984) y el llamado de *consenso directo* (Siceri et al., 2004) el establecimiento de los criterios de clasificación. Aunque el primero se utiliza preferiblemente para test de opción múltiple, como el TCE tiene secciones de opción múltiple, pero otras de respuesta abierta, por lo que no podemos acogernos a él de forma estricta.

En la *Tabla 37* se muestra la escala utilizada en la versión definitiva. Así, para clasificar a un aprendiz en el nivel A2 debe tener entre siete y doce aciertos. No obstante, para clasificarse en un nivel B2 debe acertar entre diecinueve y veinticuatro de los treinta y seis ítems. En el eje vertical de la tabla mostramos los niveles en los que se divide el TCE y las colocaciones de cada uno. En el eje horizontal tenemos los niveles de competencia colocacional. Con gris más oscuro resaltamos el número máximo de colocaciones para asignar un nivel. En la *Tabla 38* observamos la correspondencia numérica entre nivel del aprendiz, número de aciertos y porcentaje de aciertos.

NIVEL	N.º DE ACIERTOS	PORCENTAJE
A1	1 – 6	16%
A2	7 – 12	33%
B1	13 – 18	50%
B2	19 – 24	66,6%
C1	25 - 31	86%
C2	32 – 36	100%

Tabla 33. Correspondencia entre la escala numérica a porcentual del TCE

B. Gradación del conocimiento colocacional

Los criterios expuestos en el subapartado anterior nos permiten calificar los test en torno a una de las preguntas de investigación, si la competencia colocacional es índice de la competencia léxica y de la competencia lingüística. Pero esta no es la única, también nos preguntamos, desde una perspectiva de diagnóstico, cuáles son las fortalezas y debilidades del conocimiento colocacional para lo que diseñamos un test con cuatro modalidades que tratan de mostrar una imagen global del conocimiento de las colocaciones desde cuatro ángulos diferentes. Pese a que no todos los participantes tuvieron la oportunidad de demostrar su dominio de cada colocación en cada una de las modalidades, sí lo pudieron hacer en dos.

El vocabulario como constructo complejo que implica varios aspectos y niveles (Durrant et al., 2022, p. 203) es graduable. En este apartado explicaremos cómo a partir del TCE podemos graduar los niveles del conocimiento colocacional. Se asume que el conocimiento léxico no es dicotómico (saber o no saber) sino que es gradual y oscila desde el desconocimiento al conocimiento pleno (Barclay & Schmitt, 2019). Sin embargo, no contamos con una escala estandarizada gradual del mismo. En la *Figura 45* podemos

ver una representación de los grados de conocimiento de una palabra en sus diferentes componentes. Esta gradación es fácilmente visible en los típicos errores ortográficos del español: un aprendiz puede conocer varios componentes de una palabra como *baile*, pero no saber si se escribe con *b* o con *v*. En lo referente al conocimiento léxico, partimos de dos suposiciones comúnmente aceptadas: producir requiere más conocimiento que comprender, además de que suele adquirirse *a posteriori*, y el proceso cognitivo de la recuperación es más costoso que el de reconocimiento, desde un punto de vista de demanda cognitiva. En nuestra propuesta para graduar el conocimiento colocacional nos situamos en la línea de la solidez (Laufer et al., 2004) propuesta en CATSS.

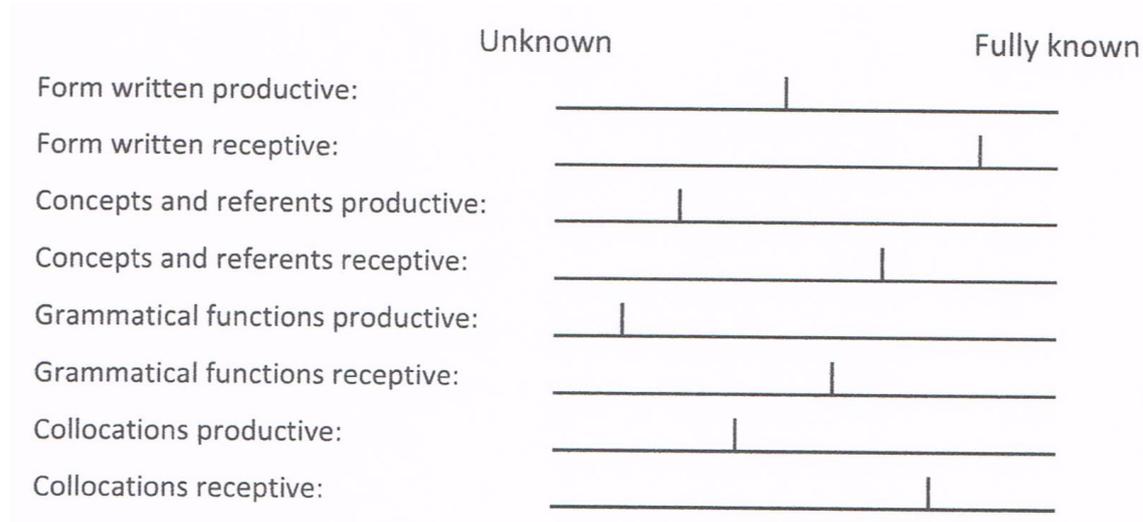


Figura 45. Representación de los grados del conocimiento componencial de una palabra (Barclay & Schmitt, 2019)

De acuerdo con esta situación y con la conceptualización que del conocimiento colocacional propusimos anteriormente (vid. 4.1.2), proponemos los siguientes criterios de clasificación. La escala propuesta consta de cuatro niveles: el *conocimiento total*, el *conocimiento parcial esperado*, el *conocimiento parcial anómalo* y el desconocimiento o *conocimiento nulo*. Esta gradación es posible porque una misma colocación meta se mide a partir de dos formatos diferentes.

Un aprendiz tendrá un nivel nulo si responde de forma incorrecta a los dos test de su modalidad. Por ejemplo, un aprendiz que haya realizado la Modalidad A y obtenga en el ítem treinta, *relación tormentosa*, tanto el test PRECO como en el test CRECU, una respuesta incorrecta, se le asignará un conocimiento nulo de dicha colocación. De forma opuesta, si un aprendiz responde de forma correcta en los dos test, el conocimiento de

dicha colocación será total, concluiremos que la colocación se conoce de forma óptima. Si falla en una de las modalidades, el conocimiento será parcial (esperable o anómalo). Para describir el conocimiento general que un aprendiz tiene de las colocaciones calcularemos el número de ítems en los que ha fallado en las dos opciones, el número de ítems en los que ha acertado en una de las dos modalidades y el número de ítems acertados en las dos modalidades. En caso de que sea más elevado el número de aciertos en las dos modalidades predominará el conocimiento total; en caso de que sea mayor el número de aciertos solo en una de las modalidades, predominará el conocimiento parcial. En caso contrario, el conocimiento será deficiente.

	DESCONOCIMIENTO	CONOCIMIENTO PARCIAL	CONOCIMIENTO TOTAL
TCE	No conoce la colocación en ninguna de las dos modalidades.	Conoce la colocación en una de las modalidades.	Conoce la colocación en ambas modalidades.

Tabla 34. Primera escala del conocimiento colocacional

Para una mayor profundización en el conocimiento se podría especificar el análisis al atender a las distintas posibilidades del conocimiento parcial. En la *Tabla 40*, podemos ver cómo sería la escala atendiendo a este aspecto. Es evidente que si no se acierta el ítem en ninguna de las modalidades no se conoce la colocación (conocimiento nulo) y que, si se acierta en ambas modalidades, el conocimiento es profundo, lo que hemos llamado *conocimiento total u óptimo*. Sin embargo, no es lo mismo demostrar conocer la colocación acertando una u otra modalidad. Por ejemplo, acertando CRECO y fallando en PRECU está claro que se puede reconocer el sentido de la combinación, pero no se puede recuperar el colocativo; pero si se acierta la colocación en la modalidad PRECU, es decir, si se puede recuperar el colocativo, parece anómalo no poder reconocer el sentido de la colocación.

	DESCONOCIMIENTO	CONOCIMIENTO PARCIAL	CONOCIMIENTO TOTAL	ANOMALÍA
MODALIDAD A	No conoce la colocación ni en PRECU ni en CRECO	Conoce la colocación en CRECO, pero no la conoce en PRECU	Conoce la colocación tanto en PRECU como en CRECO	No conoce la colocación en CRECO, pero sí en PRECO
MODALIDAD B	No conoce la colocación ni en PRECO ni en CRECU	Conoce la colocación en CRECU, pero no PRECO	Conoce la colocación tanto en PRECO como en CRECU	No conoce la colocación en CRECU, pero sí en PRECO.

Tabla 35. Segunda escala del conocimiento colocacional

En la otra modalidad (PRECO y CRECU) las conclusiones no son tan claras. Aparentemente parece habitual que, si se acierta PRECO, se debe acertar también CRECU, esto es, si se reconoce el colocativo de una combinación, también se debería recuperar su sentido. Aunque en este caso se puede chocar con las limitaciones propias de la prueba como la competencia lingüística del aprendiz y su creatividad. Para un análisis más pormenorizado que nos llevara a definir de forma más exacta la imagen global del conocimiento colocacional deberíamos tener en cuenta todas las combinaciones posibles que abren un amplio número de opciones dentro del conocimiento parcial y que cabría definir para poder así establecer una más clara relación entre los componentes del conocimiento colocacional. También es importante aclarar que es posible reconocer el colocativo porque la combinación resulte reconocible, pero no saber qué significa.

	Test de comprensión (CRECO o CRECU)	Test de producción (PRECU o PRECO)	Explicación
CONOCIMIENTO TOTAL	1	1	La respuesta del ítem x es correcta en los dos test
CONOCIMIENTO PARCIAL ESPERADO	1	0	La respuesta es correcta en uno de los test y ese test es el de comprensión.
CONOCIMIENTO PARCIAL ANÓMALO	0	1	La respuesta es correcta en uno de los test y ese test es el de producción.
CONOCIMIENTO NULO	0	0	Ninguna respuesta correcta del ítem x.

Tabla 36. Correspondencia entre la gradación del conocimiento colocacional y el TCE

4.3.3.3. *Criterios de administración*

En la parte inicial de este apartado ya se han adelantado algunos de los puntos que se tratan de forma más pormenorizada en este. Atenderemos al orden de administración y formato e instrucciones de las cuatro modalidades que integran el test, pero también a los otros test que los aprendices deben cumplimentar al participar en el estudio.

A. Orden de administración

Comprobar si la competencia colocacional de los aprendices de español es índice de la competencia léxica y de la competencia lingüística nos obliga a conocer, además de la competencia colocacional, la léxica y la lingüística. Por eso antes de realizar el TCE los

participantes deben cumplimentar un formulario, la biografía lingüística, y un test de léxico, X-LEX. El orden de administración de las diferencias pruebas se planeó para la primera prueba piloto y se mantuvo hasta la versión definitiva. En la guía del test (ANEXO III) se indica qué otros cuestionarios deben cumplimentar y cuál será el orden en el que se realizarán concretando el tiempo estimado de realización de cada uno de ellos.

En primer lugar, se presenta a los aprendices la biografía lingüística, un pequeño cuestionario que nos permite conocerlos como aprendices de español; en segundo lugar, el test de competencia léxica, X-LEX (Meara & Milton, 2003), que nos permitirá correlacionar los resultados del TCE y sus respectivas modalidades con la competencia léxica y esta, a su vez, con la competencia lingüística. Por último, el TCE en una de sus dos opciones, en la que primero realizarán el test que mide la producción y, a continuación, el test que mide la comprensión. En este caso representamos el orden y duración del proceso de administración de la Modalidad A del TCE, es decir, aquella en la que se realiza PRECO y CRECU. Consideramos, y así lo comprobamos, que los test de reconocimiento (como PRECO) se realizan en menos tiempo que los test de recuperación (como CRECU). Por lo tanto, tal y como se observa en la Tabla 41, la previsión es que la Modalidad A, en este caso PRECO, ocupa menos tiempo que la Modalidad B, en este caso CRECU. Aun así, esta división de los tiempos es orientativa, a los aprendices se les da el test van proporcionando las pruebas a medida que van acabando y ellos gestionan los tiempos, no el orden.

Biografía lingüística •Duración: 5 min
X-LEX •Duración: 10 minutos
TCE •Modalidad 1 •Duración: 10 minutos •Modalidad 2 •Duración: 20 minutos

Tabla 37. Orden y duración en la administración del TCE de la Modalidad A

La biografía lingüística es un breve formulario en el que los aprendices cubren algunos datos relacionados con su experiencia como aprendices de español. En este únicamente se solicita la nacionalidad, la profesión, la lengua materna, el centro de

estudios de español, su nivel de competencia en español, el conocimiento de otras L2 y el tiempo que lleva en situación de inmersión. En el ANEXO VII se puede ver la versión en papel de la biografía lingüística. Además, es el documento a partir del que se les asigna un número identificativo que permite mantener su anonimato en el proceso de corrección. El test de competencia léxica, X-LEX, se realiza antes de las dos modalidades que miden la competencia. Como hemos dicho, se trata de un test rápido que no supone un esfuerzo cognitivo para los participantes.

La Modalidad del TCE se asigna de forma aleatoria; sin embargo, la elección de cada una de las modalidades no. En todos los casos se realizó en primer lugar el test que implicaba la producción, con la intención de no adelantar en la medida de lo posible la respuesta de la siguiente sección y minimizar los efectos que una de las secciones pudiera tener en la otra, ya que la memoria pudiera condicionar las respuestas de las siguientes secciones. Así, los participantes que realizaron la *Modalidad A* realizaron en primer lugar PRECU y, en segundo lugar, CRECO; mientras que los que realizaron la *Modalidad B*, realizaron en primer lugar PRECO y en segundo lugar CRECU.

B. Versión en papel y versión electrónica

Para la versión electrónica, una vez descartada la opción de crear una versión adaptable, se recurrió a una herramienta gratuita y fácil de utilizar tanto para los aprendices de ELE a la hora de participar, como para los creadores del test a la hora de corregir. La herramienta elegida fue los *Formularios de Google* que nos permitía una serie de ventajas con respecto a otras herramientas. Primero, se trata de una herramienta gratuita. Segundo, permite enviar el formulario a través de un enlace y que los participantes pudieran enviar sus repuestas con tan solo hacer un clic. Además, admitía que los participantes pudieran enviar la autorización para el uso de sus respuestas como material de investigación.

C. Instrucciones

A la hora de crear las instrucciones se tuvieron en cuanto los dos agentes que las necesitaron: los administradores y los participantes. La realización de diferentes pruebas

en diferentes centros implicó la necesidad de contar con administradores externos que facilitaron la labor. Estos agentes necesitaron, además de las instrucciones de cada modalidad, la guía y especificaciones del TCE (vid. ANEXO III). Las instrucciones para los participantes, además de aparecer escritas en español en cada una de las pruebas, fueron explicadas de forma oral a los aprendices, siempre en la L2, el español y acompañadas de un ejemplo resuelto. Recordemos que el TCE está diseñado para que accedan a él participantes desde el nivel A1 hasta el C2.

4.4. Evolución del TCE

Uno de los pasos clave en desarrollo del test es el pilotaje. El objetivo de este proceso es planificar las actividades previstas de evaluación para mejorar la calidad del instrumento de evaluación antes de que se utilice con fines diagnósticos. Este proceso implica siete elementos clave (Read, 2017) que se resumen en los siguientes puntos: redactar un test en formato borrador y probarlo tanto con hablantes nativos como a profesores de la L2; administrar el test a un grupo (alrededor de treinta participantes) similar a la población que se quiere analizar y hacer las modificaciones pertinentes en función de las observaciones y conclusiones extraídas del análisis de los resultados de esta selección; además, es importante prestar especial atención a la valoración de cada uno de los ítems.

En el caso del TCE se pasó por cada una de las fases recomendadas a partir de las tres versiones del test (vid. Figura 46): pre-TCEp (*test previo de colocaciones del español piloto*), TCEp (*test de colocaciones del español piloto*) y TCE (*test de colocaciones del español*)⁸². Cada una de las versiones del test se pilotó primero en un grupo reducido de hablantes nativos: por un lado, especialistas en colocaciones y evaluación; por otro lado, hablantes de español ajenos profesionalmente al mundo de la lingüística. Todos entendieron las instrucciones de la prueba, encontraron las respuestas sin dificultad y consideraron cada uno de los ítems como construcciones naturales de la lengua. Las dos

⁸² En los ANEXOS se pueden ver los test de las tres versiones:

- preTCEp: PRECU (ANEXO XII), CRECO (ANEXO XIII), PRECO (ANEXO XIV), CRECU (ANEXO XV).
- TCPp: PRECU (ANEXO XXI), CRECO (ANEXO XXI), PRECO (ANEXO XXIII), CRECU (ANEXO XXIV).
- TCE: PRECU (ANEXO XXV), CRECO (ANEXO XXVI), PRECO (ANEXO XXVII), CRECU (ANEXO XXVIII).

primeras versiones (preTCEp y TCEp) funcionan como prueba piloto del test definitivo, TCE. El análisis de los ítems de las dos primeras versiones nos permitió eliminar y reformularlos según el grado de fiabilidad de cada uno.

La experiencia en la administración, así como el análisis estadístico, sobre todo del análisis del ítem, nos permitió modificar cada una de las versiones y reconstruirlas según los criterios de practicidad y las necesidades del análisis. Estas cuestiones se describen de forma pormenorizada en las conclusiones que se exponen en el apartado de análisis tanto del *preTCEp* (vid. 5.2.1) como del *TCEp* (vid. 5.2.2).

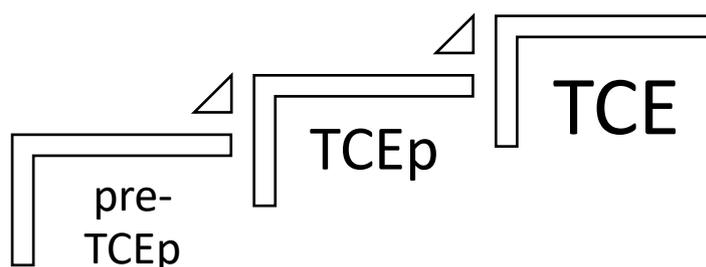


Figura 46. Evolución del test de colocaciones del español

Las modificaciones se llevaron a cabo desde una perspectiva conceptual, pero también sobre todo desde un punto de vista práctico; en ningún caso se modificó el propósito, ni los formatos elegidos, sino que se buscó la mejora hasta llegar a la versión que permitiera atender los objetivos para los que se diseñó la herramienta. Las principales variaciones están en el paso del pre-TCEp al TCEp, ya que se modificó la muestra parcialmente, con cierta variación en los criterios de selección de colocaciones. En el paso del TCEp de TCE el principal cambio fue la reducción del número de ítems del test, por lo tanto, se optó por el descarte de algunas de las colocaciones previamente elegidas. El criterio para hacerlo fue ajeno a la combinación en sí misma, y se optó por la selección según la facilidad a la hora de crear el banco de ítems y los errores detectados en los ítems de las versiones precedentes. En los apartados siguientes explicamos las decisiones que se tomaron en cada una de las pruebas.

Algunas de las cuestiones para tener en cuenta a la hora de pilotar un test dependen del tipo de formato elegido. En el TCE tenemos dos formatos diferentes: elección múltiple y rellenar el hueco; para cada uno de los cuales se deben de tener en cuenta algunas recomendaciones (Read, 2017) que enumeramos e iremos aplicando en

cada una de las versiones. En los formatos de seleccionar respuesta hay que tener en cuenta a) si solo hay una respuesta correcta por cada ítem, b) si la respuesta correcta no se puede adivinar, c) si los distractores funcionan, es decir, si atraen a un mínimo de la población; y d) si el ítem es apropiado para el nivel de dificultad. Las principales cuestiones para tener en cuenta para el pilotaje de formatos de rellenar hueco son a) si se puede dar una información adecuada de acuerdo con el contexto aportado, b) si se evita que se den varias respuestas correctas; c) la respuesta para ser correcta tener plena corrección o se pueden cometer errores ortográficos o fónicos.

4.4.1. Del pre-TCEp al TCEp

Una vez seleccionada la muestra (elegido las treinta y seis colocaciones meta) y construido el test (elaborado todos los ítems de las cuatro modalidades) se pilotó en un grupo de hablantes nativos: primero, a un grupo de expertos en fraseología y evaluación; posteriormente, a un grupo de hablantes nativos ajenos al ámbito de la lingüística. Tras obtener unos resultados positivos y valorar todo el proceso desde el diseño al análisis se llegó a una serie de conclusiones que nos permitieron afianzar algunas de las decisiones tomadas y mejorar otros aspectos hasta llegar a la versión definitiva del test.

En la administración del *pre-TCEp* se entregó a cada participante una hoja con las preguntas y otra hoja en donde debían escribir las respuestas. Esto produjo confusión en los aprendices por lo que en la siguiente versión se optó por presentar preguntas y respuestas en una misma hoja y así facilitar el trabajo a los aprendices en la siguiente versión. Tras el análisis, se comprobó que la extensión de la muestra era insuficiente para cumplir con los objetivos propuestos, entre ellos, el de comprobar si la dificultad del aprendizaje de las colocaciones estaba condicionada por la función léxica de dicha colocación. Se plantearon dos soluciones posibles: aumentar el tamaño de la muestra o reducir objetivos para futuras investigaciones. Se optó por incrementar el tamaño de la muestra para la nueva versión y reducir alguno de los objetivos, como el efecto de la función léxica de la colocación en la carga de aprendizaje de la colocación, y mantener otros como el efecto de la categoría gramatical de la colocación.

Una de las preguntas de investigación es la posible correlación entre la competencia léxica y la competencia colocacional. Para esto, uno de los requisitos era el de contar con una muestra nivelada. La dificultad de nivelar colocaciones a partir de la

frecuencia se ha explicado ya en capítulos anteriores (Vid. 2.2.3 y 2.3.3) y todavía no hay una propuesta que clasifique las colocaciones léxicas del español en niveles del MCER. Tener las colocaciones ordenadas según su frecuencia, pero no según su nivel, suponía una labor extra, proponer un método en el que la frecuencia sirviera como criterio para nivelar las colocaciones, tarea que excede los límites de esta investigación.

4.4.2. Del TCEp al TCE: reconstrucción

El proceso de reconstrucción del TCEp al TCE fue un proceso arduo, que combinó las conclusiones extraídas del análisis estadístico con la experiencia en cada uno de los pasos llevados a cabo en la construcción y diseño y, sobre todo, la administración. Pese a que las conclusiones del TPCp, en general, se ajustaron a las hipótesis planteadas, se detectaron diversos problemas a lo largo de todo el proceso de creación: elaboración, administración y corrección, lo que nos llevó a crear una nueva versión, que fue la definitiva.

Primero, en relación con la fase de construcción, los cuatro formatos están contextualizados (vid. 4.3) y se percibió en el análisis como, en algunos casos, el conocimiento previo (pragmático o incluso cultural) pudo beneficiar y condicionar la respuesta de los participantes. En estos casos, se rediseñaron los contextos intentando que el marco cognitivo de cada participante afectara lo menos posible a la elección de la respuesta correcta. Estos problemas se detectaron principalmente gracias a las medidas estadísticas de fiabilidad que nos permitieron ver en qué ítems no se obtenían los resultados esperados y qué distractores no funcionaron. Por lo tanto, los dos formatos de opción múltiple, que en este caso miden la comprensión, fueron el objeto central de las modificaciones.

Segundo, con relación a la administración, se detectó un problema de practicidad. El test es demasiado largo lo que provocaba la frustración de los testantes. Por lo tanto, decidimos modificar el test. En este punto se plantearon dos opciones: reducir el número de ítems del test o convertir el TPCp en un test adaptativo⁸³. La segunda opción es la más

⁸³ En un test adaptativo (TAI) se adecúan las respuestas al conocimiento de los aprendices, lo cual era perfecto para los objetivos de esta investigación.

atractiva para la investigación ya que nos permitía mantener todas las variables previstas. Tras una primera valoración se descartó la adaptación al TAI tanto por la dificultad que entraña el análisis estadístico, como por el amplio número de aprendices que realicen el test, lo que lo convertía en inviable para este trabajo. Por lo tanto, optamos por la reducción de ítems. La elección de las setenta y dos colocaciones que se incluyen el TPCp no es aleatoria, sino que viene determinada por unos criterios estrictos de selección, basados en diferentes aspectos que se iban a analizar (nivel de MCER, esquema sintáctico y función léxica). Así redujimos el test a la mitad con seis colocaciones por nivel y tres de cada esquema sintáctico, obteniendo así un total de treinta y seis colocaciones.

Tercero, en cuanto al proceso de corrección. El test se administró en formato papel y lápiz, de modo que tanto la respuesta como las correcciones tuvieron que pasarse a formato electrónico, lo que supuso una inversión temporal importante. Evitarlo fortalecía la idea de crear un test informatizado adaptativo. Ya que el TAI había sido descartado (por las razones anteriormente explicadas) se optó por un test informatizado, pero lineal. Aunque se llevó a cabo con la mayor parte de los aprendices, con uno de los colaboradores fue imposible llevarlo a la práctica por las condiciones del lugar en el que se imparte la enseñanza del español. Así, fue necesario crear una doble versión del TCE: una en papel y otra electrónica, para lo que utilizamos una herramienta gratuita y accesible a todos los participantes, *Google Forms*.

En el proceso de análisis de los datos también se detectaron los problemas del diseño de los test. Fue necesario eliminar algunos ítems y modificar todos tras analizar el índice de facilidad / dificultad del ítem, ya que algunos no cumplían con los criterios mínimos, que debe situarse en la horquilla entre 30 y 70. El hecho de tratarse de un test de diagnóstico en el que se presentan ítems de todos los niveles, pero contando únicamente con aprendices de dos niveles, provocaron que estos índices no se vieran alterados⁸⁴. Aun así, fue necesario cambiar algunas de las colocaciones meta por otras de

⁸⁴ Todos los participantes del TPCp tenían un nivel intermedio del MCER (entre un B1 y un B2) y el test cuenta con ítems desde el nivel A1 al C2. Por lo tanto, es esperable que los ítems de nivel bajo tengan un índice de facilidad más alto de lo adecuado y los más difíciles más bajo de lo adecuado. En el apartado dedicado al análisis de la prueba, tanto en lo relativo a su fiabilidad, como a la nivelación de los aprendices, se desarrolla con mayor detalle.

la muestra y modificar algún ítem, en algunas ocasiones el contexto, y en otras, alguno de los distractores.

4.5. Selección de la muestra del TCE

La lista de colocaciones meta del TCE ha de seguir al menos dos criterios básicos para adaptarse a los objetivos de la investigación. Primero, las colocaciones deben adecuarse a los criterios para identificación y definición de las colocaciones que se proponen desde este trabajo. Segundo, las colocaciones deben tener asignado un nivel del MCER. También, se planteó la idea de que tuviera una frecuencia asignada. El problema con este último criterio es que no existe una identificación estandarizada en los estudios en español entre la frecuencia y el nivel de las colocaciones, aunque sí distintas propuestas que tratan de establecer un criterio objetivo para nivelar las colocaciones según su frecuencia (García-Salido & Alonso-Ramos, 2018)⁸⁵. Esta problemática supone un obstáculo en la selección de la muestra. Entre los recursos con los que sí contábamos eran con diccionarios de colocaciones del español (Alonso Ramos, 2004; Bosque, 2006) y el PCIC (Instituto Cervantes, 2006). Ninguno de los recursos combinaba el criterio fraseológico con la nivelación en relación con un criterio de frecuencia que se adaptara a los niveles de MCER.

Dado que una de las motivaciones del test es comprobar la dificultad que entraña el aprendizaje de colocaciones, para la selección de las colocaciones se han tenido en cuenta diversos criterios de entre los habituales, aunque con diferente peso en la selección final, y algunos más específicos: enfoque fraseológico, nivel del MCER, el esquema sintáctico, la función léxica y el grado de transparencia u opacidad de cada colocación. El criterio principal para seleccionar las colocaciones ha sido de carácter fraseológico. Hemos identificado y seleccionado las unidades del léxico que se consideran colocación a partir de este enfoque. Descartamos la frecuencia como criterio único para seleccionar colocaciones ya que no es el aspecto identificador del concepto de colocación manejado, pero también hay estudios que revelan que la fiabilidad de la frecuencia como

⁸⁵ Ver apartado 2.3.3. en el que se hace una revisión teórica sobre los estudios vigentes sobre la selección de muestras para test de colocaciones y la asignación de nivel a estas unidades del léxico.

criterio único de selección es relativo: no coincide con el juicio de los nativos y tampoco unos estudios con otros (Alderson, 2007). Así, desde la primera muestra recogida a la definitiva, el criterio de frecuencia ha ido perdiendo peso de forma progresiva.

A continuación, explicamos cómo seleccionamos la muestra para los dos test piloto y para la versión definitiva, el TCE. Veremos cuáles han sido los criterios y el procedimiento para cada una de las versiones, que sufrieron modificaciones que conllevaron diferencias en algún caso incluso de carácter metodológico.

4.5.1. Muestra para el *preTCEp*

La compilación de las colocaciones para la elaboración del test se desarrolló en dos fases en esta primera versión del TCE, la primera versión piloto que denominamos *preTCEp*. En la primera fase se seleccionan las bases de las colocaciones y en la segunda, los colocativos. La sucesión de las fases no es aleatoria, es indispensable partir de la base, incluso en la identificación, para posteriormente elegir el o los colocativos. El criterio para la identificación es de orden fraseológico, el mismo que nos obliga a seleccionar primero las bases. Se han detectado dos problemas a la hora de crear listas de colocaciones frecuentes: la identificación de colocaciones en el corpus y la asignación de frecuencia en el corpus a las colocaciones. Por este motivo la solución es la construcción de la lista en fases y no la creación de una lista de colocaciones frecuentes sino una lista de colocaciones a partir de bases y colocativos frecuentes, siguiendo el procedimiento ya propuesto en otros trabajos que seleccionan muestras para medir la competencia colocacional (Moreno-Jaén, 2007b), en el que crea un banco de colocaciones frecuentes del inglés en un proceso de dos fases. La lista resultante no es una lista de colocaciones frecuentes, sino una lista de colocaciones a partir de bases frecuentes. En este proceso se usan dos herramientas: el *Corpus del español* (Davies, 2006); el *Diccionario Práctico* (Bosque, 2006).

4.5.1.1. Selección de las bases

Para la selección de las bases se ha usado una lista de palabras frecuentes del español. La herramienta utilizada ha sido *A frequency dictionary of Spanish. Core*

vocabulary for learners (Davies, 2006) en la que se recogen las cinco mil palabras más frecuentes del español. De esta lista se eligieron de forma aleatoria los sustantivos que van a convertirse en las bases de las colocaciones resultantes. Una de las primeras decisiones fue la de que todas las bases fueran sustantivos, limitando así considerablemente los tipos de colocaciones posibles. Las razones que no llevaron a tomar esta decisión son las siguientes: a) los sustantivos son palabras de fácil lematización, al contrario que los verbos, lo que permite hacer búsquedas en corpus de forma más sencilla; b) la mayor parte de las colocaciones tienen como base un sustantivo, por lo menos la mayoría de los tipos de colocaciones según su esquema sintáctico (Koike, 2001); y c) los aprendices aprenden con mayor facilidad los sustantivos (Daller et al., 2007).

Con la intención de que exista una relación estrecha entre las bases, el nivel y la frecuencia organizamos la lista en bandas de frecuencia. Así, de acuerdo con los niveles del MCER dividimos la lista de cinco mil palabras más frecuentes en seis bandas, una por cada nivel. Seguimos las medidas establecidas para el test de extensión X-LEX (Meara & Milton, 2003). En la siguiente tabla se puede ver la correspondencia entre el nivel y la frecuencia de cada banda. Se seleccionaron una serie de sustantivos por cada nivel o banda de forma aleatoria. Pese a necesitar tan solo doce por cada una, para restringir el número de ítems del test y no restarle practicidad, se eligió un número mayor en previsión de que muchos de ellos fueran a ser descartados. Se descartaron aquellos que no contenían significado léxico concreto. Por ejemplo, una palabra muy frecuente como *cosa*, no se mantuvo en la lista ya que su capacidad para formar colocaciones es limitada.

NIVEL DEL MCER	FRECUENCIA DE CADA BANDA
A1	< 1500
A2	1500-2500
B1	2750-3250
B2	3250-3750
C1	3750-4500
C2	4500-5000

Tabla 38. Relación entre el nivel y las bandas de frecuencia (Meara & Milton, 2003)

4.5.1.2. Selección de los colocativos

Para la selección del colocativo se hizo un pequeño análisis comparativo. Se extrajeron posibles colocativos para los diez sustantivos más frecuentes de la lista del *corpus del español* (Davies, 2006) a partir de dos herramientas diferentes: corpus⁸⁶ y diccionario (*A Frequency Dictionary of Spanish*). Las colocaciones resultantes de ambas fuentes se compararon y, tras aplicar los criterios de identificación fraseológicos, se comprobó que los resultados que ofrecía el diccionario se acercaban más a los estándares previstos y, además, ofrecían una variedad mayor para elegir la *colocación meta* que si se hacía esta operación a través del corpus. En definitiva, el diccionario resultó ser una herramienta más útil para selección de los colocativos que el corpus.

Una vez elegida la herramienta que se iba a utilizar, el *Diccionario combinatorio práctico del español contemporáneo* (*Práctico* a partir de ahora), buscamos los colocativos para cada base. Si la base no aparecía en el *Práctico*, se descartaba. Establecimos colocaciones con tres categorías sintácticas: *verbo + nombre*, *nombre + verbo* y *nombre + adjetivo*. Por lo tanto, solo buscamos verbos y adjetivos para cada base. Si no había un mínimo de tres colocativos por cada base y esquema sintáctico también se descartaba la base. Además de estos, los adjetivos y verbos seguían los siguientes criterios:

- El colocativo ha de formar colocación con la base según nuestro marco teórico. Es decir, entre las palabras se debe establecer una relación restrictiva y debe haber un vínculo fraseológico entre ellas. Debemos apuntar que, aunque el *Práctico* es un diccionario combinatorio, no todas las combinaciones que recoge se consideran colocación en nuestro marco teórico. Por ejemplo, *caer una gota* o *coser un pantalón*.
- El colocativo también ha de pertenecer a las cinco mil palabras más frecuentes de la lista de Davies. Si no es así, se descarta. Sin embargo, si el colocativo no aparece entre las 5000, pero sí una palabra relacionada morfológicamente sí la incluimos. Por ejemplo, para la base *opinión*, no

⁸⁶ <https://www.corpusdelespanol.org/xs.asp?c=6> [Último acceso 24/08/2023]

encontramos el adjetivo *imparcial*, pero sí *parcial*, así que la aceptamos como buen colocativo de la base *opinión*.

- Todos los colocativos seleccionados se deben corresponder con un mismo sentido de la base en los casos de polisemia. Por ejemplo, en un caso como *sueño*, aceptamos adjetivos como *profundo*, *placentero* o *ligero*, referidos a ‘estado que se alcanza mientras se duerme’; pero no aceptaríamos adjetivo como *premonitorio* o *terrorífico* que se referirían al sentido ‘sucesos imaginarios que ocurren mientras se duerme’.
- Enmarcar cada colocación (base + colocativo) en su correspondiente esquema sintáctico⁸⁷ y función léxica.

En un primer momento se pretendió elegir una colocación que cubriera un esquema fijo de funciones léxicas simétrico⁸⁸ para cada uno de los niveles. Este criterio no se pudo mantener por la imposibilidad de forma la lista y se decidió mantener la colocación con la función correspondiente, aunque no se adscribiera a las previstas.

- En el caso de que haya demasiadas opciones, se eligen según el criterio de opacidad/transparencia y carácter fraseológico del nivel.

Para comprender mejor el proceso, revisaremos tres casos reales que ayudan a comprender la creación de la lista y que representan cada una de las situaciones anteriormente descritas. En un primer caso, entre las bases de la segunda banda y tras la selección aleatoria, uno de los sustantivos seleccionados es *departamento*. El segundo paso es buscarlo en el *Práctico*. Al comprobar que esa palabra no está en este diccionario, se descarta automáticamente. En el segundo caso, el sustantivo seleccionado aleatorio seleccionado como base es *cola*. El siguiente paso es buscarla en el *Práctico*. Al comprobar que sí está, se sigue con el proceso. En el diccionario cuenta con tres sentidos: ‘extremidad, parte posterior’, ‘fila’, y ‘pegamento’. Se elige el segundo sentido porque es el que más opciones combinatorias muestra a nivel cuantitativo. Al no lograr tres colocaciones por cada uno de los esquemas sintácticos, también se descarta. En el caso tres, el sustantivo seleccionado como base es *pelo*. Se busca en el *Práctico* y vemos que

⁸⁷ Verbo + nombre y nombre + adjetivo.

⁸⁸ Par las funciones verbo + nombre las funciones buscadas fueron *Oper*, *Real*. Para las funciones sustantivo + adjetivo Magn, por ejemplo.

sí aparece. Buscamos las posibles combinaciones y vemos que, tras aplicar el criterio fraseológico, sí encontramos tres colocativos que se adaptan a los criterios preestablecidos; así que se incluyen como colocaciones las combinaciones resultantes

Las colocaciones resultantes de este proceso se utilizaron para una prueba experimental que tiene el nombre de *preTCEp*. De la lista de colocaciones primitiva se utilizaron treinta y seis como *colocaciones meta*. Sin embargo, este proceso de identificación de colocaciones resultó poco rentable. El principal obstáculo fue la correspondencia de estas colocaciones, seleccionadas parcialmente según la frecuencia de sus partes constituyentes, con un nivel del MCER. Esto nos impedía dar respuesta a alguna de las preguntas de investigación como la correspondencia entre la competencia colocacional, que pretende ser observada a partir de este test, y la competencia léxica y la lingüística.

4.5.1.3. *Construcción de la muestra*

Siguiendo estos pasos se formó una lista de treinta y seis colocaciones, que se puede consultar en el ANEXO X. Una vez elaborada la lista se incorporó la función léxica de cada combinación con la intención de posteriormente analizar si existe relación entre el número de aciertos y la FL de la colocación. Además, se especificó el esquema sintáctico de cada combinación, así como nivel asignado. Debemos tener en cuenta que el nivel de la combinación se concluyó a partir de la frecuencia de los elementos (base y colocativo) por separado.

4.5.2. Muestra para el TCEp

Para la construcción del TCEp, segunda versión piloto del TCE, se tomaron varias decisiones, cuyas consecuencias repercutieron en también en las características de la muestra: tanto en los criterios de selección como en el número de colocaciones meta. La principal razón de este cambio fueron los problemas detectados en la muestra al analizar los resultados, entre ellos estaba la inoperatividad del criterio de frecuencia tomado en relación con los objetivos buscados y el reducido número de la muestra, sobre todo en relación con algunos aspectos, por ejemplo, la función léxica de las colocaciones. La

nueva muestra consta de setenta y dos colocaciones, más del doble que en la fase anterior. En esta versión, hemos tenido en cuenta cuatro criterios: el nivel de la colocación en el MCER, el esquema sintáctico, la función léxica y el grado de transparencia del colocativo. La frecuencia como criterio de identificación queda relegada a un segundo plano, ya que al igual que se detectó en el análisis de COLLEX, el criterio de frecuencia en la selección de colocaciones no dio resultado al dejar fuera a muchas colocaciones que no seguían este criterio (Gyllstad, 2007b, p. 86). El aumento del tamaño de la muestra dotó al test de mayor validez y fiabilidad, de hecho, basándonos en distintitos estudios se afirma que para que una muestra sea representativa en un test la extensión del vocabulario y del nivel debe de estar en una proporción de diez por mil (Gyllstad et al., 2021).

Al relegar la frecuencia como criterio para asignar un nivel a las colocaciones de la muestra, nos decantamos por una herramienta en la que se incluyen (aunque no se identifican) colocaciones del español niveladas, esta es el PCIC, *Plan Curricular del Instituto Cervantes* (Instituto Cervantes, 2006), que como ya hemos apuntado recoge un inventario de elementos del léxico organizados según el marco nocional y en niveles de competencia, que se basa en MCER. Además, uno de los usos del MCER es el de especificar el contenido de las diferentes pruebas y exámenes (Consejo de Europa, 2002).

Algunos de los criterios que mantenemos, como el esquema sintáctico o la función léxica, en esta muestra se concretaron: solo incluimos colocaciones con el esquema sintáctico (verbo + nombre y nombre + adjetivo)⁸⁹ y la función léxica. Elegimos doce colocaciones de cada nivel del MCER: nivel inicial (A1 y A2), nivel intermedio (B1 y B2) y nivel avanzado (C1 y C2), el doble que en la muestra del *preTCEp*. De esas doce, seis se enmarcan en un esquema sintáctico y seis, en el otro. Las seis colocaciones deben representar las funciones léxicas más representativas de cada esquema. Para las colocaciones *nombre + adjetivo* se buscan dos colocaciones que denotan ‘intensificación’⁹⁰, ‘algo positivo’⁹¹ y dos que no se aplican a los parámetros habituales⁹².

⁸⁹ Eliminamos, así, la categoría sintáctica nombre + verbo por tratarse de la más difícil de encontrar.

⁹⁰ La glosa ‘intensificación’ describe la función léxica *Magn*.

⁹¹ La glosa ‘algo positivo’ se relaciona con la función léxica *Bon*.

⁹² Por último, una serie de colocaciones consideradas no estándar, es decir, que no se aplica a los parámetros habituales y que se corresponden con la función *A no estándar*.

Para las colocaciones verbo nombre buscamos colocaciones que indiquen ‘acción’⁹³, ‘realización’⁹⁴ o ‘la causa de la acción’⁹⁵. El último de los criterios, la transparencia, se sigue únicamente como tendencia para comprobar si la restricción es índice de la dificultad de la colocación. Por ejemplo, si encontramos colocaciones como *comida buena* y *comida rica*, nos inclinaremos por la segunda por ser la colocación menos transparente de las dos, es decir, un colocativo más específico que se combina con menos bases o solo con una. Este colocativo se usa solo con *comida* para expresar el sentido de ‘buen sabor’.

Los criterios utilizados para elegir la lista definitiva de setenta y dos colocaciones quedan resumidos en la siguiente tabla en las que queda claro cuáles son las colocaciones buscadas y qué criterios deben cumplir para aceptarse como muestra. En el ANEXO X aparece la muestra completa.

	COLOCACIÓN ELEGIDA	CARÁCTER RESTRICTIVO	FL	ESQUEMA SINTÁCTICO	NIVEL
1		R	Oper	V+N	A1
2		NR			
3		R	Real		
4		NR			
5		R	CausOper o CausFunc		
6		NR			
7		R	A no estándar	N+A	
8		NR			
9		R	Magn		
10		NR			
11		R	Oper		
12		NR			

Tabla 39 Esquema de colocaciones de un nivel para el TCEp

El proceso se desarrolla en las siguientes fases. En primer lugar, se hizo una lista con todas las unidades combinatorias que aparecen en el apartado de *Nociones específicas* del PCIC (vid. Figura 47), ordenadas según el nivel en el que se encuentra y su esquema sintáctico. A estas colocaciones se les asignó función léxica y se determinó su grado de transparencia o restrictividad (R [RESTRICTIVO] o NR [NO RESTRICTIVO]). El número de combinaciones de la lista no fue suficiente para cubrir la nómina de las setenta y dos colocaciones buscadas. En los casos en los que no se encontró colocación para cubrir los requisitos, elegimos una palabra alternativa en el PCIC⁹⁶ que pudiera funcionar

⁹³ La función *Oper* se aplica a los verbos de apoyo e indica ‘acción’.

⁹⁴ La función *Real* se aplica a los verbos de realización e indica ‘realización’ de una acción.

⁹⁵ Las funciones *CausOper* y *CausFunc* indican la causa de la acción.

⁹⁶ Este mismo criterio es el elegido por Rojo Mejuto (2015) en su propuesta de nivelación de las colocaciones léxicas del español.

como base y buscamos un colocativo en el *Práctico* que sí las cumpla, añadiendo esta colocación a ese nivel. Además, la colocación creada no debía de aparecer en otro nivel del PCIC. Entendiendo así que se trata de un vacío y no una reasignación del nivel, estaríamos ante colocaciones olvidadas y no de colocaciones mal niveladas.

Nociones específicas. Inventario B1-B2	
1. Individuo: dimensión física	
1.1. Partes del cuerpo [v. Nociones específicas 13.]	
B1	B2
<ul style="list-style-type: none"> ▪ músculo, hueso, piel ▪ corazón, pulmón ▪ cuello, hombros, pecho, cintura, barriga ▪ rodilla, tobillo, codo, muñeca 	<ul style="list-style-type: none"> ▪ frente, mejilla, barbilla, ceja, pestaña ▪ cana, uña ▪ cerebro, hígado, riñón, intestino ▪ esqueleto, columna, costilla ▪ nervio, articulación, arteria, tendón
1.2. Características físicas	
B1	B2
<ul style="list-style-type: none"> ▪ postura, mirada ▪ parecerse ~ a su padre/a su hermano/entre ellos ▪ tener la piel ~ clara/oscura/sensible/suave ▪ estar moreno ▪ ser ~ blanco/moreno ~ (de piel) ▪ tener ~ buen/mal ~ tipo/figura/estilo/presencia 	<ul style="list-style-type: none"> ▪ ciego, sordo, mudo ▪ saliva, lágrima, sudor ▪ gesto, expresión ▪ ser de ~ estatura/altura ~ normal/media/mediana ▪ tener ~ arrugas/una cicatriz ▪ tener ~ buen/mal ~ color/aspecto ▪ ser ~ clavados/idénticos/como dos gotas de agua

Figura 47. Organización de las Nociones del PCIC (Instituto Cervantes, 2006)

Una vez seleccionadas las setenta y dos colocaciones se buscó aplicar otro criterio, el de frecuencia. Para esto se buscaron las colocaciones en un corpus del español, Sketch Engine⁹⁷. Posteriormente también se buscaron en el *Diccionario de frecuencias del español, A Frequency Dictionary of Spanish* (Davies, 2006), aunque no todas aparecían. En este proceso no se obtuvieron los resultados esperados y la correlación entre la frecuencia de la combinación y el nivel de MCER no es significativa. Por lo tanto, descartamos definitivamente el criterio de frecuencia para la selección de la muestra. En la *Tabla 44* vemos la estadística descriptiva.

⁹⁷ Se trata de un corpus web del español con 17,553,075,259 palabras (A. Kilgarriff & Renau, 2013). Al que se puede acceder a través del siguiente enlace <https://www.sketchengine.eu/>.

	MI (SK)		NIVEL B2	
	N+A	V+N	N+A	V+N
Válidas	36	36	36	36
Perdidas	0	0	0	0
Media	8.209	7.106	3.500	3.500
Desviación estándar	2.681	2.534	1.732	1.732
Mínimo	3.190	3.240	1.000	1.000
Máximo	15.920	12.730	6.000	6.000

Tabla 40. Estadística descriptiva de la frecuencia de las combinaciones

4.5.3. Muestra para el TCE

En el TCE la muestra previa (seleccionada para el segundo piloto, TCEp) no se modificó, sino que se simplificó al reducir el número de ítems a la mitad. En lugar de utilizar las setenta y dos colocaciones de la lista, se seleccionaron treinta y seis, volviendo al número inicial de treinta y seis colocaciones, pero manteniendo la mayor parte de los criterios de selección del TCEp⁹⁸. La extensión de la muestra es uno de los puntos clave en la evolución del TCE, es un debate constante entre la fiabilidad y la practicidad o utilidad en todos los pasos de construcción del test (Véase 3.1.2.1). La simplificación, además de en número de colocaciones se produjo en número de criterios al mantener únicamente el de nivel y esquema sintáctico. Los dos criterios descartados son la restrictividad o transparencia y la función léxica. Las razones no son teóricas sino prácticas. Las colocaciones de la muestra con la función Oper⁹⁹ eran insuficientes como para realizar un análisis estadístico fiable y estadístico. En la *Tabla 46* quedan representados dos criterios básicos de la selección de la muestra (nivel y esquema sintáctico) y el número de colocaciones escogidas para representar cada uno de los niveles en el test.

	COLOCACIÓN ELEGIDA	ESQUEMA SINTÁCTICO	NIVEL
1		V+N	A1
2			
3			
4		N+A	
5			
6			

Tabla 41. Esquema de colocaciones de un nivel para el TCE

⁹⁸ Para ver la reconstrucción del test y, en concreto, la reducción de colocaciones, ir al apartado 4.4

⁹⁹ En el total de la muestra había únicamente seis colocaciones *Oper* una por cada nivel, lo que era insuficiente. Ocurría lo mismo con el resto de funciones, ya que solo contábamos con seis en toda la muestra y una solo por nivel.

Una vez decidido que el número de colocaciones meta debía pasar de nuevo a treinta y seis se establecieron los criterios para descartar las otras treinta y seis colocaciones. El análisis estadístico del ítem, en concreto el índice de facilidad, no sirvió para descartar aquellas colocaciones metas cuyos ítems tenían en las cuatro modalidades índices que no atendían a las horquillas preestablecidas de aceptación. Aun así, tuvimos en cuenta que al tratarse de un test multinivel y una población adscrita toda a un único nivel, era lógico que los ítems del nivel inicial tuvieran un índice de facilidad muy alto y los del nivel avanzado, un índice muy bajo. El punto de corte establecido del IF en cada nivel para descartarse fue el siguiente: los ítems correspondientes con el nivel inicial (A1-A2) debían tener un índice mayor a .70; los ítems pertenecientes al nivel intermedio (B1-B2) debían oscilar entre .30 y .70; y, por último, los pertenecientes al nivel avanzado (C1-C2), menos de .30. Este criterio estadístico no fue suficiente, ya que más de treinta y seis de los ítems de las colocaciones meta cumplían con estas condiciones. Por lo tanto, fue necesario establecer un segundo criterio, que en este caso tenían que ver con la adecuación del ítem y su relación con el contexto, evitando así todos aquellos que podían caer en realidades culturales ignoradas por algunos de los participantes. Es decir, mantuvimos las colocaciones meta con los ítems más válidos, fiables y adecuados. En el ANEXO XI se puede ver la lista definitiva de colocaciones meta.

En la muestra del TCE también se ha buscado la correlación de las colocaciones con la frecuencia, en este caso la herramienta utilizada ha sido el *Corpus aprendices del español como lengua extranjera*, CAES¹⁰⁰. La intención era correlacionar el nivel del PCIC con el del CAES, sin embargo, la mayoría de las combinaciones no aparecen en el Corpus de aprendices. Solo nueve de las treinta y seis colocaciones aparecen en el corpus; cuatro de A1 (*hacer frío, llevar gafas, pelo rubio y habitación doble*), dos de A2 (*vino tinto y comida rápida*) y una de B1 (*sacar fotos*), una de B2 (*familia numerosa*) y otra de C2 (*contraer matrimonio*). Algunas con una única ocurrencia, como es el caso de *comida rápida* o *contraer matrimonio*. Otras aparecen con muchas ocurrencias como *llevar gafas*

¹⁰⁰ Este corpus compila textos escritos por aprendices de ELE con diferentes lenguas maternas y adscritos a diferentes niveles del MCER. Consta de 570000 elementos lingüístico en su última versión (1.3.) de abril de 2020. En el siguiente enlace se puede acceder al corpus: <https://galvan.usc.es/caes>.

(84 ocurrencias en A1 y 24 en A2), *pelo rubio* (20 ocurrencias en A1 y 17 en A2) y *habitación doble* (133 ocurrencias en A2).

En conclusión, los cambios en la muestra han sido relevantes y han ido unidos a la evolución del TCE. En la siguiente tabla se muestran algunos de los cambios de criterio más reseñables. En los ANEXOS IX, X, XI se puede ver la muestra de cada una de las fases del test: las treinta y dos colocaciones del preTCEp, las setenta y dos colocaciones del TCEp y las treinta y seis colocaciones del TCE.

Lista de colocaciones meta			
	Muestra de preTCEp	Muestra de TCEp	Muestra del TCE
Número de colocaciones meta	36	72	36
Fuentes	Davies <i>Práctico</i>	PCIC <i>Práctico</i> Sketch Engine	PCIC <i>Práctico</i> CAES
Criterios de selección	Esquema sintáctico, frecuencia, función léxica y grado de restricción	Nivel, esquema sintáctico, frecuencia, función léxica y grado de restricción	Nivel y esquema sintáctico

Tabla 42. Comparación de las tres muestras en las tres versiones del TCE

4.6. Construcción del TCE

Una vez que la muestra estaba seleccionada, el formato elegido para cada modalidad y los criterios de diseño definidos se pasó a la construcción del test. De cada colocación meta se construyeron a la vez los cuatro ítems, y sobre todo los contextos, de cada una de las modalidades. Con este proceder se aseguraba que los cuatro ítems estuvieran correlacionados y que, además, se equiparara el nivel de dificultad entre las modalidades. Además, esta metodología era imprescindible para evitar que, en las modalidades de producción, PRECO y PRECU, se diera alguna pista para la realización de las modalidades de comprensión, CRECO y CRECU. En la *Tabla 48* podemos ver un ejemplo de los cuatro formatos para una colocación del nivel más alto.

SUEÑO LIGERO		
	PRODUCCIÓN	COMPRENSIÓN
RECUPERACIÓN	PRECU Tengo el sueño l..., me despierto con cualquier ruido.	CRECU Tengo el sueño ligero , me despierto con cualquier ...
RECONOCIMIENTO	PRECO Tengo el sueño ..., no duermo bien. a) voluble b) ligero c) profundo d) recurrente	CRECO Tengo el sueño ligero , ... a) me cuesta despertarme. b) me despierto fácilmente. c) duermo muy bien. d) nunca quiero dormir.

Tabla 43. Ejemplo de ítem en las cuatro modalidades del TCE

Otra de las observaciones detectadas en la fase de construcción de los ítems es que a medida que se incrementa el nivel de la colocación, el mayor peso de la carga de aprendizaje de la colocación deja de centrarse en la base para hacerlo en el colocativo. Lo podemos ver de forma clara en la elaboración de las glosas para el formato de comprensión reconocimiento (CRECO); por ejemplo, en dos colocaciones como *tener sueño* y *cumplir sueño*. En ambas, la base es la misma desde el punto de vista de la forma, aunque no desde el punto de vista del sentido (dos aspectos diferentes que debemos tener en cuenta en el conocimiento de la colocación). En la primera, la glosa es sobre la base ‘ganas de dormir’. Sin embargo, en la segunda, la glosa se centra en el sentido que aporta el verbo *cumplir*, sería algo como ‘conseguir algo que se desea mucho’. En el primer caso el verbo *tener* es un colocativo formado por un ver soporte; en el segundo, el verbo *cumplir* tiene un sentido mucho más restrictivo.

4.7. X-LEX en español

Con la intención cumplir con uno de los objetivos de esta investigación, la comprobación de si la competencia colocacional es índice de la competencia léxica, es decir, si las dos variables están correlacionadas, los aprendices de español que participan en este estudio, además del cuestionario (basado en la biografía lingüística del Porfolio) y el TCE, deben cumplimentar un test que mide la extensión de vocabulario. El propósito es doble: por un lado, confirmar el nivel de los participantes en el test; por otro lado, establecer una correlación positiva o negativa entre la competencia léxica y la competencia colocacional, así como con la competencia lingüística. Ya que se ha

comprobado que la extensión del vocabulario aumenta en la medida en la que lo hace la competencia lingüística (Milton, 2013), parece obvio presuponer que en este caso también la competencia colocacional aumentará al incrementarse la competencia colocacional. Esto es, que la competencia léxica es índice de la competencia lingüística en L2. Si bien es cierto, estas afirmaciones se limitan al contexto del inglés y no hay estudios que certifiquen dicha asociación en español. Dado que el conocimiento de las colocaciones es uno de los componentes del conocimiento de las palabras, cabe presuponer que el conocimiento de las colocaciones es dependiente del conocimiento del léxico y que entre ambas competencias se establecerá una correlación positiva (Nizonkiza, 2012, 2015). Un aprendiz no tendrá una buena competencia colocacional si no tiene una buena competencia léxica; no entenderá la colocación *pensamiento fugaz* si no conoce el sentido que adquiere *fugaz* en esa combinación.

A la hora de medir la extensión del vocabulario se ha usado por primera vez la versión paralela del X-LEX (Meara & Milton, 2003) en español. El test original X-LEX¹⁰¹ mide la extensión del vocabulario del inglés y hay otras versiones paralelas¹⁰² en otros idiomas. Una de las versiones publicadas es la que mide la extensión de vocabulario del griego (James Milton & Alexiou, 2010). El diseño del test tiene como base la relación entre la extensión y la frecuencia. Uno de los puntos más destacados es la vinculación entre *el perfil basado en frecuencias* (Milton, 2013) y la competencia léxica (vid. 2.2.3). El perfil representa que el número de palabras conocidas disminuye a medida que aumenta la frecuencia de las palabras en cuestión.

Este test mide la extensión del vocabulario en relación con la frecuencia. Utiliza un formato de sí / no (o *checklist*) en el que los aprendices se encuentran con una lista de palabras y deben señalar cuáles conocen, midiendo así solo la comprensión desde el reconocimiento. Las palabras meta se extraen de una lista de palabras frecuentes, en este caso de *A Frequency dictionary of Spanish: core vocabulary for learners* (Davies, 2006),

¹⁰¹ Otros test miden la extensión del vocabulario comprensivo como *Vocabulary Size Test* (VST) (Nation & Beglar, 2007). *X-LEX* tampoco es el único test que mide la extensión del vocabulario comprensivo a partir de *checklist*, Meara y sus colegas han creado una serie de test que siguen esta línea: *Eurocentres Vocabulary Size Test* (EVST), *EFL Vocabulary Test* y *V-YesNo test*. La diferencia entre *V-YesNo* y *X-LEX test* reside en el número de lemas frecuentes que se toman en cuenta. Mientras que en *V-YesNo* se tienen en cuenta los 10000 lemas más frecuentes, *X-LEX* solo tiene en cuenta los 5000 lemas más frecuentes.

¹⁰² Incluso crearon una versión oral de extensión de vocabulario receptivo: *A-LEX (Aural Lex)* que presenta una lista de ciento veinte ítems grabados que incluye veinte *pseudopalabras*.

empleando el siguiente criterio: un número de palabras de cada una de las bandas de frecuencia; así, se extraen veinte palabras lematizadas de la lista de las mil palabras más frecuentes, veinte palabras de las dos mil más frecuentes, veinte de las tres mil más frecuentes, veinte de las cuatro mil y otras veinte de las cinco mil.

Además, a estas cien palabras se añadan veinte *pseudopalabras*. Estas siguen el modelo de formación de palabras y la combinación de sonidos propios del español para evitar la sobreestimación a partir de falsas alarmas, aquellas *no palabras* que los aprendices marcan como conocidas. La fórmula para calcular el resultado es muy sencilla, se resta el número de falsas alarmas, entendido como el número de pseudopalabras marcadas como palabras conocidas, al de aciertos. Según el resultado, de un máximo de 5000, se clasifican los participantes en su nivel del MCER. En la tabla siguiente se pueden ver las franjas de equivalencia para cada resultado. De modo que, si conoce entre 2750 palabras y 3250 su competencia léxica se corresponderá con el nivel B1 del MCER. En las modalidades realizadas hasta el momento se han percibido ciertas diferencias en las diferentes versiones (inglés, griego o francés) en cuanto al límite de las franjas de cada nivel. No es necesario saber el mismo número de palabras para tener un nivel B1 en griego que en inglés, por ejemplo.

<i>CEF level</i>	<i>Cambridge exam</i>	<i>XLex score (max. 5000)</i>
A1	Starters, Movers and Flyers	<1500
A2	Kernel English Test (KET)	1500–2500
B1	Preliminary English Test (PET)	2750–3250
B2	First Certificate in English (FCE)	3250–3750
C1	Cambridge Advanced English (CAE)	3750–4500
C2	Cambridge Proficiency in English (CPE)	4500–5000

Figura 48. Relación entre niveles del MCER y resultados de XLex (Milton & Treffers-Daller, 2013)

En el caso del test realizado a aprendices de griego, al igual que ocurre en inglés y en francés, las hipótesis se confirman y la fiabilidad y validez del test también. Los resultados reflejan que el efecto de la frecuencia está presente y que la media de los resultados incrementa con cada nivel. Esto no hace más que corroborar el fuerte impacto de la frecuencia en el conocimiento léxico al observar diferencias entre aprendices de diferentes niveles y coherencia con las predicciones del perfil. Debemos señalar, sin embargo, que uno de los puntos diferenciales entre las diferentes versiones del X-LEX es

el número de palabras necesarias en cada banda de frecuencia para llegar al nivel. En griego, el número requerido para cada nivel podrá ser más que el requerido para alcanzar el nivel en inglés o en francés. Este será uno de los aspectos que debemos tener en cuenta a la hora de analizar los resultados en la versión en español de X-LEX.

CEFR Levels	XLex (5000 max)		
	English	French	Greek
A1	<1500	1160	1486
A2	1500 - 2500	1650	2237
B1	2750 - 3250	2422	3288
B2	3250 - 3750	2630	3956
C1	3750 - 4500	3212	
C2	4500 - 5000	3525	

Figura 49. Asignación niveles del MCER y los resultados de X-LEX en inglés, francés y griego (Milton & Alexiou, 2010)

La elección de este test pudiera ser contradictoria con los principios que rigen el TCE, sobre todo, pero prevaleció la adecuación a una de las propiedades de la evaluación, la practicidad. La principal ventaja de este formato es el elevado número de palabras que se pueden incluir y la rapidez con la que se realiza. Además, es fácilmente adaptable a cualquier idioma, fácil de construir y, sobre todo, cómodo para los aprendices. Además, en otras lengua se ha demostrado la correlación de la competencia léxica así medida con otras competencias (Meara & Miralpeix, 2017b). Primero, se elige un test de léxico que mide la extensión, mientras se prefiere medir la profundidad en el test de colocaciones. En el de colocaciones buscamos una forma válida y eficaz de medir la solidez del conocimiento porque se trata de un test de diagnóstico, pero no pretendemos detectar la dificultad en el aprendizaje del léxico, por lo tanto, necesitamos un tipo diferente de test. Esta elección viene determinada por el propósito del test: no se trata de diagnosticar las fortalezas y debilidades sino de comprobar lo conocido, es decir un test de nivel (*achievement*). Segundo, mientras se rechaza la frecuencia tanto como criterio definitorio de las colocaciones como para la elección de la muestra, el X-LEX tiene la frecuencia como principio general. La relación entre la frecuencia y las palabras se marca en su uso, pero en el caso de las colocaciones se marca en su definición e identificación y ahí es donde no estamos de acuerdo. Está demostrado que el uso de palabras frecuentes es correlativo con la competencia, por eso es válido para nuestro propósito este test. El uso de colocaciones frecuentes no lo está con la competencia colocacional, léxica o lingüística,

principalmente porque desde la perspectiva fraseológica todavía no está claro cómo se define una colocación frecuente y si esa colocación frecuente se aprende antes que una colocación poco frecuente. Entre las razones que nos llevaron a elegir este y no otro test para tener una referencia sobre la competencia léxica de los aprendices también es el hecho de que este formato ya está validado (Meara & Milton, 2003). Nuestra aportación es validarlo en español por primera vez. El coste de crear un nuevo test para medir la extensión de vocabulario sería demasiado ambicioso en esta investigación. Otra de las razones importantes parte de su simplicidad, la rapidez a la hora de cubrirlo supone una ventaja importante ya que una de las mayores dificultades del TCE es su extensión y, en consecuencia, la cantidad de tiempo empleado con las implicaciones que esto conlleva en la motivación de los aprendices / voluntarios que participan de forma voluntaria.

Otros test que tradicionalmente han medido la extensión del vocabulario son los test de traducción de la palabra a la L1. Cumplían con una de las ventajas del anterior, la rapidez y fácil corrección. Además, no ocasiona tantos problemas de sobreestimación y su validez es más evidente. Sin embargo, ocasionaba otros inconvenientes. Para hacer un test de traducción se necesita conocimiento de dos lenguas, la L1 y la L2. Esto es posible cuando los participantes del test comparten la misma lengua o tienen un nivel muy alto en una lengua meta, normalmente el inglés. Este no es el caso de los participantes de esta investigación. La población es heterogénea, con L1 muy diferentes desde ruso hasta portugués. El hecho de que no haya ningún otro test estandarizado que mida la extensión en español y que exista una versión en español de éste, junto con el resto de motivaciones expuestas, hace que sea la opción elegida. En el ANEXO VIII se puede ver la versión en español de X-LEX.

4.8. Conclusión

En este capítulo nos hemos centrado en la descripción del centro de esta investigación, el TCE. Hemos explicado las etapas de elaboración del test buscando la sistematicidad, a partir tanto de los modelos diseñados para la evaluación del léxico como los modelos seguidos por los que comparten un mismo propósito, el diagnóstico. Así, hemos atendido a la planificación, el diseño y la construcción del test en sus diferentes fases. Los tres puntos esenciales de este capítulo y que son la base de esta investigación

son los siguientes: la creación de argumento de validez para un test que mide el conocimiento de colocaciones; la propuesta de un modelo de procesamiento y un modelo de conocimiento de colocaciones; por último, el diseño y construcción de un test para medir la competencia colocacional. En el siguiente capítulo se centra en la administración, corrección y el análisis estadístico de los resultados del test.

5. ADMINISTRACIÓN, ANÁLISIS E INTERPRETACIÓN DE LOS RESULTADOS DEL TCE

Este apartado se centra en las tres últimas fases del ciclo de la evaluación: la administración, el análisis y la interpretación desde el primer pilotaje hasta la versión definitiva del test, lo que nos permitió implementar mejoras en las sucesivas versiones. El TCE cuenta con tres versiones (vid. 4.4). En primer lugar, se creó y diseñó el denominado preTCEp (*previo Test de Colocaciones del español piloto*) y que tenía como objetivo comprobar sobre todo la utilidad de los formatos elegidos. En segundo lugar, se adaptó el que se denominó TCEp (*Test de Colocaciones del español piloto*) que buscaba la fiabilidad de la herramienta de evaluación. Finalmente, se adaptó y se creó el TCE (*Test de colocaciones del español*). Primero, describimos el proceso de administración desde la primera versión piloto hasta la versión definitiva. El siguiente apartado se dedica al análisis, tanto cuantitativo como cualitativo, siguiendo preceptos de la Evaluación de Lenguas y del *argumento de validez*. El tercero se centra en la interpretación de resultados y en él se explica cómo la imagen global del conocimiento que los aprendices de ELE tienen de la colocación permite trazar algunas directrices para la mejora de su proceso de enseñanza-aprendizaje.

5.1. Población y administración

Pese a que presentamos tres versiones del TCE, tanto los criterios de administración como los criterios de corrección se mantienen prácticamente inalterables en todas las versiones, salvo pequeñas excepciones que ya hemos explicado. Las principales diferencias entre las tres versiones residen en el número de ítems de cada modalidad y, sobre todo, en el método de selección de las colocaciones, que ya hemos expuesto anteriormente (vid. 4.4).

El TCE se ha administrado a más de doscientos estudiantes de español como lengua extranjera en sus diferentes fases (preTCEp, TCEp y TCE). Además, como cada participante realiza una de las dos opciones, integrada por dos modalidades, cada colocación ha sido evaluada más de cuatrocientas veces. En la *Tabla 47* podemos ver el número de aprendices que ha participado en cada una de las fases del test. Todos

pertenecían al nivel intermedio, B1 y B2, y casi todos cursaban sus estudios de español en inmersión en alguno de los centros de enseñanza del español que ha colaborado en esta investigación.

	N.º de participantes (x2) ¹⁰³	N.º de ítems (de colocaciones testadas) (x4) ¹⁰⁴	Estructura del test	Centro de enseñanza	Niveles de los participantes
preTCEp	38	36	Dos modalidades, cuatro formatos	Centro de Lenguas Modernas de la Universidad de Santiago de Compostela	B1 y B2
TCEp	90	72	Dos modalidades, cuatro formatos	Centro de Lenguas de la UDC EOI de A Coruña	B1 y B2
TCE	102	36	Dos modalidades, cuatro formatos	Universidad de Nápoles Centro de Lenguas de la UDC EOI de A Coruña	B1 y B2

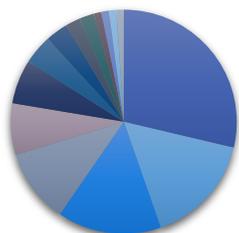
Tabla 44. Datos de participación en las tres versiones del TCE

La L1 de los participantes no es la misma para todos: en el estudio se cuenta con aprendices de español con inglés, portugués o italiano como lengua materna, pero también con ruso, polaco o árabe (vid. Gráfica 1). La gran variedad de lenguas maternas, y ausencia de una lengua franca común para todos los participantes, nos impidió valorar la L1 como factor de análisis pese a que está demostrado que la congruencia es una de las causas de la facilidad o dificultad en la adquisición y dominio de las colocaciones (Gyllstad & Wolter, 2016; Peters, 2016). Además, el nivel cultural de los participantes también es muy variable, la población se divide entre estudiantes universitarios y emigrantes en situación de inmersión, entre los que había personas con un nivel académico alto y otro bajo. Por lo tanto, en nuestro análisis nos centramos únicamente en los factores intrínsecos e intencionales de la carga de aprendizaje.

¹⁰³ El (x2) significa que el número de participantes de cada test se debe multiplicar por dos si contamos el número de evaluaciones realizadas, ya que cada participante realiza dos modalidades del test.

¹⁰⁴ El (x4) significa que el número de colocaciones de cada test se debe multiplicar por cuatro para contabilizar el número de veces que fueron testadas.

L1 DE LOS PARTICIPANTES EN EL TCE



Gráfica 1. Porcentaje de participantes según su L1 en el TCE

El *preTCEp* se administró a un grupo de treinta y ocho aprendices de español con unas características similares: estudiantes universitarios, con unas edades comprendidas entre los veinte y los veinticinco años y en situación de inmersión. Además, todos eran estudiantes del *Centro de Lenguas Modernas de la Universidad de Santiago de Compostela*: veintiuno de ellos, estudiantes del nivel B1 y diecisiete, estudiantes del nivel B2. Todos de ámbitos académicos muy diferentes (desde Medicina hasta Educación) y con diferentes lenguas maternas. Es necesario aclarar que no tienen el nivel indicado certificado ni adquirido, sino que están asistiendo a un curso para alcanzarlo. Por motivos de calendario, el test no se pudo realizar a la vez a todos los participantes, sino que se administró a cada grupo¹⁰⁵ de forma independiente, pero en las mismas condiciones: en el tiempo de clase cada profesora cedió media hora para la realización del test, en el primer cuarto de hora se realizó la primera parte del test, en el siguiente cuarto de hora, la segunda. Los participantes se dividieron aleatoriamente en dos grupos, cada grupo realiza una de las opciones (Modalidad A y B). En esta versión había una hoja de preguntas y una de respuestas (ANEXO XIII, XVI, XVIII y XX), en las posteriores versiones se eliminó.

	B1	B2	
Modalidad A	11	9	20
Modalidad B	10	7	17
	21	16	37

Tabla 45. Número de participantes por modalidad y nivel del *preTCEp*

¹⁰⁵ En el *Centro de Lenguas Modernas de la Universidad de Santiago de Compostela* no todos los alumnos pudieron realizar el test de forma simultánea, sino que lo hicieron en el horario de su clase. Con cada grupo me refiere a cada una de las clases que participaron: una de B1 y otra de B2.

Setenta y tres aprendices de ELE participaron en el TCEp a alumnado del *Centro de linguas* y de la *Escola Oficial de Idiomas da Coruña*. Treinta y siete de los alumnos pertenecían al nivel B1 y 36 al nivel B2.

	B1	B2	
Modalidad A	18	24	42
Modalidad B	19	12	31
	37	36	73

Tabla 46. Número de participantes por modalidad y nivel del TCEp

A la versión final, TCE, contribuyeron más de cien estudiantes de español como L2, pertenecientes a diferentes centros que colaboraron en esta investigación. Los centros participantes, como ya hemos avanzado, son el *Centro de Linguas da Universidade da Coruña*, la *Escola Oficial de Idiomas da Coruña* y la *Università degli studi "Suor Orsola Benincasa" di Napoli*. Cincuenta y un estudiantes realizaron la modalidad A y, cincuenta y un estudiantes realizaron la modalidad B. En la *Tabla 50* se puede ver la distribución de los participantes según dos criterios: por un lado, la modalidad realizada del TCE; y, por el otro, el nivel del MCER que están estudiando. Por problemas de gestión, la prueba se administró de forma presencial a través de papel y lápiz y en formato electrónico (más de cincuenta participantes), con una adaptación en los Formularios de Google. Sería ideal haber podido realizar todas las pruebas utilizando el mismo formato, pero las condiciones lo hicieron imposible.

	B1	B2	
Modalidad A	37	14	51
Modalidad B	30	21	51
	67	35	102

Tabla 47. Número de participantes por modalidad y nivel del TCE

5.2. Análisis de los resultados del TCE

Para la realización del análisis del TCE usamos varias herramientas estadísticas: el procesador de textos Excel, el programa JASP y, sobre todo, JAMOVI. El análisis en las tres versiones del TCE mantiene una estructura muy similar, únicamente alterada por los cambios que las conclusiones del propio análisis nos permiten plasmar en el análisis de la siguiente versión y por una progresión en la concreción del análisis. A partir de este estudio y la aplicación tanto de los criterios de corrección como de clasificación daremos

respuesta a las preguntas de investigación. Previamente se presentan los resultados del test de léxico que realizan los participantes, X-LEX¹⁰⁶.

Para el análisis cuantitativo de los resultados, tendremos en cuenta las cuatro inferencias que propone Kane (2013) para el análisis cuantitativo de los datos: *inferencia de los resultados*¹⁰⁷, *generalización*¹⁰⁸, *extrapolación*¹⁰⁹ e *implicaciones*¹¹⁰ (Cook et al., 2015; Kane, 2013) o, lo que es lo mismo, según la metáfora del puente: evaluación, generalización, explicación y utilización (Green, 2014). La distribución del apartado, sin embargo, no se estructura según estas inferencias, sino que lo hace según las preguntas de investigación, que tienen en cuenta cada una de las fases de inferencia del *argumento de validez* y, aunque no se corresponden de forma biunívoca las preguntas con los pasos de inferencia, sí lo hacen parcialmente, ya que la ejecución de cada uno de los pasos nos permite responder a cada una de las preguntas.

5.2.1. Análisis del *preTCEp*

En esta primera versión el número de participantes es muy reducido para extraer conclusiones generalizables, pero sí nos permitió tomar resoluciones importantes que aplicamos en las siguientes versiones.

Los hablantes nativos que realizaron el test no cometieron ningún error en ninguna de las modalidades ni detectaron ningún problema en la construcción de los ítems, incluidos los distractores. Como vimos anteriormente, veintiuno de los aprendices realizaron la Modalidad A y diecisiete realizaron la Modalidad B. En la *Tabla 51* podemos ver la puntuación media que los aprendices obtuvieron en cada una de las modalidades

¹⁰⁶ En la primera versión del test, *preTCEp*, no podemos comparar la competencia lingüística con la competencia léxica porque no se había decidido todavía incluir el test de competencia como instrumento de evaluación ni parte de la investigación.

¹⁰⁷ La evaluación propiamente que permite reflejar la habilidad evaluada en unos resultados, que explicamos en el apartado dedicado a los criterios generales de administración 4.3.3.3 y corrección de las diferentes versiones 4.3.3.1

¹⁰⁸ La generalización permite establecer una relación entre el test y las especificaciones, que condensamos en el apartado 4.3.3.

¹⁰⁹ El proceso de extrapolación o explicación consiste en relacionar las especificaciones con el diseño del test, que esbozamos en el 4.3.2.

¹¹⁰ Definición de implicación o utilización conecta el diseño con las decisiones, que vemos en el Capítulo cinco.

y la puntuación empírica¹¹¹ al calcular la competencia colocacional (a partir de ahora CCol), a partir del análisis de la estadística descriptiva de la muestra. De la observación de estos datos podemos concluir que la modalidad en la que peores resultados se obtienen es en PRECU, es decir, en el formato que mide la producción desde la recuperación. Esto es, una de las formas más habituales de medir la competencia colocacional y desde la que se ha venido argumentando que las colocaciones, como unidad fraseológica, son uno de los aspectos más problemáticos en el aprendizaje de segundas lenguas (Leńko-Szymańska, 2020, p. 37; Wray, 2013). También podemos observar como la sección en la que mejores resultados se obtiene es en CRECO, modalidad en la que se mide el conocimiento colocacional desde la comprensión y el reconocimiento. La diferencia entre una sección y la otra es de más de diez puntos. En la Modalidad B, por el contrario, la diferencia entre cada modalidad es mínima, de tan solo medio punto. Esta tendencia se mantiene también en la versión definitiva y que trataremos de explicar de forma más concreta en ese punto. Aun así, los resultados son mejores en CRECU, índice muy próximo a su sección homóloga.

	Modalidad A			Modalidad B		
	PRECU	CRECO	CC ¹¹²	PRECO	CRECU	CC
Válido	20	20	20	18	18	18
Media	19.250	30.050	24.850	23.833	24.333	24.083
Desviación estándar	4.315	5.083	4.283	6.981	7.978	7.158
Mínimo	12.000	18.000	15.000	10.000	10.000	10.000
Máximo	26.000	35.000	31.000	36.000	34.000	35.000

Tabla 48. Estadística descriptiva de preTCEp

Si comparamos los resultados de la competencia entre una y otra modalidad, son muy similares, casi idénticos. Lo que nos puede llevar a pensar que ambas opciones son un recurso aceptable para graduar la competencia colocacional y, quizás, que las cuatro secciones como partes de una misma modalidad describirían de la mejor forma posible la competencia colocacional. A pesar de esto, y centrándonos en los resultados de cada sección (que hemos explicado en el párrafo anterior) no dejan de ser unos resultados sorprendentes. Si observamos la desviación estándar, vemos cómo es mayor en la competencia colocacional obtenida a partir de la Modalidad B, lo que nos lleva a pensar

¹¹¹ Se opone aquí a puntuación real libre de error.

¹¹² Con CC nos referimos a la Competencia Colocacional, calculada desde una u otra modalidad. En el primer caso se calcula a partir de los resultados de PRECU y CRECO; en la segunda, a partir de los resultados de PRECO y CRECU.

que esta descompensación se deba a la mayor variedad de niveles de competencia de los participantes que los que realización las modalidades de la Modalidad A.

Desde el punto de vista del test de diagnóstico, debemos detectar cuáles son los puntos fuertes y cuáles los débiles de este test. Para ello ordenamos las secciones de forma jerárquica, al igual que se hace en CATSS, desde el que supone un mayor reto para los aprendices y cuál supone un menor desafío para los aprendices. En la *Figura 50* podemos ver como PRECU es el que supone un mayor reto, seguido de PRECO, luego CRECU y, finalmente, CRECO, como la sección en donde los aprendices encontraron menos dificultades. Uno de los objetivos es comprobar si en las sucesivas versiones esta jerarquía se mantiene y si, por lo tanto, podemos afirmar que el conocimiento colocacional implica un dominio jerárquico de una serie de aspectos. En este sentido, es también interesante comprobar cuál de las dos dualidades (comprensión / producción o reconocimiento / recuperación) ocasionan un mayor coste de aprendizaje.

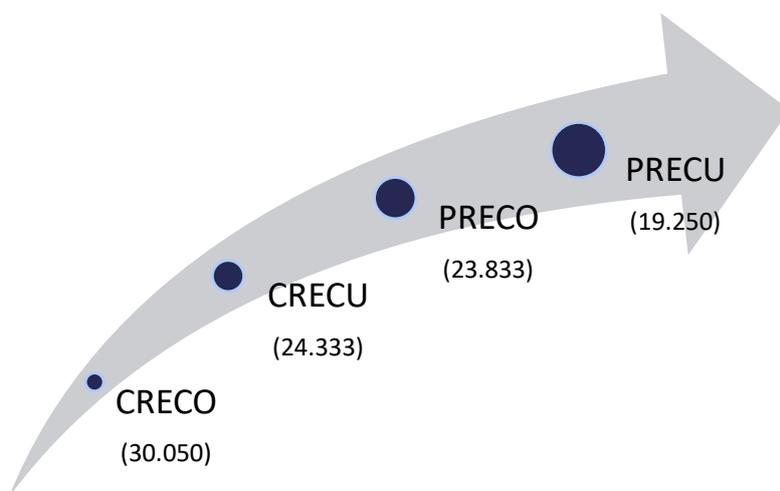


Figura 50. Jerarquía del conocimiento colocacional en preTCEp

Otro de los interrogantes que se planteaba es si la competencia colocacional es diferente en función del nivel de competencia lingüística de los aprendices. En las *Tablas 52 y 53* podemos ver los datos desglosados según este criterio. A pesar de que en la muestra solo contamos con dos grupos de aprendices, unos enmarcados en el nivel B1 según su competencia lingüística, y otros en un nivel B2, observamos claras diferencias. Los aprendices que están cursando B2 obtienen mejores resultados en todas las secciones que los de B1. Lo que demostraría que la competencia colocacional aumenta a medida que aumenta el nivel de la competencia lingüística, de forma clara en la

Modalidad A. Sin embargo, en la Modalidad B, los resultados son superiores en B1. Cabe la opción de que la competencia de los participantes no esté bien nivelada. De todas formas, el número de informantes es muy bajo como para extraer conclusiones generalizables.

	PRECU		CRECO		CC	
	B1	B2	B1	B2	B1	B2
Válidas	10	8	10	8	10	8
Media	18.500	20.500	28.800	31.000	23.900	25.875
Desviación estándar	4.428	4.690	6.286	3.703	5.216	3.399
Mínimo	12.000	13.000	18.000	26.000	15.000	20.000
Máximo	26.000	26.000	35.000	35.000	31.000	29.000

Tabla 49. Resultados de la Modalidad A en el preTCEp

	PRECO		CRECU		CC	
	B1	B2	B1	B2	B1	B2
Válidas	10	8	10	8	10	8
Media	24.600	24.1	25.500	22.9	25.050	23.5
Desviación estándar	6.769	5.54	8.489	7.586	7.59	6.26
Mínimo	11.000	14.0	10.000	10.000	11.000	12.0
Máximo	36.000	33.000	34.000	34.000	35.000	33.05

Tabla 50. Resultados de la Modalidad B en el preTCEp

Tras los cálculos relativos a la estadística descriptiva, se pasó al análisis del ítem. Al tratarse de un test de diagnóstico multinivel basado en criterio, los mecanismos estadísticos habituales para comprobar la validez y la fiabilidad no funcionan, como hemos explicado anteriormente (vid. 3.1.5). Aun así, consideramos útil realizar varios análisis teniendo en cuenta tanto la teoría clásica del test (TCT) como la teoría de respuesta del ítem (TRI). En primer lugar, calculamos la dificultad o facilidad del ítem en los treinta y seis ítems de las cuatro modalidades. Para que el ítem sea fiable el índice ha de situarse en un rango entre el 20% y el 70% de aciertos (Green, 2013). En este caso, presentar la media del índice en cada una de las modalidades reduce la información. Al observar el índice ítem a ítem comprobamos que los que se corresponden con los niveles más bajos (del 1 al 12), con relación al nivel de competencia de los participantes, superan el 70% de los aciertos y los ítems que se corresponden con los niveles superiores (del 25 al 26) no llegan al mínimo. Únicamente los ítems correspondientes al nivel intermedio (del 13 al 24) se encuentran entre los límites esperables.

Presentamos dos ejemplos en la *Tabla 54 y 55*. Recordamos que asignamos un nivel a cada uno de los ítems del test, de modo que los seis primeros se corresponderían

con el nivel A1, los seis siguientes con nivel A2 y así sucesivamente hasta los seis últimos que se corresponderían con el nivel C2. De manera que el ÍTEM 5 representa los ítems de nivel A1 y el ÍTEM 35 los de nivel C2. Pese a que el índice de facilidad se mueve en un rango concreto para considerarse válido y fiable, en el caso del test de diagnóstico los resultados son otros. Como cabría esperar, para participantes que cursan un nivel B1 y B2 el ÍTEM 5 es fácil en todas las modalidades, aunque con resultados muy diferentes al resto de modalidades en PRECU. Igualmente, los resultados del ÍTEM 35 no se adaptan al rango estándar del índice de facilidad, pero sí a los esperados en un test de diagnóstico para unos aprendices del nivel intermedio. En todas las modalidades este ítem resulta difícil para los aprendices. Aunque debemos señalar que la diferencia entre CRECO y PRECO, en las que resulta muy fácil, y en CRECU, en la que resulta bastante fácil, se contrapone con los resultados obtenidos en PRECU, en los que la dificultad es elevadísima. Esto avanza ya algunas de las conclusiones de esta investigación: la producción desde la recuperación es un desafío, pero la comprensión desde la comprensión no supone un reto para los aprendices de ELE.

Ítem 5 <i>pele rubio</i>	PRECU	CRECO	PRECO	CRECU
B1	0.36	0.91	0.80	1.00
B2	0.44	0.90	1.00	0.75

Tabla 51. Índice de facilidad de un ítem catalogado como de baja dificultad

Ítem 35 <i>envidia sana</i>	PRECU	CRECO	PRECO	CRECU
B1	0.18	0.45	0.38	0.60
B2	0.00	0.70	0.83	0.50

Tabla 52. Índice de facilidad de un ítem catalogado como de alta dificultad

Tras el análisis cuantitativo del preTCEp hemos comprobado que la selección de la muestra no cumple con lo esperado porque no nos permite comparar la competencia colocacional con la competencia lingüística. En este punto de la investigación los criterios de clasificación no estaban todavía especificados y el estándar no estaba definido, de modo que se necesitó una nueva selección de la muestra que nos permitiera trazar esos parámetros. Además, comprobamos que treinta y seis ítems eran insuficientes para atender a todos los criterios que se planteaban como el grado de transparencia de las colocaciones o su función léxica. Por tanto, en la nueva selección, como se ha explicado anteriormente, tanto la función como el grado de transparencia se eliminaron como criterios y se añadió el nivel de la colocación como criterio base. Aun así, el test sí parece

medir la imagen global del conocimiento colocacional, tanto en una modalidad como en la otra. Incluso, los resultados en las distintas modalidades nos permiten graduar el conocimiento de forma jerárquica, de modo que una modalidad (PRECU, PRECO, CRECO y CRECU) supone mayor obstáculo que la otra.

Sin embargo, sí fue necesario modificar la metodología del test en las posteriores versiones, al menos parcialmente. Aunque mantenemos las dos modalidades con sus dos opciones, comprobamos que para asignar un nivel de la competencia colocacional es imprescindible diseñar un estándar y, sobre todo unos criterios de clasificación, que nos permitieran situar la competencia colocacional dentro de un rango o sección, buscando así que la comparación con la competencia lingüística se pueda hacer efectiva. Además, se detectó que entre la competencia lingüística y la colocacional el espacio cognitivo era demasiado grande y se propuso comparar también con la competencia léxica. Mientras el dato de la competencia lingüística ya venía dado por las instituciones colaboradoras, no teníamos ningún tipo de información de la competencia léxica. De modo que para la siguiente versión se buscó un test rápido que pudiera medir dicha competencia sin perjuicio del TCE.

5.2.2. Análisis del TCEp

Antes de iniciar el análisis del TCEp debemos recordar que en esta versión el número de ítems es considerablemente superior al de las otras dos versiones (previa y definitiva): contamos con setenta y dos ítems, lo que se refleja tanto en los criterios de clasificación como en las cifras, aunque no en las interpretaciones de los resultados. Tal y como hemos hecho en el análisis del preTCEp, planteamos una serie de propósitos: a) comprobar si la selección de la muestra nivelada es válida para clasificar a los participantes y asignarles un nivel; b) evidenciar si la nueva extensión de la muestra se adapta a la practicabilidad de la prueba; c) demostrar si las conclusiones a las que llegamos con el primer piloto se mantienen en este y se pueden hacer generalizables. Recordemos que en este test el número de participantes supone una importante mejora con respecto a la anterior versión (preTCEp). Al test para medir la competencia y conocimiento colocacional se suma un test que mide la competencia léxica y que nos permite comprobar si la competencia colocacional es índice de la competencia léxica. Por

lo tanto, en esta versión ya incorporamos la competencia léxica como parámetro, así que en primer lugar analizamos los resultados obtenidos en X-LEX, el test elegido para medir la competencia léxica, y en segundo lugar analizamos los resultados del TCEp.

5.2.2.1. Resultados del test de competencia léxica

En esta versión del TCE se introdujo la prueba para medir la competencia léxica, como explicamos anteriormente, elegimos X-LEX como herramienta de evaluación. Es importante señalar, frente a lo que ocurre en otros test, que no se comprueba si se conocen las palabras que integran las colocaciones que forman parte de la muestra del TCE, sino que se mide la competencia léxica de forma general. La incorporación de esta nueva información supone una mejora considerable con relación a la versión previa. Estos datos nos permitirán establecer una posible correlación entre la CCol (Competencia Colocacional) y la CLéx (Competencia Léxica). Como ya explicamos (vid. apartado 3.2.2.2), los resultados de este test de extensión del vocabulario se pueden ordenar en bandas que se identifican con niveles del MCER¹¹³. La media de los resultados del test sitúa la muestra en un nivel intermedio en la escala de competencia léxica, más concretamente en el nivel B1, clasificando así en el mismo nivel a los participantes en la prueba según su competencia léxica y lingüística. Lo que corrobora el nivel que los aprendices tienen asignado.

	N	Media	Mediana	DE	Mínimo	Máximo
XLE	90	2850	2800	681	1350	4250

Tabla 53. Media de resultados de X-LEX

Además, cuando comparamos la media de resultados de cada uno de los grupos de participantes, según su nivel de competencia lingüística, comprobamos que la competencia léxica aumenta al aumentar el nivel de lengua. De hecho, el grupo que tiene un nivel B1 en la competencia lingüística también mantendría el nivel B1 en la competencia léxica porque la media del total de sus resultados se encuentra en la

¹¹³ Los rangos son los siguientes (Milton, 2006):

- A1, menos de 1500 palabras
- A2, entre 1500 y 2500 (2750) palabras
- B1, entre 2750 y 3250 palabras
- B2, entre 3250 y 3750 palabras
- C1, entre 3750 y 4500 palabras
- C2, entre 4500 y 5000 palabras

horquilla correspondiente a ese nivel (entre 2500 y 3250 palabras) y los participantes que parten de un nivel de lengua B2 alcanza un nivel de competencia léxica B2 también con 3274 palabras.

	CLING	N	Media	Mediana	DE	Mínimo	Máximo
XLEX	B1	55	2580	2550	682	1350	4250
	B2	35	3274	3300	416	2300	3850

Tabla 54. Media de resultados de X-LEX según el nivel de competencia lingüística

Con un análisis más pormenorizado clasificamos a cada participante en su nivel de CLéx y obtenemos los siguientes resultados. En la *Tabla 58* podemos ver el número de aprendices que se asignó a cada nivel de CLéx según los resultados de X-LEX. Casi el 70% se sitúan en el nivel intermedio distribuidos entre el nivel B1 y el nivel B2.

Descriptivas								
	CL	N	Perdidos	Media	Mediana	DE	Mínimo	Máximo
XLEX	A1	1	0	1350	1350	NaN	1350	1350
	A2	24	0	2065	2125	264	1600	2450
	B1	36	0	2811	2800	259	2350	3250
	B2	20	0	3435	3500	311	2550	3750
	C1	9	0	3967	3850	197	3700	4250

Tabla 55. Resultados de X-LEX en TCEp

Un 62% de los participantes se sitúan en un nivel intermedio de los que solo un 22% pertenece al nivel B2 y un 40% al nivel B1. Casi un treinta por ciento tiene un nivel léxico por debajo de su competencia lingüística, de los cuales la inmensa mayoría pertenece al nivel A2, lo que supone un desajuste bastante importante. El 10% restante tiene un nivel de competencia léxica superior a su competencia lingüística. Aunque los datos parecen están acorde con lo esperado, las discrepancias entre la competencia léxica y lingüística pueden deberse a diversas razones: mala adjudicación de los aprendices en un nivel de competencia por sus centros o influencia de su lengua materna.

NIVEL DE CLéx		% DE PARTICIPANTES ASIGNADOS EN CADA NIVEL	
INICIAL	A1	1%	28%
	A2	27%	
INTERMEDIO	B1	40%	62%
	B2	22%	
AVANZADO	C1	10%	10%
	C2	-	

Tabla 56. Porcentaje de participantes asignado a un nivel según los resultados de X-LEX

Comprobamos estadísticamente si las percepciones eran correctas y calculamos el coeficiente de correlación *Pearson*. El resultado muestra que existe correlación entre la competencia léxica y la competencia colocacional.

Matriz de Correlaciones			
		M CER	XLE
M CER	R de Pearson	—	
	valor p	—	
XLE	R de Pearson	0.592	—
	valor p	<.001	—

Tabla 57. Coeficiente Pearson entre el valor de la CL y el valor asignado de la Competencia lingüística

5.2.2.2. Resultados del test de competencia colocacional

Uno de los objetivos concretos de este test piloto era comprobar si el aumento del número de los ítems (pasamos de treinta y seis a setenta y dos colocaciones meta) y la presentación de las colocaciones meta niveladas, afianzaba la fiabilidad del instrumento de evaluación. Un primer acercamiento a la estadística descriptiva nos revela que la competencia, medida a partir de la Modalidad A del TCE y de la Modalidad B del TCE, son bastante similares o no están tan alejadas como cabría esperar, al igual que ocurría en el preTCEp.

	Modalidad A			Modalidad B		
	PRECU	CRECO	CCA	PRECO	CRECU	CCB
Válidos	45	45	45	45	45	45
Media	28.378	52.689	40.533	50.044	42.089	46.067
Desviación estándar	9.452	9.760	9.192	8.762	12.424	9.562
Mínimo	9.000	27.000	18.000	31.000	2.000	27.500
Máximo	44.000	69.000	53.500	69.000	70.000	69.500

Tabla 58. Estadística descriptiva TCEp

En cuanto a la fiabilidad, primero medimos el índice de facilidad¹¹⁴ (IF) de la muestra, pese a que, como ya hemos explicado (3.1.5) no es el mecanismo estadístico más adecuado para los test basados en criterio como el TCE. Los resultados revelan que hay una diferencia importante en los ítems a lo largo de la muestra, aunque la media de cada test revele que los ítems son adecuados, incluso tratándose de un test multinivel. El nivel asignado a las colocaciones meta nos permitió detectar que el índice varía con el nivel de las colocaciones y que, mientras las colocaciones meta de los niveles más bajos

¹¹⁴ Se considera un ítem adecuado cuando su índice oscila entre 30% y 70% (Green, 2013).

son demasiado fáciles, superando la horquilla establecida, las colocaciones adscritas a los niveles superiores eran demasiado difíciles para los aprendices de intermedio que realizaban el test. Esta información nos lleva a pensar que las colocaciones están bien niveladas y que esta oscilación en los índices de facilidad es normal al tratarse de un test de diagnóstico multinivel administrado a aprendices de un solo nivel, intermedio. La fiabilidad interna de los ítems se vio alterada por diferentes factores que reducen su consistencia interna como el mayor número de preguntas que de participantes.

TCE			
MODALIDAD A		MODALIDAD B	
PRECU	CRECO	PRECO	CRECU
0,39	0,71	0,67	0,59

Tabla 59. Índice de facilidad del ítem del TCEp

Para establecer relación entre los resultados del TCE y el MCER y, por lo tanto, clasificar a los participantes en diferentes niveles de competencia, establecimos un punto de corte según el nivel asignado de las colocaciones. Esto nos permite asignar un nivel de competencia colocacional a los aprendices, que comparamos con el nivel de competencia lingüística certificado de cada participante. En las especificaciones, más concretamente el apartado dedicado a los criterios de clasificación (vid.4.3.3.2), se explica de forma más pormenorizada este proceso. En la *Tabla 63* podemos ver los resultados. Entre los participantes que realizaron la Modalidad A, la mayoría se clasifica en un nivel intermedio de CCol, un 62%, del cual un 33% se adscribe al nivel B1 y un 29% al nivel B2, un porcentaje bastante similar. Al igual que los participantes de la otra modalidad, los que realizaron lo Modalidad B se clasifican mayoritariamente en el nivel intermedio, un 45%: solo un 16% en el nivel B1 y un 40% en el nivel B2. Lo más destacable es que tanto en la Modalidad A como en la Modalidad B un porcentaje importante de aprendices de ELE alcanza el nivel C1 (36% en la Modalidad A y un 38% en la Modalidad B). Esto nos lleva a pensar que la CCol no supone un reto para el aprendiz ya que su CCol está por encima de su Cling asignada. En este punto percibimos una diferencia importante entre cada una de las opciones. Mientras la diferencia entre los clasificados en nivel intermedio y avanzado en la Modalidad A es relativamente baja (entre 62% y 36%), la diferencia en la Modalidad B no es tan alta (entre un 56% y un 45%). Finalmente, el número de aprendices que se clasifica en los niveles iniciales (A1 y A2) no es significativa, así como los que se clasifican en el nivel C2.

		Modalidad A (CCol Modalidad A)		Modalidad B (CCol Modalidad B)	
Inicial	A1	2%	0%	0%	0%
	A2		2%		0%
Intermedio	B1	62%	33%	56%	16%
	B2		29%		40%
Avanzado	C1	36%	36%	45%	38%
	C2		0%		7%

Tabla 60. Asignación del nivel de la CCol según los criterios de clasificación preestablecidos

Otro de los puntos clave en el análisis es demostrar que la CCol es índice de la competencia lingüística. Comprobamos si la CLing se correlaciona con la CCol teniendo en cuenta el valor de la CLing asignada a los participantes. Tanto en la CCol calculada a partir de la Modalidad A, como en la CCol calculada a partir de la Modalidad B, la media de resultados es superior en los aprendices con un nivel B2 que B1. Lo mismo ocurre con todas las modalidades de test. De modo que podemos afirmar que la competencia colocacional aumenta a medida que se incrementa la competencia lingüística, igual que en la anterior versión del test, el *preTCEp*.

	PRECU		CRECO		CCA	
	B1	B2	B1	B2	B1	B2
Válidos	31	14	31	14	31	14
Media	26.065	33.500	51.032	56.357	38.548	44.929
Desviación estándar	8.978	8.680	9.901	8.670	9.024	8.237
Mínimo	9.000	16.000	27.000	36.000	18.000	26.000
Máximo	39.000	44.000	69.000	67.000	53.000	53.500

Tabla 61. Estadística descriptiva en la Modalidad A del TCEp

	PRECO		CRECU		CCB	
	B1	B2	B1	B2	B1	B2
Válidos	33	12	33	12	33	12
Media	48.576	54.083	39.727	48.583	44.152	51.333
Desviación estándar	9.209	6.007	12.926	8.284	9.804	6.696
Mínimo	31.000	44.000	2.000	31.000	27.500	37.500
Máximo	69.000	62.000	70.000	59.000	69.500	60.500

Tabla 62. Estadística descriptiva en la Modalidad B del TCEp

A continuación, se muestran los índices de correlación entre las tres competencias o variables del TCE. Esto nos permite determinar si existe algún tipo de relación entre las tres variables: CLing, CLéx y CCol. La CCol, como se mide a partir de dos modalidades diferentes, se relaciona con las otras variables de forma independiente. Si el índice de correlación R se sitúa entre 0.1 y 0.3 la correlación se considerará débil; si R está entre 0.4 y 0.6 será moderada; y, finalmente, si se ubica entre 0.7 y 0.9 la consideraremos fuerte (Green, 2013). Utilizamos el coeficiente de Pearson que nos permite relacionar

variables en distribución normal con variables dicotómicas, como es el caso de la CLing, en este estudio en el que solo contamos con participantes de nivel B1 y B2. En la *Tabla 66* vemos la matriz de correlación.

Como es lógico los índices R de Pearson más altos son los que se dan entre la CCol y cada una de las modalidades, pero también entre las dos modalidades que miden la CCol. Vemos que es ligeramente superior la correlación de la CCol con CRECO que con PRECU (0.958 frente a 0,955) y, por lo tanto, valoramos la correlación como fuerte. Sin embargo, y frente a nuestra hipótesis, la correlación entre la CCol y la CLéx (0.236) y entre la CCol y la CLing (0,437) es moderada. Tampoco la correlación es fuerte entre las distintas modalidades y la CLing, también sería moderada, aunque hay una diferencia de más de un punto entre una y otra: mientras entre CRECO y la CLing la relación es de 0,362, el índice R entre PRECU y la CLing es de 0,478. Incluso es sorprendente que la correlación entre la CLing y la CLéx aun siendo positiva no es fuerte, sino moderada, 0.480. Debemos señalar aquí que en X-LEX no se mide el conocimiento de las bases de las colocaciones testadas en el TCE, sino que se mide el conocimiento léxico de forma genérica. Es decir, se solicita únicamente el reconocimiento de una formal, sin asociar con un ningún significado, lo que supone un esfuerzo cognitivo menor.

Al igual que ocurre con los resultados de la Modalidad A que mide la CCol - A, existe una correlación fuerte entre cada una de las modalidades (PRECO y CRECU) con la CCol - B. La correlación es algo superior en CRECU con CCol (0,933) que en PRECO con la CCol (0.860). Aunque la diferencia es algo mayor que la que se daba entre cada modalidad y la CCol no parece significativa. Donde sí observamos una diferencia notable es en el índice de correlación entre cada una de las modalidades: mientras entre PRECO y CRECU la correlación es moderada (0.618), la correlación entre las modalidades de la modalidad anterior es superior (0.831), que se considera fuerte. En cuanto a la relación entre las otras variables (CLéx y CLing con la CCol) observamos lo siguiente: la correlación entre la CLéx y la CLing es baja (0,259), la correlación entre la CLéx y la CCol-B es también baja y la correlación entre la CLing y la CCol-B, por el contrario, es moderada. Estos resultados chocan con las investigaciones previas. Si concretamos más y nos fijamos en la correlación entre cada una de las modalidades y la CLing vemos como en ambos casos la correlación

es moderada. Pero la relación entre las modalidades y la CL es baja con CRECU y casi moderada con PRECO.

		CL		CLing		PRECU		CRECO		CC-A	
CL	R de Pearson	—									
	valor p	—									
	Rho de Spearman	—									
	valor p	—									
CLing	R de Pearson	0.480	***	—							
	valor p	<.001		—							
	Rho de Spearman	0.446	**	—							
	valor p	0.002		—							
PRECU	R de Pearson	0.242		0.476	***	—					
	valor p	0.109		<.001		—					
	Rho de Spearman	0.228		0.478	***	—					
	valor p	0.132		<.001		—					
CRECO	R de Pearson	0.210		0.362	*	0.831	***	—			
	valor p	0.166		0.014		<.001		—			
	Rho de Spearman	0.146		0.371	*	0.841	***	—			
	valor p	0.338		0.012		<.001		—			
CC-A	R de Pearson	0.236		0.437	**	0.955	***	0.958	***	—	—
	valor p	0.118		0.003		<.001		<.001		—	—
	Rho de Spearman	0.209		0.451	**	0.951	***	0.960	***	—	—
	valor p	0.169		0.002		<.001		<.001		—	—

Nota. * p < .05, ** p < .01, *** p < .001

Tabla 63. Matriz de correlación de la Modalidad A (CC-A), PRECU, CRECO, la Competencia léxica (CL) y la competencia lingüística (CLing)

		CL		CLing		PRECO		CRECU		CC-B	
CL	R de Pearson	—									
	valor p	—									
	Rho de Spearman	—									
	valor p	—									
CLing	R de Pearson	0.259		—							
	valor p	0.086		—							
	Rho de Spearman	0.276		—							
	valor p	0.067		—							
PRECO	R de Pearson	0.318	*	0.401	**	—					
	valor p	0.033		0.006		—					
	Rho de Spearman	0.343	*	0.407	**	—					
	valor p	0.021		0.005		—					
CRECU	R de Pearson	0.157		0.464	**	0.618	***	—			
	valor p	0.302		0.001		<.001		—			
	Rho de Spearman	0.203		0.490	***	0.672	***	—			
	valor p	0.181		<.001		<.001		—			
CC-B	R de Pearson	0.248		0.485	***	0.860	***	0.933	***	—	—
	valor p	0.101		<.001		<.001		<.001		—	—
	Rho de Spearman	0.270		0.498	***	0.845	***	0.949	***	—	—
	valor p	0.073		<.001		<.001		<.001		—	—

Nota. * p < .05, ** p < .01, *** p < .001

Tabla 64. Matriz de correlación de la Modalidad B (CC-B), PRECU, CRECO, la Competencia léxica (CL) y la competencia lingüística (CLing)

En conclusión, la correlación entre la CCol en cualquiera de sus versiones se correlaciona de forma moderada con la CLing, pero entre la CLéx y la CCol o la CLing, el

índice correlación es bajo, a excepción de la CLing y la CLéx en la Modalidad A. Así, podemos concluir que el formato elegido por X-LEX se aleja bastante de los formatos elegidos para la evaluación de la CCol que suponen en cualquiera de sus modalidades un proceso cognitivo más elevado. Es decir, el test de competencia léxica es más sencillo, desde el punto de vista del esfuerzo cognitivo, que cualquiera de las modalidades del TCE; entonces, es comprensible que los resultados en uno sean superiores a los otros. El hecho de que la correlación sea moderada entre la CLing y la CCol puede deberse a que para clasificar a los estudiantes no se suele tener en cuenta su competencia colocacional o a que los estudiantes no estén bien nivelados.

En cuanto a la imagen global del conocimiento colocacional y a la noción del conocimiento gradual y jerárquico, hemos ordenado los resultados obtenidos en cada modalidad desde el que les supone menos dificultades a los aprendices (CRECO, que mide la comprensión desde el reconocimiento) hasta que el que supone un mayor reto para los aprendices (PRECU, que mide la producción desde la recuperación). En esta versión PRECO supone menos problemas para los aprendices que CRECU, de forma contraria a lo que ocurría en el *preTCEp* (vid. Figura 43), aunque debemos recordar que se trata de una versión piloto previa con un número muy bajo de participantes. Aun así, esta alternancia ya avanza que la diferencia significativa entre los dos polos del conocimiento colocacional no es tan clara en los pasos intermedios.

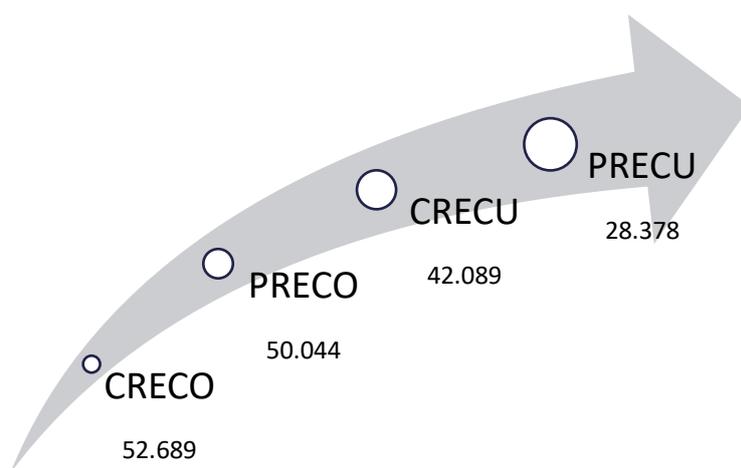


Figura 51. Jerarquía del conocimiento colocacional en el TCEp

Por último, otra de las variables que tuvimos en cuenta en el TCEp fue la categoría gramatical de las colocaciones. Recordemos que en esta versión se incluyeron setenta y

dos ítems, setenta y dos colocaciones, de las cuales la mitad seguían el esquema sintáctico N + A y la otra mitad el esquema V + N. En la *Tabla 68* se muestra el análisis descriptivo centrándonos en esta variable. Los resultados no son significativos en cada una de las modalidades.

		N	Perdidos	Media	Mediana	DE	Mínimo	Máximo
PRECU	N+A	36	0	15.3	16.0	12.07	0	39
	V+N	36	0	20.1	19.0	14.35	0	45
CRECO	N+A	36	0	33.2	37.0	11.08	8	44
	V+N	36	0	32.7	34.0	8.04	6	44
PRECO	N+A	36	0	31.3	34.0	8.75	7	43
	V+N	36	0	31.3	32.0	8.93	7	45
CRECU	N+A	36	0	29.3	29.0	9.43	5	43
	V+N	36	0	23.4	22.0	9.89	3	43

Tabla 65. Análisis descriptivo de la influencia de la categoría gramatical de las colocaciones en las distintas modalidades

Pese a que en la selección de la muestra habíamos barajado otras variables como el carácter restrictivo de la colocación, en el análisis, descartamos esta variable tanto por su carácter subjetivo como por la insuficiencia de los ítems, incluso cuando se había aumentado el número de colocaciones. Al grupo de nativos con conocimiento de diferentes L2 al que se le administró el test, también se le preguntó por el carácter restrictivo de las colocaciones y los resultados, contrarios a lo esperado, revelaban más dudas que certezas. Asimismo, la variable de función léxica también fue descartada por razones similares. Por lo tanto, se descartaron estos dos criterios como condicionantes en la reducción de la muestra que se llevó a cabo en la reconstrucción del test.

5.2.3. Análisis del TCE

En el análisis del TCE nos detendremos en dar respuesta a las preguntas de investigación que sirvan para alcanzar los objetivos propuestos. En primer lugar, analizaremos los resultados del test de léxico, X-LEX, que nos permitirá contar antes de iniciar el análisis del TCE con dos variables: nivel de CLing y CLéx. Organizamos el resto del análisis en los siguientes bloques: a) la adecuación del test según los parámetros estadísticos de fiabilidad (estadística descriptiva y análisis del ítem); b) la relación entre la competencia colocacional, el MCER y las otras competencias (lingüística y léxica); c) la descripción de la imagen global del conocimiento colocacional jerárquico y escalar; d) la gradación del conocimiento colocacional; y e) la posible influencia del tipo de colocación.

Por último, haremos algunas consideraciones desde el punto de vista cualitativo que nos permitirán trazar las debilidades y fortalezas de la competencia colocacional de los aprendices de español como L2.

A continuación, recupero los objetivos de investigación que nos planteábamos al inicio, y explicamos en este apartado:

- Objetivo 1: Medir la competencia colocacional de aprendices de español y compararla con la competencia léxica.
- Objetivo 2: Identificar las dificultades y fortalezas que entraña el dominio de la competencia colocacional del español como L2.
- Objetivo 3: Diseñar un test válido y fiable que mida la competencia colocacional de aprendices de español como lengua extranjera.

No recogemos la pregunta relacionada con el objetivo número cuatro por haber sido ya alcanzado en un apartado precedente (vid. 4.5).

5.2.3.1. *Análisis de la competencia léxica*

En las *Tablas 69 y 70* observamos el análisis estadístico descriptivo en el que diferenciamos los resultados de forma global, es decir, de todo el test sin distinción de niveles, y según el nivel de competencia lingüística asignado a los participantes en sus respectivos centros de estudios¹¹⁵. Como hemos descrito ya, solo contamos con dos grupos de aprendices según nivel de competencia y el número de participantes de B2 es inferior al de participantes de B1; en este caso comprobamos de nuevo la tendencia ascendente en tanto que el conocimiento léxico es mayor en los participantes de mayor nivel que en los de nivel inferior. Cabe apuntar que el resultado mínimo pertenece a un participante clasificado en el nivel más alto, B2, y el resultado máximo en un participante clasificado con el nivel más bajo, B1, lo que demuestra la alta variabilidad entre los participantes y anuncia posibles alteraciones en la asignación de nivel de los participantes en los centros de estudio de español.

¹¹⁵ No se debe confundir este nivel de competencia lingüística, asignado por el centro de estudios, con la competencia léxica alcanzable, asignada por X-LEX.

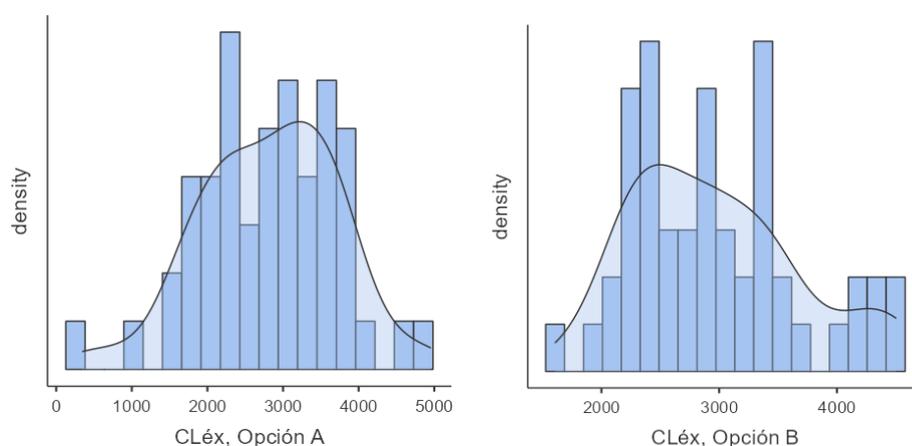
Grupo	N	Media	Mediana	DE	EE
B1	37	2586	2650	867	143
B2	14	3421	3575	751	201

Tabla 66. Estadística descriptiva de los resultados de X-LEX en la Modalidad A

Grupo	N	Media	Mediana	DE	EE
B1	30	2724	2575	624	114
B2	21	3310	3350	709	155

Tabla 67. Estadística descriptiva de los resultados de X-LEX en la Modalidad B

Para demostrar que existe una relación entre la CLing y la CLéx hemos comprobado si la comparación entre las medias es significativa. En la Modalidad A la competencia léxica tiene una distribución normal y como los participantes únicamente tienen nivel B1 o B2, una variable nominal, podemos usar la prueba T Student, para muestras independientes. La prueba nos permite comprobar nuestra hipótesis: en B2 se obtienen mejores resultados que en B1 ($H_a \mu_{B1} < \mu_{B2}$). Según el valor de p , la diferencia entre los dos grupos es significativa con un margen de error muy bajo ($p = 0,001$). No podemos utilizar la misma prueba con los datos de la Modalidad B ya que los resultados de X-LEX en la Modalidad B no muestran una distribución normal, pero sí podemos realizar la prueba no paramétrica, U de Man-Whitney, con la misma hipótesis ($H_a \mu_{B1} < \mu_{B2}$). En esta prueba el índice de significación es idéntico ($p = 0,001$). Podemos concluir, entonces, que existe una diferencia significativa entre los resultados de B1 y B2 en las dos Opciones. La *Gráfica 1* nos permite comparar la densidad entre una y otra modalidad sin tener en cuenta la diferencia de nivel.



Gráfica 2. Histogramas y gráficos de densidad de los resultados de X-LEX

Si comparamos estos resultados con los de las otras lenguas, en líneas generales observamos que los resultados del X-LEX en español están algo por debajo de los márgenes establecidos para el inglés. Los rangos de palabras por nivel en cada lengua se pueden ver en la *Tabla 71*. En B1 el rango oscila entre 2750 y 3250 palabras frecuentes y nuestros informantes se sitúan de media en un 2650 (Modalidad A) y en 2575 (Modalidad B), un poco por debajo de lo esperado. En el caso de B2, el rango esperable oscila entre las 3250 y 3750 palabras y los participantes en el TCE en la Modalidad A tienen una media de 3575 (ligeramente por encima del rango) y los de la Modalidad B, 3350. Los aprendices de español también se encuentran algo por debajo del rango esperado. Si tomamos la referencia de la media del total de la muestra y el rango de intermedio (suma de B1 y B2) vemos que sigue situándose por debajo de los rangos preestablecidos.

CEFR Levels	XLex (5000 max)		
	English	French	Greek
A1	<1500	1160	1486
A2	1500 - 2500	1650	2237
B1	2750 - 3250	2422	3288
B2	3250 - 3750	2630	3956
C1	3750 - 4500	3212	
C2	4500 - 5000	3525	

Tabla 68. Relación entre los niveles del MCER y los resultados de XLEX (Milton & Alexiou, 2010)

Según los rangos establecidos, los resultados de X-LEX nos permiten clasificar a los aprendices que han participado en este estudio en niveles del MCER de acuerdo con su competencia léxica (CLéx). Aunque la mayoría se clasifican en un nivel intermedio, igual al que tienen asignado en competencia lingüística, un porcentaje importante se sitúa en el nivel inicial, A2 de forma mayoritariamente (37,62%), como podemos observar en la *Tabla 72*. Este nivel inferior de la competencia léxica con respecto al de la competencia lingüística asignado indica que la competencia léxica pueda estar infraestimada a la hora de evaluar la competencia lingüística. Esto se contradice con las limitaciones propias de este tipo de test, de *checklist*, cuya principal crítica es la sobreestimación de los resultados. Pese a esto debemos recordar que los participantes, que clasificamos en dos grupos de intermedio, todavía no han alcanzado dicho nivel, sino que están cursando un itinerario para conseguirlo y al finalizar el curso, si aprueban, lo

obtendrán. Y esta podría ser la razón del importante número de aprendices clasificados en nivel A2. No entendemos esta conclusión como una crítica a X-LEX ni a la clasificación realizada por el centro, sino a un grado de profundidad diferente en la clasificación. Si clasificamos a los aprendices a partir de un formato que exige poco esfuerzo a los aprendices, por ejemplo, uno que implique el reconocimiento de la palabra, ni siquiera de la forma, el proceso cognitivo es más sencillo que si clasificamos a partir de un formato que exija un esfuerzo mayor, por ejemplo, la recuperación de un forma sea cual sea el estímulo de partida. En

NIVEL DE CLEX		% DE PARTICIPANTES ASIGNADOS EN CADA NIVEL	
INICIAL	A1	2,97%	40,59%
	A2	37,62%	
INTERMEDIO	B1	24,75%	44,55%
	B2	19,80%	
AVANZADO	C1	13,86%	14,85%
	C2	0,99%	

Tabla 69. Clasificación en niveles del MCER según los resultados de X-LEX

5.2.3.2. *Estadística descriptiva en el TCE*

Una primera observación tras el análisis de la estadística descriptiva (vid. *Tabla 73 y 74*), nos indica que la competencia colocacional medida a través de una y otra modalidad es muy similar: una media de 22,157 en la Modalidad A y de 22,686 en la Modalidad B de 36 posibles repuestas. Esto nos hace pensar que cualquiera de las dos podría ser válida como medida de la competencia colocacional. Sin embargo, hay una diferencia importante entre los resultados de los dos test de la Modalidad A y los dos test de la Modalidad B. Entre las dos modalidades de la primera modalidad, PRECU y CRECO, hay una diferencia notable en sus resultados, de más de diez puntos. Sin embargo, entre las dos de la Modalidad B, PRECO y CRECU, la diferencia es mínima, apenas de dos puntos. De todas formas, la desviación estándar no es tan diferente en las cuatro modalidades del test lo que indica que los participantes representarían una población bastante homogénea.

	PRECU	CRECO	CC-A
Validos	51	51	51
Media	15.137	28.118	22.157
Desviación est.	5.892	6.970	6.192
Varianza	34.7	48.6	37,6
Mínimo	3.000	10.000	8.000
Máximo	26.000	36.000	31.000

Tabla 70. Resultados del TCE en la Modalidad A

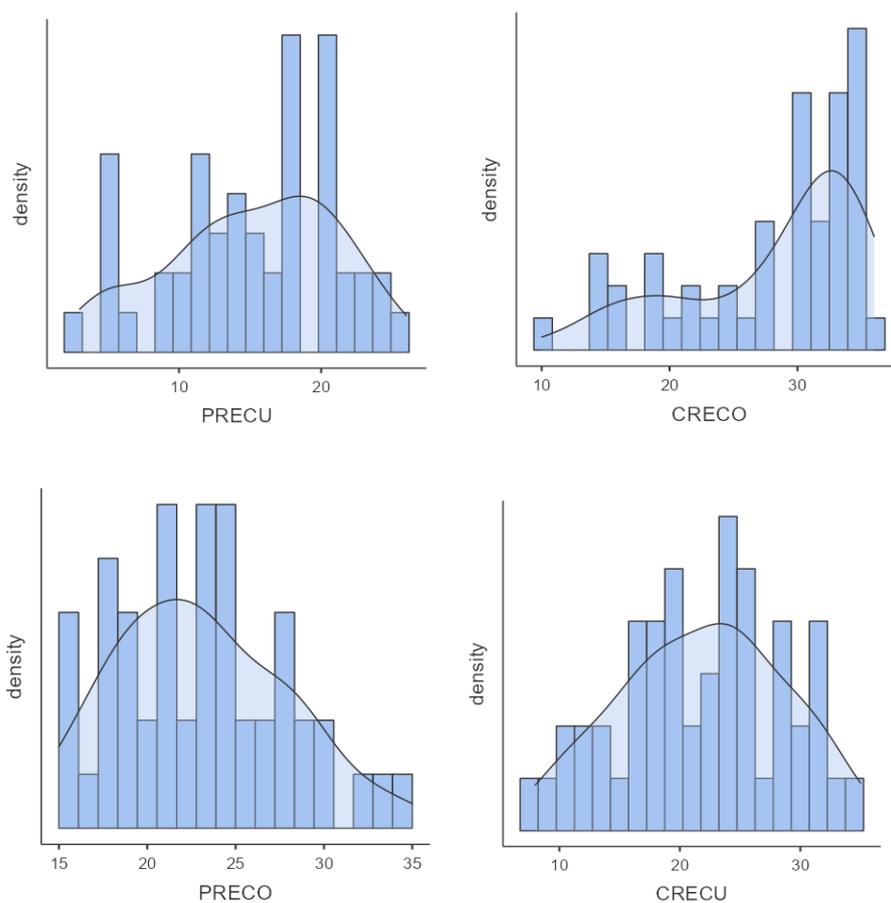
	PRECO	CRECU	CC-B
Validos	51	51	51
Media	23.039	21.843	22.686
Desviación est.	4.837	6.709	4.714
Varianza	23.4	45.0	21.7
Mínimo	15.000	8.000	14.000
Máximo	35.000	35.000	31.000

Tabla 71: Resultados del TCE en la Modalidad B

La distribución de los resultados del TCE no es normal, lo que condiciona la elección de los diferentes métodos estadísticos. En el análisis de la normalidad hemos utilizado los tres métodos habituales: método analítico, método de contraste de hipótesis y método gráfico. Según el método analítico, calculamos la asimetría y la curtosis, como vemos en la *Tabla 75* y comprobamos que no es nula. Según el método de contraste, el valor de *Shapiro-Wilk* está por debajo de 0,05. Asimismo, en los histogramas, siguiendo el método gráfico, también podemos observar el alejamiento de la normalidad. (Vid. Gráfica 2): los resultados de CRECO muestran claramente asimetría hacia la izquierda (negativa), PRECO parece mostrar una asimetría hacia la derecha (positiva). Esto quiere decir que los resultados son más negativos de los esperable en PRECO y más positivos de lo esperable en CRECO. Esto viene condicionado por la poca variedad en el nivel de competencia lingüística de los participantes, recordemos que todos pertenecen al nivel intermedio, pero que, además, viene a demostrar que los formatos que implican un proceso cognitivo de producción tienen una distribución normal y no la tienen los que implican un proceso cognitivo de reconocimiento.

	PRECU	CRECO	PRECO	CRECU
N	51	51	51	51
Perdidos	0	0	0	0
Asimetría	-0.328	-0.977	0.380	-0.115
Error est. asimetría	0.333	0.333	0.333	0.333
Curtosis	-0.730	-0.229	-0.389	-0.703
Error est. curtosis	0.656	0.656	0.656	0.656
W de Shapiro-Wilk	0.961	0.860	0.974	0.982
Valor p de Shapiro-Wilk	0.092	<.001	0.315	0.615

Tabla 72. Análisis de la normalidad de las cuatro modalidades del TCE



Gráfica 3. Gráficos de dispersión de las cuatro modalidades: PRECU, CRECO, PRECO y CRECU

En las siguientes tablas representamos el análisis estadístico descriptivo de la muestra de forma más específica, mostrando los resultados también según nivel de competencia lingüística asignado a los aprendices que participan en el estudio. Se muestran los resultados en dos tablas diferentes: en una los resultados de la Modalidad A (*Tabla 84*) y en otra los de la Modalidad B (*Tabla 85*); lo hacemos así porque cada participante solo ha realizado dos test, tal y como explicamos previamente (vid. 4.3.3.3). La comparativa entre niveles, como cabía esperar, nos permite comprobar que los

resultados siempre son mejores en el nivel más alto tanto en una modalidad como en la otra, en cualquiera de los cuatro test. De modo que aparentemente se cumplen totalmente las expectativas: la competencia colocacional es reflejo de la competencia lingüística. Es decir, los resultados del nivel B2 superan a los del nivel B1 en todas las secciones, confirmando así las expectativas y en consonancia con los resultados de otros estudios en otras lenguas que observan la misma tendencia (Brown, 2018). Además, el margen de diferencia entre los resultados de un nivel y de otro en cada una de las pruebas oscila entre un 4,233 y un 5,357, con una media de cinco aciertos (4,8) de diferencia; lo que demuestra que la diferencia entre uno y otro nivel es constante.

	PRECU		CRECU		CC-A	
	B1	B2	B1	B2	B1	B2
Válidos	37	14	37	14	37	14
Media	13.865	18.500	26.649	32.000	20.865	25.571
Desviación est.	6.111	3.653	7.554	2.542	6.516	3.546
Mínimo	3.000	12.000	10.000	27.000	8.000	18.000
Máximo	24.000	26.000	35.000	36.000	29.000	31.000

Tabla 73: Estadística descriptiva de la Modalidad A del TCE

	PRECU		CRECU		CC-B	
	B1	B2	B1	B2	B1	B2
Válidos	30	21	30	21	30	21
Media	20.833	26.190	20.100	24.333	20.733	25.476
Desviación est.	3.630	4.654	5.950	7.080	4.218	3.983
Mínimo	15.000	18.000	8.000	9.000	14.000	16.000
Máximo	28.000	35.000	32.000	35.000	29.000	31.000

Tabla 74. Estadística descriptiva de la Modalidad B del TCE

5.2.3.3. *Análisis del ítem en el TCE*

La primera pregunta de investigación de este trabajo plantea si el TCE¹¹⁶ es una herramienta de evaluación apropiada para medir la competencia y el conocimiento colocacional. Antes de iniciar el análisis pormenorizado de los resultados es necesario tener claras las características del test y establecer los tipos de variables que se van a

¹¹⁶ La población ideal del TCE sería variada, heterogénea, con relación al nivel, lo que provocaría una distribución no-normal de los resultados y, por lo tanto, no apta para TCE. Como la población real del TCE que aquí se administra está toda nivelada en el grupo intermedio, entre B1-B2, la Teoría clásica del Test nos puede ser de utilidad, aunque, como hemos apuntado, también está condicionada por la ausencia de ítems totalmente paralelos y de test paralelos. Recordemos que los test paralelos son los que miden el mismo constructo y, en este caso, miden las mismas colocaciones meta, la misma muestra, pero desde dos perspectivas diferentes, es decir, miden constructos diferentes.

tener en cuenta para elegir las medidas estadísticas adecuadas. El TCE es un test basado en criterio, por lo tanto el tratamiento de la fiabilidad requiere la utilización de unos métodos específicos (Barrios & Coscolluela, 2013; Dean-Brown, 2012; Ockey, 2012), relegando los dos enfoques básicos de estimación de la fiabilidad: la *Teoría Clásica del Test* (TCT) y la *Teoría de Respuesta al ítem* (TRI), así como la estimación de fiabilidad de la escala. En el capítulo tres (vid. 3.1.5) ya expusimos las diferencias, ventajas y condiciones de cada una de ellas. Pese a las restricciones, al igual que otros test similares, recurriremos a la TCT y a la TRI para estimar la fiabilidad de la herramienta y obtener información útil para la fiabilidad del TCE, aunque no determinante.

El análisis del ítem supone un punto imprescindible para probar la fiabilidad y consistencia del test. Comenzaremos este análisis desde la TCE para, posteriormente, hacer un análisis del *modelo Rasch*. En cuanto al análisis del ítem, utilizaremos tres procedimientos: el índice de dificultad / facilidad, el índice de discriminación, pasos claves en su observación (Aryadoust & Raquel, 2020).

El *índice de facilidad* (IF) o *índice de dificultad* (ID) mide la proporción de aciertos para cada ítem (vid. 3.1.5.1). Se considera un ítem adecuado cuando su índice oscila entre 30% y 70% de respuestas. Este es uno de los problemas con respecto a los test como este, basado en criterio, en el que toda la población se clasifica en un nivel según su competencia, intermedio, y las preguntas están ordenadas en una escala de dificultad desde el nivel inicial al avanzado. Según los criterios de selección de la muestra, se crearon seis bandas, una por cada nivel del MCER, y el test se dividió también en seis secciones, cada una de las cuales consta de seis colocaciones meta clasificadas según uno de los niveles del *Marco*. Hemos calculado el índice de dificultad¹¹⁷ de todos los ítems de cada uno de los test (vid. Tabla 76). Sin embargo, estos resultados aportan una información sesgada ya que no tienen en cuenta el carácter escalar de la muestra, de los ítems y del test. Según la estimación del IF, consideraríamos los ítems de PRECU como difíciles y CRECO, PRECO y CRECU como fáciles.

¹¹⁷ Para su interpretación se suelen emplear cinco categorías: muy fáciles (<0,75), fácil (entre 0,55 y 0,75), normal (entre 0,45 y 0,54), difícil (entre 0,35 y 0,44) y muy difíciles (>0,25).

TCE			
MODALIDAD A		MODALIDAD B	
PRECU	CRECO	PRECO	CRECU
0,34	0,66	0,62	0,64

Tabla 75. Índice de facilidad (IF) en cada una de las modalidades del TCE

Según estos resultados, comprobamos cómo el test que les resulta más difícil a los aprendices de español es PRECU, el más fácil sería CRECO; ambos realizamos con la misma modalidad. Y, aunque con una diferencia mínima, PRECO, parece más sencillo que CRECU. Sin embargo, la diferencia entre CRECO, PRECO y CRECU es mínima y, por lo tanto, poco significativa. Para una observación más detallada hemos desglosado el índice de facilidad de cada una de las secciones de los cuatro test (vid. Tabla 77) y estos datos nos aportan información relevante al contar con aprendices únicamente de nivel intermedio. El índice de dificultad más bajo es el del grupo de ítems del nivel más alto (del ítem 25 al 36) en los cuatro test, lo que indica que al aumentar el nivel de la sección aumenta su dificultad para los aprendices, aunque en cada formato en un grado distinto. Además, en la Modalidad A los índices más altos se corresponden con los ítems en el nivel inferior (del ítem 1 al ítem 6) y así sucesivamente en una escala continua. En la Modalidad B no ocurre exactamente lo mismo, el índice del grupo A2 es el que altera la escala que salvo por esta excepción sigue, al igual que en la otra modalidad, la sucesión esperada. Pese a esto, determinar si un ítem es apropiado o no según su índice de facilidad o dificultad depende de varios factores, entre ellos el tipo de test, el tipo de población o las condiciones de administración (Green, 2013, p. 27). Por ejemplo, en test de nivel un valor muy bajo en el índice indicará que los participantes no alcanzan el nivel de ese ítem de forma general.

	TCE			
	Modalidad A		Modalidad B	
	PRECU	CRECO	PRECO	CRECU
A1	0,54	0,78	0,82	0,79
A2	0,52	0,72	0,88	0,86
B1	0,37	0,69	0,65	0,60
B2	0,32	0,66	0,61	0,60
C1	0,17	0,61	0,52	0,57
C2	0,10	0,48	0,26	0,44

Tabla 76: IF por niveles de cada una de las modalidades del TCE

Continuamos el análisis del ítem con el *índice de discriminación*¹¹⁸ que nos permiten distinguir la capacidad que tiene un ítem para discernir entre un participante con un buen rendimiento en el test (Barrios et al., 2013). El índice se considera bajo en muchos de los ítems. En el caso de los resultados del TCE, el índice de discriminación no es el valor más adecuado para corroborar la fiabilidad del ítem y del test. Hay ítems demasiado fáciles e ítems demasiado difíciles, lo que reduce la fiabilidad interna ya que la mayoría de los participantes responde correctamente a los mismo ítems e incorrectamente también a los mismos (Green, 2013, p. 39). Varias razones pueden reducir la fiabilidad y el TCE cumple dos de ellas. Primero, el test está diseñado para asignar un nivel a los aprendices, contiene ítems de diferente dificultad, y los participantes únicamente pertenecen a nivel intermedio, tanto de B1 como de B2. Segundo, el hecho de que la población se concentre en dos niveles cuando el test se estructura en seis es una de causas de la falta de adecuación a los estándares del índice de discriminación. Su poca variación, igual que pasa en los test de nivel o de logro, hace que el índice sea previsiblemente bajo. Por lo tanto, el rango de la competencia de los aprendices es otro de los factores que altera el índice de discriminación y, en consecuencia, afecta a los niveles de fiabilidad.

Haremos una estimación de la fiabilidad para lo que se suele utilizar el α de *Cronbach*, pero para lo que también es común usar la ω de *McDonald's*, que tiende a sustituir a la anterior cuando las herramientas de evaluación no cumplen las condiciones para aplicarla. *Cronbach* se caracteriza por subestimar la medida cuando las opciones de respuesta son inferiores de cinco; en nuestro caso, alguno de los formatos es de respuesta libre. Las diferencias entre el índice de omega y alfa en cada uno de los formatos utilizados son mínimas, lo que aumenta la fiabilidad de los resultados. El alfa se considera fiable si supera el valor de 0,70 (Devellis, 2012) y el índice en todos los formatos del TCE lo supera, por lo tanto, podemos afirmar que son fiables. El formato más consistente es CRECO y el menos PRECO, en el que hay una mayor variabilidad entre las respuestas. Mientras CRECO se consideraría excelente, PRECO únicamente se podría

¹¹⁸ Valoraciones sobre el índice de discriminación: .40 o superior, muy buen ítem; entre .30 y .40, razonablemente buen ítem; entre .20 y .20, ítem marginal; .19 o menos, ítem pobre (Green, 2013).

considerar respetable¹¹⁹. Esto puede ser debido a diversos factores, uno de ellos la mayor o menor discriminación de los distractores. Podemos afirmar que los resultados de cada uno de los formatos del TCE muestran consistencia interna.

	McDonald's ω	Cronbach's α
PRECO	0.756	0.746
CRECU	0.847	0.837
PRECU	0.844	0.837
CRECO	0.962	0.960

Tabla 77. Fiabilidad de cada una de las opciones

5.2.3.4. *La competencia colocacional y el MCER*

En primer lugar, observaremos cómo los resultados del TCE ofrecen un índice de la CCol y cómo este se puede clasificar en niveles del MCER según los estándares previamente establecidos (vid. 4.3.3.2). En segundo lugar, compararemos el índice de la CCol con el índice de la CLéx, que hemos obtenido con el análisis del resultado de X-LEX, y con el de la CLing, índice que viene dado desde los centros que participaron en este estudio. Esto nos permitirá conocer si el TCE puede clasificar a los aprendices en un nivel de lengua, es decir, si la CCol es válida como índice de la competencia lingüística.

a) Clasificación de los resultados del TCE en niveles de CCol

Nos centraremos en la clasificación de la competencia colocacional de los aprendices en un nivel de MCER. En el punto 4.3.3.2 ya hemos explicado los criterios de clasificación que utilizamos y en este punto desarrollaremos la asignación de nivel y realizaremos un análisis de su fiabilidad a partir de diferentes métodos. En la *Tabla 79* podemos ver la distribución de aprendices en niveles según la modalidad realizada (CCol-A y CCol-B) y según la sección o formato (PRECU o CRECO y PRECO o CRECU).

	Modalidad A (CCol-A)	Modalidad B (CCol-B)
A1	4%	0%
A2	10%	0%
B1	12%	18%
B2	49%	51%
C1	24%	25%
C2	2%	6%

Tabla 78. Porcentaje de aprendices clasificados en un nivel del MCER según su CCOL

¹¹⁹) Tomamos la siguiente interpretación de los valores de Alpha: <0,60 es inaceptable, entre 0,60 y 0,65 es indeseable, entre 0,65 y 0,75 es mínimamente aceptable, entre 0,70 y 0,80 es respetable, entre 0,80 y 0,90 es muy buena y >0,90 es excelente (Devellis, 2012).

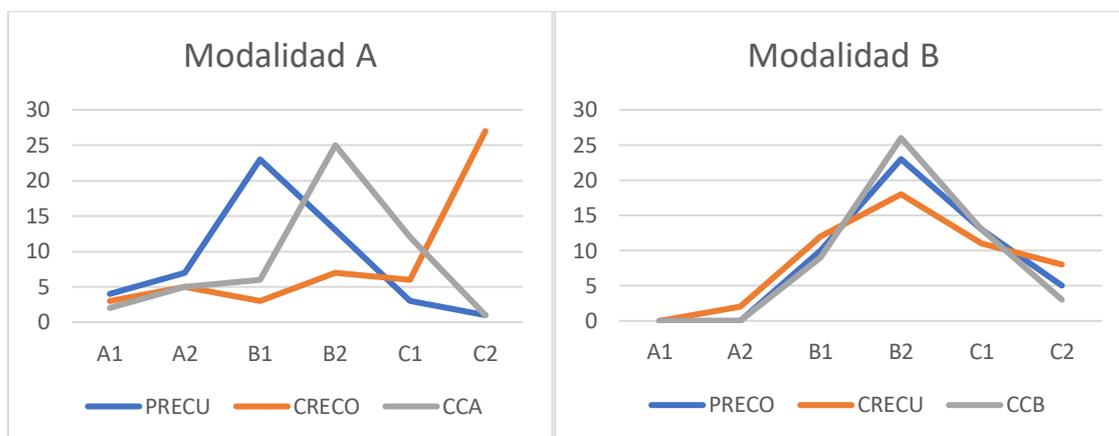
Pese a que todos los aprendices están asignados en un nivel intermedio de competencia lingüística, según su CCol, los niveles en los que se clasifica un mayor número de aprendices es en B2 y en C1, y algunos también se adscriben al nivel A1. Por lo tanto, podríamos afirmar que la competencia colocacional está ligeramente por encima de la competencia lingüística, lo que puede explicarse de diferentes maneras: porque la CCol no es un buen índice de la CLing y que, entonces, no podrá funcionar como índice de la CLéx o quizás sea un indicio de que la asignación de la CLing no tiene en cuenta de forma suficiente la CCol en su evaluación. Además, estas conclusiones se pueden extraer tanto de los resultados de la CCol en la Modalidad A como en la Modalidad B. Bien es cierto que en la Modalidad A los resultados son más dispersos, clasificando a aprendices en todos los niveles, aunque con un porcentaje de participantes muy bajo en C2. De forma contraria, en la Modalidad B la clasificación es más estable, no clasifica a ninguno de los aprendices en los niveles iniciales.

Si nos detenemos en la clasificación atendiendo a cada una de las modalidades, vemos que en PRECU la mayoría de los aprendices se clasifican en el nivel B1 y B2, un 70% del total, y que se correspondería con lo esperado. Por el contrario, CRECO clasifica a casi todos los aprendices en C2 (como podemos ver en la *Tabla 80*), con lo que dista mucho del nivel de referencia, lo que no hace más que demostrar que la competencia colocacional desde la comprensión no supone un reto para los aprendices de español como L2 y pudiera sobreestimar su competencia lingüística; solo el 20% de la población se clasificaría en un nivel intermedio con este formato. En PRECO y CRECU, aunque un 35% y un 38% se adscriben a un nivel avanzado, el 65% y el 59% respectivamente se clasifica de acuerdo con el nivel de referencia. Parece que las secciones que más relación mantienen con los niveles esperados son los que evalúan la producción.

		PRECU		CRECO		PRECO		CRECU	
Inicial	A1	8%	22%	6%	16%	0%	0%	0%	4%
	A2	14%		10%		0%		4%	
Intermedio	B1	45%	70%	6%	20%	20%	65%	24%	59%
	B2	25%		14%		45%		35%	
Avanzado	C1	6%	8%	12%	65%	25%	35%	22%	38%
	C2	2%		53%		10%		16%	

Tabla 79. Porcentaje de aprendices clasificados en un nivel del MCER según cada una de las modalidades del TCE

De forma más clara (vid. Gráfica 3) podemos observar en las siguientes gráficas la mayor homogeneidad en la Modalidad B frente a los de la Modalidad A, en la que podríamos afirmar que clasifican a los aprendices de forma opuesta.



Gráfica 4. Comparación de los niveles del MCER en las diferentes modalidades

Para comprobar que estas apreciaciones son significativas y generalizables y que el test está dotado de fiabilidad, calcularemos *el coeficiente de contingencia* (Barrios & Cosculluela, 2013). Así obtendremos la consistencia de las clasificaciones a partir de dos formas homólogas del test. El p (valor de contingencia) de los test de la Modalidad A es $p_{H-N} = 0,566$, lo que indica que el uso de los test permite mejorar más de un 50% la clasificación de los sujetos de la que se realizaría por azar. En la Tabla 81 podemos observar como el grupo más numeroso es el de aprendices que se clasifican en B1 atendiendo a los resultados de PRECU y en C2 atendiendo a los resultados de CRECO.

	CRECO						
PRECU	A1	A2	B1	B2	C1	C2	Total
A1	1	2	0	0	0	1	4
A2	0	1	1	1	0	4	7
B1	1	1	2	3	3	13	23
B2	0	1	0	2	3	7	13
C1	1	0	0	1	0	1	3
C2	0	0	0	0	0	1	1
Total	3	5	3	7	6	27	51

Tabla 80. Tabla de contingencia de las dos modalidades de la Modalidad A

Si comparamos los dos test paralelos de la Modalidad B el resultado es muy similar $p_{H-N} = 0,600$; es decir, el resultado indica que el uso de los test permite mejorar un 60% la clasificación de los sujetos de la que se realizaría por azar. En la *Tabla 82* podemos observar que ningún aprendiz se clasifica según los resultados obtenidos en PRECO en el nivel inicial y, según CRECU, ningún en el A1. El grupo mayoritario pertenece a alumnos

que se clasificaron en PRECO y en CRECU en un nivel B2. Lo que demuestra la homogeneidad de la modalidad.

	CRECU					
PRECO	A2	B1	B2	C1	C2	Total
B1	0	7	1	1	1	10
B2	1	3	13	4	2	23
C1	1	1	3	3	5	13
C2	0	1	1	3	0	5
Tota	2	12	18	11	8	51

Tabla 81. Tabla de contingencia de las dos modalidades de la Modalidad B

El índice de contingencia de la CCol-A y CCol-B es de 0,589, casi un 60% de los participantes no habrían sido clasificados por el azar. En la *Tabla 83* podemos ver la tabla de contingencia de forma global, es decir, atendiendo a la CCol en las dos modalidades. Como podemos observar la mayor parte de los aprendices se clasifica en B2 según una y otra modalidad.

	NIVEL según la CCol-B				
NIVEL según la CCol-A	B1	B2	C1	C2	Total
A1	0	2	0	0	2
A2	2	2	1	0	5
B1	2	1	3	0	6
B2	4	12	7	2	25
C1	1	9	2	0	12
C2	0	0	0	1	1
Total	9	26	13	3	51

Tabla 82. Tabla de contingencia

b) Relación entre variables: CLing y CLéx vs. CCol

Las variables que queremos comparar, con la intención de descubrir el tipo de relación existente entre ellas, son la competencia lingüística (CLing), la competencia léxica (CLéx) y la competencia colocacional (CCol). El primer paso es contar con un índice para cada una de las competencias que se van a correlacionar y definir el tipo de variable. Como hemos apuntado al describir la población del TCE (vid. 5.1) contamos con 67 estudiantes cuyo nivel asignado es B1 y 35 estudiantes cuyo nivel asignado es B2. Tenemos dos grupos pertenecientes a una escala mayor¹²⁰, por lo tanto, se trata de una

¹²⁰ Con escala mayor se hace referencia a una escala que incluye más de dos niveles (B1 y B2), se trata de una escala de seis fases (A1, A2, B1, B2, C1 y C2).

variable ordinal. Para obtener la competencia léxica, se analizaron los datos extraídos de la versión en español de X-LEX (Meara & Milton, 2003), de los que hemos hecho un análisis previo (vid. 5.2.3.1). Para esta variable tenemos dos opciones, presentarla como una variable ordinal, indicando el nivel del *Marco*, o como una variable de intervalo, teniendo en cuenta el resultado observado del test. El tercer paso es identificar la competencia colocacional, uno de los objetivos de este estudio y, por lo tanto, del TCE. En este caso, tenemos dos resultados de la competencia colocacional: de los participantes que realizaron la *Modalidad A* y de los participantes que realizaron la *Modalidad B*. También podríamos usar la variable ordinal y la de intervalo, en ambos casos nos decantamos por la segunda modalidad. Sería interesante, para tener una visión totalmente global, contar con las cuatro perspectivas, pero la imposibilidad de que un mismo grupo de participantes realizara los cuatro formatos nos obliga a calcular la competencia de esta manera.

A continuación, pasamos a establecer las correlaciones, por un lado, de la CLéx con la CCol y, por el otro, de la CLing con la CCol. Contamos con varios antecedentes de correlación entre la competencia léxica y la competencia colocacional del inglés, por ejemplo, tanto COLLEX como COLLMATCH presentan una correlación positiva con el test de vocabulario SINGLEX (Gyllstad, 2009a), concluyendo así que existe un relación biunívoca entre la competencia léxica y la competencia colocacional desde la comprensión. Incluso se establece una relación entre la competencia oral (en oposición a la escrita) y la capacidad para identificar combinaciones de palabras como colocaciones (Eyckmans, 2009). Es importante señalar que SINGLEX (Gyllstad, 2007b) tiene como vocabulario meta¹²¹ el que aparece en las colocaciones meta de COLLEX y COLLMATCH, lo que no pasa en el caso de X-LEX y el TCE. En la misma línea, en otros trabajos también se comprueba la correlación entre la competencia colocacional y la competencia léxica (Nizonkiza, 2015; Nizonkiza et al., 2014). Sin embargo, algunos detectan ausencia de relación entre la competencia colocacional y la competencia léxica (Brown, 2014); competencia asociada a partir de la correlación entre el último estudio de *LexCombi* y un test de identificación o reconocimiento de colocaciones (*sí o no*) como base para el

¹²¹ En SINGLEX se usan como palabras meta las bases de las colocaciones que se evalúan en COLLEX y en COLLMACH.

establecimiento entre la competencia léxica y la lingüística. Los estudios en los que se observa una correlación entre la competencia colocacional y el léxico suelen usar test de reconocimiento; mientras que no observan correlación los que usan formatos de test basados en la producción. Esto nos da una clave para interpretar la relación de la competencia léxica con cada una de las secciones del test.

En la misma línea, en algunos estudios de componentes múltiples (González-Fernández & Schmitt, 2020; Ishii & Schmitt, 2009; Pellicer-Sánchez, 2014), que incluyen las colocaciones como uno de esos componentes, la relación que se establece entre la colocaciones y el resto del componentes del léxico no es la más estrecha. A pesar de no ser uno de los componentes que resulta más complejo para los aprendices (aunque con variantes en función de si se mide desde la comprensión o producción) la correlación con la competencia léxica general es baja.

Para obtener el coeficiente de correlación¹²² entre la CCol y la Cléx utilizamos la medida de *Spearman* ya que estamos ante variables de intervalo y al menos una de las variables no sigue una distribución normal. La correlación entre la CCol y la CLéx es débil y muy similar en ambas opciones (Rho de 0,361 y 0,319). Al observar la correlación entre la CLéx y cada una de las modalidades de forma individual la correlación sigue siendo muy baja, aunque, en función de las escalas utilizadas, la correlación entre PRECO y la CLéx podría considerarse casi moderada (0,438). Esta diferencia con respecto a las otras modalidades podría deberse a la similitud entre este formato y el utilizado en X-LEX.

		CLEX	PRECU	CRECO	CC-A
CLEX	Rho de Spearman	—			
	valor p	—			
PRECU	Rho de Spearman	0.165	—		
	valor p	0.247	—		
CRECO	Rho de Spearman	0.330	0.734	—	
	valor p	0.018	< .001	—	
CCol-A	Rho de Spearman	0.361	0.302	0.400	—
	valor p	0.009	0.031	0.004	—

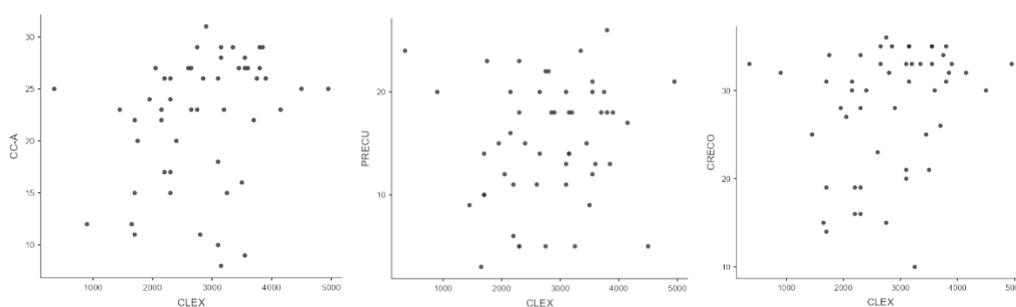
Tabla 83. Matriz de correlaciones en la Modalidad A del TCE

¹²² En la mayoría de las ocasiones se suelen aceptar las interpretaciones siguientes del coeficiente de correlación: a) muy baja (entre 0,00 y 0,20), b) baja (entre 0,21 y 0,40), c) media (0,41 y 0,70), alta (entre 0,71 y 0,90) y muy alta entre (0,91 y 1)

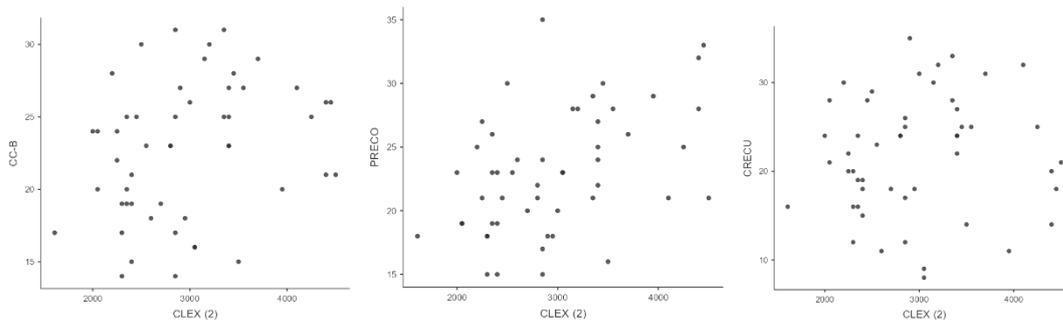
		CLEX (2)	PRECO	CRECU	CC-B
CLEX (2)	Rho de Spearman	—			
	valor p	—			
PRECO	Rho de Spearman	0.438	—		
	valor p	0.001	—		
CRECU	Rho de Spearman	0.132	0.328	—	
	valor p	0.358	0.019	—	
CCol-B	Rho de Spearman	0.319	0.709	0.882	—
	valor p	0.023	<.001	<.001	—

Tabla 84. Matriz de correlaciones en la Modalidad B del TCE

En conclusión, como X-LEX es un test que mide la CLéx desde el reconocimiento del conocimiento (ni desde la producción ni la comprensión), la modalidad que mantiene un índice de correlación R más alto, en este caso moderado, con el TCE es PRECO, una de las modalidades que también mide desde el reconocimiento y que, por lo tanto, situamos próxima cognitivamente. Quizás la suma del resultado de las cuatro secciones sí mantuviera una relación estable positiva con la CLéx, pero no podemos medirlo de forma conjunta al haber sido realizadas por participantes diferentes. También se podría haber medido la competencia léxica desde un punto de vista más amplio buscando una imagen global del conocimiento léxico que tuviera en cuenta un proceso cognitivo en el diseño de sus formatos. En los siguientes gráficos de dispersión comprobamos como la relación entre la CCol (ya sea en una u otra modalidad) y cada una de las cuatro modalidades es baja o nula.



Gráfica 5. de dispersión de la CL con la CC-A y con cada una de las modalidades (PRECO y CRECU)



Gráfica 6. dispersión de la CL con la CC-B y con cada una de las podalidades (PRECO y CRECU)

Para analizar la relación existente entre la competencia colocacional en sus dos modalidades y la competencia lingüística fue necesario hacer una regresión logística binomial. Este coeficiente se utiliza cuando las dos variables son de diferente naturaleza. En este caso, la CLing es ordinal de carácter dicotómico (solo contamos con dos niveles de aprendices, B1 y B2) y otra es una variable de intervalo. En ambas modalidades el coeficiente es muy bajo (0,205 en la Modalidad A) o bajo (0,329 en la Modalidad B), con lo cual, no podemos establecer correlación entre la CLing y la CCol.

Modelo	Desviación	AIC	R ² _{MCF}
1	47.7	55.7	0.205

Tabla 85. Medidas de ajuste del Modelo de regresión logística binomial

Los resultados de este estudio muestran que la correlación entre las tres competencias es muy débil. Sin embargo, según los criterios establecidos, el conocimiento colocacional sí se corresponde con el valor esperado. Esto se puede deber a varias razones. La competencia colocacional no se puede considerar índice de la competencia léxica por su complejidad y la diferencia entre los resultados en función de la perspectiva desde la que se analice. Si lo miramos desde un punto de la comprensión y el reconocimiento, las colocaciones son más fáciles y su índice puede ser superior al de la competencia léxica y lingüística; en gran medida por el proceso inferencial, ya sea en relación con la L1 o con la L2. Si lo miramos desde el punto de la producción y la recuperación, la competencia colocacional es difícil, se requiere un conocimiento mayor que el adquirido o demostrado en la competencia léxica o la competencia lingüística tal y como se suelen evaluar. Por el contrario, la comprensión desde la recuperación y la producción desde el reconocimiento podría ser más representativo y quizás índice de la competencia léxica o lingüística. Para comprobarlo habría que revisar el test y administrar de nuevo con la combinación de ambas modalidades.

5.2.3.5. *Dimensiones de la competencia colocacional*

Entendemos por dimensiones de la competencia colocacional cada una de las modalidades del TCE (PRECU, CRECO, PRECO y CRECU) que representan un modo de acceder al conocimiento de la colocación: 1) la producción desde la recuperación, 2) la comprensión desde el reconocimiento, 3) la producción desde el reconocimiento y 4) la comprensión desde la recuperación. Para analizar estas dimensiones, recuperamos las observaciones y consideraciones esbozadas en el análisis descriptivo y comprobamos cómo al comparar la media de los resultados obtenidos en cada una de las pruebas (PRECU, CRECO, PRECO y CRECU) unas modalidades suponen un obstáculo para los aprendices y otras parecen no presentar ningún reto. Según los resultados expuestos previamente (vid. 5.2.3.2) el test que supone un mayor reto para los aprendices es PRECU (primero de la Modalidad A), con un índice de 15.137, y el que supone un menor desafío es CRECO (el otro formato de la Modalidad A), con un índice de 28.118. El número de aciertos es similar en los dos test de la Modalidad B: 23.039 en PRECO y 21.843 en CRECU. Sin embargo, si observamos el global de la modalidad, es decir, el resultado de la CCol en cada una de las Modalidades los resultados se igual bastante: con la Modalidad A los aprendices obtiene un resultado medio de 22.157 y con la Modalidad B un resultado de 22.686.

Por lo tanto, la prueba en la que los aprendices obtienen mejores resultados es en CRECO, es decir, el test que mide la comprensión desde el reconocimiento, variables que tradicionalmente se han considerado las más sencillas (Szudarski & Carter, 2016; Zhong, 2018). En el test en el que peores resultados han obtenido es en PRECU, el test que mide la producción desde la recuperación, que, tradicionalmente, se ha considerado el que supone un mayor reto para los aprendices (Durrant et al., 2022, pp. 148-149; Zhong, 2018). Estos resultados no distan de los obtenidos en otros test que miden el conocimiento léxico (Aviad-Levitzky et al., 2019; Laufer & Goldstein, 2004). También en el ámbito del español, en un estudio que tiene por objetivo comprobar la utilidad de las listas de palabras en la presentación del léxico a los aprendices, se utiliza un test con las mismas cuatro dimensiones que utilizamos en este test con un resultado similar en el que la prueba que recoge el reconocimiento del significado es la más sencilla y la recuperación de la forma la más compleja (Hidalgo Gallardo, 2020).

Sin embargo, la diferencia entre las otras dos secciones, PRECO y CRECU, es mínima; lo que dificulta el análisis de la modalidad, ya que se debe comprobar si la diferencia es significativa. Además, esto supone un reto para comprobar cuál de los criterios que valoramos al medir la competencia colocacional tiene más peso o condiciona más la dificultad de las colocaciones. En el TCE no contamos con dos formas paralelas, sino homólogas, en el sentido de que miden el mismo propósito desde diferentes puntos de vista, esto es, las colocaciones meta son las mismas, pero el constructo es diferente: en uno se mide el conocimiento colocacional desde la producción y la recuperación; en otro, desde la comprensión y el reconocimiento¹²³. Por esto, no es posible que funcionen varias medidas, al igual que pasa en otros test, como en COLLEX y COLLMACH (Gyllstad, 2007b).

Si ordenamos las modalidades desde la que supone una mayor dificultad a la que entraña menor dificultad nos encontramos con la siguiente escala: CRECO > PRECO > CRECU > PRECU (vid. Figura 52), mostrando así la imagen global de la competencia colocacional escalar y jerarquizada, igual que en CATSS (Aviad-Levitzky et al., 2019; Laufer & Goldstein, 2004). En un extremo se sitúa la comprensión desde el reconocimiento (CRECO) y en el otro la producción desde la recuperación (PRECU). Asimismo, las dos versiones que miden el reconocimiento son las más sencillas para los aprendices y las que miden la recuperación las más difíciles. Es importante destacar que se establece el mismo orden jerárquico en el test de léxico CATSS (Aviad-Levitzky et al., 2019; Laufer & Goldstein, 2004) y con los resultados expuestos en la versión previa de pilotaje, TCEp. Lo que puede indicar que tanto la competencia colocacional como la competencia léxica siguen un mismo patrón de dominio. Además, la diferencia entre PRECO y CRECU es también mínima, manteniendo la distancia con las otras dos modalidades que se sitúan a los extremos en la escala jerárquica del conocimiento colocacional¹²⁴. La diferencia entre la competencia colocacional obtenida de la Modalidad A (CCol-A y CCol-B), que se ha calculado a partir de la media de los resultados de los dos formatos realizados en cada

¹²³ Esto en la Modalidad A. En la Modalidad B se mide de forma opuesta, con la combinación de criterios alternativa.

¹²⁴ En el preTCEp PRECO y PRECU alteran el orden.

una de las modalidades¹²⁵, es mínima. El alto resultado en CRECO viene compensado por el bajo resultado en PRECU, mientras que PRECO y CRECU mantienen unos valores similares.

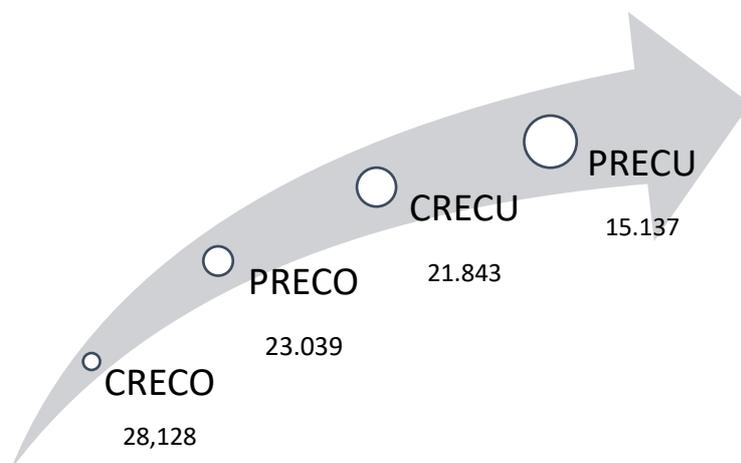


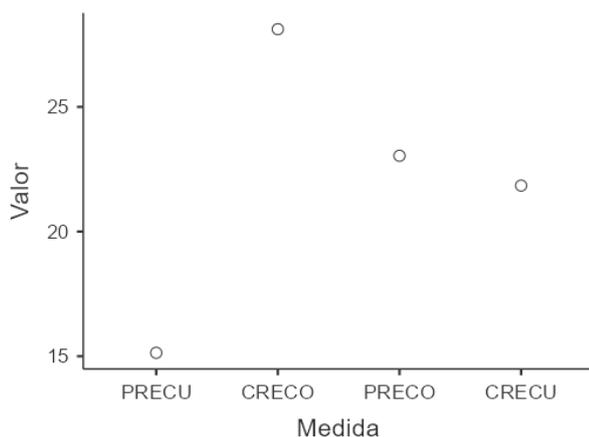
Figura 52. Imagen global y escalar de la competencia colocacional

Para comprobar que las cuatro modalidades del test representan diferentes grados del conocimiento colocacional debemos demostrar que la diferencia de la media de los resultados de las cuatro modalidades es significativa. Para este fin realizamos el análisis de la varianza a partir del estadístico Friedman de ANOVA ya que la distribución de las variables, como ya avanzamos, no es normal. Comprobamos que sí existen diferencias significativas entre al menos dos variables, ya que la p es de 0,001 (menor, por lo tanto, de 0,05). Como también queremos saber cuáles muestran una diferencia significativa entre sí, las analizamos por pares y en la *Tabla 90* podemos ver los resultados. La comparación entre parejas (Durbin-Conover) demuestra que la relación entre PRECO y CRECU no es significativa y, por lo tanto, resulta muy difícil generalizar qué dimensión supone una mayor dificultad para los aprendices. Finalmente, en la gráfica podemos observar la significación de forma más visual al comprobar como la distancia entre el valor de PRECU y CRECU es mucho mayor que la distancia entre el valor de PRECO y CRECU (vid. Gráfica 6).

¹²⁵ En la Modalidad A se ha calculado la media entre PRECU y CRECO y en la Modalidad B, entre PRECO y CRECU. En un primer momento en calcular la competencia colocacional a partir de la media de los cuatro formatos, pero el hecho de que participantes diferentes hayan realizado cada test, lo hace inviable.

			Estadístico	p
PRECU	-	CRECO	12.18	<.001
PRECU	-	PRECO	7.35	<.001
PRECU	-	CRECU	5.93	<.001
CRECO	-	PRECO	4.83	<.001
CRECO	-	CRECU	6.26	<.001
PRECO	-	CRECU	1.43	0.156

Tabla 86. Comparaciones Entre Parejas (Durbin-Conover)



Gráfica 7 descriptiva de las dimensiones de la competencia colocacional del TCE

5.2.3.6. Grados del conocimiento colocacional

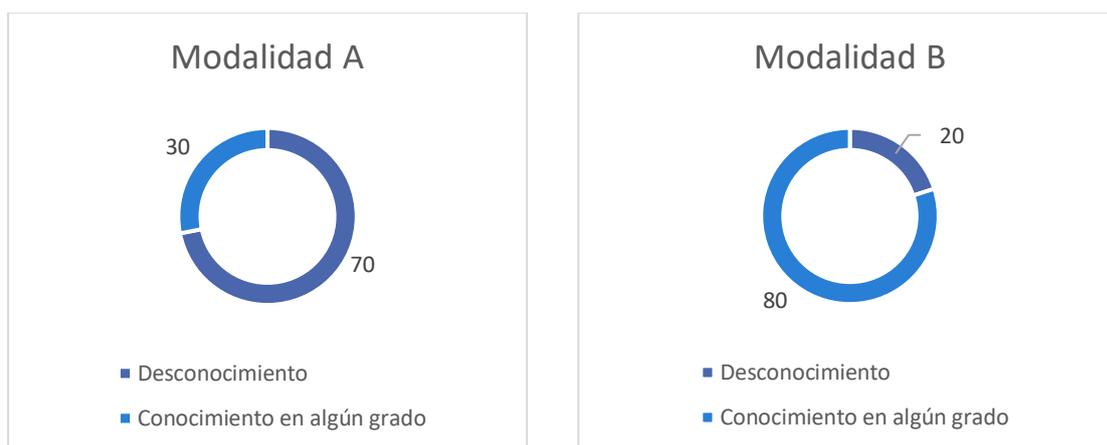
Como explicábamos en el apartado anterior, la noción jerárquica de la competencia colocacional no se puede ordenar desde una perspectiva dimensional significativa. Por lo que hemos comparado los resultados de cada aprendiz en cada uno de los formatos de cada Modalidad del TCE para conseguir una visión lo más global posible del grado del conocimiento colocacional. Siguiendo los trabajos ya realizados en otros test de vocabulario (Sasao & Webb, 2017) que también miden varios aspectos del conocimiento en diferente grado. Así, por ejemplo, de un aprendiz que realizó el TCE en la Modalidad A, comprobamos si obtuvo los mismos resultados en PRECO y en CRECU. De esta manera, debemos recoger la respuesta a cada ítem en el primer test y su homólogo: si en ambos la respuesta correcta, confirmaremos que el *conocimiento* colocacional es *total*; si no da la respuesta correcta en ninguna de las dos versiones del test, concluimos que su *conocimiento* de la colocación es *nulo*; finalmente, si acierta en una de las versiones, pero no en la otra, concluimos que se muestra un *conocimiento*

parcial de dicha colocación (vid. 4.3.3.2). En la *Tabla 91*, se expone esta relación entre aciertos y errores que da lugar a la gradación del conocimiento colocacional.

		CRECO / CRECU	
		CORRECTA	INCORRECTA
PRECU / PRECO	CORRECTA	Conocimiento total	Conocimiento parcial esperado
	INCORRECTA	Conocimiento parcial anómalo ¹²⁶	Conocimiento nulo

Tabla 87. Modelo de gradación del conocimiento colocacional

En un primer análisis (vid. Gráfica 7), según los resultados obtenidos en la Modalidad A, los aprendices conocen en algún grado alrededor de un 70% de las colocaciones y las desconocen algo más del 30%. Sin embargo, en la Modalidad B, un 20% de las colocaciones se desconocen y un 80% se conocen en algún grado. Solo hay un diez por ciento de diferencia entre una modalidad y otra, lo que concuerda bastante con los resultados de competencia colocacional en una y otra modalidad.



Gráfica 8. Proporción del grado de conocimiento en cada Opción del TCE

En cambio, si concretamos más la descripción y discriminamos entre conocimiento *total* o *parcial* (además del desconocimiento o conocimiento nulo), el resultado sí es más llamativo. Mientras en la Modalidad A el grupo mayor es del del conocimiento parcial; en la Opción B el grupo más amplio es el del conocimiento total. Podemos relacionar esto con los datos obtenidos en las distintas modalidades; mientras en la Modalidad A muchos aprendices contestan correctamente solo a CRECO y mal a PRECU, en la Modalidad B el grupo más amplio responde bien tanto a PRECO como a CRECU. Por lo tanto, parece que discrimina mejor la Modalidad A que la Modalidad B y confirma de nuevo los polos del conocimiento colocacional: comprensión desde el

¹²⁶ Una explicación más detalla la podemos encontrar en la sección B) del apartado 4.3.3.2

reconocimiento como el menor reto y producción desde la recuperación como el mayor obstáculo.



Gráfica 9. Proporción del grado de conocimiento en cada Modalidad del TCE: CN, conocimiento nulo; CP, conocimiento parcial; CT, conocimiento total:

De forma aún más detallada, observamos qué porcentaje de colocaciones de cada uno de los grados de conocimiento conocen los aprendices y cómo las conocen. Como era esperable, el porcentaje de aciertos que demuestran el conocimiento de colocaciones es mayor en las bandas que representan el conocimiento inicial y decrece a medida que aumenta el nivel. Y el porcentaje de colocaciones que demuestran algún grado del conocimiento ocurre lo contrario: desde un porcentaje mínimo va incrementado. Además, se percibe una diferencia notable en los porcentajes de una y otra opción: en la opción B la diferencia es bastante mayor.

		Bandas de nivel de las colocaciones meta					
		A1	A2	B1	B2	C1	C2
Grado de conocimiento	CONOCIMIENTO EN ALGÚN GRADO	84,22	78,22	71,78	67,78	62,89	47,33
	CONOCIMIENTO NULO	15,78	21,78	28,22	32,22	37,11	52,67

Tabla 88. Porcentaje de grado de conocimiento general en la Modalidad A por niveles del MCER

		Bandas de nivel de las colocaciones meta					
		A1	A2	B1	B2	C1	C2
Grado de conocimiento	CONOCIMIENTO EN ALGÚN GRADO	95,45	97,88	81,52	79,70	76,36	53,64
	CONOCIMIENTO NULO	4,55	2,12	18,48	20,30	23,64	46,36

Tabla 89. Porcentaje de grado de conocimiento general en la Modalidad B por niveles del MCER

Si atendemos a las diferencias entre el conocimiento total y parcial de las colocaciones encontramos que se mantiene un porcentaje superior del conocimiento total en las bandas de nivel inicial (A1 y A2). Mientras, el porcentaje de colocaciones conocidas solo parcialmente supera al de conocidas totalmente en las bandas de nivel

avanzado (C1 y C2). En las bandas de nivel intermedio, la Modalidad A y la Modalidad B funcionan de forma diferente, mientras en la primera el conocimiento parcial es mayor que el total; en la segunda el conocimiento total es mayor al parcial.

		Bandas de nivel de las colocaciones meta					
		A1	A2	B1	B2	C1	C2
Grado de conocimiento	TOTAL	46,22%	46,67%	34,00%	20,67%	15,11%	8,89%
	PARCIAL	38,00%	31,56%	37,78%	47,11%	47,78%	38,44%
	NULO	15,78%	21,78%	28,22%	32,22%	37,11%	52,67%

Tabla 90. Porcentaje de grado de conocimiento concreto en la Opción A por niveles del MCER

		Bandas de nivel de las colocaciones meta					
		A1	A2	B1	B2	C1	C2
Grado de conocimiento	TOTAL	66,36%	76,36%	44,85%	41,52%	32,42%	16,97%
	PARCIAL	29,09%	21,52%	36,67%	38,18%	43,94%	36,67%
	NULO	4,55%	2,12%	18,48%	20,30%	23,64%	46,36%

Tabla 91. Porcentaje de grado de conocimiento general en la Opción B por niveles del MCER

El conocimiento parcial, que implica el acierto en una de las modalidades de la opción, puede manifestarse de dos maneras en cada opción. Esta dualidad dio lugar a diferenciar entre dos grados del conocimiento parcial, lo que nos permitirá ahondar en la descripción del conocimiento global. De ese modo, diferenciamos entre el *conocimiento parcial esperable* y el *conocimiento parcial anómalo*. Esta anomalía tiene su origen en la creencia de que la producción supone un reto mayor a la comprensión y que es un paso un paso posterior en el proceso de aprendizaje. De modo que clasificamos las respuestas de los aprendices teniendo en cuenta cuál de las dos modalidades habían fallado (para una explicación más detallada, vid. 4.3.3.2). Las diferencias entre una opción y otra son notables. Mientras en la Modalidad A la mayor parte de los participantes tiene un conocimiento parcial, en la Modalidad B tiene un conocimiento nulo. Si ordenamos los grados del conocimiento en un orden descendente:

- En la Modalidad A: conocimiento parcial (40%), conocimiento nulo (31%) y conocimiento total (29%).
- En la Modalidad B: conocimiento nulo (46%), conocimiento parcial (34%) y conocimiento total (20%).

EL único punto en común sería el conocimiento total o pleno como grado menos popular. Es decir, muy pocos aprendices conocen las colocaciones de forma total. Aun así, las principales diferencias residen en los tipos de conocimiento parcial. La Modalidad A sigue

la tendencia esperable: los participantes responden de forma correcta al test de comprensión y de forma incorrecta al de producción. Sin embargo, en la Modalidad B los aprendices que solo conocen la colocación de forma parcial cometen un porcentaje de errores en el test de comprensión (CRECO) similar al del test de comprensión (PRECU). En la Tabla 98 se pueden ver los porcentajes en función del grado de conocimiento.

Conocimiento		Modalidad A		Modalidad B	
		Total			
Parcial	Esperable	40%	37%	34%	16%
	Anómalo		3%		18%
	Nulo	31%		46%	

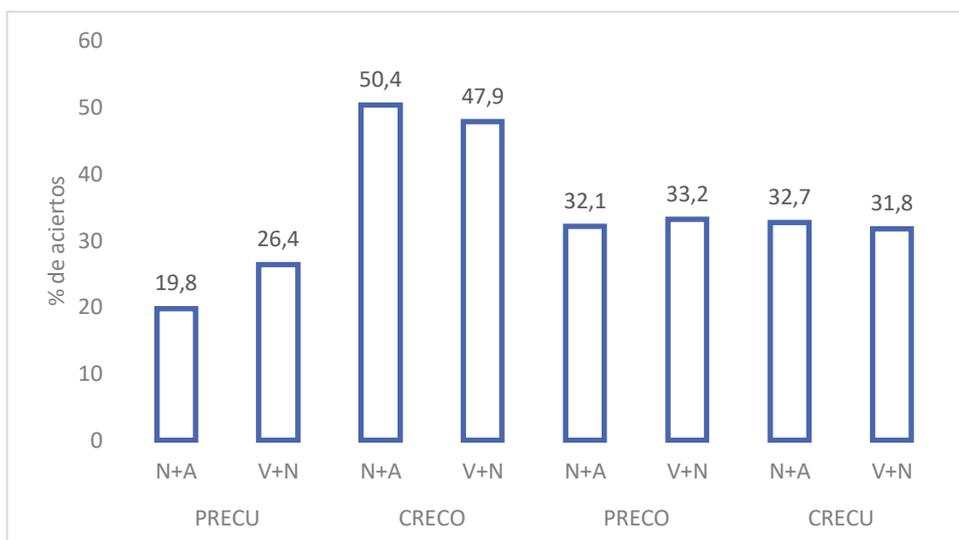
Tabla 92. Comparación de los grados del conocimiento colocacional en las dos Modalidades del TCE

En la búsqueda de las causas de esta diferencia encontramos dos que pueden resultar determinantes: el proceso cognitivo y el tipo de formato. Entre los participantes que conocen la colocación de forma parcial, el formato más errado es el de recuperación frente al de reconocimiento (PRECU y CRECU). La otra razón es la novedad del formato, aunque esta solo vale para CRECU. El formato utilizado en CRECU, C-Test, es uno de los más habituales para medir la producción de colocaciones. En cambio, el formato de CRECU, que consiste en la formulación de una continuación natural a una oración en la que se demuestra el conocimiento del significado de la colocación, pudo no ser totalmente obvio para el aprendiz de ELE. Aun así, los resultados son muy próximos.

5.2.3.7. *Influencia de la estructura gramatical de la colocación*

La última variable que tenemos en cuenta en el análisis del conocimiento colocacional es la estructura gramatical de la colocación. Entendemos por estructura gramatical de la colocación la combinación de las categorías gramaticales implicadas en la composición. En los criterios de selección de la muestra explicamos que incluimos dos tipos, V + N y N + A, y las razones que nos llevaron a restringir la selección a estos dos esquemas gramaticales. En este punto buscamos una relación entre el tipo de colocación y su mayor o menor carga de aprendizaje. Aunque las diferencias en el porcentaje de aciertos entre un tipo y otro de colocación son muy pequeños en todas las modalidades, en PRECU es en la modalidad en la que hay mayor diferencia. Al comparar los resultados de cada tipo de colocación en cada modalidad detectamos que en las dos modalidades que miden la comprensión, tanto desde la recuperación como desde el reconocimiento,

las colocaciones N+A son más fáciles para los aprendices y que en los formatos que miden la producción, en los dos procedimientos cognitivos, se obtienen mejores resultados en las colocaciones V+N. Estos resultados los podemos ver en la *Gráfica 9* de forma gráfica y en la *Tabla 99* de forma más detallada.



Gráfica 10. de barras que muestra el número de aciertos del esquema gramatical de la colocación en cada modalidad del TCE

	TIPO	N	Media	Mediana	DE
PRECU	N+A	18	19.8	16.0	13.67
	V+N	18	26.4	32.0	16.54
CRECO	N+A	18	50.4	50.0	10.24
	V+N	18	47.9	49.0	9.63
PRECO	N+A	18	32.1	32.0	11.08
	V+N	18	33.2	38.0	14.06
CRECU	N+A	18	32.7	33.0	10.44
	V+N	18	31.8	33.5	13.26

Tabla 93. Influencia de la categoría gramatical en la competencia y conocimiento colocacional

En el test ADELEX (Moreno-Jaén, 2007) también se ha tenido en cuenta el esquema sintáctico de la colocación como condicionante para el aprendizaje de las colocaciones con un resultado similar, aunque en este caso, del inglés. Pese a que en este estudio se tuvieron en cuenta cuatro esquemas sintácticos (A+N, N+V, V+N y N+N), la desigual proporción de colocaciones, con un muy bajo número de colocaciones tipo N+V y N+N, hace que las observaciones de los resultados se centren en las colocaciones A+N y V+N, que representan alrededor del 90% de la muestra, y son los esquemas utilizados en el TCE. También en el test de asociación LexCombi2 (Brown, 2018) se tiene en cuenta el tipo de colocación, pero el bajo grado de significación impide hacer generalizaciones.

El t-test nos dice si las diferencias entre dos muestras independientes son lo suficientemente grandes como para asumir que provienen de diferentes grupos o que las diferencias en muestras relacionadas son también significativas (Herrera Soler et al., 2011). La media de aciertos con uno y otro esquema muestran mayor facilidad en las colocaciones V+N (59,39% de aciertos en colocaciones V+N y 54,1% en colocaciones A+N), pero con una diferencia aparentemente poco relevante. La prueba T-TEST muestra que la diferencia entre ambas no es significativa y, por lo tanto, al igual que en LexCombi2, el esquema sintáctico no parece condicionar el aprendizaje de las colocaciones. Podemos concluir que la diferencia entre tipos de colocación según su categoría gramatical no es significativa.

		Estadístico	gl	p
PRECU	T de Student	-1.318	34.0	0.196
CRECO	T de Student	0.771	34.0	0.446
PRECO	T de Student	-0.263	34.0	0.794
CRECU	T de Student	0.237	34.0	0.814

Tabla 94. Prueba T para Muestras Independientes

5.2.4. Consideraciones finales

Podemos considerar que los resultados del TCE son estables, así lo atestigua la continuidad de las tendencias marcadas en los estudios piloto hasta la prueba definitiva. Además, con las peculiaridades del test (multinivel, diagnóstico, etc.), los resultados de las pruebas estadísticas de fiabilidad son positivas. En las tres versiones realizadas, dos piloto y una definitiva del TCE, hemos mantenido la estructura del test con sus cuatro modalidades y mismo formato. Esto nos permite observar los resultados de forma longitudinal y comprobar que en los tres casos la dimensión en la que mayores errores cometen los aprendices es PRECU, es decir, la que supone un mayor reto para los aprendices; la dimensión en la que menos errores cometen es CRECO, lo que indica que es la que menores problemas les supone a los aprendices. En la *Tabla 101* podemos comparar la media y los porcentajes de aciertos. En el TCEp la media es diferente de la de preTCEp y TCE ya que esa versión contaba con el doble de ítems. Además, la proporción de aciertos en cada dimensión es bastante similar en las cuatro competencias en las tres versiones del test. La mayor diferencia está en el TCE con 24% de aciertos frente al 29% y 39% de las versiones anteriores.

Versiones piloto						Versión definitiva		
preTCEp			TCEp ¹²⁷			TCE		
Modalidad	Media de aciertos	Porcentaje de aciertos	Modalidad	Media de aciertos	Porcentaje de aciertos	Modalidad	Media de aciertos	Porcentaje de aciertos
CRECO	29	80%	CRECO	52.689	73%	CRECO	28.128	78%
PRECO	22,26	61%	PRECO	50.044	69%	PRECO	23.039	63%
CRECU	20,27	56%	CRECU	42.089	58%	CRECU	21.843	60%
PRECU	17,92	49%	PRECU	28.378	39%	PRECU	15.137	24%

Tabla 95. Orden jerárquico de las cuatro modalidades del TCE a lo largo de todas sus versiones

Al disponer las dimensiones de forma jerárquica (vid. x) de forma ascendente desde la dimensión más sencilla a la más compleja podemos detectar la principal diferencia entre las tres versiones del test: el orden inverso entre las dos modalidades de la Opción B. En el preTCEp se obtienen mejores resultados en PRECO que en CRECU y en las otras dos versiones (TCEp y TCE) CRECU resulta más sencillo a los aprendices que PRECO. Aun así, la distancia entre estas dos modalidades es mínima, no significativa. Esto mismo ocurre con la clasificación jerárquica en CATSS (Aviad-Levitzky et al., 2019; Laufer & Goldstein, 2004).

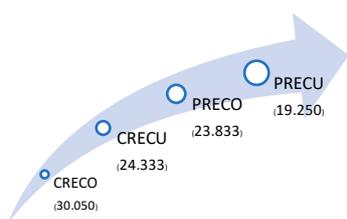


Figura 53: PreTCEp

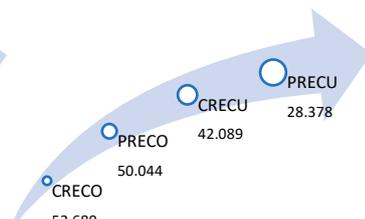


Figura 54: TCEp

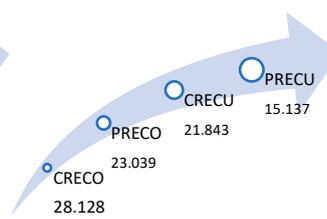


Figura 55: TCE

La finalidad de construir un test de diagnóstico es identificar las fortalezas y debilidades de un constructo, en el caso del TCE, del conocimiento colocacional. Pese a asociarse el dominio colocacional como uno de los retos del proceso de aprendizaje de una lengua, entre las fortalezas de los aprendices, podemos destacar la facilidad para comprender colocaciones aun cuando estas se clasifican en un nivel de competencia superior al de competencia del aprendiz. Entre las debilidades, el gran reto que supone su producción, sobre todo, la selección del colocativo. La estructura cuadrangular del TCE nos permite profundizar más y descubrir que la producción desde el reconocimiento

¹²⁷ La media es superior en el TCEp porque el número de ítems también lo es. Recordemos que tanto en el preTCEp como en el TCE el número de ítems es de 36 y en el del TCE es de 72. Al comparar el porcentaje de aciertos en lugar de la media de los resultados podemos obtener una comparación de forma más clara.

puede producir problemas, aunque no es el mayor reto. Esta situación es artificial y no se da en el discurso real, pero es indicador de que el aprendizaje se encuentra en un grado intermedio. La comprensión desde la recuperación sí se puede identificar con situaciones cotidianas ya que la paráfrasis forma parte de los recursos conversacionales habituales, entre otras funciones, por lo que es habitual en el discurso oral o escrito. Sin embargo, tampoco se puede identificar con las fortalezas, ya que sí supone problemas para los aprendices. Estas generalizaciones, aunque fiables y válidas, no están sujetas a otros factores, por ejemplo, los interlinguales.

Uno de los puntos más conflictivos es la falta de correlación entre la competencia colocacional, la competencia léxica y la competencia lingüística. No obstante, debemos tener en cuenta que la competencia léxica se midió a partir de un test en el que se valora únicamente el autoconocimiento, es decir, el reconocimiento de la existencia de la forma; quizás, si midiéramos el léxico también de forma global podríamos establecer la correlación entre una y otra competencia. También, al igual que en otros test donde sí se correlacionan la competencia colocacional con la léxica, usan las bases de las colocaciones meta como palabras meta en los test de vocabulario. De modo similar, en lugar de tomar como referencia del nivel de competencia el curso o el nivel en el que estaban inscritos los aprendices, sería conveniente asignarles nivel a partir de una herramienta válida como pudiera ser el DELE en español.

Una de las cuestiones más confusas de esta investigación reside en la diferencia no significativa entre el formato que mide la producción desde el reconocimiento, PRECO, y la comprensión desde el reconocimiento, CRECU, algo no exclusivo de esta investigación. Es también una de las limitaciones que se presenta en las distintas versiones de CATSS. La cuestión radica en conocer qué aspectos condiciona más la carga de aprendizaje de las colocaciones: la habilidad lingüística (comprensión o producción) o el proceso cognitivo mediante el que se accede a la información requerida (reconocimiento o recuperación). En este punto omitimos otros factores que ya se ha demostrado sí tienen una incidencia en la carga de aprendizaje de las colocaciones: la frecuencia, los cognados o el tipo de aprendizaje (Pellicer-Sánchez, 2015; Peters, 2016; Wolter & Gyllstad, 2013)

5.3. Aplicación de los resultados del TCE a la didáctica

Tras el análisis estadístico, podemos aplicar las conclusiones obtenidas al campo de la didáctica a través de una serie de directrices metodológicas. La naturaleza del TCE, de carácter diagnóstico, nos permite abordar algunos de los retos y mostrar algunas de las facilidades con las que los aprendices se encuentran al adquirir colocaciones y, de este modo, buscar las soluciones pertinentes a las dificultades encontradas. En este apartado concretamos los problemas detectados, identificamos las necesidades y proponemos soluciones para desarrollar el proceso de enseñanza-aprendizaje efectivo. Partimos de la noción de colocación adquirida ya explicada y las conclusiones extraídas del análisis del resultado del TCE.

La diagnosis es un proceso complejo que consta de diferentes fases que algunos sintetizan en tres componentes principales: diagnosis, retroalimentación y aprendizaje de recuperación o intervención (Fan et al., 2021; Lee, 2015, p. 308). El primer componente lo hemos desarrollado ya en los apartados precedentes al describir el instrumento de evaluación diagnóstica y analizar los resultados obtenidos. En este apartado nos dedicaremos a los otros dos componentes. El siguiente apartado nos centraremos en la retroalimentación al especificar los problemas y necesidades detectados en la actuación de los aprendices de ELE en el TCE. Finalmente, nos ocuparemos del último de los componentes: el aprendizaje de recuperación, cuyo objetivo es la mejora de la competencia evaluada, en este caso, la colocacional. En el apartado siguiente, propondremos algunas posibles soluciones para un proceso de aprendizaje colocacional eficaz.

5.3.1. Problemas y necesidades

Las fortalezas y debilidades detectadas en la actuación de los participantes y que vamos a desglosar en este apartado se extraen del análisis de datos previo. Estas conclusiones están condicionadas de forma general por el marco teórico de este estudio y de manera más concreta por los cuatro formatos elegidos que condicionan la información obtenida y son reflejo de las limitaciones de sus características (Durrant et al., 2022, p. 161). En el TCE todos los formatos están contextualizados, dos son de opción múltiple y dos de respuesta libre controlada en los que buscamos un equilibrio

entre la practicabilidad (sobre las condiciones de administración y corrección) y utilidad (sobre los datos obtenidos). El uso de cada uno de ellos tiene una serie de ventajas y desventajas a las que ya nos referimos en la descripción de la estructura del test (vid. 4.3.2) y que determinan la información analizada. En este punto, siguiendo la evaluación diagnóstica, presentamos, por un lado, las fortalezas y, por el otro, las debilidades y necesidades de los aprendices de ELE y las colocaciones.

La extendida y tajante afirmación sobre la dificultad del aprendizaje de colocaciones ya ha sido ampliamente rebatida. Algunos de los argumentos esgrimidos en el ámbito del inglés como lengua extranjera son los mismos que extraemos del TCE. El argumento central reside en la facilidad que los aprendices muestran en la comprensión contextualizada de colocaciones. En el TCE se ha medido la comprensión desde el reconocimiento, al seleccionar el sentido que aportaba la colocación, y desde la recuperación, recuperando el sentido de la combinación. Pese a que los resultados son mejores en el primer caso, cuando pueden elegir entre varias opciones, en ambos casos los aprendices muestran un dominio más que aceptable.

Esto está muy en línea con la definición de colocación que seguimos en esta investigación. La mayor parte de las definiciones de colocación se hacen desde un punto de vista de la comprensión, de ahí que se basen en criterios de frecuencia o que se aluda a su grado de transparencia. Sin embargo, si la definición se hace desde el punto de vista de la producción siempre partimos del sentido de la base y la selección restringida del colocativo para expresar el sentido buscado (Alonso Ramos, 2010, 2016).

A pesar de estas fortalezas, el dominio de la competencia y conocimiento colocacional no se encuentra exento de dificultades. La más evidente es la que se muestra en PRECU, la dimensión que mide la producción de colocaciones desde la recuperación. Los datos muestran un dominio deficiente desde esta perspectiva, incluso cuando el formato elegido está contextualizado e incluye la inicial del colocativo como pista. Este es, sin duda, el principal reto. La explicación se puede dar también desde una perspectiva teórica. La producción de colocaciones desde el reconocimiento, formato que habitualmente se suele tratar como de comprensión, ya no supone tantos problemas a los aprendices. La dificultad que implica la restricción léxica y conocimiento del vínculo fraseológico pierde peso al mostrarse varias opciones para el colocativo.

El efecto positivo de la estrategia de inferencia para comprender una colocación cuando se conoce una combinación parcialmente no aporta el mismo resultado en la producción. Por lo tanto, el último componente del test de diagnóstico, el aprendizaje de recuperación tiene su lugar. En esta investigación lo desarrollamos en el apartado siguiente dedicada al esbozo de propuestas y aporte de soluciones.

5.3.2. Propuestas y soluciones

Las propuestas de aprendizaje, que pretenden ser una posible solución para las dificultades en el proceso de enseñanza aprendizaje de colocaciones y que esbozamos en este apartado, tienen como objetivo profundizar en el conocimiento de colocaciones, descartar la idea dual de conocimiento frente a desconocimiento y pasar a la de grado de conocimiento de una colocación (conocimiento nulo, conocimiento parcial y conocimiento pleno). El punto de partida de la propuesta reside en el marco teórico, sobre todo, de la definición de colocación, y en la interpretación de los resultados del TCE. Relacionaremos los grados del conocimiento colocacional con alguna de las estrategias de aprendizaje de vocabulario más comunes (Gu, 2019; Nation, 2013b).

Al igual que el proceso de aprendizaje de vocabulario, el de aprendizaje de las colocaciones debe ser gradual, continuado y flexible (Durrant et al., 2022). Es gradual porque avanza desde el desconocimiento de la colocación hasta su conocimiento pleno; continuado en el sentido de que no tiene una meta más allá del uso correcto y el mantenimiento en la memoria; por último, flexible según el punto de partida del aprendiz, de sus necesidades y recursos (Laufer & Hulstijn, 2001). Organizamos estas propuestas en una serie de pasos que atienden a la cantidad mínima de información que un aprendiz necesita saber en cada grado de conocimiento. No tenemos en cuenta en esta propuesta las estrategias de selección de las colocaciones que se deben aprender. Sin embargo, y a modo de paso previo, haremos unas consideraciones generales. La primera sobre la presentación de la colocación. En ningún de los formatos del TCE se presentó como un todo, sino que partimos de la base para acceder al colocativo. La razón de esta consideración radica en la contraposición entre la definición de colocación de partida y las ideas de que las colocaciones se aprenden en bloque (Higuera García, 2006; Hoey, 2012; Lewis, 2000)

El modelo de aprendizaje de colocaciones del español (MACE) propuesto consta de cuatro fases, cada una de las cuales relacionamos con el paso de un grado del conocimiento al otro. Primero, denominamos *aproximación* al paso del conocimiento nulo al conocimiento parcial. Segundo, *afianzamiento* a la evolución desde un conocimiento parcial al total. Tercero, *consolidación* al mantenimiento de unos conocimientos en el tiempo. Esta propuesta está justificada por el análisis en el que analizamos las respuestas que cada aprendiz dio a las dos modalidades del TCE que realizó. En el proceso de aproximación se trabajará la comprensión desde el reconocimiento aplicando estrategias de aprendizaje como la inferencia; en el de consolidación, la comprensión desde la recuperación, con el uso de glosas o paráfrasis, y la producción desde la recuperación a partir de la memorización y correspondencia con la L1; en el de afianzamiento, la producción desde la recuperación, la memorización y la repetición. Es decir, las estrategias usadas en cada uno de los formatos seleccionados para el TCE.

Fases del modelo metodológico	Grado de conocimiento	Dimensiones del conocimiento colocacional	Estrategias de aprendizaje del vocabulario utilizadas
APROXIMACIÓN	Adquisición del conocimiento parcial	Comprensión desde el reconocimiento	Inferencia Interferencia
ESTABLECIMIENTO	Acercamiento al conocimiento parcial	Comprensión desde la recuperación. Producción desde el reconocimiento.	Glosa Memorización
CONSOLIDACIÓN	Mantenimiento del conocimiento total	Producción desde la recuperación.	Memorización Repetición

Tabla 96. Fases del MACE en relación con el grado y las dimensiones del conocimiento colocacional

En un capítulo anterior se hace una breve revisión sobre la situación actual del aprendizaje de colocaciones (vid. 2.3.4.). En el ámbito de ELE, algunos trabajos se centran en la valoración de las tareas utilizadas en los manuales para aprender colocaciones (Fernández Lázaro, 2014; Ferrando Aramo, 2012; Pérez Serrano, 2015).

6. CONCLUSIÓN

En este apartado recogeremos las principales aportaciones, limitaciones y posibles líneas futuras de esta investigación. Esta tesis ha tenido como eje central la construcción del TCE, un instrumento de evaluación que viene a llenar un vacío en el marco de la enseñanza y evaluación del español como lengua extranjera. Además, este instrumento de evaluación original cumple con todas y cada una de las premisas que se buscaban y se explicaron en los capítulos precedentes (vid.0 y 4.1): un instrumento de evaluación de las colocaciones del español de nivel y de diagnóstico y siguiendo la validez basada en argumentos.

A lo largo de los capítulos anteriores se ha presentado la construcción, administración y uso del test y se ha ido respondiendo a las preguntas de investigación al cumplir con los objetivos que la han guiado: la creación de un instrumento de evaluación válido y fiable que mida la competencia colocacional de aprendices de español (clasificación de su competencia) y la detección sus fortalezas y debilidades en este dominio (gradación de su conocimiento). Aparte de esto, destacamos tres principales aportaciones de esta tesis: a) la definición operativa del conocimiento colocacional desde una perspectiva fraseológica, b) la selección de una lista de colocaciones organizadas en niveles del MCER según los estándares del PCIC, y, finalmente, c) la adaptación de un test de léxico, CATSS, a un test de colocaciones de ELE. En el capítulo final presentamos las limitaciones observadas en la investigación en relación con las etapas de ciclo, en uno de los principios de la evaluación. La mayor parte de las limitaciones expuestas se solventarían con la adaptación del TCE a un test adaptativo informatizado, que presentamos como principal línea futura de trabajo.

6.1. Aportaciones e implicaciones

Dado que el Objetivo 1 apuntaba a medir la competencia colocacional de aprendices de español y a compararla con su competencia léxica y su competencia lingüística, podemos decir que con el TCE y las correlaciones estadísticas que se derivaron de él, este objetivo ha sido plenamente alcanzado.

Por una parte, el diseño del TCE implicó descripción de un modelo de conocimiento de colocaciones que tiene como base la aplicación de las teorías componenciales del conocimiento léxico (Nation, 2013b; Nation & Webb, 2011; Schmitt & Schmitt, 2020) y su adaptación de los conceptos teóricos formulados a partir de la TST (Mel'čuk, 2012b; Mel'čuk et al., 1995), así como su aplicación a las colocaciones del español (Alonso Ramos, 2010, 2016). Con estas fuentes propusimos una definición operativa de la colocación, destinada a su aplicación en los procesos de evaluación y aprendizaje del léxico, basada en cuatro componentes interdependientes. La definición operativa de colocación, que nos llevó a la conceptualizar el modelo del conocimiento colocacional, incluye un taxonomía de componentes, siguiendo la tendencia multicomponencial establecida en el ámbito del conocimiento léxico (González-Fernández & Schmitt, 2020). Con esta base fue posible la descripción de los criterios de gradación del conocimiento colocacional.

Por otra parte, se administró un test que mide la competencia léxica, X-LEX (Meara & Milton, 2003), del que existen distintas versiones en otros idiomas (Milton & Alexiou, 2010), pero es la primera vez que se administra en español. Esto nos permitió establecer la comparativa entre la competencia colocacional, a partir de diferentes métodos estadísticos, y comprobar la correlación entre las diferencias competencias: colocacional, léxica y lingüística. Esta última no se midió en esta investigación, sino que tomamos la que los centros de enseñanza-aprendizaje nos proporcionó de cada uno de los alumnos. Pese a que una de las hipótesis de partida era la de que la competencia colocacional, por su relevancia y estrecha relación con la fluidez, era índice de la competencia léxica y la competencia lingüística, los resultados de esta investigación no corroboran completamente esta hipótesis. Según los resultados obtenidos, extensión y el nivel de competencia colocacional se sitúa algo por debajo de los niveles asignados a las otras dos competencias. Esto no hace más que confirmar que en algún punto las colocaciones suponen un reto y se aprenden después de otros aspectos en la L2. Los matices a esta afirmación proceden del análisis de los resultados del TCE con los que comprobamos que la producción desde el reconocimiento no supone dificultades y se sitúa bastante por encima de los niveles asignados en las otras competencias, mientras que en los componentes en los que entra en juego la recuperación el dominio de las

colocaciones sí supone un problema. Creemos que la principal razón de esta disociación estriba en la diferencia de los procesos cognitivos elegidos para la evaluación.

Nuestro segundo objetivo, que apuntaba a identificar las dificultades y fortalezas que entraña el dominio de la competencia colocacional, ha sido también completamente alcanzado con el TCE que fue creado bajo las condiciones de los test de diagnóstico. Se buscó una herramienta que cumpliera con los requisitos específicos fijados. El análisis pormenorizado de algunos de los test de léxico y de los test que medían las colocaciones nos reveló que no existían ninguna herramienta previa que cumpliera con todas las exigencias, de modo que resultó imprescindible crear una nueva. Además, detectamos algunos parámetros confusos en la revisión de test, tanto por la forma como por las conclusiones que se extraen de su aplicación, que podrían alterar los resultados. En primer lugar, la identificación como fórmula de medida de las colocaciones presuponiendo un bagaje metalingüístico que no demuestra el dominio de las colocaciones de la lengua. En segundo lugar, la asimilación de la producción con la recuperación y la comprensión con el reconocimiento. En tercer lugar, la utilización de las mismas palabras meta para la elaboración del test de léxico y del de colocaciones. Esto hace que se asentara que el dominio colocacional es bajo, que supone un reto para los aprendices y que la competencia colocacional correlaciona con la competencia léxica. Sin embargo, al igual que otros estudios ya avanzaron, en esta investigación no se sostienen de forma rígida estas afirmaciones.

La detección de las dificultades y facilidades del aprendizaje de colocaciones nos obligó a buscar una herramienta de evaluación que ofreciera una visión panorámica del conocimiento de las colocaciones, que asociamos con la idea de sofisticación frente a los conceptos habituales de extensión y profundidad (vid. 2.1.2). Por lo tanto, seguimos la estructura de CATSS (Laufer & Goldstein, 2004) que adaptamos manteniendo los cuatro formatos de test y teniendo en cuenta las destrezas básicas (comprensión y producción) y dos procesos cognitivos (reconocimiento y recuperación). De esta forma el TCE está formado por una batería de test: PRECO, CRECU, PRECU y CRECO que ofrecen una visión global (y multicomponential) del conocimiento colocacional. De esta forma, medimos el conocimiento de las colocaciones desde cuatro puntos de vista. Esta visión panorámica nos permitió concluir que los aprendices encuentran PRECU como principal reto, es decir,

la producción desde la recuperación, y no encuentran mayores obstáculos en la realización de la modalidad CRECO, que mide la comprensión desde el reconocimiento.

La sistematización de este análisis nos llevó a la gradación del conocimiento colocacional que consiste en situar en una escala los diferentes modos de conocer una colocación: desde el conocimiento nulo al conocimiento total pasando por el conocimiento parcial esperable y anómalo. Consideramos esta gradación un punto de inicio para la elaboración de modelos de aprendizaje de colocaciones de ELE más efectivos. En los aprendices de nivel intermedio, que han realizado el TCE, predomina el conocimiento parcial de las colocaciones. Sin embargo, esta conclusión se ve alterada si observamos los datos en cada una de las modalidades del TCE, lo que da cuenta de la influencia que los formatos, destrezas y procesos cognitivos ejercen en los resultados y en las conclusiones extraídas de estas. Esto confirma la idea de que las colocaciones solo suponen un reto si es necesario recuperarlas, pero no si la meta es su reconocimiento.

Para alcanzar el objetivo 3, la asignación de un nivel de competencia colocacional a los aprendices de ELE, nos encontramos con un desafío: la nivelación de las colocaciones. La clasificación de aprendices en un nivel de competencia y la correlación con la competencia léxica y lingüística exigía que las colocaciones meta estuvieran niveladas según los criterios del MCER. La ausencia de una herramienta en español que cumpliera con los requisitos que necesitábamos nos obligó a crear nuestra propia lista de colocaciones meta, que era nuestro objetivo 5. De hecho, tal y como hemos explicado previamente (vid. 4.5), construimos diferentes listas hasta adaptarnos a las necesidades requeridas por el test. La selección de la muestra de colocaciones meta a partir de combinaciones niveladas nos permitió clasificar a los aprendices en niveles correspondientes al MCER. La nivelación de colocaciones es un vacío en el ámbito de la de ELE; primero por la falta de un método estandarizado y, segundo, por la falta de recursos, lo que supone un hándicap en la enseñanza-aprendizaje de colocaciones, sobre todo, al descartar la frecuencia como criterio para identificar las colocaciones.

De la clasificación de aprendices de ELE en niveles del MCER según su competencia colocacional concluimos que los aprendices con mayor competencia lingüística obtienen mejores resultados en los test de competencia colocacional; también que según el formato utilizado podemos sobreestimar los resultados, si predomina el

reconocimiento, o subestimarlos, si predomina la recuperación. Por lo que creemos que la medida de la extensión de la competencia colocacional y, en consecuencia, su nivelación, debe realizarse de forma combinada, es decir, a partir de más de un formato que aporte una imagen lo más global posible.

Detectamos varios factores como causas de la sobreestimación y subestimación de la competencia colocacional que derivan en la falta de correlación con las otras competencias. Primero, la competencia léxica suele medirse únicamente a partir de la comprensión y el reconocimiento o simplemente reconocimiento con lo que estamos correlacionando una visión global (TCE) con una parcial (X-LEX). Segundo, la asignación de nivel realizada por los centros de enseñanza de ELE no suele tener en cuenta ni la competencia léxica ni muchos menos la competencia colocacional. Creemos que estos dos factores pueden ser condicionantes en las correlaciones entre las competencias.

La gradación del conocimiento colocacional y la clasificación de la competencia colocacional son dos aspectos innovadores en el ámbito del español como lengua extranjera que suponen un avance en la profundización en los estudios de enseñanza-aprendizaje de vocabulario. La gradación del conocimiento colocacional es la fase de la investigación que aporta mayor cantidad de información y que realmente permite identificar las fortalezas y debilidades de los aprendices en el dominio de las colocaciones del español. La clasificación de los aprendices, según la actuación mostrada en el test, permitió comparar el nivel de competencia colocacional con el nivel de la competencia léxica y de la competencia lingüística. Esta tarea resultó especialmente ardua por los problemas derivados de la selección de la muestra.

Todo los objetivos mencionados han sido conseguidos gracias al diseño de un test válido y fiable que era nuestro objetivo 4. Para asegurar la validez y fiabilidad del TCE usamos el argumento de validez (vid. 3.1.2.1) así como el resto de principios de la evaluación (3.1.2) que han guiado el proceso de diseño y construcción; es decir, hemos establecido una serie de premisas que nos han permitido determinar y generalizar las conclusiones a las que hemos llegado tras el análisis cuantitativo y cualitativo de los resultados del TCE. Además, hemos aplicado los métodos estadísticos tradicionales y específicos para el tipo de test, lo que nos ha permitido afirmar que se trata de un test válido, fiable y práctico, en consonancia con las circunstancias, adaptándonos así a los

requerimientos de la evaluación de lenguas y del léxico de forma específica (vid. 2.2.1 y 3.1.2).

El TCE abarca la evaluación tanto de la competencia colocacional como el conocimiento colocacional así que el test tiene una doble función: de dominio y de diagnóstico, al igual que otros test que miden el léxico e, incluso, las colocaciones (Aviad-Levitzky et al., 2019; Gyllstad, 2007a). Si atendemos a la competencia estamos ante un test de dominio; si atendemos al conocimiento, ante un test de diagnóstico. Sin embargo, pensamos que la principal ventaja del TCE de colocaciones es la fusión de tres factores: la combinación de las cuatro dimensiones que muestre el test, el reflejo de una imagen global del conocimiento de la colocación y la facilidad de administración.

6.2. Limitaciones y líneas futuras

Pese a las aportaciones que se han enumerado en el apartado anterior, el TCE no está libre de dificultades. Para algunas de ellas, se encontró una solución y otras se desarrollarán en futuras investigaciones. Las principales tienen que ver con uno de los principios de la evaluación de lenguas, la practicabilidad que busca adecuar el proceso de administración a las necesidades y circunstancias del proceso, así como facilitar el proceso de administración y la comodidad de los aprendices a la hora de realizar el test. Organizaremos este apartado siguiendo las etapas del ciclo de evaluación desde la proyección del proceso de evaluación hasta la interpretación de los resultados, con especial hincapié en el proceso de construcción y administración.

Una vez el diseño se adecuó a los requisitos y se adaptó la estructura de CATSS a las colocaciones del español, surgieron varios problemas asociados a la construcción y a los primeros pasos de construcción del test. Para asignar un nivel a los aprendices que participaron en la investigación era necesario contar con una muestra de colocaciones también niveladas. Lo habitual en los test de léxico, es recurrir a listas de palabras frecuentes. Como ya hemos indicado, el principal problema es que no hay acuerdo sobre qué es una colocación frecuente, ni la frecuencia es el criterio definitorio de la colocación según la definición teórica tomada. De modo que fue necesario crear una muestra a partir de los recursos operativos en ELE. En esta investigación no creamos una lista de

colocaciones frecuentes, pero sí esbozamos una serie de criterios para crear una lista de colocaciones niveladas en la línea de las listas de vocabulario basadas en el conocimiento (Schmitt et al., 2021), *Knowledge-based vocabulary list* (KVL): partimos de bases niveladas, seleccionamos los colocativos según criterio fraseológico y presentamos combinaciones lematizadas.

En el proceso de administración uno de los problemas fue la consecución de aprendices de español como lengua extranjera que pudieran participar en la investigación. Consecuentemente, tanto el tamaño como la distribución de la población a la que se le administró el TCE es una de las principales problemáticas del estudio. Con una población más amplia y, sobre todo, distribuida en todos los niveles del MCER de forma homogénea mejoraría tanto la validez como la fiabilidad del test (Gyllstad et al., 2021). Recordemos que únicamente contamos con participantes de nivel intermedio (B1 y B2) en situación de inmersión. Una población representativa de todos los niveles del MCER nos permitiría, además, comprobar si en todos los niveles las dificultades son las mismas, si el grado de conocimiento predominante es el mismo o si el TCE es una buena herramienta para clasificar a los aprendices según su nivel de competencia. Esto implicaría describir una imagen más amplia de la competencia y conocimiento colocacional a todos los niveles, no solo en el intermedio.

Otra de las limitaciones es la estructura condicionada del TCE. Con la intención de que la imagen que los resultados de test proyectaron fuera panorámica se diseñaron cuatro modelos que miden cada una de las perspectivas del conocimiento determinadas por la solidez. Sin embargo, pensando en la administración y en facilitar la participación a los aprendices de español de los centros colaboradores, se decidió que cada aprendiz realizará solo dos de las cuatro modalidades del test creando así dos versiones del TCE (Modalidad A y Modalidad B). Para esta decisión tuvimos en cuenta el efecto de saturación que la repetición del test podría causar en los participantes, que se relaciona con el principio de las consecuencias beneficiosas o el impacto que tanta carga podría causar en los aprendices. Además, usamos estas opciones como un recurso para frenar el efecto de memorización producido al evaluar más de una vez la misma colocación meta. Esta decisión ocasionó una serie de inconvenientes como la necesidad de encontrar una población similar para cada una de las opciones y la necesidad de realizar

el test de competencia léxica a ambos. Asimismo, originó una serie de repercusiones detectadas en el análisis de los resultados (vid. 5.2.3.5).

Por todo esto, una buena solución sería la adaptación del TCE a una versión informatizada y adaptativa que permitiera evaluar a un mismo aprendiz de las cuatro modalidades, así como agilizar tanto la realización como la corrección del test. Esto proporcionaría, además, una información esclarecedora como la distinción en la escala jerárquica entre PRECO y CRECU. La versión TAI del test neutralizaría la limitación de las dos opciones, anteriormente expuesta, con lo que el test se podría administrar a un mayor número de participantes. La herramienta de evaluación se podría distribuir así con mayor facilidad y ofrecer una retroalimentación más eficaz. Así, además de la relación de respuestas del test sería posible mostrarle al aprendiz de forma inmediata su nivel de competencia colocacional. En cuanto a la mejora en la información recogida, esto nos permitiría definir un perfil de la competencia del participante de modo que se mostrara, por un lado, el nivel de competencia y por el otro, el grado de conocimiento. Además, se podría indicar en cada una de las dimensiones del conocimiento el porcentaje de aciertos, mostrando así las fortalezas y debilidades de cada aprendiz, de tal manera que esta información le resultara de utilidad para su competencia y proceso de dominio, mejorando así notablemente la retroalimentación y sus condiciones de la administración (Doe, 2015). Incluso se agilizaría de forma determinante el proceso de corrección que se realizó de forma manual. Pese a estas ventajas, de nuevo la distancia entre los recursos necesarios y disponibles fue insalvable, pero es una clara línea de trabajo futura.

Con la aplicación de un test adaptativo, además de aumentar el número de participantes, de obtener un perfil del conocimiento y competencia de cada aprendiz, se podrían aumentar el número de preguntas que integran el test, incrementando así su validez y fiabilidad. Luego también sería viable recuperar condicionantes descartados en el proceso de pilotaje del test como la función léxica de la colocación y la transparencia. Incluso nos permitiría tener en cuenta la L1 de los participantes, atendiendo así a uno de los criterios que más veces se ha dado como razón de la facilidad o dificultad de este elemento del léxico (Lemhöfer & Dijkstra, 2004; Peters, 2016; Wolter & Gyllstad, 2011). Recientemente se ha profundizado en la determinante influencia que la L1 tiene en el aprendizaje de las colocaciones (Gyllstad & Wolter, 2016; Peters, 2016a; Pulido & Dussias,

2020). Las conclusiones de estos estudios nos llevan a poner en valor la realización del TCE separando a la población según su L1 para comprobar cuál es el grado de influencia de la congruencia según la L1. La duda consiste en si una lengua más próxima traerá más problemas o menos. La hipótesis de partida tendría como punto de inicio la idea de que las lenguas más próximas acarrearían más problemas en la recuperación que en el reconocimiento, pero podría no ocurrir lo mismo en las más alejadas ya que el error o la dificultad deseable podría evitar la influencia de la congruencia. Ver esto en las cuatro formatos del TCE supondría un salto importante en la descripción de la imagen global del conocimiento colocacional.

Aunque ya hemos avanzado algunos puntos relacionados con la interpretación de resultados, mostrando así que la creación de un test es un proceso cíclico y que no es posible pasar de una etapa a otra de forma independiente, en el análisis nos encontramos con dos dificultades que no han quedado resueltas. Primero, la escala jerarquizada del conocimiento colocacional, cuyo punto bajo y más alto han quedado ampliamente constatados. Al igual que en el léxico, los pasos intermedios no han llegado a mostrar una diferencia significativa al mostrar en cada versión del test cierta variabilidad de posición y poco intervalo de diferencia. Por lo tanto, no es fácil constatar si la destreza o el proceso cognitivo tienen mayor influencia. Contando con la versión TAI del TCE podríamos realizar un análisis de escala implicativa y modelo ecuativo estructural, para los que necesitaríamos que un mismo participante realizara las cuatro modalidades del test. Por último, el análisis estadístico de los datos se ha realizado atendiendo a técnicas de análisis fundamentales (Bachman, 2004; Bachman & Kunnan, 2005; Green, 2013, 2019). Un conocimiento más profundo de técnicas más avanzadas permitiría profundizar más en la interpretación de los datos (Aryadoust & Raquel, 2020). Los métodos específicos para los test basados en criterio, así como las técnicas multinivel hubieran elevado la calidad del estudio.

Otro de los puntos observados es la posible influencia de CRECU por la realización previa del test PRECO. Pese a tratarse de un test de producción, los participantes pudieron leer previamente la colocación en contexto y esto pudo influir positivamente en sus respuestas. Por esto se proponen dos alternativas en la administración que pudiera evitar esta posible influencia: realizar cada modalidad del test a un grupo

diferente de aprendices con unas características similares o que los participantes de cada modalidad realicen los test con el mismo componente cognitivo, así una modalidad haría PRECO y CRECO y otra haría PRECU y CRECU. Lo esperable sería que los resultados en reconocimiento fueran mucho más altos que en recuperación. Serían problemas que se evitarían con la aplicación del TAI.

Con el TCE hemos dado un paso en el campo de la evaluación de español como lengua extranjera. Hemos podido confirmar en español varios de los puntos clave del aprendizaje del vocabulario; y también trasladar y adaptar algunos puntos clave del léxico al ámbito de las colocaciones. Dadas las características principales del TCE (colocación como unidad fraseológica, test de competencia, test de diagnóstico y argumento de validez), este test se puede convertir en la base para realización de futuras adaptaciones, pero en su estado actual cumple el cometido para el que fue diseñado: asignar un nivel de competencia colocacional a partir de los niveles del MCER y comprobar el grado de conocimiento colocacional, así como detectar las dificultades del proceso de aprendizaje.

REFERENCIAS BIBLIOGRÁFICAS

- Ackermann, K., & Chen, Y. (2013). Developing the Academic Collocation List (ACL) – A corpus-driven and expert-judged approach. *Journal of English for Academic Purposes, 12*, 235-247.
- Ainciburu, C., & Santana, M. E. (2022). *Estrategias para la adquisición del léxico*. Edelsa.
- Ainciburu, M. (2012). Recursos para el estudio del léxico disponible. Presentación del portal Dispolex. *Revista Neb, 11*.
- Alderson, J. C. (2007). Judging the Frequency of English Words. *Applied Linguistics, 28*(3), 383-409.
- Alderson, J. C., Brunfaut, T., & Harding, L. (2015). Towards a theory of diagnosis in second and foreign language assessment: Insights from professional practice across diverse fields. *Applied Linguistics, 36*(2), 236-260.
- Almela, M., Cantos, P., & Sánchez, A. (2013). Collocation, Co-collocation, Constellation... Any Advances in Distributional Semantics? *Procedia - Social and Behavioral Sciences, 95*, 231-240.
- Alonso-Ramos, M. (1994-1995). Hacia una definición del concepto de colocación: de J.R. Firth a I.A. Mel'cuk. En *Revista de Lexicografía, 1*, 9-28.
- Alonso-Ramos, M. (2002). Colocaciones y contorno de la definición lexicográfica. *Lingüística española actual, 24*(1), 63-96.
- Alonso-Ramos, M. (2004). *Diccionario de colocaciones del español*. <http://www.dicesp.com/paginas>
- Alonso-Ramos, M. (2004). *Las construcciones con verbo de apoyo*. Visor Libros.
- Alonso-Ramos, M. (2006). Glosas para las colocaciones en el Diccionario de Colocaciones del Español. En *Diccionario y fraseología. Anejo de la Revista de Lexicografía* (pp. 59-88). Servicio de Publicaciones de la Universidad de La Coruña.
- Alonso-Ramos, M. (2010). No importa si la llamas o no colocación, descríbela. En Frank & Timme (Ed.), *La fraseografía del S. XXI: Nuevas propuestas para el español y el*

alemán, 55-80.

Alonso-Ramos, M. (2012). Explorando la frecuencia léxica para el Diccionario de colocaciones del español. *Cum corde et in nova grammatica: Estudios ofrecidos a Guillermo Rojo*, 19-40.

Alonso-Ramos, M. (2016). Can collocations be deduced? A lexically driven analysis from the perspective of language production. En S. Torner-Castells & E. Bernal (Eds.), *Collocations and Other Lexical Combinations in Spanish. Theoretical, Lexicographical and Applied Perspectives*, 21-40. Routledge.

Alonso-Ramos, M. (2017). Dicionarios combinatorios. *Estudios de Lingüística del Español*, 38, 173-201.

Alonso-Ramos, M., & Wanner, L. (2007). Collocation Chains: How to Deal with Them? *Wiener Slawistischer Almanach, Sonderband*, 69.

Alonso-Ramos, M., Wanner, L., Vázquez Veiga, N., Vincze, O., Mosqueira Suárez, E., & Prieto González, S. (2010). Tagging Collocations for Learners. En S. Granger & M. Paquot (Eds.), *Lexicography in the 21st century: new challenges, new applications. Proceedings of ELEX2009*, 01, 375-380. Presses Univeristaires de Louvain.

Alonso-Ramos, M., Wanner, L., Vincze, O., Casamayor, G., Vazquez Veiga, N., Mosqueira Suárez, E., & Prieto González, S. (2010). Towards a Motivated Annotation Schema of Collocation Errors in Learner Corpora Introduction: The Problem. En N. Calzolari, K. Choukri, B. Maegaard, J. Mariani, J. Odijk, S. Piperidis, & M. Rosner (Eds.), *Proceedings of the Seventh conference on International Language Resources and evaluation (LREC'10)*, 19-21. Language Resources Evaluation.

ALTE. (2011). *Manual for Language Test Development and Examining For use with the CEFR Produced by ALTE on behalf of the Language Policy Division*. Council of Europe.

ALTE. (2020). Principles of good practice. En *Elderly Care*.

Anderson, R. C., & Freebody, P. (1981). Vocabulary knowledge. En J. T. Guthrie (Ed.), *Comprehension and teaching: Research reviews*, 77–117. International Reading Association.

- Angoff, W. (1984). Scales, Norms, and Equivalent Scores. En *Educational Measurement*. Educational Testing Service.
- Aryadoust, V., & Raquel, M. (2020). *Quantitative data analysis for language assessment Volume 1, Fundamental techniques: Vol. I*. Routledge.
- Aviad-Levitzky, T., Laufer, B., & Goldstein, Z. (2019). The New Computer Adaptive Test of Size and Strength (CATSS): Development and Validation. *Language Assessment Quarterly*, 16(0), 1-24.
- Ávila Muñoz, A. M. (2016). El léxico disponible y la enseñanza del español. Propuesta de selección léxica basada en la teoría de los conjuntos difusos. *Journal of Spanish Language Teaching*, 3(1), 31-43.
- Bachman, L. (1990). *Fundamental Considerations in Language Testing*. Oxford University Press.
- Bachman, L. (2004). *Statistical Analyses for Language Assessment*. Cambridge University Press.
- Bachman, L. (2005). Building and Supporting a Case for Test Use. *Language Assessment Quarterly*, 2(1), 1-34.
- Bachman, L., & Domböck, B. (2017). *Language Assessment for Classroom Teachers | Oxford University Press*. Oxford University Press.
- Bachman, L., & Palmer, A. (2010 [1996]). *Language assessment in practice: Developing language assessments and justifying their use in the real world*. Oxford University Press.
- Bachman, Lyle, & Kunnan, A. (2005). *Statistical Analyses for Language Assessment Workbook*. Cambridge University Press.
- Barclay, S., & Schmitt, N. (2019). *Current Perspectives on Vocabulary Teaching and Learning* (pp. 799-819). Springer, Cham.
- Barcroft, J., Muñoz-Basols, J., & Muñoz-Basols, J. (2022). Spanish Vocabulary Learning in Meaning-Oriented Instruction. En J. Barcroft & J. Muñoz-Basols (Eds.), *Spanish Vocabulary Learning in Meaning-Oriented Instruction*. Routledge.

- Barfield, A. (2009). Exploring Productive L2 Collocation Knowledge. En T. Fitzpatrick & A. Barfield (Eds.), *Lexical Processing in Second Language Learners*, 95-110. Multilingual Matters.
- Barfield, A. (2013). Lexical Collocations. *Encyclopedia of Applied Linguistics, 1909*, 17-18.
- Barfield, A., & Gyllstad, H. (2009). Introduction: Researching L2 Collocation Knowledge and Development. En *Researching Collocations in Another Language*, 1-18. Palgrave Macmillan UK.
- Barrios, M., Bonillo, A., Cosculluela, A., Lozano, L. M., Turbany, J., & Valero, S. (2013). *La psicometría*. UOC.
- Barrios, M., & Cosculluela, A. (2013). Fiabilidad. En et al. Meneses (Ed.), *Psicometría*. UOC.
- Barrios Rodríguez, M. A. (2015). *Las colocaciones del español*. Arco Libros.
- Barrios Rodríguez, M. A. (2012). El aprendizaje de las colocaciones por medio de las glosas de las funciones léxicas. *Estudios e investigaciones, 35*, 123-130.
- Bartol Hernández, J. A. (2010). Disponibilidad léxica y selección del vocabulario. En R. Castañer Martín & V. Lagüens Gracia (Eds.), *De moneda nunca usada: Estudios dedicados a José M^o Enguita Utrilla*, 95-107.
- Batista, R., & Horst, M. (2016). A new receptive vocabulary size test for French. *Canadian Modern Language Review, 72(2)*, 211-233.
- Beglar, D. (2010). A Rasch-based validation of the vocabulary size test. *Language Testing, 27(1)*, 101-118.
- Benson, M. (1990). Collocations and General-purpose Dictionaries. *International Journal of Lexicography, 3(1)*, 23-34.
- Benson, M. (2010). *The BBI Combinatory Dictionary of English* (M. Benson, E. Benson, & R. F. Ilson (Eds.). John Benjamins Publishing Company.
- Berton, M. (2020). *Riqueza léxica y expresión escrita en aprendices suecos de ELE. Proficiencia general, competencia léxica pasiva, tipo y complejidad de la tarea*.

Stockholm University.

- Bisht, R. K., Dhimi, H. S., & Tiwari, N. (2006). An evaluation of different statistical techniques of collocation extraction using a probability measure to word combinations. *Journal of Quantitative Linguistics*, 13(2-3), 161-175.
- Boers, F., Demecheleer, M., Coxhead, A., & Webb, S. (2014). Gauging the effects of exercises on verb-noun collocations. *Language Teaching Research*, 18(1).
- Boers, F., & Webb, S. (2018). Teaching and learning collocation in adult second and foreign language learning. *Language Teaching*, 51(1), 77-89.
- Bonify, W. (2022). *Multidimensional Item Response Theory*. SAGE.
- Bonk, W. J. (1996). *Testing ESL learners' knowledge of collocations*.
- Bordag, D., Gor, K., & Opitz, A. (2021). Ontogenesis Model of the L2 Lexical Representation. *Bilingualism: L*, 2.
- Bosque, I. (2004). Combinatoria y significación. Algunas reflexiones. En *REDES. Diccionario combinatorio del español contemporáneo*.
- Bosque, I. (2006). *Diccionario combinatorio práctico del español contemporáneo: las palabras en su contexto*. SM.
- Bosque, I. (2011). Deducing collocations. En I. Boguslavsky, L. Wanner (Eds.), *Proceedings of the 5th International Conference on Meaning-Text Theory*, 8-9. 1006-1023
- Bosque, I. (2017). On the conceptual bases of collocations. Restricted adverbs and lexical selection. En S. Torner & E. Bernal (Eds.), *Collocations and Other Lexical Combinations in Spanish. Theoretical, Lexicographical and Applied Perspectives* (pp. 9-20). Routledge.
- Bosque, I. (2011b). Deducing collocations. En I. Boguslavsky & L. Wanner (Eds.), *Proceedings of the 5th International Conference on Meaning-Text Theory* (pp. 1006-1023).
- Bouma, G. (2009). Normalized (Pointwise) Mutual Information in Collocation Extraction. *Proceedings of the Biennial GSCL Conference*, 31-40.

- Bowles, M. A. (2022). Using instructor judgment, learner corpora, and DIF to develop a placement test for Spanish L2 and heritage learners. *Language Testing*, 39(3), 355-376.
- Brezina, V., & Gablasova, D. (2015). Is There a Core General Vocabulary? Introducing the New General Service List. *Applied Linguistics*, 36 (1), 1–22.
- Brown, D. (2013). Knowledge of Collocations. En James Milton & T. Fitzpatrick (Eds.), *Dimensions of vocabulary knowledge*. Macmillan international.
- Brown, D. (2014). Knowledge of Collocations. En James Milton & T. Fitzpatrick (Eds.), *Dimensions of vocabulary knowledge*, 123-139. Palgrave Macmillan.
- Brown, D. (2018). *Developing a measure of L2 learners's productive knowledge of English collocations*. Cardiff University.
- Brok Antle, J. (2014). Frequent Adjective + Noun Collocations for Intermediate English Language Learners. *Jalt2013 Conference Proceedings*, 297-307.
- Bruton, A. (2009). The Vocabulary Knowledge Scale: A Critical Analysis. *Language Assessment Quarterly*, 6(4), 288-297.
- Brysbaert, M., Keuleers, E., & Mandera, P. (2021). Which words do English non-native speakers know? New supranational levels based on yes/no decision. *Second Language Research*, 37(2), 207-231.
- Bulté, B., Housen, A., Pierrad, M., & Van Daele, S. (2008). Investigating lexical proficiency development over time - the case of Dutch-speaking learners of French in Brussels. *Journal of French Language Studies*, 18, 277-298.
- Cantos-Gómez, P., & Almela-Sánchez, M. (Eds.). (2018). *Lexical Collocation Analysis*. Springer.
- Capel, A. (2010). A1–B2 vocabulary: insights and issues arising from the English Profile Wordlists project. *English Profile Journal*, 1.
- Capel, A. (2012). Completing the English Vocabulary Profile: C1 and C2 vocabulary. *English Profile Journal*, 3.

- Carr, N. (2011). *Designing and Analyzing Language Test*. Oxford University Press.
- Carton, A. S. (1971). Inferencing: a process in using and learning language. En P. Pimsleur & T. Quinn (Eds.), *The psychology of second language learning* (pp. 45–58). Cambridge University Press.
- Cervatiuc, A. (2007). Assessing Second Language Vocabulary Knowledge. *International Forum of Teaching and Studies*, 3(3), 40-47.
- Chamorro Guerrero, M. D. (2018). Las colocaciones en el español académico: diseño de una investigación experimental sobre los efectos del Enfoque léxico. *RILEX. Revista sobre investigaciones léxicas*, 1(1), 113-134.
- Chandler, P. (2022). New Vocabulary Levels Test for L2 Spanish. En J. Barcroft & J. Muñoz-Bassols (Eds.), *Spanish Vocabulary Learning in Meaning-Oriented Instruction*, 150-170. Routledge.
- Chapelle, C. A. (2012a). Conceptions of validity. En *The Routledge Hand book of Language Testing*, 21-33. Routledge Handbooks.
- Chapelle, C. A. (2012b). Construct definition and validity inquiry in SLA research. *Interfaces between Second Language Acquisition and Language Testing Research*, 32-70.
- Chapelle, C. A. (2012c). Validity argument for language assessment: The framework is simple... *Language Testing*, 29(1), 19-27.
- Chapelle, C. A. (2021). Validity in Language Assessment. En P. Winke & T. Brunfaut (Eds.), *The Routledge Handbook of Second Language Acquisition and Language Testing*, 11-21. Routledge Handbooks.
- Chapelle, C. A., & Douglas, D. (2006). *Assessing Language through Computer Technology*. Cambridge University Press.
- Chapelle, C. A., Enright, M. K., & Jamieson, J. (2010). Does an argument-based approach to validity make a difference? *Educational Measurement: Issues and Practice*, 29(1), 3-13.
- Chui. (2006). A study of the English vocabulary knowledge of university students in Hong

- Kong. *Asian Journal of English Language Teaching*, 16, 1-23.
- Cizek, G. (2012). *Setting Performance Standards. Foundations, Methods, and Innovations*. En G. J. Cizek (Ed). Routledge.
- Cobb, T. (2007). *Why and how to use frequency lists to learn words*. LEXTUTOR. <http://www.lex tutor.ca/research/>
- Cohen, A. D. (2017). Test-taking strategies and task design. En G. Fulcher & F. Davidson (Eds.), *The Routledge Handbook of Language Testing* (pp. 262-278). Routledge Handbooks.
- Consejo de Europa. (2002). Marco común europeo de referencia para las lenguas. En *Strasburgo: Consejo de Europa, Ministerio de ...*
- Contreras-Izquierdo, N., & Martos-Eliche, F. (2020). Las secuencias formulaicas en la adquisición de español L2. *Porta Linguarum*, 33, 111-127.
- Cook, D. A., Brydges, R., Ginsburg, S., & Hatala, R. (2015). A contemporary approach to validity arguments: A practical guide to Kane's framework. *Medical Education*, 49(6), 560-575.
- Corpas-Pastor, G. (1996). *Manual de fraseología española*. Gredos.
- Corpas Pastor, G. (2001). Apuntes para el estudio de la colocación. *LEA: Lingüística española actual*, 23 (1), 41-56.
- Cowie. (1981). The Treatment of Collocations and Idioms in Learners' Dictionaries. *Applied Linguistics*, 11(3), 223-235.
- Coxhead, A. (2000). A New Academic Word List. *TESOL Quarterly*, 34(2), 213-238.
- Crossley, S., Salsbury, T., & McNamara, D. (2010). The development of polysemy and frequency use in english second language speakers. En *Language Learning* (60, 3).
- Crosson, A. C., McKeown, M. G., & Ward, A. K. (2019). An Innovative Approach to Assessing Depth of Knowledge of Academic Words. *Language Assessment Quarterly*, 16(2), 196-216.
- Cuberos Vicente, R., Rosado Villegas, E., & Mañas Navarrete, I. (2024). Towards a graded

- lexical inventory of multi-word combinations The distribution of collocations across Spanish CEFR levels. *ITL - International Journal of Applied Linguistics*.
- Cumming, A. (2012). Validation of Language Assessments. *The Encyclopedia of Applied Linguistics*. Wiley.
- Daller, H., Milton, J., & Treffers-Daller, J. (2007). Modelling and Assessing Vocabulary Knowledge. *Language Teaching*, 46(01), 41-52.
- David Hirsh, & Nation, P. (1992). What Vocabulary Size is Needed to Read Unsimplified Text for Pleasure? *Reading in a Foreign Language*, 8.
- Davies, M. (2006). *A Frequency Dictionary of Spanish*. Routledge.
- Davies, M. (2008). *The Corpus of Contemporary American English (COCA)*. <https://www.english-corpora.org/coca/>
- De Alba Quiñones, V. (2016). Variable de género y competencia colocacional léxica en el aula de ELE. *Porta Linguarum*, 2016(26), 201-215.
- De Jong, J., & Zheng, Y. (2016). Linking to the CEFR: validation using a priori and a posteriori evidence. En J. Banerjee & D. Tsagari (Eds.), *Contemporary Second Language Assessment*, 83-100. Bloomsbury Academic.
- De Miguel, E. (2006). Tensión y equilibrio entre nombres y verbos: el reparto de la tarea de predicar. *Actas del XXXV Simposio Internacional de la Sociedad Española de Lingüística*, 2, 1289-1313.
- Dean-Brown, J. (2012). Classical test theory. En G. Fulcher & F. Davidson (Eds.), *The Routledge Hand book of Language Testing of Language Testing*, 323-335.
- Devellis, R. (2012). *Scale Development. Theory and Applications*. SAGE.
- Díez-Ortega, M., & Kyle, K. (2023). Measuring the development of lexical richness of L2 Spanish: A longitudinal learner corpus study. *Studies in Second Language Acquisition*, 1-31.
- Dóczy, B. (2019). An Overview of Conceptual Models and Theories of Lexical Representation in Mental Lexicon. En S. Webb (Ed.), *The Routledge Hand book of*

- Vocabulary Studies* (pp. 46-65). Routledge Handbooks.
- Dóczi, B., & Kormos, J. (2016). *Longitudinal Developments in Vocabulary Knowledge and Lexical Organization*. Oxford University Press.
- Doe, C. (2015). Student interpretations of diagnostic feedback. *Language Assessment Quarterly*, 12(1), 110-135.
- Dunn, K. (2013). *What makes L2 words difficult to know? Using explanatory Item Response Theory to Model the Difficulty of Vocabulary Test for Learners of English as a Second Language*. Lancaster University.
- Durrant, P. (2008). *High frequency collocations and second language learning*. PhD thesis, University of Nottingham.
- Durrant, P. (2014). Corpus frequency and second language learners' knowledge of collocations: A meta-analysis. *International Journal of Corpus Linguistics*, 19(4), 443-477.
- Durrant, P., Siyanova-Chanturia, A., Kremmel, B., & Sonbul, S. (2022). *Research Methods in Vocabulary Studies*. John Benjamins.
- Eguchi, M., Suzuki, S., & Suzuki, Y. (2022). Lexical Competence Underlying Second Language Word Association Tasks. En *Studies in Second Language Acquisition*, 44 (1).
- Elder, C. (2012). Bias in Language Assessment. En *The Encyclopedia of Applied Linguistics*.
- Eyckmans, J. (2009). Toward an Assessment of Learners' Receptive and Productive Syntagmatic Knowledge. En A. Barfield & H. Gyllstad (Eds.), *Researching Collocations in Another Language* (pp. 139-152). Palgrave Macmillan.
- Ezquerro, A. (2011). *El uso del diccionario en el aula de E/LE: problemas y posibles soluciones*. 1993, 1-10.
- Fan, J., & Bond, T. (2019). Applying Rasch measurement in language assessment. Unidimensionality and local independence. En V. Aryadoust & M. Raquel (Eds.), *Quantitative data analysis for language assessment volume I. Fundamental Techniques* (pp. 81-102). Routledge.

- Fan, T., Song, J., & Guan, Z. (2021). Integrating diagnostic assessment into curriculum: a theoretical framework and teaching practices. *Language Testing in Asia*, 11(1).
- Fernández Lázaro, G. (2014). Enseñanza-aprendizaje de las colocaciones en nivel inicial (A1-A2). *MarcoELE: Revista de Didáctica Español Lengua Extranjera*, 19, 6.
- Ferrando Aramo, V. (2012). *Aspectos teóricos y metodológicos para la compilación de un diccionario combinatorio destinado a los estudiantes de E/LE*. Universitat Rovira i Virgili.
- Ferrando, P. J., Lorenzo-Seva, U., Hernández-Dorado, A., & Muñiz, J. (2022). Decalogue for the Factor Analysis of Test Items. *Psicothema*, 34(1), 7-17.
- Ferrando, V. (2018). El papel de las colocaciones en la enseñanza y el aprendizaje del español. En *Enseñar léxico en el aula. El poder de las palabras*. Difusión.
- Firth, J. R. (1958). *Papers in linguistics 1934-1951*. Oxford University Press.
- Fitzpatrick, T. (2013). Word Associations. *Encyclopedia of Applied Linguistics*, 10, 6193-6199.
- Fitzpatrick, T., & Meara, P. (2004). Exploring the validity of a test of productive vocabulary. *Vigo International Journal of Applied Linguistics*, 1, 55-74.
- Fitzpatrick, T., & Munby, I. (2014). Knowledge of Word Associations. En James Milton & T. Fitzpatrick (Eds.), *Dimensions of vocabulary knowledge*, 92-105. Palgrave Macmillan.
- Fitzpatrick, T., & Thwaites, P. (2020). Word association research and the L2 lexicon. *Language Teaching*, 237-274.
- Forsberg Lundell, F., & Lindqvist, C. (2014). Lexical aspects of very advanced L2 french. *Canadian Modern Language Review*, 70(1), 28-49.
- Forsberg Lundell, F., Lindqvist, C., & Edmonds, A. (2018). Productive collocation knowledge at advanced CEFR levels: Evidence from the development of a test for advanced L2 French. *Canadian Modern Language Review*, 74(4), 627-649.
- Frankenberg-Garcia, A. (2018). Investigating the collocations available to EAP writers.

Journal of English for Academic Purposes, 35, 93-104.

Fulcher, & Davidson, F. (Eds.). (2017). *The Routledge Handbook of Language Testing*. Routledge Handbooks.

Fulcher, G. (2010). *Practical Language Testing*. Hodder Education.

Fulcher, G., & Davidson, F. (2007). *Language Testing and Assessment*. Cambridge University Press.

Gablasova, D., Brezina, V., & McEnery, T. (2017). Collocations in Corpus-Based Language Learning Research: Identifying, Comparing, and Interpreting the Evidence. *Language Learning*, 67(1), 155-179.

García-Salido, Marcos. (2016). Error Analysis of Support Verb Constructions in Written Spanish Learner Corpora. *The Modern Language Journal*, 100(1), 362-376.

García-Salido, M., & Alonso-Ramos, M. (2018). Asignación de niveles de aprendizaje al Diccionario de Colocaciones del Español. *Signos*, 51(97), 153-174

García-Salido, Marcos, & Alonso-Ramos, M. (2019a). Testing the Use of a Collocation Retrieval Tool Without Prior Training by Learners of Spanish. *International Journal of Lexicography*, 32(4), 480-497.

García-Salido, Marcos, & Alonso-Ramos, M. (2019b). A comparison of statistical association measures for identifying dependency-based collocations in various languages. *Proceedings of the Joint Workshop on Multiword Expressions and WordNet (MWE-WN 2019)*, 49-59.

García-Salido, M. G., Garcia, M., & Alonso-Ramos, M. (2019c). Towards a graded dictionary of Spanish collocations. *Proceedings of Electronic Lexicography in the 21st Century Conference*, 849-864.

García-Salido, M. G., Vincze, O., Orol González, A., & Ramos, M. A. (2016). A corpus study of Spanish as a foreign language learners' collocation production. En M. Alonso-Ramos (Ed.), *Spanish Learner Corpus Research: Current trends and future perspectives*, 299-331. John Benjamins.

- García, Á. H., & Rull, A. N. (2019). Hacia una propuesta de selección léxica en manuales de español como lengua extranjera de nivel A1. *Estudios de Lingüística Aplicada*, 0(68), 59-98.
- Gardner, D., & Davies, M. (2013). A New Academic Vocabulary List. *Applied Linguistics*, 1-24.
- Garnier, M., & Schmitt, N. (2015). The PHaVE List: A pedagogical list of phrasal verbs and their most frequent meaning senses. *Language Teaching Research*, 19(6), 645-666.
- Gitsaki, C. (1996). *The Development of ESL Collocational Knowledge*. The University of Queensland.
- Godfroid, A. (2020). Sensitive Measure of Vocabulary Knowledge and Processing. En S. Webb (Ed.), *The Routledge Handbook of Vocabulary Studies* (pp. 433-453). Routledge Handbooks.
- González-Fernández, B. (2021). Vocabulary Knowledge: Lexical Depth and Its Relationship with Out-of-Class Exposure. En Pawel Szudarski & S. Barclay (Eds.), *Vocabulary Theory, Patterning and Teaching* (pp. 21-47). SLA. Second Language Acquisition.
- González-Fernández, B., & Schmitt, N. (2020). Word Knowledge: Exploring the Relationships and Order of Acquisition of Vocabulary Knowledge Components. *Applied Linguistics*, 41 (4), 481–505.
- González Fernández, B., & Schmitt, N. (2015). *How much collocation knowledge do L2 learners have? The effects of frequency and amount of exposure*. 166(1), 94-126.
- Gouverneur, G. U. (2008). Evaluación Del Léxico En La Enseñanza De Ele. En S. Pastor Cesteros & S. Roca Marín (Eds.), *La evaluación en el aprendizaje y la enseñanza del español como lengua extranjera / segunda lengua XVIII Congreso Internacional de la Asociación para la Enseñanza del Español como lengua Extranjera (ASELE)*. Alicante, 19-22 de septiembre de 2007, 590-594.
- Granger, Sylviane. (2018). Formulaic sequences 1 in learner corpora. En *Understanding Formulaic Language*, 228-247. Routledge.
- Green, A. (2014). *Exploring Language Assessment and Testing*. Routledge.

- Green, A. (2017). Linking Tests of English for Academic Purposes to the CEFR: The Score User's Perspective, *Language Assessment Quarterly*, 15(1), 59–74.
- Green, A., & Fulcher, G. (2021). Test Design Cycle. En P. Winke & T. Brunfaut (Eds.), *The Routledge Handbook of Second Language Acquisition and Language Testing*, 69-77. Routledge.
- Green, R. (2013). *Statistical Analyses for Language Testers*. Palgrave Macmillan.
- Green, R. (2019). Item analysis in language assessment. En V. Aryadoust & M. Raquel (Eds.), *Quantitative data analysis for language assessment volume I. Fundamental Techniques*, 15-29. Routledge.
- Gries, S., & Durrant, P. (2019). *Analyzing co-occurrence data*. En S. Gries & M. Paquot (Eds.). A practical handbook of corpus linguistics. Springer.
- Gries, S. T. (2013). 50-something years of work on collocations. *International Journal of Corpus Linguistics*, 18(1), 137-166.
- Gries, S. T. (2021). What do (most of) our dispersion measures measure (most)? Dispersion? *Journal of Second Language Studies*, 5(2), 171 - 205
- Gu, P. Y. (2019). Strategies for learning vocabulary. *The Routledge Handbook of Vocabulary Studies*, 271-287.
- Guzzi, E., & Alonso-Ramos, M. A. (2023). Descripción y usabilidad de HARTA, una herramienta de ayuda para la redacción de textos académicos en español. *TEISEL. Tecnologías para la investigación en segundas lenguas*, 2.
- Gyllstad, H. (2007). *Testing English Collocations. Developing Receptive Test for Use with Advanced Swedish Learners*. Lund University.
- Gyllstad, H. (2009). Designation and Evaluating Test of Receptive Collocations Knowledge: COLLEX and COLLMACH. En A. Barfield & H. Gyllstad (Eds.), *Researching Collocations in Another Language*. Palgrave Macmillan.
- Gyllstad, H. (2011). Researching Vocabulary: A Vocabulary Research Manual. *System*, 39(4), 567-569.

- Gyllstad, H. (2013). Looking at L2 vocabulary knowledge dimensions from an assessment perspective – challenges and potential solutions. En C. L. Camilla Bardel, Batia Laufer (Ed.), *L2 Vocabulary acquisition, knowledge, and use: New perspectives on assessment and corpus analysis*, 11-28. Eurosla Monographs Series.
- Gyllstad, H. (2020). Measuring Knowledge of Multiwords Items. En S. Webb (Ed.), *The Routledge Handbook of Vocabulary Studies*, 387-405. Routledge Handbooks.
- Gyllstad, H., McLean, S., & Stewart, J. (2021). Using confidence intervals to determine adequate item sample sizes for vocabulary tests: An essential but overlooked practice. *Language Testing*, 38(4), 558-579.
- Gyllstad, H., & Schmitt, N. (2019). Testing Formulaic Language. En A. Siyanova & A. Pellicer-Sánchez (Eds.), *Understanding Formulaic Language. A Second Language Acquisition perspective*, 174-191. Routledge.
- Gyllstad, H., Vilkaitė, L., & Schmitt, N. (2015). Assessing vocabulary size through multiple-choice formats Issues with guessing and sampling rates. *International Journal of Applied Linguistics*, 166(2), 278-306.
- Gyllstad, H., & Wolter, B. (2016). Collocational Processing in Light of the Phraseological Continuum Model: Does Semantic Transparency Matter? *Language Learning*, 66(2), 296-323.
- Haladyna, T. M., Downing, S. M., & Rodriguez, M. C. (2002). A Review of Multiple-Choice Item-Writing Guidelines for Classroom Assessment. En *Applied Measurement in Education*, 15(3), 309-333.
- Haladyna, T. M., & Rodriguez, M. C. (2013). *Developing and validating test items*. Routledge.
- Halliday, M. A. K. (1966). Lexis as a linguistic level. En M. A. K. H. C. E. Bazell, C. Catford & R. H. Robbins (Eds.), *In memory of J. R. Firth*, 148-162. Longman.
- Harrington, M. (2018). Lexical Facility. Size as a Dimension of L2 Vocabulary Skill. En *Palgrave Macmillan*. Palgrave Macmillan UK.
- Hausmann, F. . J. (1998). O diccionario de colocacións. Criterios de organización. En X.

- Ferro Ruibal (Ed.), *Actas do I Coloquio Galego de Fraseoloxía* (pp. 63-83). Centro Ramón Piñeiro, Xunta de Galicia.
- Hausmann, F. J. (1991). Collocations in monolingual and bilingual English dictionaries. En V. Ivir & D. Kalogjera (Eds.), *Languages in Contact and Contrast: Essays in Contact Linguistics*.
- Henriksen. (2013). Research on L2 learners' collocational competence and development - a progress report. *EUROSLA MONOGRAPHS SERIES 2 L2 vocabulary acquisition, knowledge and use*, 2, 29-56.
- Henriksen, B. (1999). Three Dimensions of Vocabulary Development. *Studies in Second Language Acquisition*, 21(2), 303-317.
- Herrera Soler, H., Martínez Arias, R., & Amengual Pizarro, M. (2011). *Estadística aplicada a la investigación lingüística*. Editorial EOS.
- Hidalgo Gallardo, M. (2020). Efectos de las listas de palabras en el aprendizaje del vocabulario en L2. *marcoELE. Revista de Didáctica Español Lengua Extranjera*, 30, 1-13.
- Higueras García, M. (2006). *Las colocaciones y su enseñanza en la clase de ELE*. ARCO/LIBROS.
- Higueras García, M. (2017). Pedagogical principles for the teaching of collocations in the foreign language classroom. En S. Torner & E. Bernal (Eds.), *Collocations and Other Lexical Combinations in Spanish*, 250-266. Routledge.
- Higueras, M. (1997). Las unidades léxicas y la enseñanza del léxico a extranjeros. *REALE*, 8, 35-49.
- Hoey, M. (2012). Lexical Priming. *The Encyclopedia of Applied Linguistics*. Wiley.
- Hoshino, Y. (2013). Relationship between types of distractor and difficulty of multiple-choice vocabulary tests in sentential context. *Language Testing in Asia*, 3(1), 16.
- Instituto Cervantes. (2006). *Plan Curricular del Instituto Cervantes*. Instituto Cervantes - Biblioteca Nueva.

- Ishii, T., & Schmitt, N. (2009). Developing an integrated diagnostic test of vocabulary size and depth. *RELC Journal*, 40(1), 5-22.
- Izquierdo Gil, M. C. (2004). *La selección del léxico en la enseñanza del español como lengua extranjera. Su aplicación en el nivel elemental en estudiantes francófonos*. Universitat de Valencia.
- Jang, E. E. (2017). Diagnostic assessment in language classroom. En G. Fulcher & F. Davidson (Eds.), *The Routledge Handbook of Language Testing*, 120-133.
- Jang, E. E., & Wagner, M. (2013). Diagnostic Feedback in the Classroom. En *The Companion to Language Assessment* (pp. 693-711). John Wiley & Sons, Inc.
- Jarvis, S., & Daller, M. (Eds.). (2013). *Vocabulary Knowledge. Human ratings and automated measures*. John Benjamins.
- Jiménez-Catalá, R., & Moreno-Espinosa, S. (2004). Using Lex30 to measure the L2 productive vocabulary of Spanish primary learners of EFL. *Vigo International Journal of Applied Linguistics VIAL*, 27-44.
- Jiménez-Catalá, R. M. (2017). Estudios de disponibilidad léxica en español y en inglés: revisión de sus fundamentos empíricos y metodológicos. *Revista Nebrija de Lingüística aplicada a la enseñanza de Lenguas*, 22, 16-31.
- Kachergis, G., Marchman, V. A., Dale, P. S., Mankewitz, J., & Frank, M. C. (2022). Online Computerized Adaptive Tests of Children's Vocabulary Development in English and Mexican Spanish. *Journal of Speech, Language, and Hearing Research*, 65(6), 2288-2308.
- Kalantari, R., & Gholami, J. (2017). Lexical complexity development from dynamic systems theory perspective: Lexical density, diversity, and sophistication. *International Journal of Instruction*, 10(4), 1-18.
- Kane. (2017). Articulating a validity argument. En *The Routledge Handbook of Language Testing*, 34-61. Routledge.
- Kane, M. (2010). Validity and fairness. *Language Testing*, 27(2), 177-182.
- Kane, M. (2012a). Articulating a validity argument. En *The Routledge Hand book of*

- Language Testing* (pp. 34-47). Routledge.
- Kane, M. (2012b). Articulating validity argument. En G. Fulcher & F. Davidson (Eds.), *The Routledge Handbook of Language Testing*, 34-47. Routledge.
- Kane, M. (2013). Validating the Interpretations and Uses of Test Scores. *Journal of Educational Measurement*, 50(1), 1-73.
- Kenyon, D. (2012). Standard Setting on Language Tests. En *The Encyclopedia of Applied Linguistics*.
- Kilgarriff, A., & Renau, I. (2013). esTenTen, a Vast Web Corpus of Peninsular and American Spanish. *Procedia - Social and Behavioral Sciences*, 95(12-19).
- Kilgarriff, Adam, Charalabopoulou, F., Gavrilidou, M., Johannessen, J. B., Khalil, S., Johansson Kokkinakis, S., Lew, R., Sharoff, S., Vadlapudi, R., & Volodina, E. (2013). Corpus-based vocabulary lists for language learners for nine languages. *Language Resources and Evaluation*.
- Knoch, U., & Chapelle, C. A. (2018). Validation of rating processes within an argument-based framework. *Language Testing*, 35(4).
- Koike, K. (2001). *Colocaciones léxicas en el español actual. Estudio formal y lexico-semántico*. Universidad de Alcalá de Henares. Servicio de Publicaciones.
- Koike, K. (2012). Colocaciones complejas metafóricas. En *Phraseology and Discourse Crosss Linguistic and Corpus-based Approaches*. Pamies.
- Kremmel, B. (2020). Measuring Vocabulary Learning Progress. En S. Webb (Ed.), *The Routledge Handbook of Vocabulary Studies*, 406-418. Routledge Handbooks.
- Kremmel, B. (2022). Does One Size Fit All? Comparing Two Computer-Adaptive Algorithms for a Diagnostic Vocabulary Test. En Paweł Szudarski & S. Barclay (Eds.), *Vocabulary Theory, Patterning and Teaching* (pp. 48-64). SLA. Second Language Acquisition.
- Kremmel, B., & Schmitt, N. (2016). Interpreting Vocabulary Test Scores: What Do Various Item Formats Tell Us About Learners' Ability to Employ Words? *Language Assessment Quarterly*, 13, 377 - 392.

- Kremmel, B., & Schmitt, N. (2017). Vocabulary Levels Test. En *The TESOL Encyclopedia of English Language Teaching* (pp. 1-7). John Wiley & Sons, Inc.
- Kunnan, A., & Carr, N. (2012). Statistical analysis of results. *Encyclopedia of Applied Linguistics*, 31(1), 75.
- Kunnan, A. J., & Jang, E. E. (2009). Diagnostic Feedback in Language Assessment. En M. H. Long & C. J. Doughty (Eds.), *The Handbook of Language Teaching*, 610-627. Blackwell Publishing Ltd.
- Laufer, B. (1997). What's in a word that makes it hard or easy? Intralexical factors affecting the difficulty of vocabulary acquisition. En Schmitt & M. McCarthy (Eds.), *Vocabulary: Description, Acquisition and Pedagogy* (pp. 140-155). Cambridge University Press.
- Laufer, B. (1998). The Development of Passive and Active Vocabulary in a Second Language: Same or Different? *Applied Linguistics*, 19, 255-271.
- Laufer, B. (2012). Second Language Word Difficulty. En *The Encyclopedia of Applied Linguistics*. Blackwell Publishing Ltd.
- Laufer, B., Elder, C., Hill, K., & Congdon, P. (2004). Size and strength: Do we need both to measure vocabulary knowledge? *Language Testing*, 21(2), 202-226.
- Laufer, B., & Goldstein, Z. (2004). Testing vocabulary knowledge: Size, strength, and computer adaptiveness. *Language Learning*, 54(3), 399-436.
- Laufer, B., & Hulstijn, J. (2001). Incidental vocabulary acquisition in a second language: The construct of task-induced involvement. *Applied Linguistics*, 22(1), 1-26.
- Laufer, B., & Nation, P. (1995). Vocabulary Size and Use: Lexical Richness in L2 Written Production. *Applied Linguistics*, 16 (3), 307-322.
- Laufer, B., & Nation, P. (1999). A vocabulary-size test of controlled productive ability. *Language Testing*, 16, 33-51.
- Laufer, B., & Waldman, T. (2011). Verb-Noun Collocations in Second Language Writing: A Corpus Analysis of Learners' English. *Language Learning*, 61(2), 647-672.

- Laya Gómez, L. (2014). *Nivelación de las colocaciones en ELE*. Universidade da Coruña.
- Lee Revier, R. (2009). Evaluating a New Test of Whole English Collocations. En A. Barfield & H. Gyllstad (Eds.), *Researching Collocations in Another Language. Multiple Interpretations*, 125-138. Palgrave Macmillan.
- Lee, Y. W. (2015). Diagnosing diagnostic language assessment. *Language Testing*, 32(3), 299-316.
- Lemhöfer, K., & Dijkstra, T. (2004). Recognizing cognates and interlingual homographs: Effects of code similarity in language-specific and generalized lexical decision. En *Memory and Cognition*, 32(4), 533-550.
- Leńko-Szymańska, A. (2020). *Defining and Assessing Lexical Proficiency*. Routledge.
- Levitzky-Aviad, T., & Laufer, B. (2013). Lexical properties in the writing of foreign language learners over eight years of study: Single words and collocations. En *L2 vocabulary acquisition, knowledge and use. New perspectives on assessment and corpus analysis*, 2, 127-148.
- Lewis, D., Mitzel, H., Mercado, R., & Schulz, M. (2012). The Bookmark Standard Setting Procedure. En Gregory Cizek (Ed.), *Setting Performance Standards. Foundations, Methods, and Innovations*. Routledge.
- Lewis, M. (1993). *The Lexical Approach: The State of ELT and a Way Forward*. Language Teaching Publications.
- Lewis, M. (1997). *Implementing the Lexical Approach*. Language Teaching Publications.
- Lewis, M. (2000). *Teaching collocation. Further Developments in the Lexical Approach*. Teacher Training.
- Llorián-González, S. (2015). Exámenes certificativos de ELE referenciados a sistemas externos como el del MCER. *Revista Internacional de Lenguas Extranjeras / International Journal of Foreign Languages*, 0(4), 97-122.
- López-Mezquita Molina, M. T. (2005). *La evaluación de la competencia léxica: tests de vocabulario. Su fiabilidad y validez*. Universidad de Granada.

- Macis, M., & Schmitt, N. (2017). Not just «small potatoes»: Knowledge of the idiomatic meanings of collocations. *Language Teaching Research*, 21(3), 321-340.
- Martín Peris, H. (2008). *Diccionario de términos clave de ELE*. SGEL. https://cvc.cervantes.es/ensenanza/biblioteca_ele/diccio_ele/
- Martinez, R. (2013). A framework for the inclusion of multi-word expressions in ELT. *ELT Journal*, 67(2), 184-198.
- Martínez, R. (2011). *The development of a corpus-informed list of formulaic sequences for language pedagogy*. Nottingham.
- Martinez, R., & Schmitt, N. (2012). A Phrasal Expressions List. *Applied Linguistics*, 33(3), 299-320.
- Martinez, Ron, & Murphy, V. a. (2011). Effect of Frequency and Idiomaticity on Second Language Reading Comprehension. *TESOL Quarterly*, 45(2), 267-290.
- Masrai, A., & Milton, J. (2018). Measuring the contribution of academic and general vocabulary knowledge to learners' academic achievement. *Journal of English for Academic Purposes*, 31, 44-57.
- McKay, T., & Plonsky, L. (2021). Reliability Analyses: Estimating Error. En P. Winke & T. Brunfaut (Eds.), *The Routledge Handbook of Second Language Acquisition and Language Testing*, 468-482. Routledge.
- Mcnamara, T. (2008). Language Testing. En *The Handbook of Applied Linguistics* (pp. 763-783). Blackwell Publishing Ltd.
- Meara, P. (1992). *EFL Vocabulary Test*. University College, Centre for Applied Language Studies.
- Meara, P. (1996). The dimensions of lexical competence. En J. W. Gillian Brown, Kirsten Malmkjaer (Ed.), *Performance & Competence in second lanaguage acquisiton*. Cambridge University Press.
- Meara, P., & Fitzpatrick, T. (2000). Lex30: an improved method of assess-ing productive vocabulary in an L2. *System*, 28(1), 19-30.

- Meara, P., & Milton, J. (2003). *X-Lex, The Swansea Levels Test*. Express.
- Meara, P., & Miralpeix, I. (2017a). Measuring Lexical Access. Q_Lex v4.o. En P. Meara & I. Miralpeix (Eds.), *Tools for Researching Vocabulary* (pp. 185-214). SLA Second Language Acquisition.
- Meara, P., & Miralpeix, I. (2017b). *Tools for Researching Vocabulary*. SLA. Second Language Acquisition.
- Meara, P., & Wolter, B. (2004). V-Links: Beyond Vocabulary depth. *Angles on the English speaking world*, 4, 85-96.
- Mel'čuk, I. (1998). Collocations and Lexical Functions. En *Phraseology. Theory, Analysis, and Applications*, 22-53. Claredon Press.
- Mel'čuk, I. (2012a). Phraseology in the language, in the dictionary, and in the computer. *Yearbook of Phraseology*, 3, 31-56.
- Mel'čuk, I. (2012b). Phraseology in the language, in the dictionary, and in the computer. *YoP*, 3, 31-56.
- Mel'čuk, I. (2017). Clichés, an Understudied Subclass of Phrasemes. *Yearbook of Phraseology*, 6(1), 55-86.
- Mel'čuk, I., Clas, A., & Polguère, A. (1995). *Introduction à la lexicologie explicative et combinatoire*. Duculot.
- Mel'čuk, I., & Mel'čuk, I. (2012). Phraseology in the language, in the dictionary, and in the computer. *YoP*, 3, (pp. 31-56).
- Melka, F. (1997). Receptive vs. Productive Aspects of Vocabulary. En *Vocabulary: Description, Acquisition and Pedagogy*, 84-102.
- Men, H. (2018). Vocabulary Increase and Collocation Learning. En *Vocabulary Increase and Collocation Learning*. Springer Singapore.
- Miller, D., & Biber, D. (2015). Evaluating reliability in quantitative vocabulary studies: The influence of corpus design and composition. *International Journal of Corpus Linguistics*, 20(1), 30-53

- Milton, J. (2006). X-Lex: the Swansea Vocabulary Levels Test. *Proceedings of the 7th and 8th Current Trends in English Language Testing (CTELT) Conference vol 4*, 29-39.
- Milton, J. (2013). *Measuring Second Language Vocabulary Acquisition*. Multilingual Matters.
- Milton, J., & Alexiou, T. (2010). *Developing a vocabulary size test in Greek as a foreign language*.
- Milton, James, & Fitzpatrick, T. (2014). *Dimensions of vocabulary knowledge*. Palgrave Macmillan.
- Milton, & Treffers-Daller. (2013). Vocabulary size revisited: the link between vocabulary size and academic achievement. *Applied Linguistics Review*, 4(1), 151_172.
- Mizumoto, A., Sasao, Y., & Webb, S. A. (2019). Developing and evaluating a computerized adaptive testing version of the Word Part Levels Test. *Language Testing*, 36(1), 101-123.
- Moreno-Jaén, M. (2007a). A Corpus-driven Design of a Test for Assessing the ESL Collocational Competence of University Students. *International Journal of English Studies*, 7(2), 127-147.
- Moreno-Jaén, M. (2007b). *Collocational Competence of University Students*. 7(2), 127-147.
- Moreno-Jaén, M. (2009). *Recopilación, desarrollo pedagógico y evaluación de un banco de colocaciones frecuentes de la lengua inglesa a través de la lingüística de corpus y computacional*. Universidad de Granada.
- Muñiz, J. (2010). Las Teoría de los Tests: TCT y TRI. *Papeles del Psicólogo*, 31(1), 57-66.
- Nation, P. (2013). *Learning vocabulary in another language*. Cambridge University Press.
- Nation, P. (1983). *Testing and Teaching Vocabulary*. Guidelines, 5(1), 12-25.
- Nation, P. (1990). *Teaching and Learning Vocabulary*. Heinle and Heinle.
- Nation, P. (2006). How Large a Vocabulary Is Needed For Reading and Listening ? *The Canadian Modern Journal Review*, 63(1), 59-82.

- Nation, P. (2012). Vocabulary Size in a Second Language. *The Encyclopedia of Applied Linguistics*.
- Nation, P. (2013). Learning vocabulary in another language. En *Learning Vocabulary in Another Language* (second). Cambridge University Press.
- Nation, P. (2020). The Different Aspects of Vocabulary Knowledge. En S. Webb (Ed.), *The Routledge Hand book of Vocabulary Studies* (pp. 15-29). Routledge Handbooks.
- Nation, P., & Beglar, D. (2007). A vocabulary size test. *JALT 2007*, 31(7), 9-12.
- Nation, P., & Webb, S. (2011). *Researching and analyzing vocabulary*. HEINLE CENGAGE Learning.
- Nation, P., & Webb, S. (2013). Teaching Vocabulary. En *Encyclopedia of Applied Linguistics*. Wiley.
- National Academies of Sciences, Engineering, and M. (2020). A principled approach to language assessment: Considerations for the U.S. foreign service institute. En *A Principled Approach to Language Assessment: Considerations for the U.S. Foreign Service Institute*. The National Academies Press.
- Nesselhauf, N. (2003). The Use of Collocations by Advanced Learners of English and Some Implications for Teaching. *Applied Linguistics*, 24(2), 223-242.
- Nesselhauf, N. (2005). *Collocations in Learner Corpus*. John Benjamins Publishing Company.
- Ngoc, T., & Dang, Y. (2020). Corpus-Based Word Lists in Second Language Vocabulary Research, Learning, and Teaching. En S. Webb (Ed.), *The Routledge Handbook of Vocabulary Studies* (pp. 288-302). Routledge Handbooks.
- Nguyen, T. M. H., & Webb, S. (2017). Examining second language receptive knowledge of collocation and factors that affect learning. *Language Teaching Research*, 21(3), 298-320.
- Nissen, H. B., & Henriksen, B. (2006). Word class influence on word association test results. *International Journal of Applied Linguistics*, 16(3), 389-408.
- Nizonkiza, D. (2011). The relationship between lexical competence, collocational

- competence, and second language proficiency. *English Text Construction*, 4(1), 113–145.
- Nizonkiza, D. (2012). Quantifying controlled productive knowledge of collocations across proficiency and word frequency levels. *Studies in Second Language Learning and Teaching*, 2(1), 67-92.
- Nizonkiza, D. (2015). Measuring receptive collocational competence across proficiency levels. *Stellenbosch Papers in Linguistics*, 44(0), 125-146.
- Nizonkiza, D. (2017). Predictive power of controlled productive knowledge of collocations over L2 proficiency. En *Usage-Based Approaches to Language Acquisition and Language Teaching*, 265-286. Walter de Gruyter GmbH.
- Nizonkiza, D., & Van den Berg, K. (2015). The dimensional approach to vocabulary testing: What can we learn from past and present practices? *Stellenbosch Papers in Linguistics*, 43(0), 45-61.
- Nizonkiza, D., Van Dyk, T., & Louw, H. (2014). First-year university students' productive knowledge of collocations. *Stellenbosch Papers in Linguistics Plus*, 42.
- Nontasee, W., & Sukying, A. (2021). The learnability of word knowledge aspects in Thai EFL high school learners. *Journal of Language and Linguistic Studies*, 17(1), 34-55.
- Ockey, G. J. (2012). Item response theory. En Glen Fulcher & F. Davidson (Eds.), *Handbook of Language Testing*, 336-349.
- Orol-González, A., & Alonso-Ramos, M. A. (2013). A Comparative Study of Collocations in a Native Corpus and a Learner Corpus of Spanish. *Procedia -Social and Behavioral Sciences*, 96 (25), 563-570.
- Palapanidi, K. (2012). La aplicación de la disponibilidad léxica a la didáctica del léxico de LE. *Revista Nebrija de Lingüística Aplicada a la Enseñanza de Lenguas*, 21.
- Palapanidi, Kiriakí. (2012). La aplicación de la disponibilidad léxica a la didáctica del léxico de LE. *Revista Nebrija de Lingüística Aplicada*, 21.
- Papageorgiou, S., & Tannenbaum, R. J. (2016). Situating Standard Setting within Argument-Based Validity. *Language Assessment Quarterly*, 13(2), 109-123.

- Pellicer-Sánchez, A. (2014). Incidental L2 vocabulary acquisition incidental L2 from and while reading: An Eye-Tracking Study. *Studies in Second Language Acquisition*, 38(1), 97-130.
- Pellicer-Sánchez, A. (2015). Learning L2 collocations incidentally from reading. *Language Teaching Research*, 21(3), 381-402.
- Pellicer-Sánchez, A. (2020). Learning Single Words vs. Multiword Items. En S. Webb (Ed.), *The Routledge Handbook of Vocabulary Studies*, 158-173. Routledge Handbooks.
- Pellicer-Sánchez, A., & Schmitt, N. (2012). Scoring Yes–No vocabulary tests: Reaction time vs. nonword approaches. *Language Testing*, 29(4), 489-509.
- Pellicer-Sánchez, A., Siyanova-Chanturia, A., & Parente, F. (2022). The effect of frequency of exposure on the processing and learning of collocations: A comparison of first and second language readers' eye movements. *Applied Psycholinguistics*, 43(3), 727-756.
- Pérez-Serrano, M. (2015). Tratamiento de la combinatoria léxica en documentos de referencia y curriculares: el caso del MCER y del PCIC. *Revista de Investigación Lingüística*, 18, 213-232.
- Pérez Serrano, M. (2014). *Análisis de errores colocacionales en un corpus de aprendices de ELE*. MarcoELE, 19.
- Pérez Serrano, M. (2015). *Un enfoque léxico a prueba: efectos de la instrucción en el aprendizaje de las colocaciones léxicas*. Universidad de Salamanca.
- Peters, E. (2014). The effects of repetition and time of post-test administration on EFL learners' form recall of single words and collocations. *Language Teaching Research*, 18(1), 75-94.
- Peters, E. (2016). The learning burden of collocations: The role of interlexical and intralexical factors. *Language Teaching Research*, 20(1), 113-138.
- Peters, E., Velghe, T., & Van Rompaey, T. (2019). The VocabLab tests: The development of an English and French vocabulary test. *ITL - International Journal of Applied Linguistics (Belgium)*, 170(1), 53-78.

- Pintard, A., & François, T. (2020). Combining Expert Knowledge with Frequency Information to Infer CEFR Levels for Words. *Proceedings of the 1st Workshop on Tools and Resources to Empower People with Reading Difficulties, May*, 85-92.
- Polguère, A. (2010). *Lexicologie et sémantique lexicale : notions fondamentales*. Presses de l'Université de Montréal.
- Pulido, D. (2004). The effect of cultural familiarity on incidental vocabulary acquisition through reading. *The Reading Matrix*, 4(2), 20-53.
- Pulido, D. (2007). The effects of topic familiarity and passage sight vocabulary on L2 lexical inferencing and retention through reading. *Applied Linguistics*, 28(1), 66-86.
- Pulido, M. F., & Dussias, P. E. (2020). Desirable difficulties while learning collocations in a second language: Conditions that induce L1 interference improve learning. *Bilingualism*, 23(3), 652-667.
- Qian, D., & Lin, L. (2020). The Relationship Between Vocabulary Knowledge and Language Proficiency. En S. Webb (Ed.), *The Routledge Handbook of Vocabulary Studies* (pp. 66-80). Routledge Handbooks.
- Read, J. (1998). Validating a test to measure depth of vocabulary knowledge. En Kunnan (Ed.), *Validation in Language Assessment*. Lawrence Erlbaum.
- Read, J. (2004). Plumbing the depths: How should the construct of vocabulary knowledge be defined? En P. Bogaards & B. Laufer (Eds.), *Vocabulary in a Second Language* (pp. 209-227). John Benjamins Publishing Company.
- Read, J. (2011). *Assessing Vocabulary*. Cambridge University Press.
- Read, J. (2012). Assessment of Vocabulary. *The Encyclopedia of Applied Linguistics*.
- Read, J. (2017). Piloting vocabulary test. En Glenn Fulcher & F. Davidson (Eds.), *The Routledge Handbook of Language Testing* (pp. 307-320).
- Read, J., & Chapelle, C. A. (2001). A framework for second language vocabulary assessment. *Language Testing*, 18(1), 1-32.
- Read, J., & Dang, T. N. Y. (2022). Measuring depth of academic vocabulary knowledge.

Language Teaching Research.

- Richards, J. C. (1976). The role of vocabulary teaching. *TESOL Quarterly*, 10, 77-89.
- Rodríguez-Sánchez, I. (2022). Testing L2 Vocabulary Knowledge. En J. Barcroft & J. Muñoz-Basols (Eds.), *Spanish Vocabulary Learning in Meaning-Oriented Instruction* (pp. 130-149). Routledge.
- Rojo-Mejuto, N. (2015). *Hacia una clasificación de las colocaciones léxicas por niveles de aprendizaje*. Universidad de Salamanca.
- Rufat, A. S. (2010). *Apuntes sobre las combinaciones léxicas y el concepto de colocación*. Anuario de Estudios Filológicos, vol. XXXIII, 291-306.
- Sánchez, A., Cantos, P., & Simón, J. (2001). *Corpus CUMBRE del español contemporáneo de España e Hispanoamérica. Extracto de dos millones de palabras*.
- Salazar García, V. (2004). Acercamiento crítico a la selección objetiva de contenidos léxicos en la enseñanza de E/LE. *Estudios de lingüística*, 18, 243-273.
- Santos Díaz, I. C., Trigo Ibáñez, E., & Romero Oliva, M. F. (2020). La activación del léxico disponible y su aplicación a la enseñanza de lenguas. *Porta Linguarum*, 33, 75-93.
- Sasao, Y., & Webb, S. (2017). The Word Part Levels Test. *Language Teaching Research*, 21(1), 12-30.
- Sawaki, Y., & Xi, X. (2019). Univariate generalizability theory in language assessment. En V. Aryadoust & M. Raquel (Eds.), *Quantitative data analysis for language assessment volume I. Fundamental Techniques*, 30-53. Routledge.
- Schmidgall, J. E. (2017). Articulating and Evaluating Validity Arguments for the TOEIC® Tests. *ETS Research Report Series*, 2017(1), 1-9.
- Schmitt, N., Schmitt, D., & Clapham, C. (2001). Developing and exploring the behaviour of two new versions of the Vocabulary Levels Test. *Language Testing*, 18(1), 55-88.
- Schmitt, N. (2004). *Formulaic sequences. Acquisition, processing and use*. John Benjamins Publishing Company.
- Schmitt, N. (2010). *Researching Vocabulary. A Vocabulary Research Manual*. Palgrave

Macmillan.

Schmitt, N. (2014). Size and Depth of Vocabulary Knowledge: What the Research Shows. *Language Learning, 64*(4), 913-951.

Schmitt, N. (2019). Understanding vocabulary acquisition, instruction, and assessment: A research agenda. *Language Teaching, 52*(2), 261-274.

Schmitt, N., Cobb, T., Horst, M., & Schmitt, D. (2017). How much vocabulary is needed to use English? Replication of van Zeeland & Schmitt (2012), Nation (2006) and Cobb (2007). *Language Teaching, 50*(2), 212-226.

Schmitt, N., Dunn, K., O'Sullivan, B., Anthony, L., & Kremmel, B. (2021). Introducing Knowledge-based Vocabulary Lists (KVL). *TESOL Journal, 12*(4).

Schmitt, N. & Meara, P. (1997). Researching vocabulary through a word knowledge framework. *Studies in Second Language Acquisition, 19*(1), 17-36.

Schmitt, N., Nation, P., & Kremmel, B. (2020). Moving the field of vocabulary assessment forward: The need more rigorous test development and validation. *Language Teaching, 53*(4).

Schmitt, N., & Schmitt, D. (2014). A reassessment of frequency and vocabulary size in L2 vocabulary teaching. *Language Teaching, 47*(4), 484-503.

Schmitt, N., Schmitt, D., & Clapham, C. (2001). Developing and exploring the behaviour of two new versions of the Vocabulary Levels Test. *Language Testing, 18*(1), 55-88.

Schmitt, N., Schmitt, D., & Schmitt, D. (2020). *Vocabulary in Language Teaching*. Cambridge University.

Shin, D., & Nation, P. (2007). Beyond single words: the most frequent collocations in spoken English. *ELT Journal, 62*(4), 339-348.

Shrock, S. A., & Coscarelli, W. C. (2008). *Criterion-referenced Test Development Technical and Legal Guidelines for Corporate Training* (Tercera).

Siceri, S., Hambleton, R., & Pitoniak, M. (2004). Setting Passing Scores on Licensure Examinations Using Direct Consensus. *CLEAR Exam Review, 14*, 21-25.

- Simpson-Vlach, R., & Ellis, N. C. (2010). An Academic Formulas List: New Methods in Phraseology Research. *Applied Linguistics*, 31(4), 487-512.
- Sinclair, J. (1966). Beginning the study of lexis. En & C. E. Bazell, C. Catford, M. A. K. Halliday & R. H. Robbins (Eds.), *In memory of J. R. Firth* (pp. 410-430). Longman.
- Sinclair, J. (1991). *Corpus, Concordance, Collocation*. Oxford University Press.
- Sonbul, S., & El-Dakhs, D. (2020). Timed versus untimed recognition of L2 collocations: Does estimated proficiency modulate congruency effects? *Applied Psycholinguistics*, 41(5), 1197-1222.
- Sonbul, S., El-Dakhs, D. A. S., & Masrai, A. (2022). Second language productive knowledge of collocations: Does knowledge of individual words matter? *Studies in Second Language Acquisition*, 1-23.
- Sonbul, S., & Schmitt, N. (2013). Explicit and Implicit Lexical Knowledge: Acquisition of Collocations Under Different Input Conditions. *Language Learning*, 63(1), 121-159.
- Sonbul, S., & Siyanova-Chanturia, A. (2019). Research on the on-line processing of collocations: Replication of Wolter and Gyllstad (2011) and Millar (2011). *Language Teaching*, 1-9.
- Sorell, C. J. (2013). Zipf's Law and Vocabulary. En *Encyclopedia of Applied Linguistics* Wiley.
- Stengers, H., Boers, F., Housen, A., & Eyckmans, J. (2011). Formulaic sequences and L2 oral proficiency: Does the type of target language influence the association? *IRAL - International Review of Applied Linguistics in Language Teaching*, 49(4), 321-343.
- Stewart, J. (2014). Do Multiple-Choice Options Inflate Estimates of Vocabulary Size on the VST? *Language Assessment Quarterly*, 11(3).
- Stoeckel, T., McLean, S., & Nation, P. (2021). Limitations of siez and levels test of written receptive vocabulary knowledge. *Studies in Second Language Acquisition*, 43(1), 181-203.
- Stubbe, R., & Stewart, J. (2012). Optimizing scoring formulas for yes/no vocabulary checklists using linear models. *Shiken Research Bulletin*, 16(2), 2-7.

- Stubbe, Raymond. (2012). Do pseudoword false alarm rates and overestimation rates in Yes/No vocabulary tests change with Japanese university students' English ability levels? *Language Testing*, 29(4), 471-488.
- Sunderman, G., & Forcelini, J. (2022). Quality of lexicosemantic representations in L2 Spanish. En J. Barcroft & J. Muñoz-Basols (Eds.), *Spanish Vocabulary Learning in Meaning-Oriented Instruction*, 112-129. Routledge.
- Szudarski, Paweł, & Carter, R. (2016). The role of input flood and input enhancement in EFL learners' acquisition of collocations. *International Journal of Applied Linguistics (United Kingdom)*, 26(2), 245-265.
- Tavares-Pinto, P.; Rees, G.; Frankeberg-Garcia, A. (2021). Identifying collocations issues in English L2 research article writing. En *Corpora in ESP/EAP Writing Instruction* (p. 24). Routledge.
- Trizano-Hermosilla, I., & Alvarado, J. M. (2016). Best alternatives to Cronbach's alpha reliability in realistic conditions: Congeneric and asymmetrical measurements. *Frontiers in Psychology*, 7, 1-8.
- Tseng, W.-T. (2022). The Effects of Item Exposure Control on English Vocabulary Size Estimates in Computerized-Adaptive Testing. En Paweł Szudarski & S. Barclay (Eds.), *Vocabulary Theory, Patterning and Teaching*, 65-85.
- Vázquez Veiga, N. (2014). Marcas de uso en el Diccionario de colocaciones del español. *Zeitschrift für romanische Philologie*, 130(3), 698-724.
- Vilkaite, L., & Schmitt, N. (2019). Reading collocations in an L2: Do collocation processing benefits extend to non-adjacent collocations? *Applied Linguistics*, 40(2), 329-354.
- Vincze, O., Alonso Ramos, M., & Ramos, M. A. (2013). Incorporating frequency information in a collocation dictionary: Establishing a methodology. *Procedia - Social and Behavioral Sciences*.
- Vincze, O., García-Salido, M., Orol-González, A., & Alonso-Ramos, M. (2016). A corpus study of Spanish as a Foreign Language learners' collocation production. En *Spanish Learner Corpus Research*, 299-331.

- Vitevitch, M. S., Goldstein, R., Siew, C. S. Q., & Castro, N. (2015). Using complex networks to understand the mental lexicon. *Yearbook of the Poznan Linguistic Meeting*, 1(1), 119-138.
- Wall, D. (2012). Washback in Language Assessment. *The Encyclopedia of Applied Linguistics*, 1987.
- Webb, S. (2012). Depth of Vocabulary Knowledge. En *The Encyclopedia of Applied Linguistics*.
- Webb, S., & Nation, P. (2017). *How Vocabulary Is Learned*. Oxford University Press.
- Webb, S., Newton, J., & Chang, A. (2013). Incidental Learning of Collocation. *Language Learning*, 63(1), 91-120.
- Webb, S., Sasao, Y., & Ballance, O. (2017). The updated Vocabulary Levels Test. *ITL - International Journal of Applied Linguistics*, 168(1), 33-69.
- Wesche, M, & Paribakht, T. S. (1994). *Enhancing Vocabulary Acquisition through Reading: A Hierarchy of Text-Related Exercise Types*.
- Wesche, Marjorie, & Paribakht, T. S. (1996). Assessing Second Language Vocabulary Knowledge: Depth Versus Breadth. *Canadian Modern Language Review*, 53(1).
- Willis, M., & Ohashi, Y. (2012). A model of L2 vocabulary learning and retention. *The Language Learning Journal*, 40(1), 125-137.
- Wolter, B. (2002). Assessing proficiency through word associations: Is there still hope? *System*, 30(3), 315-329.
- Wolter, B., & Gyllstad, H. (2011). Collocational Links in the L2 Mental Lexicon and the Influence of L1 Intralexical Knowledge. *Applied Linguistics*, 32(4), 430-449.
- Wolter, B., & Gyllstad, H. (2013). Frequency of Input and L2 Collocational Processing. *Studies in Second Language Acquisition*, 35(03), 451-482.
- Wolter, B., & Yamashita, J. (2014). Processing collocations in a second language: A case of first language activation? *Applied Psycholinguistics*, 36(5), 1193-1221.
- Wolter, B., & Yamashita, J. (2018). Word Frequency, Collocational Frequency, L1

- Congruency, and Proficiency in L2 Collocational Processing: What Accounts for L2 Performance? *Studies in Second Language Acquisition*.
- Wray, A. (2013). Formulaic language. *Language Teaching*, 46(3), 316-334.
- Yanagisawa, A., & Webb, S. (2020). Measuring Depth of Vocabulary Knowledge. En S. Webb (Ed.), *The Routledge Handbook of Vocabulary Studies* (pp. 371-386). Routledge Handbooks.
- Yang, Y., & Green, S. B. (2011). Coefficient alpha: A reliability coefficient for the 21st century? *Journal of Psychoeducational Assessment*, 29(4), 377-392.
- Zaabalawi, R. S., & Gould, A. M. (2017). English collocations: A novel approach to teaching the language's last bastion. *Ampersand*, 4, 21-29.
- Zareva, A. (2010). Multicompetence and L2 users' associative links: Being unlike nativelike. *International Journal of Applied Linguistics*, 20(1), 2-22.
- Zareva, A., & Wolter, B. (2012). The «promise» of three methods of word association analysis to L2 lexical research. *Second Language Research*, 28(1), 41-67.
- Zhang, D., & Koda, K. (2017). Assessing L2 vocabulary depth with word associates format tests: issues, findings, and suggestions. *Asian-Pacific Journal of Second and Foreign Language Education*, 2(1), 1-30.
- Zhang, H., & Pei, Z. (2022). Word Knowledge Dimensions in L2 Lexical Inference: Testing Vocabulary Knowledge and Partial Word Knowledge. *Journal of Psycholinguistic Research*, 51(1), 151-168.
- Zhang, M., Li, W., & Zhang, H. (2014). *Paraphrase Collocations Extraction Based on Concept Expansion* (pp. 191-199). Springer, Berlin, Heidelberg.
- Zhong, H. F. (2018). The relationship between receptive and productive vocabulary knowledge: a perspective from vocabulary use in sentence writing. *Language Learning Journal*, 46(4), 357-370.

ANEXOS

ÍNDICE DE ANEXOS

ANEXO I. Recopilación de test de léxico	329
ANEXO II. Recopilación de test de colocaciones	330
ANEXO III. ESPECIFICACIONES DEL TCE	331
Primer documento: GUÍA DE ADMINISTRACIÓN DE LA PRUEBA.....	331
Segundo documento: ESTRUCTURA DEL TCE	331
ANEXO IV. CRITERIOS DE CORRECCIÓN: ESCALA DE CLASIFICACIÓN (O NIVELACIÓN)	332
ANEXO V. CRITERIOS DE CORRECCIÓN: Rúbrica de gradación del conocimiento colocacional comprensivo.....	333
ANEXO VI. CRITERIOS DE CORRECCIÓN: Rúbrica de gradación del conocimiento colocacional prductivo	334
ANEXO VII. BIOGRAFÍA LINGÜÍSTICA.....	335
ANEXO VIII. X-LEX.....	336
ANEXO IX. Muestra del preTCEp	1
ANEXO X. Muestra TCEp	2
ANEXO XI. Muestra TCE.....	3
ANEXO XII. preTCEp PRECU.....	4
ANEXO XIII. Hoja de respuesta de PRECU en el preTCEp.....	5
ANEXO XV. preTCEp CRECO	6
ANEXO XVI. Hoja de respuesta de CRECO en el preTCEp.....	10
ANEXO XVII. preTCEp PRECO	11
ANEXO XVIII. Hoja de respuestas preTCEp PRECO	15
ANEXO XIX. preTCE CRECU	16
ANEXO XX. Hoja de respuestas preTCEp CRECU.....	17
ANEXO XXI. TCEp PRECU	18
ANEXO XXII. TCEp CRECO	21
ANEXO XXIII. TCEp PRECO	25
ANEXO XIV. TCEp CRECU	29
ANEXO XXV. TCE PRECU	31
ANEXO XXVI. TCE CRECO	32
ANEXO XXVII. TCE PRECO	35
ANEXO XXVIII. TCE CRECU.....	37
ANEXO XXIX. INFORME FAVORABLE DEL COMITÉ DE ÉTICA	38
ANEXO XXX. DOCUMENTO INFORMATIVA PARA LOS APRENDICES	39
ANEXO XXXI. HOJA DE CONSENTIMIENTO.....	41

ANEXO I. RECOPIACIÓN DE TEST DE LÉXICO

TEST		AUTOR	P/C	R/R	FORMATO
Vocabulary Level Test (VLT)		Nation	COMPRESIÓN	RECONOCIMIENTO	RELACIONAR
Vocabulary Size Test (VST)		Nation	COMPRESIÓN	RECONOCIMIENTO	RELACIONAR
Productive Vocabulary Levels Test (PVLТ)		Laufer	PRODUCCIÓN	RECUPERACIÓN	COMPLETAR
X-Lex y P-Lex		Meara & Milton	COMPRESIÓN	RECONOCIMIENTO	LISTA DE VERIFICACIÓN
Word Associates Test (WAT)		Read	PRODUCCIÓN	RECUPERACIÓN	RELACIÓN (DESDE LA RECUPERACIÓN)
LEX 30		Zareva & Wolter	PRODUCCIÓN	RECUPERACIÓN	ASOCIACIÓN SINTAGMÁTICA Y PARADIGMÁTICA
Vocabulary Knowledge Scale (VKS)		Wesche & Paribakht	COMPRESIÓN	RECONOCIMIENTO Y RECUPERACIÓN	DIFERENTES TAREAS
The Four Diagnostic Vocabulary Test	<i>Vocabulary Size Test</i>	Ishii & Schmitt	COMPRESIÓN		RELACIONAR (TRADUCCIÓN)
	<i>Test of Multiple Meaning Senses for Words</i>		COMPRESIÓN		ELECCIÓN MÚLTIPLE
	<i>Test of Derivative Word Forms</i>		PRODUCCIÓN		CURBRI UNA TABLA
	<i>Test of Lexical Choice between Near-Synonyms</i>		COMPRESIÓN		RELLENAR HUECO
CATSS	Producción desde la recuperación	Laufer & Goldstein	PRODUCCIÓN Y COMPRESIÓN	RECOGNITION Y RECALL	
	Comprensión desde la recuperación				SINÓNIMO
	Producción desde el reconocimiento				ELECCIÓN MÚLTIPLE
	Comprensión desde el reconocimiento				ELECCIÓN MÚLTIPLE
Word class		Nissen y Henriksen	COMPRESIÓN	RECALL	RELACIONAR (TRADUCCIÓN)

ANEXO II. RECOPIACIÓN DE TEST DE COLOCACIONES

		AÑO	COMPRENSIÓN / PRODUCCIÓN	RECONOCIMIENTO / RECUPERACIÓN	CONTEXTUALIZADO / DESCONTEXTUALIZADO	ESQUEMA GRAMATICAL	N.º DE ÍTEMS	L1	L2	DEFINICIÓN DE COLOCACIÓN	Correlaciones
Primera generación	COLLEX (Gyllstad)	2009 (2007)	COMPRENSIÓN	RECONOCIMIENTO	DESCONTEXTUALIZADO	V + N	50	SUECO	INGLÉS	ENFOQUE COMBINADO	CC vs. CLéx COLLEX VS COLLMACH
	COLLMACH (Gyllstad)	2009 (2007)	COMPRENSIÓN	RECONOCIMIENTO	DESCONTEXTUALIZADO	V + N	100	SUECO	INGLÉS		
	CONTRIX (Lee Revier)	2009	PRODUCCIÓN	RECONOCIMIENTO (DE LA BASE) Y RECUPERACIÓN (DEL COLOCATIVO)	CONTEXTUALIZADO	V + N	45	DANÉS	INGLÉS		-
	DISCO (Ayckmans)	2009	COMPRENSIÓN	RECONOCIMIENTO	DESCONTEXTUALIZADO	V + N	25	HOLANDÉS, ALEMÁN, FRANCÉS	INGLÉS	ENFOQUE COMBINADO	-
	Lex Combi (Boers)	2009	PRODUCCIÓN	RECUPERACIÓN	DESCONTEXTUALIZADO	N + V / A	30	JAPONÉS	INGLÉS	NO HAY DEFINICIÓN	-
	ADELEX (Moreno Jaén)	2007	PRODUCCIÓN	RECUPERACIÓN	CONTEXTUALIZADO	V+N y N+A	300	ESPAÑOL	INGLÉS	ENFOQUE COMBINADO	CC vs. CLing
PRODUCCIÓN	RECUPERACIÓN		DESCONTEXTUALIZADO								
PRODUCCIÓN (INVERSA) ¹²⁸	RECONOCIMIENTO		DESCONTEXTUALIZADO								
PRODUCCIÓN	RECONOCIMIENTO		CONTEXTUALIZADO								
Segunda generación	LexCombi2 (Brown)	2014, 2018	PRODUCCIÓN	RECUPERACIÓN	DESCONTEXTUALIZADO	Variado		JAPONÉS	INGLÉS	ENFOQUE BASADO EN FRECUENCIAS	Identificación estadística
	Nizonkiza	2014	COMPRENSIÓN	RECONOCIMIENTO	CONTEXTUALIZADO	V + N	40	DIFERENTES L1 ¹²⁹	INGLÉS	ENFOQUE BASADO EN FRECUENCIAS	CC vs. CLing CC VS frecuencia
		2015	PRODUCCIÓN	RECUPERACIÓN	CONTEXTUALIZADO	V + N	40				
	González Fernández & Schmitt	2015	PRODUCCIÓN	RECUPERACIÓN	CONTEXTUALIZADO	-	50	ESPAÑOL	INGLÉS	ENFOQUE COMBINADO	CC vs. grado de exposición CC VS frecuencia
	Nguyen y Webb	2017	COMPRENSIÓN	RECONOCIMIENTO	DESCONTEXTUALIZADO	V+N A+N	180	VIETNAMITA	INGLÉS	ENFOQUE BASADO EN FRECUENCIAS	CC vs. CLéx CC vs. frecuencia
TCPCT (Frankenberg- García, Sonbul)	2018, 2022	PRODUCCIÓN	RECUPERACIÓN	CONTEXTUALIZADO	V+N	31	DIFERENTES L1	INGLÉS	ENFOQUE BASADO EN FRECUENCIAS	CCvs. MCER	

¹²⁸ Denominamos producción inversa porque es la colocación errónea la que tienen que identificar los aprendices.

¹²⁹ Los participantes son procedentes de Burundi (kirundi, francés, etc.) y de Bélgica, países en los que se hablan varias lenguas. En el test de comprensión, solo con alumnado de Burundi.

ANEXO III. ESPECIFICACIONES DEL TCE

Primer documento: GUÍA DE ADMINISTRACIÓN DE LA PRUEBA

Guía de administración de la prueba			
	VERSIÓN PRESENCIAL	VERSIÓN DIGITAL	TEMPORALIZACIÓN
Primer paso	Explicación del docente	Lectura e la Guía e participación en el TCE	
Segundo paso	Biografía lingüística		5 MINUTOS
Tercer paso	X-LEX		10 MINUTOS
Cuarto paso	Primera modalidad del TCE: PRECU o PRECO		10 / 20 MINUTOS
Quinto paso	Segunda modalidad del TCE: CRECO o CRECU		10 / 20 MINUTOS
Sexto paso	Firma de la hoja e consentimiento	Cubrir la hoja de consentimiento y enviarla	5 MINUTOS



Segundo documento: ESTRUCTURA DEL TCE

Estructura del TCE				
	Opción A		Opción B	
Modalidad	PRECU	CRECO	PRECO	CRECU
N.º de preguntas	36	36	36	36
Puntuación máxima	36	36	36	36
Tipo de tarea	Completar	Opción múltiple	Opción múltiple	Completar
Foco de la evaluación: destreza y proceso cognitivo	Producción desde la recuperación	Comprensión desde el reconocimiento	Producción desde el reconocimiento	Comprensión desde la recuperación
Tiempo de duración de cada prueba	10 MINUTOS	20 MINUTOS	20 MINUTOS	10 MINUTOS

ANEXO IV. CRITERIOS DE CORRECCIÓN: ESCALA DE CLASIFICACIÓN (O NIVELACIÓN)

		INICIAL		MEDIO		SUPERIOR	
MCER	ESCALA DEL TCE	A1	A2	B1	B2	C1	C2
C2	36						
C1	30						
B2	24						
B1	18						
A2	12						
A1	6						

ANEXO V. CRITERIOS DE CORRECCIÓN: Rúbrica de gradación del conocimiento colocacional comprensivo

	DESCONOCIMIENTO	CONOCIMIENTO		
		PARCIAL		TOTAL
		GRADO 0	GRADO 1	GRADO 2
BASE	No conoce la base No reconoce ni la forma ni la estructura morfológica de la palabra que funciona como base.	Conoce la base parcialmente. Reconoce la forma oral o escrita de la palabra que funciona como base. No tiene por qué conocer la estructura morfológica.	Conoce la base	Conoce la base Conoce el colocativo
VÍNCULO	No se conoce el vínculo (no se puede inferir) El marco de conocimiento no permite sacar conclusiones.	Infiere el vínculo El vínculo se puede inferir al leer o escuchar la colocación contextualizada. Esta inferencia se puede realizar	Infiere o deduce el vínculo	Conoce el vínculo
COLOCATIVO	No se conoce el colocativo No se reconoce la forma del colocativo. También podría reconocerse la forma del colocativo, pero tratarse de un colocativo X, de modo que al no conocer ni poder inferir el vínculo, no se puede extraer el sentido de la colocación.	Conoce el colocativo parcialmente	Conoce el colocativo parcialmente	Conoce el colocativo
	No conoce la colocación	Conoce la colocación parcialmente		Conoce la colocación en su grado óptimo

ANEXO VI. CRITERIOS DE CORRECCIÓN: Rúbrica de gradación del conocimiento colocacional productivo

	DESCONOCIMIENTO	CONOCIMIENTO		
		PARCIAL		TOTAL
	GRADO 0	GRADO 1	GRADO 2	GRADO 3
BASE	No conoce la base	Conoce la base parcialmente	Conoce la base	Conoce la base
VÍNCULO	No se conoce el vínculo (no se puede inferir)	Infiere o deduce el vínculo	Infiere o deduce el vínculo	Conoce el vínculo
COLOCATIVO	No se conoce el colocativo	Conoce el colocativo parcialmente	Conoce el colocativo parcialmente	Conoce el colocativo
	No conoce la colocación	Conoce la colocación parcialmente		Conoce la colocación en su grado óptimo

ANEXO VII. BIOGRAFÍA LINGÜÍSTICA

Responde a las siguientes preguntas de la forma más clara posible

¿De dónde eres?

(Escribe el nombre del país donde naciste)

¿A qué te dedicas?

- a) Estudias
- b) Trabajas
- c) Otro

¿Cuál es tu lengua materna?

(Escribe el nombre de la lengua que hablas con tu familia)

¿Dónde estudias español?

(Escribe el nombre de la escuela o centro de estudios)

¿Qué nivel de español estudias?

- a) A1
- b) A2
- c) B1
- d) B2
- e) C1
- f) C2

¿Qué otras lenguas hablas?

¿Cuánto tiempo llevas en España?

- a) Menos de un mes
- b) Menos de un año
- c) Un año o más

ANEXO VIII. X-LEX

X-Lex Vocabulary Test

Please look at these words. Some of these words are **real** Spanish words and some are **invented** but are made to look like real words. Please tick the words that **you know or can use**. Here is an example.

perro ✓

agua	prueba	destacar	adecuado	ceñir	tricodir
afilar	gusano	emplícito	cuando	gabinete	galería
clase	recio	encender	chisco	contener	curva
astro	besar	parecer	trofeo	cohete	segarno
cerca	invierno	enamorado	barca	heredar	torreroso
ejeste	grueso	corbata	trasladar	cifra	verdad
triumfo	programa	disfrazar	manchar	humo	alcadernal
meter	flecha	amanecer	matiz	anochece	planchete
chocolate	salud	gitano	patada	ampallar	incorporar
acostar	cinturón	custodionar	delicioso	letrero	chimenea
taurete	violeta	almendra	inicial	sutil	habitación
mangual	tocador	sartén	eficaz	pieza	fiel
hospital	encuentro	distraer	fatal	secreto	cobrosamente
abarcarse	estancioso	martillo	garganta	perdonar	guisar
nada	afuera	mendigo	conformar	gallardo	emprestar
solemne	semejante	mudo	arpa	pequeño	eloaje
escuela	sapo	disparar	sordo	tentar	postismo
vaca	torre	globo	chorro	absurdo	rebodondo
temblor	remolchete	hueso	aire	creyente	consolar
desesperar	sorpende	mayorador	aspiración	vinagre	cinco

ANEXO IX. Muestra del preTCEp

Muestra <i>preTCEp</i>	
1	HACER LOS DEBERES
2	TENER BARBA
3	HACER FRÍO
4	TELÉFONO FIJO
5	PELO RUBIO
6	CAFÉ CORTADO
7	COMETER ERRORES
8	TENER SED
9	TENER CITA
10	PELO LISO
11	COMIDA RÁPIDA
12	CIUDAD COSMOPOLITA
13	TENER UNA DUDA
14	DAR PENA
15	PONER LA MESA
16	CIELO DESPEJADO
17	CALOR HORRIBLE
18	ZAPATO PLANO
19	SUFRIR UNA LESIÓN
20	HACER UN BRINDIS
21	PONER UN ENCHUFE
22	VIEJA AMIGA
23	LECHE ENTENERA
24	QUESO FRESCO
25	TENER ASPIRACIONES
26	INSPIRAR CONFIANZA
27	DICTAR SENTENCIA
28	AMOR IMPOSIBLE
29	BARBA POBLADA
30	ÉXITO APOTEÓSICO
31	DAR GRIMA
32	MONTAR BARULLO
33	PILLAR UN CATARRO
34	COMPLEXIÓN FUERTE
35	ENVIDIA SANA
36	PERSONALIDAD DESBORDANTE

ANEXO X. Muestra TCEp¹³⁰

Muestra TCEp			
1	HACER FRÍO	37	PONER UN ENCHUFE
2	TENER BARBA	38	HACER UN BRINDIS
3	HACER EL EXAMEN	39	BORRAR LAS HUELLAS
4	LLEVAR GAFAS	40	PONER UN HUEVO
5	PONER EL NOMBRE	41	NORMA ESTABLECIDA
6	HACERSE VIEJO	42	CUMPLIR UN SUEÑO
7	PELO RUBIO	43	LECHE ENTERA
8	CAFÉ CORTADO	44	QUESO FRESCO
9	HABITACIÓN DOBLE	45	FAMILIA NUMEROSA
10	DESAYUNO FUERTE	46	OLOR INTENSO
11	BUEN TIEMPO	47	VIEJA AMIGA
12	TELÉFONO FIJO	48	MAL OLOR
13	COMETER UN ERROR	49	TENER ASPIRACIONES
14	TENER SED	50	DICTAR SENTENCIA
15	PAGAR LA MATRÍCULA	51	TENER CONTROL
16	DAR PROPIA	52	SACAR EL CARNÉ
17	PEDIR EL MENÚ	53	INSPIRAR CONFIANZA
18	TENER CITA	54	ABRIR FUEGO
19	VINO TINTO	55	BARBA POBLADA
20	COMIDA RÁPIDA	56	OJO RASGADO
21	TALLA GRANDE	57	ÉXITO APOTEÓSICO
22	COLOR CLARO	58	DEPRESIÓN PROFUNDA
23	BUNEA LETRA	59	SOL RADIANTE
24	PLATO RICO	60	RELACIÓN TORMENTOSA
25	TENER UNA DUDA	61	PILLAR UN CATARRO
26	HACER COLA	62	MONTAR BARULLO
27	SACAR UNA FOTO	63	CONTRAER MATRIMONIO
28	ATENDER LLAMADA	64	PRESENTAR SOLICITUD
29	DAR PENA	65	PROVOCAR UN INCENDIO
30	PONER LA MESA	66	AGRIAR EL CARÁCTER
31	ZAPATO PLANO	67	EXTREMIDAD INFERIOR
32	RELOJ ANALÓGICO	68	LUNA NUEVA
33	TOS AGUDA	69	COMPLEXIÓN FUERTE
34	HACER RUIDO	70	PERSONALIDAD DESBORDANTE
35	CIELO DESPEJADO	71	ENVIDIA SANA
36	CALOR HORRIBLE	72	SUEÑO LIGERO

¹³⁰ Marcamos en gris las casillas con las colocaciones que proceden de la lista anterior.

ANEXO XI. Muestra TCE

Muestra TCE	
1	HACER FRÍO
2	TENER BARBA
3	LLEVAR GAFAS
4	PELO RUBIO
5	HABITACIÓN DOBLE
6	CAFÉ CORTADO
7	TENER SED
8	DAR PROPINA
9	TENER CITA
10	VINO TINTO
11	BUENA LETRA
12	COMIDA RÁPIDA
13	HACER COLA
14	SACAR FOTOS
15	PONER LA MESA
16	ZAPATOS PLANOS
17	CALOR HORRIBLE
18	RUIDO FUERTE
19	CUMPLIR UN SUEÑO
20	HACER UN BRINDIS
21	BORRAR LAS HUELLAS
22	LECHE ENTERA
23	FAMILIA NUMEROSA
24	VIEJA AMIGA
25	DICTAR SENTENCIA
26	PASAR EL CONTROL
27	INSPIRAR CONFIANZA
28	ÉXITO APOTEÓSICO
29	SOL RADIANTE
30	RELACIÓN TORMENTOSA
31	PILLAR UN CATARRO
32	MONTAR BARULLO
33	CONTRAER MATRIMONIO
34	PERSONALIDAD DESBORDANTE
35	ENVIDIA SANA
36	SUEÑO LIGERO

ANEXO XII. preTCEp PRECU

INSTRUCCIONES: Escribe la palabra que falta en las siguientes secuencias.

1. Aunque Manolo saca buenas notas, nunca **h... los** deberes.
2. Mi hermano ya **t... barba**, necesita afeitarse.
3. **H... frío**, por favor, cierra la ventana.
4. Juan, tu hermano no está en casa así que no llames al **teléfono f...**
5. Mi hermana tiene el **pelo r...**, pero a ella le gusta negro.
6. Pepe siempre pide un **café c...**, le gusta con poca leche.
7. María siempre **c... errores**, se equivoca todos los días.
8. ¡Ay!, **t... sed**, necesito un vaso de agua.
9. Esta tarde Lucía **t... cita** en el médico.
10. Sandra tiene el **pelo l...** . A ella le gustaría tenerlo rizado.
11. No es saludable, pero me encanta la **comida r...**
12. Tiene **b... letra**, todo el mundo la entiende.
13. Cuando un alumno **t... una duda**, levanta la mano para preguntar.
14. A Javier le **d... pena** la situación de mi hermano. Le gustaría ayudarlo.
15. Jaimito, **p... la mesa**, le manda su abuela antes de comer.
16. El **cielo** está **d...**, no hay ninguna nube.
17. Hace un **calor h...**, necesito un abanico.
18. María lleva todos los días **zapato p...**, nunca usa tacones.
19. No puede jugar la final, ha **s... una lesión** en la rodilla.
20. Vamos a **h... un brindis**, levantad las copas y bebed.
21. No sé **p... un enchufe**, por favor, llama a un electricista.
22. Lucía es una **v... amiga**, la conozco desde hace mucho tiempo.
23. María siempre toma **leche e...**, a ella no le importa engordar.
24. Prefiere el **queso f...**: se preocupa por su dieta y este tiene menos calorías.
25. Alfredo **t... unas aspiraciones altas**, desea algo imposible.
26. Tu nuevo novio no me **i... confianza**, no me fío nada de él.
27. El juez **d... sentencia**, condenó al preso a tres años de cárcel.
28. Tu relación con Pepe es un **amor i...**, no es correspondido.
29. Pedro tiene la **barba p...**, con mucho pelo.
30. Su primer libro fue un **éxito a...**, fue uno de los más vendidos del trimestre.
31. Hay sonidos que me **d... grima**, por ejemplo la tiza en el encerado.
32. Parad de **m... barullo**, estáis gritando mucho.
33. **P... un catarro**, creo que voy a tomar una aspirina.
34. Los nadadores tienen **complexión f...**, son grandes.
35. Pedro siente **envidia s...** por María. Se alegra por ella, pero le gustaría tener su mismo éxito.
36. Mario tiene una **personalidad d...**, es extrovertido y habla mucho.

ANEXO XIII. Hoja de respuesta de PRECU en el preTCEp

Ejemplo

0. Llevo una **colonia f...**, me gustan los olores suaves.

0.	F	R	E	S	C	A							
----	---	---	---	---	---	---	--	--	--	--	--	--	--

1.	H												
2.	T												
3.	H												
4.	F												
5.	R												
6.	C												
7.	C												
8.	T												
9.	T												
10.	L												
11.	R												
12.	B												
13.	T												
14.	D												
15.	P												
16.	D												
17.	H												
18.	P												
19.	S												
20.	H												
21.	P												
22.	V												
23.	E												
24.	F												
25.	T												
26.	I												
27.	D												
28.	I												
29.	P												
30.	A												
31.	D												
32.	M												
33.	P												
34.	F												
35.	S												
36.	D												

ANEXO XV. preTCEp CRECO

INSTRUCCIONES: Elige la opción más adecuada para completar la secuencia.

1. Manolito nunca **hace los deberes**, _____.
 - a. hace las actividades que le manda su profesora
 - b. hace lo que le mandan sus padres
 - c. no le hace caso a la profesora
 - d. no cumple con las obligaciones de sus padres
2. Mi hermano ya **tiene barba**, _____.
 - a. necesita afeitarse
 - b. no necesita afeitarse
 - c. tiene poco pelo en la cara
 - d. no se deja pelo en la cara
3. **Hace frío**, por favor _____.
 - a. cierra la ventana
 - b. abre las cortinas
 - c. apaga la calefacción
 - d. pon el aire acondicionado
4. Juan, no llames al **teléfono fijo** que mis padres _____.
 - a. están en casa
 - b. no están en casa
 - c. hablan mucho
 - d. hablan todos los días
5. Mi hermana tiene el **pelo rubio**, _____.
 - a. pero le gustaría tenerlo negro
 - b. de color blanco
 - c. de color negro
 - d. de color rojo
6. Pepe siempre pide un **café cortado**, _____.
 - a. la mitad de la taza le llega
 - b. no le echa azúcar
 - c. le gusta con poca leche
 - d. le añada agua
7. María **comete errores** todos los días, _____.
 - a. siempre se equivoca
 - b. nunca se equivoca
 - c. hace todo bien
 - d. no falla nunca
8. **Tengo sed**, _____.
 - a. quiero emborracharme
 - b. bebo dos litros de agua al día
 - c. solo me gusta el zumo de naranja
 - d. quiero un vaso de agua
9. Lucía **tiene cita** a las cinco, _____.
 - a. va a visitar al médico
 - b. va a verse con unos amigos
 - c. quedó con unos amigos
 - d. no quedó con nadie

10. Sandra tiene el **pelo liso**, _____.
- a. con muchas canas
 - b. de color negro
 - c. sin rizos
 - d. de un solo color
11. Me gusta mucho la **comida rápida**, aunque _____.
- a. no es sana
 - b. es sana
 - c. es casera
 - d. sea de dieta
12. Tiene **buena letra**, _____.
- a. nadie entiende lo que escribe
 - b. todo el mundo entiende lo que escribe
 - c. es muy bonita
 - d. es muy fea
13. Cuando un alumno **tiene un duda**, _____.
- a. sospecha que algo va a salir mal
 - b. elige una de las opciones posibles
 - c. desconfía de su profesora
 - d. levanta la mano para preguntar
14. A Javier le **da pena** su hermano, _____.
- a. piensa en como poder ayudarlo
 - b. no piensa ayudarlo
 - c. le parece una condena
 - d. se alegra por él
15. Jaimito, **pon la mesa**, _____.
- a. le ordena su madre antes de comer
 - b. aquí tienes las instrucciones
 - c. aquí tienes los instrumentos
 - d. antes de que todos se levanten de la mesa
16. El **cielo** está **despejado**, _____.
- a. va a empezar a llover
 - b. va a empezar a nevar
 - c. no hay nubes
 - d. hace mucho calor
17. Hace un **calor horrible**, _____.
- a. necesito un paraguas
 - b. necesito un abrigo
 - c. necesito un radiador
 - d. necesito un abanico
18. María usa todos los días **zapato plano**: _____.
- a. nunca usa tacones
 - b. es bajita e intenta parecer más alta
 - c. va siempre con tacones
 - d. tiene los pies pequeños
19. Messi **ha sufrido una lesión** en la rodilla, _____.
- a. el entrenador lo castiga en el banquillo
 - b. está sancionado
 - c. va a jugar el próximo partido
 - d. no podrá disputar los próximos partidos

20. Vamos a **hacer un brindis**, _____.
- a. por el éxito de la empresa
 - b. por lástima
 - c. por necesidad
 - d. por el fracaso de la empresa
21. El cable no llega y no sé **poner un enchufe**, así que _____.
- a. llama al electricista
 - b. llama al carpintero
 - c. llama al fontanero
 - d. llama al albañil
22. Lucía es una **vieja amiga**, _____.
- a. tiene más de sesenta años
 - b. hace muchos años que es mi amiga
 - c. la conocí hace dos meses
 - d. todo el mundo la ve más vieja de lo que es
23. María siempre toma **leche entera**: _____.
- a. quiere adelgazar y siempre toma todo sin grasa
 - b. a ella no le importa engordar
 - c. le gusta la comida dietética
 - d. le gustan los productos enteros, no en partes
24. Prefiere el **queso fresco**: _____.
- a. tiene un sabor fuerte y le gusta que la comida tenga mucho sabor
 - b. mira la calidad y solo compra productos caros
 - c. se preocupa por su dieta y este queso tiene menos calorías
 - d. nunca come productos congelados, prefiere los de reciente elaboración
25. Alfredo **tiene aspiraciones altas**, _____.
- a. cree que todo le va a salir mal
 - b. cree que va a subir hasta el último piso
 - c. tiene objetivos difíciles de conseguir
 - d. desea algo imposible
26. Tu nueva novia no me **inspira confianza**, _____.
- a. no me fío nada de ella
 - b. me causa muy buena impresión
 - c. es mi musa
 - d. me cae bien
27. El juez **dictó sentencia**, _____.
- a. aplazó el juicio
 - b. la secretaria copió todo lo que decía
 - c. sus palabras son sabias
 - d. condenó al preso a tres años de cárcel
28. Tu relación con Pepe es un **amor imposible**, _____.
- a. es un amor correspondido
 - b. seguro que se enamora de ti
 - c. es un amor no correspondido
 - d. te da poco cariño
29. Pedro tiene una **barba poblada**, _____.
- a. igual que la del resto de la población
 - b. con mucho pelo
 - c. con una finca llena de árboles
 - d. recién afeitada

30. Su primer libro fue un **éxito apoteósico**, _____.
- a. nadie lo compró
 - b. le gustó a mucha gente
 - c. nadie pudo publicar otro libro
 - d. no se acabó la primera edición
31. Hay sonidos que me **dan grima**, por ejemplo _____.
- a. la tiza en el encerado
 - b. la música de una guitarra
 - c. la lluvia en la ventana
 - d. las olas del mar
32. Parad de **montar barullo** que _____.
- a. estáis gritando mucho
 - b. estáis muy cansados
 - c. está muy desordenado
 - d. nunca habláis tan bajo
33. **Pillé un catarro**, creo que _____.
- a. voy a tomar una aspirina
 - b. tengo muy buena salud
 - c. voy a salir de fiesta
 - d. ya estoy bien
34. Los nadadores son de **complexión fuerte**, _____.
- a. tienen poco músculo
 - b. tienen mucho músculo
 - c. son delgados
 - d. están gordos
35. Pedro siente **envidia sana** por María, _____.
- a. se cambiaría por ella
 - b. no le gustaría estar en su lugar
 - c. se alegra por María, pero quiere su éxito
 - d. no se alegra por María, pero valora su éxito
36. Mario tiene una **personalidad desbordante**, _____.
- a. es tímido y habla poco
 - b. es extrovertido y habla mucho
 - c. es tímido, pero gracioso
 - d. es extrovertido, pero pesado

ANEXO XVI. Hoja de respuesta de CRECO en el preTCEp

Ejemplo

0. Todos los días me echo **colonia fresca**,
- a. me gustan los olores fuertes.
 - b. me gustan las colonias de hombre.
 - c. me gustan los olores suaves.
 - d. no me gusta echarme colonia.

	a.	b.	c.	d.
0.			X	

	a.	b.	c.	d.
1.				
2.				
3.				
4.				
5.				
6.				
7.				
8.				
9.				
10.				
11.				
12.				
13.				
14.				
15.				
16.				
17.				
18.				
19.				
20.				
21.				
22.				
23.				
24.				
25.				
26.				
27.				
28.				
29.				
30.				
31.				
32.				
33.				
34.				
35.				
36.				

ANEXO XVII. preTCEp PRECO

INSTRUCCIONES: Elige la opción correcta para completar las secuencias.

1. Aunque Manolo tiene buenas notas, nunca _____.
 - a. hace los deberes
 - b. descuida los deberes
 - c. abandona sus deberes
 - d. incumple sus deberes
2. Mi hermano necesita afeitarse, ya _____.
 - a. tiene el pelo largo
 - b. tiene barba
 - c. es calvo
 - d. tiene melena
3. Por favor, ponte el abrigo _____.
 - a. hace calor
 - b. ya no voy a salir
 - c. voy a salir ahora
 - d. hace frío
4. Mis padres no están casa, Juan no los llames _____.
 - a. al teléfono portátil
 - b. al teléfono móvil
 - c. al teléfono fijo
 - d. al teléfono inalámbrico
5. A mi hermana le gustaría tener el pelo rizo, pero tiene el _____.
 - a. pelo canoso
 - b. pelo negro
 - c. pelo liso
 - d. pelo rubio
6. A Pepe le gusta el café con un poco de leche. Siempre pide un _____.
 - a. café cortado
 - b. café con leche
 - c. café solo
 - d. café americano
7. María se equivoca siempre, todos los días _____.
 - a. induce a error
 - b. corrige errores
 - c. comete errores
 - d. comete delitos
8. Quiero un vaso de agua, _____.
 - a. tengo hambre
 - b. tengo sueño
 - c. tengo sed
 - d. tengo necesidades
9. María va al médico a las cinco, _____.
 - a. escribe una cita
 - b. solicita una cita
 - c. tiene cita
 - d. lee la cita
10. Aunque le gusta el pelo rizado, ella tiene el _____.
 - a. pelo encrespado
 - b. pelo castaño

- c. pelo rubio
 - d. pelo liso
11. No es saludable, pero me gusta mucho la _____.
- a. comida casera
 - b. comida ecológica
 - c. comida mediterránea
 - d. comida rápida
12. Todo el mundo entiende lo que escribe, tiene _____.
- a. buena letra
 - b. mala letra
 - c. letra bonita
 - d. letra fea
13. Cuando un alumno levanta la mano para preguntar, quiere decir que _____.
- a. tiene una sospecha
 - b. tiene una duda
 - c. no sabe qué decir
 - d. se estira
14. A Javier, la triste situación de mi hermano _____.
- a. le da pena
 - b. le parece graciosa
 - c. le da miedo
 - d. le parece simpática
15. Antes de comer, todos colocaron los platos, vasos y cubiertos, _____.
- a. todos pusieron la mesa
 - b. todos recogieron la mesa
 - c. todos montaron la mesa
 - d. todos movieron la mesa
16. No hay ninguna nube, _____.
- a. el cielo está nublado
 - b. el cielo está despejado
 - c. el cielo está cerrado
 - d. el cielo está negro
17. Necesito un abanico, _____.
- a. hace un calor agradable
 - b. hace un calor horrible
 - c. hace calor ligero
 - d. hace fresco
18. María nunca usa zapatos altos, prefiere _____.
- a. zapato de tacón
 - b. zapato plano
 - c. zapato sencillo
 - d. zapato de cordones
19. Messi pidió el cambio porque _____.
- a. sufrió una lesión
 - b. se recuperó de una lesión
 - c. sufrió un daño moral
 - d. recuperó la posición
20. Bebamos por el éxito de tu empresa, dijo mientras _____.
- a. hacían unas copas
 - b. hacían un brindis
 - c. tenían un choque
 - d. daban un brindis

21. No llega el cable hasta ahí. Hay que llama para _____.
- poner una puerta
 - poner un ladrillo
 - poner un enchufe
 - poner un grifo
22. Conozco a Lucía desde hace muchos años. Es una _____.
- joven amiga
 - vieja amiga
 - amiga vieja
 - amiga joven
23. A María no le importa engordar. Le gusta beber _____.
- leche frita
 - leche entera
 - leche desnatada
 - leche corporal
24. Juan se preocupa por su dieta. Prefiere _____.
- el queso manchego
 - el queso curado
 - el queso fresco
 - el queso para untar
25. Alfredo quiere algo imposible, _____.
- tiene aspiraciones altas
 - desea aspiraciones altas
 - son aspiraciones altas
 - quiere aspiraciones altas
26. No me fío de tu nuevo novio, _____.
- no me inspira nada
 - no me inspira confianza
 - me inspira
 - confío en él
27. El juez condenó al preso a tres años de cárcel. Nunca había sido tan rápido en _____.
- dictar la lección
 - decir su sentencia
 - dictar la sentencia
 - impugnar sentencia
28. Tu relación con Pepe es muy difícil, es un _____.
- amor correspondido
 - cariño mutuo
 - amor imposible
 - cariño imposible
29. Pedro tiene una barba con mucho pelo, _____.
- su barba es poblada
 - tiene barba de tres días
 - tiene una barba escasa
 - tiene perilla
30. Se vendieron millones de ejemplares de su libro, fue un _____.
- éxito apoteósico
 - éxito discreto
 - éxito pasajero
 - fracaso estrepitoso

31. El sonido de la tiza en el encerado _____.
- me da asco
 - me da grima
 - me entra grima
 - me agrada
32. Estáis gritando mucho, parad de _____.
- montar barullo
 - producir barullo
 - guardar silencio
 - huir del barullo
33. Creo que voy a tomar un aspirina porque _____.
- he pillado un catarro
 - he curado un catarro
 - se me ha pasado el catarro
 - he tenido un catarro
34. Los nadadores tienen una espalda musculada, suelen tener _____.
- complexión fuerte
 - complexión débil
 - complexión estrecha
 - complexión ancha
35. Pedro se alegra por el éxito de María pero le gustaría tener el mismo éxito. Le tiene _____.
- envidia fugaz
 - envidia sana
 - envidia cochina
 - envidia atroz
36. Mario es extrovertido y habla mucho, tiene una _____.
- personalidad tímida
 - personalidad desbordante
 - personalidad oculta
 - personalidad impredecible

ANEXO XVIII. Hoja de respuestas preTCEp PRECO

Ejemplo

0. Todos los días me echo **colonia fresca**,
- me gustan los olores fuertes.
 - me gustan las colonias de hombre.
 - me gustan los olores suaves.
 - no me gusta echarme colonia.

	a.	b.	c.	d.
0.			X	

	a.	b.	c.	d.
1.				
2.				
3.				
4.				
5.				
6.				
7.				
8.				
9.				
10.				
11.				
12.				
13.				
14.				
15.				
16.				
17.				
18.				
19.				
20.				
21.				
22.				
23.				
24.				
25.				
26.				
27.				
28.				
29.				
30.				
31.				
32.				
33.				
34.				
35.				
36.				

ANEXO XIX. preTCE CRECU

INSTRUCCIONES: Completa la secuencia con la palabra que falta.

1. **Hace los deberes** cuando hace las ... que le manda la profesora.
2. Mi hermano ya **tiene barba**, tiene que ...
3. Cuando **hace frío**, la ... es baja.
4. El **teléfono fijo** está conectado a la pared por un ...
5. Tenía el pelo ..., fue a la peluquería y ahora tiene el **pelo rubio**.
6. Un **café cortado** es una café con un poco de ...
7. María **cometió** errores en el examen, no ... ninguna pregunta.
8. Si **tienes sed**, tienes que ...
9. Lucía está enferma pero ya **tiene cita** con el ...
10. Si tienes el **pelo liso**, entonces no tienes ...
11. La **comida rápida** se cocina en ... tiempo.
12. En una **ciudad cosmopolita**, la gente viene de ...
13. **Tengo una duda**, no estoy ... de mi decisión.
14. Cuando a Javier le **da pena** algo se siente ...
15. Cuando **pongo la mesa** a mediodía la preparo para poder ...
16. El **cielo** está **despejado**, no se ve ninguna ...
17. Hoy el **calor** es **horrible**, la ... ha subido mucho.
18. Un **zapato plano** es un zapato sin ...
19. Si un jugador **sufre una lesión** es que tiene un ... físico.
20. Cuando **haces un brindis**, levantas la ...
21. No sé **poner un enchufe**, hay que llamar a un ...
22. Una **vieja amiga** es alguien que conoces desde hace ... tiempo.
23. A la **leche entera** no le han quitado la ...
24. Cuando compras **queso fresco**, te preocupas por tu ...
25. Cuando **tienes aspiraciones** en tu trabajo, ... un puesto mejor.
26. Si alguien **te inspira confianza**, entonces te ... de él.
27. La persona que **dicta las sentencia** es el ...
28. Un **amor imposible** es un amor sin ...
29. Una **barba poblada** es una barba con ... pelo
30. Un **éxito apoteósico** es lo contrario a un ... absoluto.
31. Si algo te **da grima**, tienes una sensación
32. Cuando alguien **monta barrullo**, hacen ruido y ...
33. **Pillar un catarro** es ... a tener tos.
34. Si alguien es de **complexión fuerte**, su constitución es ...
35. Cuando le tienes **envidia sana** a alguien, no le deseas ningún ...
36. Con una **personalidad desbordante**, se tiene un ... extrovertido.

ANEXO XXI. TCEp PRECU

INSTRUCCIONES: Rellena los huecos (...) en las siguientes SECUENCIAS INCOMPLETAS, debes escribir tu respuesta en la columna de RESPUESTAS y en letras mayúsculas (tal y como se hace en el ejemplo). Solo necesitas una palabra; para ayudarte, la primera letra de la palabra que falta ya viene dada. Intenta contestar el mayor número de preguntas, pero no te preocupes si no las puedes contestar todas.

Ejemplo:

0	Hay sonidos que me d... grima , por ejemplo la tiza en el encerado.	DAN
---	--	-----

	SECUENCIAS INCOMPLETAS	RESPUESTAS
1	No te olvides la bufanda, fuera h... frío .	
2	Mi hermano necesita afeitarse, ya t... barba .	
3	Para aprobar necesitar h... el examen bien.	
4	No veo bien, así que ll... gafas desde que era pequeña.	
5	El nombre que se le p... al niño puede condicionar su carácter.	
6	Todos nos h... viejos al cumplir años.	
7	Creo que las mujeres morenas son más guapas que las de pelo r...	
8	A Pepe le gusta el café con poca leche, siempre pide un café c...	
9	Quería reservar una habitación d... , para mi marido y para mí.	
10	Los expertos recomiendan tomar el desayuno f... y la cena ligera.	
11	En los próximos días predominará el b... tiempo : sol y temperaturas altas.	
12	Mucha gente ya no tiene teléfono f... en sus casas, con el móvil es suficiente.	
13	No puedes aprobar: has c... un error grave en la redacción.	
14	¡Ay!, necesito un vaso de agua, t... sed .	
15	Tienes que pagar la matrícula para ... la inscripción en el curso.	
16	En algunos países es obligatorio d... propina a los camareros.	
17	Puedes escoger lo que quieras, pero te aconsejo p... el menú , es más barato.	
18	Recuerde, t... cita con el doctor a las cinco y media.	
19	Por su color oscuro diría que es un vino ...	
20	La prisa hace que elijamos la comida r... antes que la comida tradicional.	
21	Solo vendo tallas g... así que todas mis clientas tienen unos kilos de más.	
22	Las habitaciones de los bebés se suelen pintar de colores c... para dar calma.	
23	Tiene b... letra , todo el mundo la entiende.	
24	La tortilla le gusta a todo el mundo, es un plato r...	

25	El juez t... dudas sobre la inocencia del sospechoso.	
26	Todos quieren ir a ver la película: hay que h... cola para coger las entradas.	
27	Los modelos no s... fotos , ellos posan para la foto.	
28	Necesito a alguien que me ayude, no puedo a... todas las llamadas.	
29	Como le d... pena su amigo, él intenta ayudar en todo lo que puede.	
30	Ayúdame a p... la mesa antes de que lleguen los invitados para comer.	
31	María usa todos los días zapato p... , nunca usa tacones.	
32	El niño usa el reloj a... con dificultad porque está acostumbrado al digital.	
33	No para de toser, tiene una tos a...	
34	El testigo no sabe si oyó un disparo, pero sí un ruido f...	
35	No hay ninguna nube, el cielo está d...	
36	Las temperaturas superarán los treinta grados, se espera un calor h...	
37	Llama al electricista, por favor, no sé p... un enchufe .	
38	Levantad las copas y bebed, vamos a h... un brindis .	
39	El asesino b... todas sus huellas de la escena del crimen.	
40	Voy a ver si hoy las gallinas han p... algún huevo .	
41	Para que no haya más problemas, debemos e... una normas .	
42	Ver un partido de Champions es directo, es c... un sueño .	
43	A la leche e... no le quitan la nata.	
44	El queso f... debe consumirse rápido.	
45	¿Tienes muchos hermanos? A mí también me gustaría tener una familia n...	
46	Prefiero un aroma suave, no me gustan los olores i...	
47	Lucía es una v... amiga , la conozco desde hace mucho tiempo.	
48	Pareces enfermo, tienes m... color , deberías meterte en cama y descansar.	
49	Alfredo t... unas aspiraciones altas , desea algo imposible.	
50	El juez d... sentencia , condenó al preso a tres años de cárcel.	
51	La azafata comprueba tu documentación cuando p... el control	
52	No se puede conducir hasta s... el carné de conducir.	
53	Tu nuevo novio no me i... confianza , no me fío nada de él.	
54	El general da la orden, a... fuego , para empezar a disparar.	
55	Pedro tiene la barba p... , con mucho pelo.	
56	Los ojos r... son grandes y alargados.	
57	Su primer libro fue un éxito a... , es uno de los más vendidos del año.	
58	Le han diagnosticado algo grave: tiene una depresión p...	
59	Entra mucha luz por la ventana, fuera hace un sol r...	

60	Las relaciones t... suelen acabar mal.	
61	Tengo que ir al médico, creo que he p... un catarro.	
62	Parad de m... barullo , estáis gritando mucho.	
63	Se han prometido y en pocos meses c... matrimonio.	
64	Debes p... la solicitud en la oficina de empleo.	
65	Si alguien p... un incendio , causa que el fuego exista.	
66	Antes era muy amable, pero se le ha a... el carácter con la edad.	
67	Hay que estirar las piernas y las extremidades s...	
68	No la vas a ver, hoy hay luna n...	
69	Los nadadores tienen complexión f... , tienen una constitución atlética.	
70	Mario tiene una personalidad d... , es extrovertido y habla mucho.	
71	Pedro siente envidia s... por María. Se alegra por ella, pero le gustaría tener su mismo éxito.	
72	Tengo el sueño l... , me despierto con cualquier ruido.	

ANEXO XXII. TCEp CRECO

INSTRUCCIONES: Elige la opción más adecuada para completar la secuencia. Rodea con un círculo la opción correcta, tal y como aparece en el ejemplo. Intenta contestar el mayor número de preguntas, pero no te preocupes si no las puedes contestar todas.

Ejemplo:

0	Hay sonidos que me dan grima , por ejemplo...			
	a) la tiza en el encerado.	b) la música de una guitarra	c) la lluvia en la ventana	d) las olas del mar

1	Hace frío , por favor, ...			
	a) cierra la ventana.	b) abre las cortinas.	c) apaga la calefacción.	d) pon el aire acondicionado.
2	Mi hermano ahora tiene barba ,			
	a) no necesita afeitarse.	b) se afeita diariamente.	c) necesita una peluca.	d) ya no se depila.
3	Necesita hacer bien el examen para ...			
	a) suspender.	b) aprobar.	c) pasar.	d) repetir.
4	En verano siempre llevo gafas de sol, ...			
	a) me molesta la oscuridad.	b) me molesta la poca visibilidad.	c) me gusta la oscuridad.	d) me molesta el sol.
5	¿Sabes qué nombre le vas a poner al bebé?			
	a) Va a ser niña.	b) Va a ser Capricornio.	c) Va a ser María.	d) Va a ser alta.
6	Todos nos hacemos viejos , ...			
	a) al tener más sabiduría.	b) al tener más años.	c) al tener menos años.	d) al tener más trabajo.
7	Mi hermana tiene el pelo rubio , pero le gustaría tenerlo ...			
	a) rizado.	b) amarillo.	c) negro.	d) liso.
8	Pepe siempre pide un café cortado , ...			
	a) la mitad de la taza es suficiente.	b) no le echa azúcar.	c) le gusta con poca leche.	d) le añade agua.
9	Quería reservar una habitación doble , para ...			
	a) mí sola.	b) mi familia.	c) mí y mi marido.	d) para mis tres hijos.
10	Los expertos recomiendan tomar un desayuno fuerte :			
	a) un café que da energía.	b) café y fruta que dan energía.	c) comida que da energía.	d) abundante comida que da energía
11	Mañana hará buen tiempo :			
	a) brillará el sol.	b) bajarán las temperaturas.	c) soplará un viento fuerte.	d) se instalará un anticiclón.
12	Juan, no llames al teléfono fijo que			
	a) mis padres están en casa.	b) mis padres no están en casa.	c) el teléfono está apagado.	d) el teléfono está encendido.
13	María comete errores todos los días,...			
	a) siempre se equivoca.	b) nunca se equivoca.	c) hace todo bien.	d) no falla nunca.
14	Tengo sed , ...			
	a) quiero emborracharme.	b) bebo dos litros de agua al día.	c) solo me gusta el zumo.	d) quiero un vaso de agua.
15	La matrícula se paga...			
	a) cuando estás excluido.	b) cuando estás admitido.	c) cuando puedas.	d) cuando reclamas.

16	En algunos países es obligatorio dar propina a los camareros ...			
	a) antes de consumir.	b) después de consumir.	c) cuando pagas.	d) antes de pagar la cuenta.
17	Cuando pides el menú ...			
	a) el precio es fijo.	b) el precio es caro.	c) el precio varía.	d) el precio es barato.
18	Lucía tiene cita a las cinco, ...			
	a) va a verse con unos amigos.	b) va a visitar al médico.	c) quedó con unos amigos.	d) no quedó con nadie.
19	El vino tinto tiene un color			
	a) rosado claro.	b) amarillo verdoso.	c) rojo oscuro.	d) negro intenso.
20	Me gusta mucho la comida rápida aunque ...			
	a) sea casera.	b) sea ecológica.	c) sea dietética.	d) sea poco sana.
21	Necesito ir a tiendas de tallas grandes , especializadas en ropa para gente ...			
	a) con muchos quilos,	b) de mucha estatura.	c) con mucha edad.	d) con mucha importancia.
22	Prefiero los colores claros para pintar la casa porque dan más...			
	a) intimidad.	b) oscuridad.	c) luz.	d) calor.
23	Ese médico tiene buena letra , ...			
	a) todos entienden lo que escribe.	b) nadie entiende lo que escribe.	c) es muy bonita.	d) es muy fea.
24	La tortilla es un plato rico , ...			
	a) tiene mal sabor.	b) tiene buen sabor.	c) cuesta mucho dinero.	d) da mucho trabajo hacerla.
25	Cuando un alumno tiene una duda , ...			
	a) sospecha que algo va a salir mal.	b) elige una de las opciones posibles.	c) desconfía de su profesora.	d) no entiende algo.
26	Hubo que hacer cola para coger las entradas,			
	a) todos esperamos de forma ordenada.	b) no tuvimos que esperar.	c) la gente se fue directa a la taquilla.	d) llegamos y las compramos.
27	Sacar una foto es lo mismo que...			
	a) no salir en una fotografía.	b) hacer una fotografía.	c) revelar una fotografía.	d) enmarcar una fotografía.
28	No puedo atender todas las llamadas , ...			
	a) el teléfono no suena nunca.	b) el teléfono no para de sonar.	c) todos los días me quiere vender un teléfono.	d) me dedico a vender teléfonos.
29	A Javier le da pena mi hermano, ...			
	a) piensa en cómo poder ayudarle.	b) no piensa ayudarle.	c) no se alegra por él.	d) se alegra por él.
30	Pedro, pon la mesa , ...			
	a) antes de que todos se levanten de la mesa.	b) aquí tienes las instrucciones.	c) le ordena su madre antes de comer.	d) pasarlos por la sartén.
31	María usa todos los días zapato plano : ...			
	a) nunca usa tacones.	b) es bajita e intenta parecer más alta.	c) va siempre con tacones.	d) tiene los pies pequeños.
32	El reloj analógico ...			
	a) funciona con electricidad.	b) se ve en la oscuridad.	c) tiene agujas.	d) marca veinticuatro horas.
33	Juan tuvo un ataque de tos aguda , ...			
	a) le duró mucho tiempo.	b) le duró poco tiempo.	c) parecía seca.	d) no era molesto.
34	El testigo oyó un ruido fuerte , no sabe si fue ...			
	a) un disparo.	b) una mosca.	c) la televisión.	d) el canto de un pájaro.
35	El cielo está despejado , ...			
	a) va a empezar a llover.	b) va a empezar a nevar.	c) no hay nubes.	d) la temperatura es estable.
36	Se espera un calor horrible , las temperaturas superarán los			

	a) cinco grados.	b) diez grados.	c) veinte grados.	d) treinta grados.
37	El cable no llega y no sé poner un enchufe , así que ...			
	a) llama al carpintero.	b) llama al electricista.	c) llama al fontanero.	d) llama al albañil.
38	Vamos a hacer un brindis ...			
	a) por el éxito de la empresa.	b) por lástima.	c) por necesidad.	d) por el fracaso de la empresa.
39	Antes de escapar, el ... borró las huellas del asesinato.			
	a) policía	b) detective	c) testigo	d) asesino
40	Las granjas en las que las gallinas ponen muchos huevos son ...			
	a) muy productivas.	b) muy ordenadas.	c) poco productivas.	d) desordenadas.
41	Ahora han establecido unas normas , así que ...			
	a) nadie ha decidido qué se puede hacer.	b) no tenemos ni idea de qué podemos hacer.	c) sabemos qué podemos hacer.	d) podemos hacer lo que queramos.
42	Cumplir un sueño es...			
	a) sumar otro logro a mi lista.	b) hacer lo que nunca quise.	c) hacer lo que siempre quise.	d) hacer lo que quiere otro.
43	María siempre toma leche entera : ...			
	a) toma todo sin grasa.	b) le gusta el sabor verdadero de la leche.	c) le gusta la comida dietética.	d) le gustan los productos enteros, no en partes.
44	Prefiere el queso fresco : ...			
	a) tiene un sabor fuerte y le gusta que la comida tenga mucho sabor.	b) mira la calidad y solo compra productos caros.	c) se preocupan de su dieta y ese queso no tiene calorías.	d) Nunca come productos curados, prefiere los de reciente elaboración.
45	Me gustaría tener una familia numerosa , ...			
	a) de muchos hijos.	b) de matemáticos.	c) de pocos hijos.	d) de actores.
46	No me gustan los olores intensos , prefiero...			
	a) los fuertes.	b) que no huela.	c) los suaves.	d) los agradables.
47	Lucía es una vieja amiga , ...			
	a) tiene más de sesenta años.	b) hace muchos años que la conozco.	c) la conocí hace dos meses.	d) todo el mundo la ve más vieja de lo que es.
48	Tienes mal color , ...			
	a) estás muy moreno.	b) te maquillas demasiado.	c) usas solo ropa negra.	d) pareces enfermo.
49	Alfredo tiene unas aspiraciones altas ,			
	a) cree que todo le va a salir mal.	b) cree que va a subir hasta el último piso.	c) tiene objetivos fáciles de conseguir.	d) desea algo imposible.
50	El juez dictó sentencia : ...			
	a) aplazó el juicio.	b) la secretaria copió todo lo que decía.	c) condenó al preso a tres años de cárcel.	d) sus palabras son sabias.
51	Hay que pasar el control ...			
	a) al cruzar la frontera.	b) al cambiar de carretera.	c) para saber qué pasa.	d) al entrar en la ciudad.
52	Hasta que saques el carné no puedes ...			
	a) aparcar.	b) conducir.	c) ir por autopista.	d) conducir sola.
53	Tu nueva novia no me inspira confianza , ...			
	a) no me fío nada de ella.	b) me causa muy buena impresión.	c) es mi musa.	d) me cae bien.
54	"Abran fuego , es hora de atacar al enemigo", ...			
	a) dice el bombero.	b) dice el soldado.	c) dice la víctima.	d) dice el pirómano.

55	Pedro tiene la barba poblada , ...			
	e) igual que la del resto de la población.	f) con mucho pelo.	g) con una finca llena de árboles.	h) recién afeitada.
56	Me gustan los ojos rasgados : ...			
	a) cortos y muy abiertos.	b) pequeños y muy abiertos.	c) grandes y poco abiertos.	d) largos y poco abiertos.
57	Su primer libro fue un éxito apoteósico , ...			
	a) nadie lo compró.	b) le gustó a mucha gente.	c) nadie pudo publicar otro libro.	d) no se acabó ni la primera edición.
58	La depresión profunda suele darse en casos de tristeza ...			
	a) dulce.	b) contagiosa.	c) reprimida.	d) extrema.
59	Hace un sol radiante , hoy es un día ...			
	a) nublado.	b) desapacible.	c) luminoso.	d) oscuro.
60	Tienen una relación tormentosa , ...			
	a) no pueden seguir juntos.	b) deben seguir juntos.	c) están hechos el uno para el otro.	d) siempre que salen, hay tormenta.
61	Creo que he pillado un catarro , ...			
	a) me encuentro bien.	b) me encuentro mal.	c) me voy a encontrar mal.	d) me encontraba mal.
62	Parad de montar barullo que ...			
	a) estáis muy cansados.	b) todo está muy desordenado.	c) nunca habláis tan bajo.	d) estáis gritando mucho.
63	En pocos meses contraerán matrimonio : ...			
	a) se acaban de divorciar.	b) se acaban de contagiar.	c) se acaban de prometer.	d) acaban de encogerse.
64	Cuando presentas una solicitud , ...			
	a) introduces tus datos personales en un formulario.	b) entregas un formulario cumplimentado.	c) enseñas la solicitud al encargado.	d) entregas tu documentación y muestras la solicitud.
65	Han provocado un incendio que acabó con la reserva natural. No se sabe quién...			
	a) causó ese fuego descontrolado.	b) apagó el fuego.	c) alimenta ese fuego.	d) le da fuego.
66	Se le ha agriado el carácter con la edad, ...			
	a) antes era muy amable.	b) ahora es muy amable.	c) tiene un nuevo sabor.	d) escribe peor.
67	Necesitas fortalecer las extremidades superiores , así que...			
	a) usa siempre el ascensor.	b) evita realizar esfuerzos.	c) intenta ejercitar los brazos.	d) sube por las escaleras.
68	Hoy hay luna nueva : ...			
	a) tiene forma de c.	b) tiene forma circular.	c) no se ve.	d) la vamos a estrenar.
69	Los nadadores tienen complexión fuerte , ...			
	a) tienen poco músculo.	b) tienen mucho músculo.	c) son delgados.	d) están gordos.
70	Mario tiene una personalidad desbordante ,			
	a) es tímido y habla poco.	b) es sociable, pero habla poco.	c) es tímido, pero gracioso.	d) es extrovertido y habla mucho.
71	Pedro siente envidia sana por María, ...			
	a) no se cambiaría por ella.	b) no le gustaría estar en su lugar.	c) se alegra por ella, pero quiere su éxito.	d) no se alegra por ella, pero valora su éxito.
72	Tengo el sueño ligero , ...			
	a) siempre tengo sueño.	b) me despierto fácilmente.	c) duermo muy bien.	d) nunca tengo sueño.

ANEXO XXIII. TCEp PRECO

INSTRUCCIONES: Elige la opción correcta para completar las secuencias. Rodea con un círculo la opción correcta, tal y como aparece en el ejemplo. Intenta contestar el mayor número de preguntas, pero no te preocupes si no las puedes contestar todas.

Ejemplo:

0	El sonido de la tiza en el encerado me ... grima.			
	a) entra.	b) da.	c) hace.	d) quita.

1	No te olvides la bufanda, fuera ... frío .			
	e) tiene	f) hace	g) es	h) está
2	Mi hermano ya ... barba , necesita afeitarse.			
	e) tiene	f) es	g) hace	h) esconde
3	Para aprobar hay que ... bien el examen .			
	a) tener	b) corregir	c) hacer	d) pasar
4	En verano siempre... gafas de sol, me molesta mucho la claridad.			
	e) cuelgo	f) visto	g) gradúo	h) llevo
5	¿Ya sabes qué nombre le vas a ... al bebé? A mí me gustan los nombres clásicos.			
	e) decir	f) tomar	g) poner	h) tener
6	Todos nos ... viejos al cumplir años.			
	e) llevamos	f) hacemos	g) tenemos	h) causamos
7	Creo que las mujeres de pelo oscuro son más guapas que las mujeres de pelo ...			
	e) rubio.	f) amarillo.	g) encrespado.	h) rizo.
8	Pepe siempre pide café ..., le gusta con poca leche.			
	e) cortado	f) pelado	g) solo	h) con leche
9	Quería reservar una habitación ... para mi marido y para mí.			
	a) colectiva	b) contigua	c) individual	d) doble
10	Los expertos recomiendan tomar un desayuno ... y comer menos para cenar.			
	e) pesado	f) grande	g) fuerte	h) ligero
11	Cuando hace ... tiempo , apetece salir.			
	e) buen	f) mal	g) adverso	h) suficiente
12	Mis padres no están en casa, no los llames al teléfono ...			
	e) móvil.	f) portátil.	g) estático.	h) fijo.
13	Has ... un error grave en la redacción de inglés, no puedes aprobar.			
	e) corregido	f) cometido	g) causado	h) forzado
14	¡Ay!, ... sed , necesito un vaso de agua.			
	a) estoy	b) quiero	c) tengo	d) padezco
15	Estoy en el banco, voy a ... la matrícula antes de ir a clase.			
	e) cambiar	f) tener	g) pagar	h) apuntar
16	En algunos países es obligatorio ... propina a los camareros.			
	e) dar	f) sacar	g) ofrecer	h) meter
17	Puedes escoger lo que quieras, pero te aconsejo ... el menú , es más barato.			
	e) preparar	f) pedir	g) ofrecer	h) ordenar
18	María ... cita con el médico a las cinco.			
	e) escribe	f) solicita	g) tiene	h) lee
19	Por su color oscuro diría que es un vino ...			

	e) dulce.	f) blanco.	g) tinto.	h) rosado.
20	La prisa hace que elijamos la comida ... antes que la comida tradicional.			
	e) casera	f) ecológica	g) mediterránea	h) rápida
21	Deberían abrir más tiendas de tallas ... para las personas con quilos de más.			
	e) intelectuales	f) grandes	g) únicas	h) anchas
22	Prefiero los colores ... porque dan mucha luz a las casas.			
	e) fríos	f) primarios	g) claros	h) fosforescentes
23	El profesor tiene ... letra , todo el mundo la entiende.			
	e) mala	f) fea	g) buena	h) bonita
24	A mucha gente la tortilla le parece un plato ...			
	e) rico	f) caro	g) pobre	h) roto
25	Cuando un alumno ... una duda , levanta la mano para preguntar.			
	e) tiene	f) da	g) despierta	h) resuelve
26	Hay que ... cola para coger las entradas, todos quieren ver la película.			
	e) tener	f) fabricar	g) cumplir	h) hacer
27	La catedral siempre está llena de turistas que ... fotos.			
	e) sacan	f) tienen	g) pasan	h) velan
28	No puedo ... todas las llamadas , necesito a alguien que me ayude a contestar el teléfono.			
	e) hacer	f) sonar	g) descolgar	h) atender
29	Como le ... pena su amigo, él intenta ayudar en todo lo que puede.			
	e) da	f) hace	g) tiene	h) alivia
30	Coge los platos, los vasos y los cubiertos para ... la mesa antes de comer.			
	e) mover	f) recoger	g) montar	h) poner
31	María usa todos los días zapato ..., nunca usa tacones.			
	e) sin cordones	f) sencillo	g) plano	h) liso
32	El niño usa el reloj ... con dificultad porque está acostumbrado al digital.			
	e) analógico	f) puntual	g) de mesa	h) de pared
33	Tengo todos los síntomas de una gripe: frío, dolor de cabeza y tos ...			
	e) afilada.	f) aguda.	g) blanca.	h) suave.
34	El testigo no sabe si oyó un disparo, pero sí un ruido ...			
	a) constante.	b) acompasado	c) fuerte.	d) suave.
35	No hay ninguna nube, el cielo está ...			
	e) despejado.	f) cerrado.	g) negro.	h) nublado.
36	Se espera un calor ..., las temperaturas superarán los treinta grados.			
	e) alto	f) feo	g) ligero	h) horrible
37	No sé ... un enchufe , por favor, llama a un electricista.			
	e) aprovechar	f) poner	g) poner	h) usar
38	Vamos a ... un brindis , levantad las copas y bebed.			
	e) hacer	f) dedicar	g) chocar	h) dar
39	El asesino ... todas sus huellas en la escena del crimen.			
	e) rastreó	f) dejó	g) borró	h) encontró
40	Voy a ver si hoy las gallinas han ... algún huevo .			
	e) sacado	f) incubado	g) cascado	h) puesto
41	El gobierno ya ha ... la norma : no se puede circular a más de 50 km / h.			
	e) saltado	f) establecido	g) vulnerado	h) seguido
42	Ver un partido de Champions en directo es ... un sueño .			
	e) cumplir	f) conceder	g) mantener	h) conciliar
43	María siempre bebe leche ..., a ella no le importa engordar.			
	e) limpiadora	f) entera	g) desnatada	h) corporal

44	Prefiere el queso ...: se preocupa por su dieta y éste tiene menos calorías.			
	e) manchego	f) curado	g) fresco	h) para untar
45	¿Tienes seis hermanos? A mí también me gustaría tener una familia ...			
	e) real	f) cuantiosa	g) numerosa	h) efímera
46	Prefiero un aroma ligero, no me gustan los olores ...			
	e) intensos	f) suaves	g) agradables	h) vehementes
47	Lucía es una ... amiga , la conozco desde hace mucho tiempo.			
	e) joven	f) vieja	g) acabada	h) nueva
48	Pareces enfermo, tienes ... color , deberías meterte en cama y descansar.			
	e) buen	f) oscuro	g) mal	h) claro
49	Alfredo ... unas aspiraciones altas , desea algo imposible.			
	e) desea	f) trunca	g) quiere	h) tiene
50	El juez ... sentencia , condenó al preso a tres años de cárcel.			
	e) dijo	f) impugnó	g) dictó	h) dio
51	Antes de cruzar la frontera, hay que ... un control .			
	e) pasar	f) andar	g) cruzar	h) saltar
52	Hasta los dieciocho años no puedes ... el carne de conducir .			
	e) hacer	f) suspender	g) sacar	h) renovar
53	Tu nuevo novio no me ... confianza , no me fío nada de él.			
	i) inspira	j) gana	k) tiene	l) abusa de
54	Un general a sus tropas: "... fuego, es hora de atacar al enemigo".			
	a) apaguen	b) abran	c) empiecen	d) cierren
55	Pedro tiene la barba ..., con mucho pelo.			
	e) habitada	f) escasa	g) afeitada	h) poblada
56	Los ojos ... son más bonitos que los redondos.			
	e) saltones	f) rasgados	g) tristes	h) vidriosos
57	Su primer libro fue un éxito ..., fue uno de los más vendidos del año.			
	e) pasajero	f) discreto	g) apoteósico	h) flagrante
58	Esa tristeza tan gran solo puede ser síntoma de una depresión ...			
	e) montañosa	f) plena	g) pasajera	h) profunda
59	Es la luz perfecta para la foto, ahora hay un sol ...			
	e) radiante	f) naciente	g) flojo	h) ardiente
60	No pueden seguir juntos, tiene una relación ...			
	e) abierta	f) envidiable	g) tormentosa	h) estable
61	He ... un catarro , creo que voy a tomar una aspirina.			
	e) pillado	f) notado	g) curado	h) pasado
62	Parad de ... barullo , estáis gritando mucho.			
	e) componer	f) deshacer	g) montar	h) dar
63	Se han prometido y en pocos meses ... matrimonio .			
	e) celebrarán	f) contraerán	g) anularán	h) harán
64	Primero cubres los datos, después ... la solicitud .			
	e) rellenas	f) presentas	g) desestimas	h) compulsas
65	No creo que nadie ... un incendio de forma voluntaria, es algo horrible.			
	e) pinche	f) encienda	g) sofoque	h) provoque
66	Antes era muy amable, pero se le ha ... el carácter con la edad.			
	e) agriado	f) endulzado	g) saboreado	h) enfadado
67	Hay que estirar tanto las extremidades inferiores como las ...			
	e) subidas	f) superiores	g) delanteras	h) avanzadas
68	No la vas a ver, hoy hay luna ...			

	e) creciente	f) llena	g) nueva	h) trasera
69	Los nadadores tienen complexión ... : con una espalda ancha y músculos definidos.			
	e) ancha	f) débil	g) estrecha	h) fuerte
70	Mario tiene una personalidad ..., es extrovertido y habla mucho.			
	e) tímida	f) desbordante	g) oculta	h) impredecible
71	Pedro siente envidia ... por María. Se alegra por ella, pero le gustaría tener su mismo éxito.			
	e) fugaz	f) sana	g) cochina	h) atroz
72	Tengo el sueño ... , no duermo nada bien.			
	a) voluble	b) ligero	c) profundo	d) recurrente

ANEXO XIV. TCEp CRECU

INSTRUCCIONES: Completa los huecos (...) en las siguientes SECUENCIAS INCOMPLETAS, debes escribir tu respuesta en la columna de RESPUESTAS y en letras mayúsculas (tal y como se hace en el ejemplo). Solo necesitas una palabra. Intenta contestar el mayor número de preguntas, pero no te preocupes si no las puedes contestar todas.

Ejemplo:

0	Si algo te da grima , tienes una sensación...	DESAGRADABLE
---	--	--------------

	SECUENCIAS INCOMPLETAS	RESPUESTAS
1	Fuera hace frío , las temperaturas son ...	
2	Mi hermano tiene barba , ya tiene ... en la cara.	
3	Sólo hay posibilidad de ... si haces bien el examen .	
4	En verano llevo gafas de sol porque me molesta la..., hay demasiada claridad.	
5	Poner el nombre del bebé es siempre decisión de los ...	
6	Hacerse viejo implica haber vivido ...	
7	Su mujer tiene el pelo rubio , un ... parecido al del trigo.	
8	Pepe siempre pide un café cortado , le gusta con ... leche.	
9	Quiero una habitación doble , por favor, vamos ... personas.	
10	Los expertos recomiendan tomar un desayuno fuerte y cenar ...	
11	Hace buen tiempo cuando hace ... y calor.	
12	El teléfono fijo está conectado a la pared por un ...	
13	Cometimos un error al venir hoy, no se puede ... siempre.	
14	Si tienes sed , tienes que ...	
15	El único requisito para pagar la matrícula , después de ser aceptado, es tener ...	
16	En algunos países es obligatorio dar propina a los ...	
17	Te aconsejo pedir el menú , si no el precio es más ...	
18	Lucía está enferma y ya tiene cita con el ...	
19	Sabes que es vino tinto por su color ...	
20	La comida rápida se cocina en ... tiempo.	
21	Vendo ropa de tallas grandes , todas mis clientas tienen unos ... de más.	
22	Me gustan los colores claros más que los ...	
23	Tiene buena letra , es ... de entender.	
24	La tortilla es un plato rico , tiene ... sabor.	
25	Si tienes una duda , es difícil ... entre dos opciones.	
26	La profesora les explicó cómo hacer cola y dijo "Poneos uno ... del otro".	
27	¿Me sacas una foto , por favor? Toma la ...	
28	La secretaria está atendiendo llamadas . Está todo el día ... por teléfono.	
29	Cuando algo le da pena , pone cara ...	
30	Cuando pongo la mesa a mediodía la preparo para ...	
31	Un zapato plano es un zapato sin ...	

32	El niño usa el reloj analógico con dificultad porque está acostumbrado al reloj...	
33	La tos aguda no tiene más importancia, es un síntoma breve, pero ...	
34	Por favor, para de hacer ese ruido fuerte que me ...	
35	El cielo está despejado , no hay ninguna ...	
36	Decimos que hace un calor horrible cuando hace ... calor.	
37	No sé poner un enchufe , por favor, llama a un ...	
38	Vamos a hacer un brindis , levantad las ...	
39	El asesino borró las huellas para no ser ... por la policía.	
40	Voy a ver si hoy las ... han puesto algún huevo .	
41	Hay unas normas establecidas así que ya no podemos hacer lo que...	
42	Se cumple un sueño cuando se consigue algo que se ...	
43	A la leche entera no le han quitado la ...	
44	Cuando compras queso fresco , te preocupas por tu ...	
45	¿Tienes ... hermanos? A mí también me gustaría tener una familia numerosa .	
46	No me gustan los olores intensos , prefiero los olores ...	
47	Una vieja amiga es alguien que conoces desde hace ... tiempo.	
48	Pareces ..., tienes mal color , deberías meterte en cama y descansar.	
49	Cuando tienes aspiraciones en tu trabajo, quieres un puesto...	
50	El responsable de dictar sentencia es el ...	
51	Cuando pasas el control , la azafata comprueba tu	
52	Hasta que te saques el carné no vas a poder ...	
53	Cuando alguien te inspira confianza , te ... de él.	
54	El general ordena abrir fuego para empezar a ...	
55	Una barba poblada es una barba con ... pelo.	
56	Los ojos rasgados son grandes y ...	
57	Un éxito apoteósico es lo contrario a un ... absoluto.	
58	Le han diagnosticado una enfermedad ...: tiene una depresión profunda .	
59	Fuera hace un sol radiante , entra ... luz por la ventana.	
60	Las relaciones tormentosas se caracterizar por tener situaciones...	
61	José nunca está ..., pero este año pilló un catarro .	
62	Parad de montar barullo , estáis haciendo...	
63	Si contraes matrimonio , te...	
64	Para presentar la solicitud hay que cubrir los datos personales y ... en plazo.	
65	Si alguien provoca un incendio , ... que el fuego exista.	
66	Juan pasó de ser simpático a agriársele el carácter y ser ...	
67	Hay que estirar tanto las extremidades ... como las extremidades superiores.	
68	Cuando hay luna nueva , la luna no se puede ...	
69	Los deportistas de complexión fuerte tienen una ... atlética.	
70	Con una personalidad desbordante , se tiene un ... extrovertido.	
71	Cuando le tienes envidia sana a alguien, no le deseas nada ...	
72	Tengo un sueño muy ligero , me despierto con cualquier ...	

ANEXO XXV. TCE PRECU

Rellena los huecos en las siguientes SECUENCIAS INCOMPLETAS. Debes contestar en la línea de abajo. Solo necesitas una palabra; para ayudarte, la primera letra de la palabra que falta ya está escrita. Intenta contestar al mayor número de preguntas, pero no te preocupes si no puedes contestar todas.

Ejemplo

0. Los platos r... tienen mucho sabor.	Ricos
---	-------

1. No te olvides del abrigo, fuera h... frío .	H
2. Mi hermano necesita afeitarse, ya t... barba .	T
3. En verano siempre l... gafas de sol , me molesta mucho la luz.	L
4. Dicen que las mujeres morenas son más guapas que las de pelo r...	R
5. Quiero reservar una habitación d... , para dos.	D
6. Pepe siempre quiere café c... , le gusta con poca leche.	C
7. ¡Ay! Necesito un vaso de agua, t... sed .	T
8. En algunos restaurantes hay que d... propina a los camareros.	D
9. Recuerda, t... cita con el doctor a las cinco y media.	T
10. Por su color rojo diría que es vino t...	T
11. El médico tiene b... letra . Todo el mundo la entiende.	B
12. La comida r... no es buena para la salud.	R
13. Hay mucha gente esperando, tenemos que h... cola .	H
14. Ahora con los móviles puedes s... fotos muy buenas.	S
15. Yo cocino, tú p... la mesa .	P
16. María usa todos los días zapato p... , nunca usa tacones.	P
17. Prefiero pasar frío polar a un calor h...	H
18. Las máquinas de la fábrica hacen un ruido f...	F
19. Conseguir lo que deseas sería c... un sueño .	C
20. Levantad las copas, vamos a h... un brindis .	H
21. El ladrón b... todas las huellas de la escena del crimen.	B
22. La leche e... no tiene grasa.	E
23. Eres de una familia n... si tienes muchos hermanos.	N
24. Lucía es una v... amiga , la conozco desde hace mucho.	V
25. El juez d... sentencia , condenó al preso a ir a la cárcel.	D
26. Esperamos dos horas para p... el control en el aeropuerto.	P
27. Su nuevo novio no me i... confianza , no me gusta.	I
28. Su primer libro fue un éxito a... , es el más vendido.	A
29. Necesito gafas de sol, hace un sol r...	R
30. Las relaciones t... suelen acabar mal.	T
31. Tengo que ir al médico, creo que he p... un catarro .	P
32. Parad de m... barullo , gritáis demasiado.	M
33. Se han prometido y pocos meses c... matrimonio .	C
34. Mi prima tiene una personalidad d... , llama la atención de todos.	D
35. Él siente envidia s... por ella. Se alegra, pero le gustaría tener su éxito.	S
36. Tengo el sueño l... , me despierto con cualquier ruido.	L

ANEXO XXVI. TCE CRECO

Elige la opción correcta para completar las secuencias. Intenta contestar el mayor número de preguntas, pero no te preocupes si no puedes contestar todas. Intenta contestar el mayor número de preguntas posible, pero no te preocupes si no las puedes contestar todas.

Ejemplo:

Un plato rico es plato que tiene ...

- a. Mal sabor **b. buen sabor** c. ingredientes caros d. pocos admiradores.

1. Hace frío , por favor, lleva...			
a. bañador	b. guantes	c. paraguas	d. chanclas
2. Tener barba es tener pelo en...			
a. el pie	b. la mano	c. los brazos	d. la cara
3. En verano siempre llevo gafas de sol , ...			
a. me molesta la oscuridad.	b. me molesta la poca visibilidad.	c. me gusta la oscuridad.	d. me molesta el sol.
4. El pelo rubio es de color...			
a. rojo	b. negro	c. marrón	d. amarillo
5. Quería reservar una habitación doble , para...			
a. una persona.	b. dos personas.	c. tres personas.	d. más de cuatro personas.
6. Un café cortado es un café con...			
a. un poco de azúcar.	b. un poco de leche.	c. mucho café.	d. mucha leche.
7. Tengo sed , ...			
a. quiero emborracharme.	b. bebo mucha agua.	c. solo bebo zumo.	d. quiero agua.
8. En algunos países es obligatorio dar propina ...			
a. a los camareros.	b. al dueño.	c. a los otros clientes.	d. al que come contigo.
9. Lucía tiene cita a las cinco, ...			
a. va a verse con unos amigos.	b. va a visitar al médico.	c. quedó con unos amigos.	d. no quedó con nadie.
10. El vino tinto tiene un color...			
a. rosado claro.	b. amarillo verdoso.	c. rojo oscuro.	d. negro intenso.
11. Ese médico tiene buena letra , ...			
a. todos entienden lo que escribe.	b. nadie entiende lo que escribe.	c. no tiene faltas de ortografía.	d. su gramática es perfecta.
12. La comida rápida promueve la alimentación...			
a. perjudicial.	b. saludable.	c. rica.	d. equilibrada.

13. Hacer cola es...			
a. esperar en fila.	b. hacer una línea.	c. estar al final.	d. formar un círculo.
14. Sacar una foto es lo mismo que...			
a. hacer una foto.	b. tener una foto.	c. poner una foto.	d. colgar una foto.
15. Pedro, pon la mesa . Vamos a...			
a. recoger.	b. montarla.	c. comer.	d. levantarnos.
16. Un zapato plano es un zapato...			
a. para la lluvia.	b. sin tacón.	c. para hacer deporte.	d. con cordones.
17. Un calor horrible es ... calor.			
a. poco	b. demasiado	c. pegajoso	d. ligero
18. Un ruido fuerte es un sonido...			
a. pegadizo.	b. desagradable.	c. inconfundible.	d. electrónico.
19. Cumplir un sueño es...			
a. conseguir lo que deseas.	b. pedir algo importante.	c. hacer lo que debes.	d. perder lo que deseas.
20. Vamos a hacer un brindis , ...			
a. coged las servilletas.	b. levantad las copas.	c. sentaos en las sillas.	d. servid las tazas.
21. Antes de escapar, el ... borró las huellas para que la policía no lo identificara.			
a. asesino	b. detective	c. testigo	d. ladrón
22. La leche entera ...			
a. no tiene azúcar.	b. tiene grasa.	c. es vegetal.	d. no es natural.
23. Si tienes una familia numerosa , tienes...			
a. muchos hermanos.	b. cuatro abuelos.	c. dos padres.	d. algunos primos.
24. Lucía es una vieja amiga , ...			
a. tiene más de sesenta años.	b. hace mucho que la conozco.	c. la conocí hace dos meses.	d. parece mayor.
25. El juez dictó sentencia por la que ...			
a. se aplazó el juicio.	b. se condenó al acusado a la cárcel.	c. se pidió silencio.	d. se dijo algo inteligente.
26. Para pasar el control en el aeropuerto, ...			
a. sacas tema.	b. te pones el cinturón.	c. sacas los líquidos.	d. te pones guapo/a.
27. Esta empresa me inspira confianza , ...			
a. tiene demasiadas deudas.	b. vamos a denunciarla.	c. podemos invertir en ella.	d. no le va bien.
28. Su primer libro fue un éxito apoteósico , ...			
a. nadie lo compró.	b. fracasó rotundamente.	c. fue el más vendido.	d. no se vendió bien.
29. Hace un sol radiante , el cielo está...			

a. nublado.	b. desapacible.	c. luminoso.	d. rojizo.
30. Una relación tormentosa implica tener...			
a. peleas continuas.	b. momentos lluviosos.	c. argumentos importantes.	d. gestos cariñosos.
31. Creo que he pillado un catarro , ...			
a. no paro de estornudar.	b. ya no estornudo.	c. todas las mañanas estornudo.	d. acabo de estornudar.
32. Parado de montar barullo , ...			
a. estáis muy cansados.	b. todo está muy desordenado.	c. nunca habláis tan bajo.	d. gritáis mucho.
33. En pocos meses contraerán matrimonio : se acaban de ...			
a. despedir.	b. contagiar.	c. prometer.	d. encoger.
34. Alguien que tiene una personalidad desbordante , ...			
a. oculta	b. desbordante	c. desmesurada	d. anormal
35. Pedro siente envidia sana por María, ...			
a. no se cambiaría por ella.	b. no le gustaría estar en su lugar.	c. se alegra por ella, pero quiere su éxito.	d. no se alegra por ella, pero valora su éxito.
36. Tengo el sueño ligero , ...			
a. me cuesta despertarme.	b. me despierto fácilmente.	c. duermo muy bien.	d. nunca quiero dormir.

ANEXO XXVII. TCE PRECO

Elige la opción correcta (a., b., c. o d.) para completar las secuencias. Selecciona la opción correcta rodeándola.

Intenta contestar el mayor número de preguntas posible, pero no te preocupes si no las puedes contestar todas.

Ejemplo

A mucha gente la tortilla de patata le parece un plato ...

- a. caro b. pobre c. rico d. roto

1. No te olvides el abrigo, fuera ... frío .			
e. tiene	f. hace	g. es	h. está
2. Mi hermano ya ... barba , necesita afeitarse.			
e. tiene	f. es	g. hace	h. pone
3. En verano siempre ... gafas de sol, me molesta mucho la claridad.			
e. cuelgo	f. visto	g. gradúo	h. llevo
4. Dice que las mujeres de pelo oscuro son más guapas que las mujeres de pelo ...			
e. rubio	f. amarillo	g. encrespado	h. rizo
5. Quería reservar una habitación ... para dos.			
e. colectiva	f. contigua	g. individual	h. doble
6. A Pepe le gusta el café con un poco de leche, siempre bebe café ...			
e. cortado	f. con leche	g. largo	h. blanco
7. ¡Ay!, ... sed , necesito un vaso de agua.			
e. estoy	f. hago	g. tengo	h. pongo
8. En algunos países es obligatorio ... propina a los camareros.			
e. dar	f. sacar	g. ofrecer	h. meter
9. María ... cita con el médico a las cinco para una revisión.			
e. hace	f. queda	g. tiene	h. apunta
10. Por su color rojo diría que es un vino ...			
e. dulce	f. blanco	g. tinto	h. rosado
11. El profesor tiene ... letra , todo el mundo la entiende.			
e. mala	f. fea	g. buena	h. bonita
12. La comida ... es mala para la salud.			
e. casera	f. ecológica	g. mediterránea	h. rápida
13. Hay mucha gente esperando, tenemos que ... cola			
e. tener	f. fabricar	g. cumplir	h. hacer
14. La catedral siempre está llena de turistas que ... fotos .			
e. sacan	f. tienen	g. velan	h. pasan
15. Coge los platos, los vasos y los cubiertos para ... la mesa antes de comer.			
e. meter	f. colocar	g. montar	h. poner
16. María usa todos los días zapato ..., nunca usa tacones.			
e. sin cordones	f. sencillo	g. plano	h. liso

17. Prefieren frío polar a calor ...			
e. malo	f. llevadero	g. horrible	h. veraniego
18. Las máquinas hacen un ruido ...			
e. pegadizo	f. fuerte	g. amplio	h. metálico
19. Conseguir lo que deseas sería ... un sueño .			
e. cumplir	f. hacer	g. rellenar	h. pisar
20. Vamos a ... un brindis , levantad las copas y bebed.			
e. hacer	f. dedicar	g. chocar	h. dar
21. El ladrón ... todas las huellas de la escena del crimen.			
e. rastreó	f. dejó	g. borró	h. encontró
22. María siempre bebe leche ... , a ella no le importa engordar.			
e. gorda	f. llena	g. entera	h. cremosa
23. ¿Tienes seis hermanos? A mí también me gustaría tener una familia ...			
e. real	f. cuantiosa	g. numerosa	h. larga
24. Lucía es una ... amiga , la conozco desde ha mucho tiempo.			
e. joven	f. vieja	g. anciana	h. nueva
25. El juez ... sentencia , condenó al ladrón a tres años de cárcel.			
e. dijo	f. impugnó	g. dictó	h. dio
26. Esperamos dos horas para ... el control en el aeropuerto.			
e. tener	f. chequear	g. perder	h. pasar
27. Su nuevo novio no me ... confianza , no me gusta.			
e. inspira	f. gana	g. tiene	h. abusa de
28. Su primer libro fue un éxito ..., fue uno de los más vendidos del año.			
e. pasajero	f. discreto	g. apoteósico	h. malo
29. Entre mucha luz por la ventana, fuera hace sol ...			
e. naciente	f. flojo	g. ardiente	h. radiante
30. Tienen muchos conflictos, su relación es...			
e. fugaz	f. tormentosa	g. estable	h. intempestiva
31. He ... un catarro , creo que voy a tomar una aspirina.			
e. pillado	f. notado	g. curado	h. pasado
32. Parad de ... barullo , estáis gritando mucho.			
e. componer	f. deshacer	g. montar	h. dar
33. Se han prometido y en pocos meses ... matrimonio .			
e. celebrarán	f. contraerán	g. anularán	h. harán
34. Alguien que tiene una personalidad ... tiene confianza en sí misma.			
e. oculta	f. desbordante	g. desmesurada	h. anormal
35. Pedro siente envidia ... por María. Se alegra por ella, pero le gustaría tener su mismo éxito.			
e. fugaz	f. sana	g. cochina	h. atroz
36. Tengo el sueño ..., no duermo bien.			
e. voluble	f. ligero	g. profundo	h. recurrente

ANEXO XXVIII. TCE CRECU

INSTRUCCIONES: Completa los huecos (...) en las siguientes SECUENCIAS INCOMPLETAS, debes escribir tu respuesta en la columna de RESPUESTAS y en letras mayúsculas (tal y como se hace en el ejemplo). Solo necesitas una palabra. Intenta contestar el mayor número de preguntas, pero no te preocupes si no las puedes contestar todas.

Ejemplo:

La tortilla de patata es un plato rico, es decir, *CON BUEN SABOR*.

ÍTEMS	RESPUESTAS
1. Hace frío , por favor, lleva...	
2. Tener barba es tener pelo en...	
3. En verano siempre llevo gafas de sol , ...	
4. El pelo rubio es de color...	
5. Quería reservar una habitación doble , para...	
6. Un café cortado es un café con...	
7. Tengo sed , ...	
8. En algunos países es obligatorio dar propina ...	
9. Lucía tiene cita a las cinco, ...	
10. El vino tinto tiene un color...	
11. Ese médico tiene buena letra , ...	
12. La comida rápida promueve la alimentación...	
13. Hacer cola es...	
14. Sacar una foto es lo mismo que...	
15. Pedro, pon la mesa . Vamos a...	
16. Un zapato plano es un zapato...	
17. Un calor horrible es ... calor.	
18. Un ruido fuerte es un sonido...	
19. Cumplir un sueño es...	
20. Vamos a hacer un brindis ,	
21. Antes de escapar, el ... borró las huellas para que la policía no lo identificara.	
22. La leche entera ...	
23. Si tienes una familia numerosa , tienes...	
24. Lucía es una vieja amiga , ...	
25. El juez dictó sentencia por la que ...	
26. Para pasar el control en el aeropuerto, ...	
27. Esta empresa me inspira confianza , ...	
28. Su primer libro fue un éxito apoteósico , ...	
29. Hace un sol radiante , el cielo está...	
30. Una relación tormentosa implica tener...	
31. Creo que he pillado un catarro , ...	
32. Parada de montar barullo , ...	
33. En pocos meses contraerán matrimonio : se acaban de ...	
34. Alguien que tiene una personalidad desbordante ,	
35. Pedro siente envidia sana por María, ...	
36. Tengo el sueño ligero , ...	

ANEXO XXIX. INFORME FAVORABLE DEL COMITÉ DE ÉTICA

Informe

Comité de Ética da Investigación e a Docencia da Universidade da Coruña

O Comité de Ética da Investigación e a Docencia da Universidade da Coruña (CEID-UDC), reunido en sesión ordinaria o mércores once de setembro de dous mil dezanove, unha vez estudada a documentación presentada por Ana Orol González en relación co proxecto de investigación "Avaliación da competencia colocacional de aprendices de español como lingua estranxeira", número de expediente 2019-0013

CONSIDERA que, de acordo coa documentación achegada, o proxecto cumpre de forma suficiente os aspectos ética e xuridicamente relevantes da investigación.

En razón do anterior, e sen prexuízo de futuras suxestións para a mellora do desenvolvemento da investigación, ACORDA por unanimidade emitir un

INFORME FAVORABLE

Para que conste aos efectos oportunos, na Coruña, a once de setembro de dous mil dezanove.

CORRALES
LORENZO
MERCEDES - DNI
32435901J

Firmado digitalmente por CORRALES LORENZO
MERCEDES - DNI 32435901J
NÚMERO DE IDENTIFICACIÓN DNI 32435901J
O UNIVERSIDADE DA CORUÑA, O CERTIFICADO
ELECTRÓNICO EMPAQUADO PUBLICO O NÚMERO
DE INDICADA UNIVERSIDADE, O DNI
MÉRQUEDES MERCEDES - DNI 32435901J
CORRALES LORENZO MERCEDES - DNI 32435901J
CORRALES LORENZO MERCEDES - DNI 32435901J
FECHA 2019.09.11 11:58:40 +0200

Asdo.: Mercedes Corrales Lorenzo

Secretaria do CEID-UDC

ANEXO XXX. DOCUMENTO INFORMATIVA PARA LOS APRENDICES

La investigación para la cual le pedimos su participación se titula: “Evaluación de la competencia colocacional” y recibió el informe favorable del Comité de Ética de la Investigación y la Docencia de la Universidade da Coruña.

a) *Equipo investigador*

La estudiante de doctorado que realiza la investigación es Ana Orol González, quien también se encargará de administrar los test a los participantes. La tesis doctoral la dirige Margarita Alonso Ramos

b) *Objetivo y utilidad del estudio*

El objetivo de esta investigación es conocer la competencia que los aprendices tienen de las colocaciones del español. El instrumento de medida elegido es un test diseñado exclusivamente para este fin. A nivel científico, se busca confirmar si el conocimiento colocacional está vinculado con la competencia lingüística del aprendiz. Algo que se ha estudiado ya en otras lenguas, pero todavía no en español. También se pretende detectar las posibles dificultades que estas unidades del léxico presentan en el aprendizaje del español como lengua extranjera y así facilitar su aprendizaje.

c) *Selección de las personas participantes*

Para la selección de los participantes, en el estudio se han tenido en cuenta únicamente dos criterios: no tener como lengua materna el español y estar actualmente estudiándolo. Es decir, responder afirmativamente a las siguientes preguntas

¿Tu lengua materna es una lengua diferente al español?	Si <input type="checkbox"/>	Non <input type="checkbox"/>
¿Estás estudiando español?	Si <input type="checkbox"/>	Non <input type="checkbox"/>

d) *Metodología que se utilizará para el estudio, tipo de colaboración de la persona participante y duración de dicha colaboración*

Para la recogida de información se utilizará un test, dividido en cuatro secciones. En la primera sección se piden una serie de datos personales relacionados exclusivamente con la biografía lingüística de los participantes. En la segunda sección, un test para medir el conocimiento léxico. En la tercera y cuarta sección, se mide el conocimiento de las colocaciones desde dos perspectivas. La duración estimada es de treinta a cuarenta minutos.

e) *Tipo de información*

En ningún momento se solicita nombre o documento identificativo, se trata de una participación de carácter anónimo. Los datos personales están únicamente vinculados a la biografía lingüística de los participantes. El resto de información recogida atiende al conocimiento léxico y al conocimiento colocacional de los participantes.

f) *Decisión de no participar*

Antes de iniciar el test los participantes pueden abstenerse de hacerlo. Además, durante el proceso también podrán dejar el test inacabado.

g) *Datos de contacto do investigador principal para aclaraciones o consultas*

ana.orol.gonzalez@udc.es

2. COMPROMISO DE CONFIDENCIALIDAD

a) *Medidas para asegurar el respeto a la intimidad y a la confidencialidad de los datos personales*

Se adoptaron las medidas oportunas para garantizar la completa confidencialidad de sus datos personales, conforme a lo que dispone la LO 3/2018, del 5 de diciembre, de protección de datos personales y garantiza los derechos digitales y el Reglamento (UE) 2016/679 del Parlamento Europeo y del Consejo, del 27/04/2016, relativo a la protección de las personas físicas en lo que respecta al tratamiento de los datos personales y a la libre circulación de estos datos y por lo que se deroga la Directiva 95/46CE (Reglamento general de protección de datos).

Los datos necesarios para llevar a cabo este estudio serán recogidos y conservados del siguiente modo:

Anónimos, es decir, que usted no podrá ser identificado si siquiera por el equipo investigador. En ningún caso se juntarán los consentimientos otorgados, donde si se identifica al sujeto, con los instrumentos utilizados en el estudio.

En el uso que se realice de los resultados del estudio con fines de docencia, investigación, publicación y/o divulgación se respetarán siempre la debida confidencialidad de los datos de carácter personal, de modo que las personas participantes no resultarán identificadas o identificables.

b) *Derechos de acceso, rectificación, supresión, limitación del tratamiento, portabilidad y oposición dos sus datos*

Si en algún momento decide que sus datos no se pueden utilizar para la investigación, debe ponerse en contacto con la investigadora a través de la dirección de correo electrónico: ana.orol.gonzalez@udc.es

c) *Cesión, reutilización e período de retención dos datos*

Los datos de la investigación se conservarán de forma anónima para poder utilizarlas en futuras investigaciones, siempre y cuando se cuente con la autorización explícita de los participante

ANEXO XXXI. HOJA DE CONSENTIMIENTO

Don/doña _____,
 mayor de edad, con DNI _____ y domicilio en _____,
 _____,

DECLARO que

Fui informado/a de las características del estudio	Si <input type="checkbox"/>	Non <input type="checkbox"/>
Leí la hoja de información que me entregaron	Si <input type="checkbox"/>	Non <input type="checkbox"/>
Pude realizar observaciones o preguntas y me aclararon las dudas	Si <input type="checkbox"/>	Non <input type="checkbox"/>
Comprendí las explicaciones que se me facilitaron y en las que consiste mi participación en el estudio	Si <input type="checkbox"/>	Non <input type="checkbox"/>
Sé cómo y a quién dirigirme para realizar preguntas sobre el estudio en el presente o en el futuro	Si <input type="checkbox"/>	Non <input type="checkbox"/>
Confirmando que mi participación es voluntaria	Si <input type="checkbox"/>	Non <input type="checkbox"/>
Comprendo que puedo revocar el consentimiento en cualquier momento sin tener que dar explicaciones y sin que repercuta negativamente en mi persona	Si <input type="checkbox"/>	Non <input type="checkbox"/>

CONSIENTO

Participar en el estudio	Si <input type="checkbox"/>	Non <input type="checkbox"/>
Que se utilicen los datos facilitados para la investigación	Si <input type="checkbox"/>	Non <input type="checkbox"/>
Que se utilicen los datos facilitados en publicaciones científicas	Si <input type="checkbox"/>	Non <input type="checkbox"/>
Que se utilicen los datos facilitados en reuniones y congresos	Si <input type="checkbox"/>	Non <input type="checkbox"/>
Que se utilicen los datos facilitados para la docencia	Si <input type="checkbox"/>	Non <input type="checkbox"/>
Que se conserven los datos de forma anónima al finalizar el estudio para su uso en futuras investigaciones	Si <input type="checkbox"/>	Non <input type="checkbox"/>

SOLICITO

Acceder a los resultados generales del estudio	Si <input type="checkbox"/>	Non <input type="checkbox"/>
--	-----------------------------	------------------------------

Y en prueba de la conformidad, firmo el presente documento en el lugar y en la fecha que se indican a continuación.

_____, _____ de _____ de ____.

<p><i>Nombre y apellidos del/de la participante:</i></p> <p>Firma:</p>	<p><i>Nombre y apellidos del representante (en el caso de menores o incapaces):</i></p> <p>Firma:</p>	<p><i>Nombre y apellidos del/de la investigador/a principal:</i></p> <p>Firma:</p>
---	--	---

La persona menor o sin capacidad para tomar por sí misma la decisión de participar ha de ser informada con mayor claridad posible, y el consentimiento de sus representantes ha de tener en cuenta sus deseos y objeciones (Ley 14/2007, artículo 20).

La persona menor de edad madura (con capacidad intelectual y emocional) tiene derecho a ser oída e ha de consentir su participación. Se presume a madu

