

## LA SELECCIÓN TEMÁTICA DEL VOCABULARIO PARA FINES DIDÁCTICOS: EVALUACIÓN DE UN ACERCAMIENTO CUANTITATIVO

Jasper Degraeuwe

Patrick Goethals

Ghent University, Belgium

**Resumen:** El presente estudio tiene por objetivo evaluar los resultados de un acercamiento cuantitativo a la selección temática del vocabulario con fines didácticos. Describimos en detalle cómo se configuran y se combinan tres medidas cuantitativas (la frecuencia absoluta, el *keyness* y la dispersión) a fin de automatizar la selección del vocabulario específico de un corpus especializado. A continuación evaluamos si la selección automática se ve confirmada por el juicio de profesores ELE. Hemos podido comprobar, en efecto, que en más del 85% de los casos el resultado del método cuantitativo es confirmado por al menos la mitad de los profesores. Esta observación también se evidencia estadísticamente, con un test de *interrater reliability* que demuestra un acuerdo sustancial (Cohen's kappa = 0,61) entre el juicio de los profesores y la selección automática.

**Palabras clave:** lingüística de corpus, aprendizaje de vocabulario, selección del vocabulario automática, selección del vocabulario temática, frecuencia absoluta, *keyness*, dispersión, español como lengua extranjera (ELE).

### THEMATIC VOCABULARY SELECTION FOR DIDACTIC PURPOSES: EVALUATION OF A QUANTITATIVE APPROACH

**Abstract:** *The aim of this study is to evaluate the results of a quantitative approach to the thematic selection of vocabulary for didactic purposes. We describe in detail how three quantitative measures (absolute frequency, keyness and dispersion) are configured and combined to automate the selection of specific vocabulary from a specialized corpus. We then evaluate whether the automatic selection is confirmed by the judgements of SFL teachers. The results of this evaluation experiment show that in more than 85% of the cases the output of the quantitative selection method is accepted by at least half of the teachers. This observation is also backed from a statistical angle, with the outcome of an interrater reliability test indicating that there is a substantial agreement (Cohen's kappa = 0.61) between the judgements of the teachers and the automatic selection.*

**Keywords:** *corpus linguistics, vocabulary learning, automatic vocabulary selection, thematic vocabulary selection, absolute frequency, keyness, dispersion, Spanish as a foreign language (SFL).*

## 1. INTRODUCCIÓN

En la adquisición de una lengua extranjera, el aprendizaje del vocabulario es uno de los procesos más lentos y costosos (Schmitt, 2008). Cualquier forma de enseñanza del vocabulario, sea implícita, mediante la lectura o el visionado de fuentes auténticas, sea explícita, mediante actividades de aula o de autoestudio, requiere una importante inversión de tiempo y esfuerzo por parte del aprendiz, sobre todo debido a la dificultad de formular principios generalizadores y el consiguiente carácter repetitivo o enumerativo del proceso (Vincze, 2015). Por eso, uno de los mayores desafíos para los autores de materiales didácticos enfocados en el aprendizaje del vocabulario es el de realizar una acertada selección y secuenciación de los ítems léxicos (Laufer et al., 2005; Nation, 2016). Se trata de uno de los problemas clásicos de la enseñanza del vocabulario y existen varias propuestas didácticas de selección del vocabulario, sobre todo para los niveles iniciales y/o intermedios (véanse p.ej. los diccionarios de español como lengua extranjera Buyse et al. [2004] y Davies [2006], y los resúmenes bibliográficos en Izquierdo Gil [2005] y García Salido y Alonso Ramos [2018]).

La selección del vocabulario con fines didácticos debe tener en cuenta al menos dos principios: el nivel de competencia del aprendiz, y el dominio semántico que se quiere trabajar. Aunque existe cierto solapamiento entre los dos principios, por ejemplo cuando los niveles de competencia se definen en función de los tipos de situaciones comunicativas en las que un hablante deba poder desenvolverse, en principio se trata de dos preguntas conceptualmente distintas. Salvo en casos excepcionales, un dominio semántico consta de unidades

**To cite this article:** Degraeuwe, J., Goethals, P. (2020). "Thematic vocabulary selection for didactic purposes: evaluation of a quantitative approach". *Revista de Lingüística y Lengas Aplicadas*, 15, 1-14. <https://doi.org/10.4995/ryla.2020.11969>

**Correspondence authors:** [jasper.degraeuwe@ugent.be](mailto:jasper.degraeuwe@ugent.be), [Patrick.Goethals@UGent.be](mailto:Patrick.Goethals@UGent.be)



léxicas que pueden clasificarse según sean más básicas o más avanzadas, y, viceversa, los vocablos relevantes para un nivel de competencia pueden clasificarse según el dominio semántico al que pertenecen. En Goethals, Tezcan y Degraeuwe (2019) hemos descrito un sistema automático que clasifica los lexemas según el nivel de competencia basado en un sistema de aprendizaje automático supervisado que vincula los juicios de dificultad de los aprendices con características como la frecuencia de uso en corpus o el grado de similitud formal (cognados). En esta contribución, en cambio, indagaremos en la selección temática o semántica.

Es muy común destacar la importancia y la utilidad de los corpus en este contexto (Bowker y Pearson, 2002; Sinclair, 2005), ya que los corpus permiten calcular la frecuencia de uso y otras medidas derivadas de la frecuencia, como la distribución a través de los textos (Okamoto, 2015) o el carácter específico (*keyness*) de los ítems para un dominio temático. El conjunto de estas medidas permite sistematizar y objetivar la selección. A pesar de que los principios generales son relativamente bien conocidos, existen muy pocos estudios que explican en detalle cómo se pueden configurar y combinar estas medidas exactamente, y que intentan evaluar la relevancia didáctica de esta selección cuantitativa y automatizada.

En esta contribución queremos contribuir a llenar este hueco. En la sección 2 describimos la exacta configuración de tres medidas cuantitativas que reflejan la frecuencia absoluta, la diferencia de frecuencia entre corpus, y las diferencias de frecuencia en el interior de un corpus. La combinación de los tres criterios pretende ser capaz de extraer el “léxico típico” de un dominio semántico. En segundo lugar, se plantea la pregunta de saber en qué medida la selección cuantitativa y automatizada se corresponde con la intuición de profesores expertos (sección 3). No en vano, el Instituto Cervantes decidió dar prioridad a la “apreciación intuitiva basada en la experiencia docente” (*Plan Curricular, Sección 9*, citado en García Salido y Alonso Ramos [2018:42]): un acercamiento cuantitativo puede perfeccionar, tal vez corregir y en todo caso facilitar la selección intuitiva, pero no debería contradecirla. Es lo que verificaremos de manera experimental en la parte exploratoria de este estudio, al comparar el resultado del análisis basado en corpus con el juicio de varios profesores experimentados. Por último, formulamos nuestras conclusiones en la sección 4.

Sin embargo, antes de presentar los resultados de la investigación, cabe destacar que nuestro acercamiento parte de algunos presupuestos didácticos específicos que conviene explicitar porque establecen las pautas para el diseño del experimento empírico. Nos dirigimos a un público de aprendices de nivel intermedio-alto y avanzado, con un nivel B2 adquirido, que desean adquirir conocimientos léxicos sobre un dominio semántico particular, lo cual diferencia nuestro enfoque de otros procedimientos de selección léxica enfocados en los niveles de aprendizaje iniciales. Además, trabajamos desde la perspectiva del aprendizaje basado en datos (Boulton, 2017), donde los corpus no solo se usan con fines de investigación sino también como fuente de actividades pedagógicas. Nos dirigimos a estudiantes que quieren estudiar de forma autónoma y haciendo uso de corpus, a quienes queremos dar el máximo grado posible de libertad para seleccionar el dominio temático o incluso los elementos léxicos de su interés. La plataforma que hemos diseñado (Goethals et al., 2017; Goethals, 2018) permite consultar un abanico de corpus especializados (y en el futuro permitirá subir corpus nuevos), facilitando procedimientos de extracción automática del léxico más típico para cierto conjunto de textos.

## 2. LA SELECCIÓN LÉXICA: UN ACERCAMIENTO CUANTITATIVO

### 2.1 Estudio de caso: las juntas de accionistas

El corpus utilizado para este estudio de caso es un corpus especializado de discursos y presentaciones pronunciados en el ámbito específico de una junta general (ordinaria) de accionistas, en el período 2015-2017. El aprendiz que tenemos en mente puede ser un estudiante especializado en interpretación, comunicación o traducción que ya tiene un nivel B2 en español y que quiere familiarizarse eficazmente con el vocabulario de este ámbito especializado. El corpus, dividido en 5 subcorpus según sector empresarial, contiene 74 transcripciones de discursos, y refleja en principio el español peninsular porque solo hemos incluido empresas españolas. La Tabla 1 recoge la información más relevante sobre la composición del corpus.

Tabla 1. El corpus de discursos pronunciados en Juntas Ordinarias de Accionistas.

Subcorpus	Nº palabras	Porcentaje	Empresas (número de textos)
Comunicación	61.302	26,05	Atresmedia (6); Mediaset España (6); Prisa (6); Telefónica (2)
Construcción	44.146	18,76	Acciona (3); Grupo ACS (6); Técnicas Reunidas (4)
Energía	54.965	23,36	Gamesa (4); Gas Natural Fenosa (2); Red Eléctrica de España (5); Saeta Yield (3)
Finanzas	58.548	24,88	Bankia (2); Bolsas y Mercados Españoles (12); Mapfre (4); Santander (2)
Transporte	16.334	6,94	Construcciones y Auxiliar de Ferrocarriles (1); International Airlines Group (6)
<b>Total</b>	<b>235.295</b>	<b>100</b>	<b>17 (74)</b>

Para el procesamiento de nuestro corpus usamos la herramienta SCAP (Spanish Corpus Annotation Project; Goethals et al., 2017; Goethals, 2018), que etiqueta de forma automática las palabras del corpus, las lematiza, y computa las métricas presentadas en la sección 2.2.

## 2.2. Medidas de frecuencia, carácter clave y dispersión

Conviene destacar que en este estudio trabajamos con la forma lematizada de las palabras del corpus, y que nos centramos exclusivamente en elementos léxicos de una sola palabra (*unigrams*). En lo que sigue presentamos en detalle tres medidas cuantitativas que tienen por objetivo poder automatizar la selección temática del léxico. Las medidas, respectivamente, filtran los lemas por ser frecuentes en el corpus (2.1.1), más frecuentes que en un corpus de lenguaje general (2.1.2) y típicos para el género en su conjunto y no para unas instancias particulares (2.1.3).

### 2.2.1. La frecuencia absoluta, representada en percentiles de frecuencia

Los percentiles de frecuencia son una manera sintética de ordenar los lemas en base a datos de frecuencia absoluta. Facilitan el manejo de los datos de frecuencia porque permiten agrupar lemas con una frecuencia similar y filtrarlos de manera progresiva, seleccionando por ejemplo el 10% o el 20% de los lemas más frecuentes. Sin embargo, la forma más lineal de calcular los percentiles, que es la de ordenar todos los ítems según su frecuencia y de calcular luego su posición relativa en la escala, es más problemática de lo que aparece a primera vista. En concreto, surge un problema porque a menudo hasta el 40% de los lemas solo ocurren una vez en corpus de tamaño relativamente reducido (lo cual es a menudo el caso con corpus especializados). La consecuencia es que todos los otros lemas pertenecen casi automáticamente a, aproximadamente, los percentiles 41100, lo cual distorsiona la imagen, y complica la comparabilidad de los datos. Por eso, la herramienta SCAP otorga seis diferentes puntuaciones de percentil (Tabla 2), colocando los ítems con una sola ocurrencia en una categoría aparte, y luego subdividiendo los lemas con más de una ocurrencia en 5 grupos de menor a mayor frecuencia. En Goethals et al. (2019) hemos podido comprobar que este método aumenta significativamente el valor de los datos de frecuencia al entrenar un modelo de predicción automática.

Tabla 2. Distribución de los percentiles según el sistema SCAP.

Categoría	Detalles
0	1 ocurrencia en el corpus
1	percentiles 1 – 20 de los lemas con más de una ocurrencia
2	percentiles 21 – 40 de los lemas con más de una ocurrencia
3	percentiles 41 – 60 de los lemas con más de una ocurrencia
4	percentiles 61 – 80 de los lemas con más de una ocurrencia
5	percentiles 81 – 100 de los lemas con más de una ocurrencia

### 2.2.2. Carácter clave (*Keyness*)

Un ítem léxico se considera un ítem clave (*key item* o *keyword*) para un corpus determinado cuando presenta una frecuencia inusualmente alta en este corpus en comparación con un corpus de referencia (Scott, 1996; 1997). El carácter clave (*keyness*) es, por tanto, un indicador de la importancia de un elemento léxico clave en un corpus determinado (Biber et al., 2007). Las métricas más comunes para determinar el *keyness* están basadas en la significación estadística, es decir, pruebas como logverosimilitud o de chi cuadrado, que permiten determinar si una diferencia de frecuencia es estadísticamente significativa o no (p.ej. Oakes y Farrow, 2007). Sin embargo, Gabrielatos y Marchi (2011) argumentan que los tests de significación estadística no son las medidas más adecuadas, al superponerse en un valor el efecto del tamaño del corpus y la magnitud de la diferencia. Por esta razón, Gabrielatos y Marchi (2011) sugieren dividir el método en dos métricas separadas. Por un lado, miden la magnitud de la diferencia mediante la fórmula %DIFF (Figura 1), y, por otro lado, determinan (con la prueba del chi cuadrado) si la diferencia es estadísticamente significativa o no.

$$\frac{(frec. normalizada corpus estudio - frec. normalizada corpus referencia) * 100}{frec. normalizada corpus referencia}$$

Figura 1. %DIFF (Gabrielatos & Marchi, 2011).

Gracias a este método dual, se llega a una interpretación más sutil de las diferencias entre el corpus de estudio y el corpus de referencia, que es particularmente interesante porque permite sopesar el efecto del tamaño de los corpus utilizados. Concretamente, si se trabaja con corpus muy grandes, se pueden rechazar las diferencias estadísticamente significativas pero de tamaño reducido (para reducir los falsos positivos), o si se trabaja con corpus pequeños, es posible considerar los casos que presentan diferencias absolutas posiblemente interesantes pero no significativas y, así, reducir el número de falsos negativos.

Para calcular la métrica %DIFF usamos un corpus de literatura juvenil de aproximadamente 7,5 millones de palabras que se ha compilado y anotado dentro del proyecto SCAP como corpus de referencia. A modo de ilustración, aplicamos la fórmula al sustantivo *inversión* (véase la Tabla 3). Para determinar la frecuencia normalizada se usa el número de palabras de contenido (*content words*) como punto de referencia, omitiendo palabras funcionales como artículos, preposiciones, etc. El valor final (46.128) se debe leer como “a la frecuencia de 23 que tiene el lema *inversión* en el corpus de referencia, hay que sumar el 46.128% de 23 para alcanzar la frecuencia que tiene este lema en el corpus de estudio”. Así, obtenemos el valor de 10.632, es decir, la frecuencia normalizada de *inversión* en el corpus de estudio. Por último, averiguamos con una fórmula de chi cuadrado (usando los datos no normalizados) si la diferencia de frecuencia es estadísticamente significativa ( $p = 0,01$ ), lo cual efectivamente es el caso.

Tabla 3. Ilustración del cálculo del valor %DIFF.

Criterio	Corpus de estudio	Corpus de referencia
#palabras	235.295	7.528.422
#palabras de contenido	131.049	4.050.528
# <i>inversión</i>	344	23
Frecuencia normalizada	$(344 / 131.049) \times 4.050.528 = 10.632$	23
%DIFF	$[(10.632-23) \times 100] / 23 = 46.128$	/

### 2.2.3. Dispersión

Siguiendo a Gries (2008), optamos por completar la métrica del *keyness* por la métrica de la dispersión, que refleja la distribución de un elemento léxico por las partes del corpus. Gries (2008) arguye, en concreto, que el hecho de solo considerar la frecuencia de ocurrencia puede resultar problemático, ya que una frecuencia de uso muy alta de un determinado elemento léxico se puede deber a su presencia en un número limitado de textos, y no ser representativa para el corpus entero. Por esta razón, Gries (2008) sugiere la medida *DP*, una métrica que genera un resultado entre cero y uno y que se calcula de la siguiente manera (véase también la Tabla 4 para un ejemplo):

1. Determinar los tamaños  $s_{1-n}$  de las  $n$  partes del corpus, expresados en un *porcentaje esperado* (= el tamaño de la parte dividido por el tamaño total del corpus);
2. Determinar las frecuencias relativas reales  $v_{1-n}$  con las cuales  $a$  ocurre en las  $n$  partes del corpus, expresadas en un *porcentaje observado* (= la frecuencia que presenta un lema en la parte dividida por su frecuencia total en el corpus);
3. Computar, en pares, todas las  $n$  diferencias absolutas de los porcentajes esperados y observados, sumarlas y dividir el resultado por dos a fin de obtener un valor final entre cero y uno.

Valores cercanos a cero indican que  $a$  presenta una distribución uniforme por las  $n$  partes del corpus, mientras que valores cercanos a uno apuntan a una distribución muy desigual. Dicho de otro modo, la medida *DP* permite identificar los ítems léxicos que pertenecen al núcleo del género, es decir, los que, en el caso de haber subgrupos, ocurren en muchos subgrupos y con una frecuencia normalizada similar. En cambio, se pueden filtrar los ítems que por su frecuencia total podrían considerarse típicos del género, pero que en realidad solo lo son para una(s) instancia(s) del género en particular.

Para ilustrar su calculación y utilidad, aplicamos la métrica *DP* a un ejemplo concreto, el sustantivo *endeudamiento*.

Tabla 4. Ilustración del cálculo del valor *DP*.

Subcorpus	#palabras	# <i>endeudamiento</i>	% esperado	% observado	Diferencia
Comunicación	61.302	9	0,26	0,24	0,02
Construcción	44.146	15	0,19	0,41	0,22
Energía	54.965	7	0,23	0,19	0,04
Finanzas	58.548	6	0,25	0,16	0,09
Transporte	16.334	0	0,07	0	0,07
Total	235.295	37	1	1	0,44
Valor <i>DP</i>					0,22

En la Tabla 5 ilustramos la funcionalidad de *DP*. Se observa que *consolidar*, *endeudamiento* y *neto* representan un valor relativamente bajo, lo cual significa que están repartidos por el corpus entero. *Aerolínea*, *petroquímico* y *transportar*, en cambio, tienen un valor mucho más alto de *DP*, lo cual nos permite concluir que esos elementos léxicos son más bien típicos de una de las cinco partes del corpus. El caso de *aerolínea* es particularmente interesante, porque muestra que una frecuencia elevada (duplicando, por ejemplo, la de *endeudamiento*) no significa necesariamente que sea un ítem típico para el género en su conjunto.

Tabla 5. Ejemplos de valores *DP*.

PoS (Part of Speech)	LEMA	FREC	DP
V (verbo)	consolidar	101	0,09
ADJ (adjetivo)	neto	185	0,18
NC (nombre común)	endeudamiento	37	0,22
V	transportar	5	0,74
ADJ	petroquímico	10	0,81
NC	aerolínea	72	0,93

#### 2.2.4. Combinación de los tres criterios

Finalmente, combinamos los tres criterios a fin de crear una lista de candidatos supuestamente “idóneos” para entrar en un glosario reducido. En el marco de este experimento nos limitamos a una selección de 100 lemas. Con “candidatos supuestamente idóneos” nos referimos a los lemas específicos (basándonos en *%DIFF*) y frecuentes (basándonos en el percentil de frecuencia) que, además, presentan una distribución relativamente homogénea en el corpus entero (basándonos en *DP*). Para llegar a la lista de 100 ítems, aplicamos los siguientes criterios:

Tabla 6. Configuración y combinación de las métricas para seleccionar 100 ítems típicos.

Medida	Criterio	Detalles
Percentil	categorías 4 y 5	Al manejar este umbral solo mantenemos los lemas frecuentes en el corpus de estudio.
<i>%DIFF_SIGN</i>	$p \leq 0,01$	Solo mantenemos los lemas cuya frecuencia en el corpus de estudio difiere significativamente de su frecuencia en el corpus de referencia.
<i>DP</i>	$\leq 0,5$	Solo mantenemos los lemas que tienen una distribución relativamente homogénea a través de los textos incluidos en el corpus de estudio.
<i>%DIFF</i>	ordenar	Ordenamos la lista de ítems restantes de mayor a menor valor <i>%DIFF</i> , para poder seleccionar los 100 candidatos más específicos.

El proceso de selección se realiza independientemente de las categorías morfológicas, por lo cual es posible que se incluyan dos lemas derivados de la misma raíz (p.ej. *sostenible* y *sostenibilidad*). Para evitar que así disminuya la riqueza semántica de la lista, llevamos a cabo una intervención manual y solo mantenemos el lema más frecuente. De esta manera, obtenemos una lista incluyendo 100 candidatos relevantes para un primer acercamiento al léxico propio de este género (Tabla 7). Excluyendo la fase en la cual eliminamos manualmente las mismas raíces léxicas, la creación de esta lista de ítems clave es un proceso completamente automático, que puede repetirse para cualquier otro tipo de corpus previamente etiquetado y lematizado, que es un paso que también se realiza de manera automática en la plataforma SCAP.

Tabla 7. Selección de 100 lemas típicos.

#	PoS	Lema	#	PoS	Lema	#	PoS	Lema
1	NC	dividendo	35	NC	filial	68	NC	balance
2	NC	ratio	36	NC	implantación	69	NC	coste
3	ADJ	sostenible	37	NC	crecimiento	70	NC	activo
4	ADJ	regulatorio	38	NC	integración	71	ADJ	organizativo
5	NC	diversificación	39	NC	incremento	72	ADJ	diferencial
6	NC	endeudamiento	40	ADJ	climático	73	NC	comisión
7	ADJ	normativo	41	NC	solvenia	74	NC	bono

Tabla 7, continúa en la página siguiente

Tabla 7, continúa de la página anterior

#	PoS	Lema	#	PoS	Lema	#	PoS	Lema
8	NC	liquidez	42	NC	contratación	75	ADJ	cualitativo
9	NC	capitalización	43	NC	financiación	76	ADJ	impositivo
10	ADJ	macroeconómico	44	NC	reducción	77	NC	gestión
11	NC	digitalización	45	NC	accionista	78	ADJ	global
12	NC	sinergia	46	V	maximizar	79	NC	ampliación
13	NC	adjudicación	47	NC	reestructuración	80	NC	cumplimiento
14	NC	volatilidad	48	NC	competitividad	81	ADJ	precedente
15	NC	revalorización	49	NC	liderazgo	82	NC	reglamento
16	NC	vicepresidente	50	NC	inversión	83	NC	semestre
17	V	optimizar	51	NC	remuneración	84	ADJ	anual
18	NC	discapacidad	52	NC	consolidación	85	NC	contribución
19	NC	internacionalización	53	ADJ	atribuible	86	NC	posicionamiento
20	NC	auditoría	54	NC	record	87	NC	productividad
21	NC	devaluación	55	NC	cotización	88	ADJ	bursátil
22	NC	apalancamiento	56	NC	mejora	89	ADJ	emergente
23	ADJ	sectorial	57	NC	segmento	90	NC	tasa
24	NC	ralentización	58	NC	déficit	91	NC	consecución
25	NC	disrupción	59	NC	retribución	92	NC	desarrollo
26	NC	reelección	60	V	priorizar	93	NC	asignación
27	ADJ	coordinador	61	NC	ratificación	94	NC	trimestre
28	ADJ	credificio	62	NC	vencimiento	95	NC	euro
29	ADJ	mayorista	63	ADJ	institucional	96	NC	adquisición
30	ADJ	geopolítico	64	NC	desempeño	97	NC	distribución
31	ADJ	porcentual	65	ADJ	consejero	98	NC	innovación
32	ADJ	neto	66	NC	proveedor	99	ADJ	ejecutivo
33	NC	rentabilidad	67	NC	eficiencia	100	NC	junta
34	ADJ	corporativo						

### 3. COMPARACIÓN CON EL JUICIO DE PROFESORES ELE

#### 3.1. Metodología

A fin de averiguar la pertinencia didáctica de la selección cuantitativa automatizada llevamos a cabo una investigación exploratoria, en la cual comparamos los resultados de dicha selección con los juicios de diez profesores expertos en la enseñanza ELE en el ámbito universitario. En concreto, el experimento consiste en presentar a los profesores ELE un único listado de los 100 lemas “idóneos” (véase la Tabla 7), junto con otros 100 supuestamente “no idóneos” (véase más abajo), pidiéndoles que seleccionen los 100 ítems que estiman más adecuados para incluir en un glosario que recoge el vocabulario típico de una junta general de accionistas. De esta manera, pretendemos averiguar en qué medida la selección automática de 100 lemas corresponde a la selección más intuitiva y subjetiva de profesores ELE.

##### 3.1.1. Participantes

Dado que el principal objetivo del experimento no es relacionar los resultados con determinadas características de los participantes, sino más bien captar el juicio de los profesores ELE en general, intentamos llegar a una muestra de participantes de carácter diverso (véase la Tabla 8 para información sobre el sexo, la edad, la nacionalidad y la experiencia [en años] de los participantes). Por el mismo motivo, no definimos criterios de selección rígidos: los participantes solo han de ser profesores ELE en el ámbito universitario con al menos tres años de experiencia, además de estar familiarizados con la enseñanza del vocabulario.



Tabla 8. Características sociodemográficas y académicas de los participantes.

Participante	Sexo	Edad	Nacionalidad	Experiencia
1	Fem	30 – 40	ES (español)	5 – 10
2	Fem	> 60	BE (belga)	> 10
3	Fem	40 – 50	BE	> 10
4	Fem	> 60	BE	> 10
5	Fem	50 – 60	BE	> 10
6	Mas	30 – 40	BE	5 – 10
7	Fem	50 – 60	BE	> 10
8	Fem	25 – 30	BE	3 – 5
9	Mas	30 – 40	ES	> 10
10	Fem	25 – 30	BE	3 – 5

### 3.1.2. Experimento

Como ya hemos mencionado, para el experimento agregamos 100 lemas “no idóneos” a la lista “idónea”. Se trata de ítems que sí ocurren en el corpus especializado, pero que no pasarían por los filtros usados en la selección automática. Primero, añadimos 50 lemas frecuentes pero no específicos, que corresponden a los criterios presentados en la Tabla 9.

Tabla 9. Criterios aplicados para seleccionar 50 ítems frecuentes pero no específicos.

Medida	Criterio	Detalles
Percentil	categorías 4 y 5	Al manejar este umbral, solo mantenemos los lemas frecuentes.
%DIFF_SIGN	p > 0,01	Solo entran en consideración los lemas cuyo valor %DIFF no muestra ninguna diferencia estadísticamente significativa (p = 0,01). Así, solo mantenemos los lemas menos/no específicos.
Percentil	ordenar	Ordenamos la lista de ítems restantes de mayor a menor frecuencia y seleccionamos los 50 primeros candidatos.

Tabla 10. Lista de 50 ítems frecuentes pero no específicos.

#	PoS (Part of Speech)	Lema	#	PoS	Lema
1	V	seguir	26	NC	entrada
2	ADJ	bueno	27	V	recibir
3	NC	mundo	28	NC	papel
4	NC	persona	29	NC	respuesta
5	NC	lugar	30	V	apoyar
6	ADJ	propio	31	V	compartir
7	V	conseguir	32	V	señalar
8	NC	manera	33	V	ganar
9	NC	caso	34	ADJ	enorme
10	ADJ	posible	35	V	partir
11	ADJ	fuerte	36	V	confiar
12	V	convertir	37	NC	orden
13	NC	duda	38	NC	centro
14	V	mostrar	39	NC	detalle
15	NC	final	40	NC	dirección
16	NC	historia	41	ADJ	real
17	V	lograr	42	ADJ	profundo
18	ADJ	capaz	43	NC	peso
19	NC	hecho	44	V	sufrir

Tabla 10, continúa en la página siguiente

Tabla 10, continúa de la página anterior

#	PoS (Part of Speech)	Lema	#	PoS	Lema
20	NC	resto	45	NC	estado
21	ADJ	claro	46	NC	motivo
22	ADJ	bajo	47	NC	circunstancia
23	ADJ	difícil	48	V	cubrir
24	NC	atención	49	NC	cuestión
25	V	asegurar	50	V	existir

A continuación, añadimos 50 lemas específicos (aunque no necesariamente del corpus entero) pero menos frecuentes, para lo cual aplicamos los criterios recogidos en la Tabla 11.

Tabla 11. Criterios aplicados para seleccionar 50 ítems específicos pero no frecuentes.

Medida	Criterio	Detalles
Percentil	categorías 1, 2 y 3	Al manejar este umbral solo mantenemos los lemas menos frecuentes.
%DIFF_SIGN	$p \leq 0,01$	Rechazamos los lemas cuya frecuencia en el corpus de estudio no difiere de manera significativa de su frecuencia en el corpus de referencia.
DP	$> 0,66$	Rechazamos los lemas con una distribución relativamente homogénea en el corpus entero.
%DIFF	ordenar	Ordenamos la lista de ítems restantes de mayor a menor valor %DIFF y seleccionamos los 50 primeros candidatos.

Tabla 12. Lista de 50 ítems específicos pero no frecuentes.

#	PoS	Lema	#	PoS	Lema
1	NC	referéndum	26	ADJ	radioeléctrico
2	NC	pyme	27	ADJ	supervisor
3	NC	recompra	28	NC	abaratamiento
4	NC	fortalecimiento	29	ADJ	actuarial
5	NC	biodiversidad	30	NC	autoconsumo
6	NC	certificación	31	ADJ	bituminoso
7	ADJ	multiplataforma	32	NC	brasil
8	ADJ	nominal	33	ADJ	curricular
9	NC	megavatio	34	NC	desaladora
10	NC	formulación	35	ADJ	desfasador
11	ADJ	generalista	36	NC	exclusivista
12	NC	kilovoltio	37	NC	gasificación
13	NC	multicanalidad	38	ADJ	hidroeléctrico
14	NC	petrolera	39	NC	homologación
15	NC	petroquímica	40	ADJ	metanero
16	ADJ	reasegurador	41	ADJ	minorista
17	NC	trillón	42	NC	negociado
18	NC	viabilidad	43	NC	otorgamiento
19	ADJ	comercializador	44	ADJ	poblacional
20	ADJ	consultivo	45	NC	reforzamiento
21	NC	formalización	46	NC	repositorio
22	NC	hidrocarburo	47	ADJ	secuencial
23	ADJ	hipotecario	48	ADJ	biocombustible
24	NC	inconformismo	49	NC	cristalización
25	ADJ	operacional	50	NC	liberalización



Por último, juntamos los 100 lemas supuestamente idóneos (véase la Tabla 7) y los 100 supuestamente no idóneos en un único listado, poniéndolos en orden alfabético.

En cuanto al procedimiento del experimento, optamos por una cita individual con cada participante, en la cual:

El participante lee una hoja de instrucciones (en formato digital; con nosotros como investigadores presentes; sin limitaciones de tiempo), en la cual esbozamos brevemente el contexto y el objetivo del experimento, pero sin revelar detalles sobre la composición del listado alfabético de 200 lemas;

El participante tiene la posibilidad de hacer preguntas a los investigadores, a las cuales solo contestamos si no revelamos nada sobre la composición de la lista;

El participante hace el experimento (en formato digital; con nosotros como investigadores ausentes; sin limitaciones de tiempo), que consiste en seleccionar de la lista alfabética de 200 ítems los 100 lemas que considera los más típicos del contexto de la junta general de accionistas.

### 3.2. Resultados

#### 3.2.1. Análisis descriptivo

La Tabla 13 recoge los resultados del test: la columna *Seleccionado\_correcto* incluye el número de lemas aceptados por el participante que figuran efectivamente en la lista de lemas “idóneos” generada de forma automática; *Seleccionado\_incorrecto* visualiza el número de lemas aceptados que no fueron seleccionados por el método cuantitativo. Ya que el diseño del experimento requiere que los participantes seleccionen 100 lemas en total, los valores para *Seleccionado\_correcto* y *Seleccionado\_incorrecto* equivalen por definición al número de, respectivamente, ítems “no idóneos” rechazados correctamente e incorrectamente. Los resultados muestran una media del 71,9% de coincidencia entre los profesores y la selección automática.

Tabla 13. Resultados de la comparación del juicio de profesores expertos y la selección cuantitativa.

Participante	Seleccionado_correcto	Seleccionado_incorrecto
1	81	19
2	73	27
3	71	29
4	81	19
5	73	27
6	57	43
7	58	42
8	74	26
9	83	17
10	68	32
<b>Promedio</b>	<b>71,9/100</b>	<b>28,1/100</b>

A fin de analizar los resultados más en profundidad, estudiamos la medida en la cual los diferentes juicios de los profesores expertos coinciden o divergen (Tabla 14). Al utilizar un umbral mínimo del  $\geq 5$ , observamos una coincidencia del 86% en los casos “idóneos”, y del 82% en los casos supuestamente “no idóneos”, pero también con un umbral mucho más severo del  $\geq 8$  sigue habiendo coincidencia en el 53% y el 63% de los casos, respectivamente. En otras palabras, 86 de los 100 lemas “idóneos”, que hemos recogido en la Tabla 15, se identifican como típicos del contexto especializado de la junta general de accionistas por al menos la mitad de los participantes (para la lista de los 14 lemas rechazados incorrectamente véase la Tabla 18).

Tabla 14. Grado de acuerdo entre los profesores expertos.

#participantes	Total de casos “idóneos” seleccionados correctamente	Total de casos “no idóneos” rechazados correctamente
10/10	21	19
$\geq 9/10$	41	40
$\geq 8/10$	53	63
$\geq 7/10$	65	67

Tabla 14, continúa en la página siguiente

Tabla 14, continúa de la página anterior

#participantes	Total de casos "idóneos" seleccionados correctamente	Total de casos "no idóneos" rechazados correctamente
≥ 6/10	73	75
≥ 5/10	86	82
≥ 4/10	89	89
≥ 3/10	94	91
≥ 2/10	97	95
≥ 1/10	100	98
≥ 0/10	100	100

Tabla 15. Ítems seleccionados por las métricas cuantitativas y los profesores expertos.

#	PoS	Lema	.../10	#	PoS	Lema	.../10
1	NC	bono	10/10	44	NC	balance	8/10
2	NC	capitalización	10/10	45	NC	comisión	8/10
3	ADJ	consejero	10/10	46	NC	coste	8/10
4	NC	cotización	10/10	47	NC	internacionalización	8/10
5	NC	déficit	10/10	48	ADJ	mayorista	8/10
6	NC	dividendo	10/10	49	V	priorizar	8/10
7	ADJ	ejecutivo	10/10	50	NC	productividad	8/10
8	NC	endeudamiento	10/10	51	NC	ratio	8/10
9	ADJ	impositivo	10/10	52	ADJ	regulatorio	8/10
10	NC	incremento	10/10	53	NC	sinergia	8/10
11	NC	junta	10/10	54	NC	activo	7/10
12	V	optimizar	10/10	55	NC	consolidación	7/10
13	NC	proveedor	10/10	56	NC	distribución	7/10
14	NC	ratificación	10/10	57	NC	diversificación	7/10
15	NC	remuneración	10/10	58	NC	liderazgo	7/10
16	NC	rentabilidad	10/10	59	ADJ	neto	7/10
17	NC	revalorización	10/10	60	ADJ	organizativo	7/10
18	NC	solvencia	10/10	61	NC	ralentización	7/10
19	NC	tasa	10/10	62	NC	reducción	7/10
20	NC	vencimiento	10/10	63	NC	reestructuración	7/10
21	NC	volatilidad	10/10	64	ADJ	sectorial	7/10
22	NC	adjudicación	9/10	65	NC	trimestre	7/10
23	NC	auditoría	9/10	66	NC	asignación	6/10
24	ADJ	bursátil	9/10	67	ADJ	atribuible	6/10
25	NC	competitividad	9/10	68	NC	crecimiento	6/10
26	NC	contratación	9/10	69	NC	desempeño	6/10
27	ADJ	corporativo	9/10	70	NC	disrupción	6/10
28	ADJ	crediticio	9/10	71	NC	mejora	6/10
29	NC	devaluación	9/10	72	NC	reglamento	6/10
30	NC	filial	9/10	73	NC	semestre	6/10
31	NC	financiación	9/10	74	NC	ampliación	5/10
32	NC	gestión	9/10	75	NC	apalancamiento	5/10
33	NC	inversión	9/10	76	NC	contribución	5/10
34	NC	liquidez	9/10	77	ADJ	coordinador	5/10

Tabla 15, continúa en la página siguiente

Tabla 15, continúa de la página anterior

#	PoS	Lema	.../10	#	PoS	Lema	.../10
35	ADJ	macroeconómico	9/10	78	NC	cumplimiento	5/10
36	V	maximizar	9/10	79	ADJ	diferencial	5/10
37	ADJ	porcentual	9/10	80	NC	eficiencia	5/10
38	NC	posicionamiento	9/10	81	ADJ	emergente	5/10
39	NC	retribución	9/10	82	NC	implantación	5/10
40	ADJ	sostenible	9/10	83	NC	innovación	5/10
41	NC	vicepresidente	9/10	84	ADJ	normativo	5/10
42	NC	accionista	8/10	85	NC	reelección	5/10
43	NC	adquisición	8/10	86	NC	segmento	5/10

La Tabla 16 presenta un resumen de los resultados basados en el umbral de coincidencia mínimo fijado en el 50%. Como acabamos de mencionar, 86 de los 100 lemas “idóneos” en la selección automática también fueron seleccionados por los profesores. No obstante, en el contexto de nuestro experimento, también es pertinente analizar más en detalle los casos donde el juicio de los profesores ELE no coincide con la selección automática.

Tabla 16. Resumen del experimento con umbral de aceptación del 50%

100 lemas idóneos	100 lemas no idóneos
Correcto: 86	50 lemas frecuentes pero no específicos Correcto: 49 Incorrecto: 1
Incorrecto: 14	50 lemas específicos pero no frecuentes Correcto: 33 Incorrecto: 17

La Tabla 17 visualiza los elementos “no idóneos” que han sido seleccionados. Como se desprende de la Tabla 16, solo 1 de los 18 ítems (*dirección*) proviene de la lista que incluye lemas frecuentes pero no específicos (*frec\_noEsp*). Los demás ítems de la lista (*noFrec\_esp*) son, pues, específicos (aunque no necesariamente del corpus entero) pero menos frecuentes, lo cual significa que una frecuencia relativamente baja no impide *per se* que se incluya el lema en una lista de ítems clave, siempre y cuando el lema tenga un valor *%DIFF* alto y estadísticamente significativo. Estos resultados apuntan a que lemas con un valor *DP* alto (> 0,66) a veces sí pueden considerarse como relevantes para el estudio de un género, a pesar de ser más bien específicos para un determinado subgénero. Sin embargo, cabe señalar que el valor *DP* alto de, por ejemplo, *certificación*, *minorista*, *nominal*, *pyme* y *viabilidad* también se podría explicar en parte por el tamaño medio de nuestro corpus (aproximadamente 230.000 palabras). Si bien es cierto que en la fórmula *DP* las frecuencias son normalizadas, en los corpus de tamaño pequeño o mediano formados por varios subcorpus es posible que lemas más bien generales con una frecuencia relativamente baja ocurran por casualidad más en una parte del corpus que en otra.

Tabla 17. Ítems no seleccionados por las métricas cuantitativas pero aceptados por los expertos.

#	PoS	Lema	.../10	Grupo	#	PoS	Lema	.../10	Grupo
1	NC	pyme	10/10	<i>noFrec_esp</i>	10	ADJ	operacional	7/10	<i>noFrec_esp</i>
2	NC	viabilidad	10/10	<i>noFrec_esp</i>	11	ADJ	supervisor	7/10	<i>noFrec_esp</i>
3	ADJ	consultivo	9/10	<i>noFrec_esp</i>	12	NC	abaratamiento	6/10	<i>noFrec_esp</i>
4	ADJ	hipotecario	9/10	<i>noFrec_esp</i>	13	NC	dirección	6/10	<i>frec_noEsp</i>
5	ADJ	nominal	9/10	<i>noFrec_esp</i>	14	NC	exclusivista	6/10	<i>noFrec_esp</i>
6	NC	certificación	8/10	<i>noFrec_esp</i>	15	NC	homologación	6/10	<i>noFrec_esp</i>
7	NC	fortalecimiento	8/10	<i>noFrec_esp</i>	16	NC	liberalización	6/10	<i>noFrec_esp</i>
8	ADJ	minorista	8/10	<i>noFrec_esp</i>	17	NC	negociado	6/10	<i>noFrec_esp</i>
9	NC	recompra	8/10	<i>noFrec_esp</i>	18	NC	otorgamiento	6/10	<i>noFrec_esp</i>

Por último, la Tabla 18 recoge los lemas “idóneos” que se han rechazado. Primeramente, llama la atención que la mitad son adjetivos, cuya relevancia posiblemente solo sea evidente en combinación con uno de los sustantivos que suele acompañar. Asimismo, la mayoría de los lemas parece ser terminología que se puede utilizar perfectamente en un contexto económico pero que es demasiado general para incluir en un glosario sobre la junta general de accionistas (p.ej. *anual*; *cualitativo*; *desarrollo*; *euro*; *integración*).

Tabla 18. Ítems seleccionados por las métricas cuantitativas pero rechazados por los expertos.

#	PoS	Lema	.../10	#	PoS	Lema	.../10
1	ADJ	climático	9/10	8	NC	consecución	7/10
2	ADJ	cualitativo	9/10	9	NC	discapacidad	7/10
3	ADJ	precedente	9/10	10	NC	euro	7/10
4	NC	digitalización	8/10	11	ADJ	institucional	7/10
5	ADJ	geopolítico	8/10	12	NC	desarrollo	6/10
6	NC	record	8/10	13	ADJ	global	6/10
7	ADJ	anual	7/10	14	NC	integración	6/10

### 3.2.2. InterRater Reliability

Para complementar el análisis descriptivo del apartado 3.2.1 desde una perspectiva estadística, aplicamos tests de *interrater reliability* (IRR), que permiten calcular el acuerdo entre los *raters* (participantes) entre sí, así como el acuerdo entre el juicio de los *raters* y la selección automática, cuyos valores presentamos en la Tabla 19.

Tabla 19. Resultados de los tests de *interrater reliability*.

Acuerdo entre	IRR
<i>Raters</i> entre sí	0,40 (Krippendorff's alfa)
<i>Raters</i> (individual) – selección automática	0,44 (Cohen's kappa)
<i>Raters</i> (como grupo) – selección automática	0,61 (Cohen's kappa)

Primero, el valor de 0,40 para los *raters* entre sí indica que, a pesar de que existe cierto grado de acuerdo, la tarea de seleccionar los ítems léxicos parece prestarse a divergencias e interpretaciones subjetivas, por lo cual debemos utilizar las respuestas individuales de los profesores con cautela (Krippendorff, 2004). Sin embargo, el IRR relativamente limitado también podría deberse en parte a las respuestas divergentes de algunos participantes. De hecho, al calcular los *outlier scores* ( $O^+$ ; Zijlstra et al., 2007), una métrica que indica en qué medida las respuestas de cada participante difieren de lo que piensa la mayoría, resulta que los participantes 5, 6 y 7 muestran valores bastante elevados (Tabla 20), aunque no lo suficientemente altos para poder ser clasificados como verdaderos *outliers* (según el método de *Tukey's fences*, véase Zijlstra et al. [2007]). A continuación, el test de IRR para estimar el acuerdo entre cada participante individual y la selección automática da un valor medio de 0,44, lo cual representa un grado de acuerdo moderado (Landis y Koch, 1977).

Tabla 20. *Outlier scores* ( $O^+$ ) por participante.

Participante	1	2	3	4	5	6	7	8	9	10
$O^+$	32	38	26	26	55	59	55	47	27	39

De lo anterior se desprende que existe solapamiento entre el juicio de los profesores y la selección automática, pero que conviene interpretar esta observación con precaución, dado el IRR relativamente limitado entre los profesores como *raters* individuales. No obstante, también se pueden juntar las respuestas para llegar a un solo juicio para el grupo de profesores en su conjunto, en lugar de trabajar con los juicios individuales. Este tipo de generalización ya quedó reflejado en la Tabla 16, con el umbral de aceptación fijado en el 50%. Si ahora recalculamos el acuerdo entre los profesores y la selección automática, el resultado del test de Cohen's kappa asciende a 0,61, el valor límite a partir del cual se puede hablar de un acuerdo sustancial (Landis & Koch, 1977). En otras palabras, la selección automática parece corresponderse relativamente bien con el juicio de profesores expertos ELE como grupo.

## 4. CONCLUSIÓN

En esta contribución hemos descrito en detalle cómo se pueden configurar y combinar diferentes métricas relativas a la frecuencia de uso de ítems léxicos en un corpus especializado. El resultado nos proporciona una forma

de seleccionar de forma automática los ítems considerados más típicos para un determinado dominio temático que podrían ser objeto de ejercicios de aprendizaje. Las métricas respectivas reflejan la frecuencia absoluta en el corpus de estudio (descrita en una escala de percentiles que otorga un estatuto especial a los casos que ocurren solo una vez), su frecuencia de uso en el corpus especializado comparada con la frecuencia de uso en un corpus de lenguaje general (distinguiendo entre la magnitud absoluta de la diferencia y la significatividad estadística, que también es un resultado del tamaño de los corpus) y la dispersión, que distingue entre ítems que son típicos para el género en su conjunto e ítems típicos para instancias particulares. La combinación de los tres criterios permite definir el núcleo del léxico clave que un aprendiz necesita estudiar para familiarizarse con un dominio temático nuevo. A este núcleo se pueden ir añadiendo ítems en función de su frecuencia o especificidad.

A fin de comprobar la relevancia didáctica de la metodología hemos comparado de manera exploratoria los resultados de las métricas con los juicios de diez profesores expertos. Hemos podido comprobar, en efecto, que en más del 85% de los casos el resultado del método cuantitativo se ve confirmado por al menos la mitad de los profesores. Esta observación también se evidencia estadísticamente, con un test de *interrater reliability* que demuestra un acuerdo sustancial (Cohen's kappa = 0,61) entre el juicio de los profesores y la selección automática. El análisis descriptivo de los juicios divergentes nos enseña, finalmente, que un resultado positivo (i.e. un valor alto) en el criterio de la especificidad puede invalidar un menor resultado en los criterios de la frecuencia absoluta o de la dispersión (i.e. un valor bajo y un valor alto, respectivamente).

## 5. BIBLIOGRAFÍA

- Biber, D., Connor, U. y Upton, T. A. (2007). *Discourse on the move: using corpus analysis to describe discourse structure*. Ámsterdam: John Benjamins. <https://doi.org/10.1075/scl.28>
- Boulton, A. (2017). "Data-Driven Learning and Language Pedagogy", en S. L. Thorne & S. May (eds.), *Language, Education and Technology, Encyclopedia of Language and Education*. Berlín & Heidelberg: Springer International Publishing, 181-192. [https://doi.org/10.1007/978-3-319-02237-6\\_15](https://doi.org/10.1007/978-3-319-02237-6_15)
- Bowker, L. y Pearson, J. (2002). *Working with specialized language: a practical guide to using corpora*. Londres & Nueva York: Routledge. <https://doi.org/10.4324/9780203469255>
- Buyse, K., Delbecq, N. y Speelman, D. (2004). *Portavoces. Thematische woordenschat Spaans*. Malinas: Wolters Plantyn.
- Davies, M. (2006). *A frequency dictionary of Spanish: Core vocabulary for learners*. Nueva York: Routledge. <https://doi.org/10.4324/9780203415009>
- Gabrielatos, C. y Marchi, A. (2011). "Keyness: Matching metrics to definitions" (Contribución presentada en the Corpus Linguistics in the South), Portsmouth, NH.
- García Salido, M. y Alonso Ramos, M. (2018). "Asignación de niveles de aprendizaje a las colocaciones del Diccionario de Colocaciones del español", *Revista signos*, 51/97, 153-174. <https://doi.org/10.4067/S0718-09342018000200153>
- Goethals, P. (2018). "Customizing vocabulary learning for advanced learners of Spanish", en T. Read, B. Sedano Cuevas y S. Montaner-Villalba (Eds.), *Technological innovation for specialized linguistic domains* (pp. 229-240). Berlín: Éditions Universitaires Européennes.
- Goethals, P., Tezcan, A. y Degraeuwe, J. (2019). "Vocabulary selection for didactic purposes: report on a machine learning approach". *Argentinian Journal of Applied Linguistics*, 7/2, 34-51.
- Goethals, P., Lefever, E. y Macken, L. (2017). "SCAP\_tur: Tagging and lemmatising Spanish tourism discourse, and beyond". *Ibérica*, 33, 279-288.
- Gries, S. T. (2008). "Dispersions and adjusted frequencies in corpora", *International Journal of Corpus Linguistics*, 13, 403-437. <https://doi.org/10.1075/ijcl.13.4.02gri>
- Izquierdo Gil, M. d. C. (2005). *La selección de léxico en la enseñanza del español como lengua extranjera. Su aplicación al nivel elemental en estudiantes francófonos*. Málaga: ASELE Colección Monografías.
- Krippendorff, K. (2004). *Content analysis: An introduction to its methodology*. Sage, California: Thousand Oaks.
- Landis, J.R. y Koch, G.G. (1977). "The measurement of observer agreement for categorical data", *Biometrics*, 33, 159-174. <https://doi.org/10.2307/2529310>
- Laufer, B., Meara, P. y Nation, P. (2005). "Ten best ideas for teaching vocabulary", *The Language Teacher*, 29/7, 36.
- Nation, P. (2016). *Making and Using Word Lists for Language Learning and Testing*. John Benjamins. <https://doi.org/10.1075/z.208>
- Oakes, M. P. y Farrow, M. (2007). "Use of the chi-squared test to examine vocabulary differences in English-language corpora representing seven different countries", *Literary and Linguistic Computing*, 22/1, 85100. <https://doi.org/10.1093/lc/fql044>

- Okamoto, M. (2015). "Is corpus word frequency a good yardstick for selecting words to teach? Threshold levels for vocabulary selection", *System*, 51, 1-10. <https://doi.org/10.1016/j.system.2015.03.004>
- Schmitt, N. (2008). "Review article: Instructed second language vocabulary learning", *Language Teaching Research*, 12/3, 329-363. <https://doi.org/10.1177/1362168808089921>
- Scott, M. (1996). *WordSmith Tools Manual*. Oxford: Oxford University Press.
- Scott, M. (1997). "PC analysis of key words - and key key words", *System*, 25/2, 233-245. [https://doi.org/10.1016/S0305-0483\(97\)90012-8](https://doi.org/10.1016/S0305-0483(97)90012-8)
- Sinclair, J. (2005). "Corpus and texts – Basic principles", en M. Wynne (ed.) *Developing linguistic corpora: a guide to good practice*. Oxford & Oakville: Oxbow Books, 116.
- Vincze, O. (2015). "Learning multiword expressions from corpora and dictionaries" (tesis de doctorado), Universidade Da Coruña.
- Zijlstra, W.P., van der Ark, A. y Sijtsma, K. (2007). "Outlier Detection in Test and Questionnaire Data". *Multivariate Behavioral Research*, 42/3, 531-555. <https://doi.org/10.1080/00273170701384340>