CONSTRUCCIÓN DE UNA LISTA DE COLOCACIONES PARA MEDIR LA COMPETENCIA COLOCACIONAL

ANA OROL GONZÁLEZ
Universidad de A Coruña, A Coruña

RESUMEN

El objetivo de nuestro trabajo es crear una lista de colocaciones para la elaboración de test. Para la construcción de la lista hemos establecido unos criterios basándonos en las listas de palabras frecuentes y de otras construcciones fraseológicas como locuciones. Además, hemos utilizado la lista para la creación de un test piloto de colocaciones del que describimos su diseño y analizamos brevemente los resultados obtenidos.

Palabras clave: colocación, lista, test, ELE

ABSTRACT

The aim of this work is to create a list of Spanish collocations with assessment purpose. For the creation of this list we have followed a set of previously established criteria which are based on lists of frequent words and other phraselogical constructions like idioms. Moreover, the list is used to design a pilot test which describes and also analyzes its results.

Keywords: collocation, list, test, ELE

1. INTRODUCCIÓN

El papel que ocupan las colocaciones en el proceso de adquisición de la competencia lingüística en una lengua extranjera ha quedado ya demostrado (Higueras, 2006; Lewis, 1993). Sin embargo, todavía quedan campos en los que las colocaciones no han sido estudiadas en profundidad; por ejemplo, el de la evaluación. Aunque

ha habido algunos estudios en inglés (Gyllstad, 2007; Moreno Jaén, 2009), todavía no se ha desarrollado ninguno en español. Para evaluar o medir dicha competencia, es necesario partir de una lista de colocaciones que proporcione el material para poder construir los test. Este será nuestro objetivo: la creación de una *Lista de colocaciones del español* (LCE) que sirva como herramienta de selección de ítems en el diseño de un test para medir la competencia colocacional.

Según nuestro marco teórico, la Teoría Sentido Texto (TST) (Mel'čuk, Clas, & Polguère, 1995), entendemos la colocación como una combinación de palabras en la que la base elige un colocativo para expresar un sentido concreto (Alonso Ramos, 2010; Hausmann, 1998). La base *sol*, por ejemplo, selecciona el colocativo *ponerse* para indicar que 'el sol desaparece' y el colocativo *salir* para indicar que 'el sol aparece'. En nuestro marco teórico, la frecuencia no es un criterio definitorio frente a otros enfoques en los que sí se define la colocación siguiendo este criterio (Sinclair, 1991); sin embargo, consideramos que es un información relevante, como veremos más adelante.

En el campo de los test de vocabulario se usan las listas de palabras frecuentes como herramienta para seleccionar los ítems de un test (Read, 2011). En los estudios sobre evaluación del léxico se parte de que las palabras más frecuentes son más útiles, y más fáciles de aprender por los hablantes en comparación con las palabras menos frecuentes (Laufer y Goldstein, 2004; Milton, 2007). Las listas de palabras, de larga tradición en el ámbito de la lingüística aplicada (Cobb 2007; Coxhead 2000; Gardner y Davies 2013), pueden tener dos finalidades: una como herramienta pedagógica y otra como instrumento para la elaboración de test (Martinez 2011). El objetivo de este trabajo es crear una lista de colocaciones con esta segunda finalidad. En inglés contamos con diferentes listas de unidades fraseológicas construidas con fines pedagógicos (Shin y Nation, 2007; Simpson-Vlach y Ellis, 2010; Martinez y Schmitt, 2012). Sin embargo, aunque sus autores se refieran a algunas de ellas como colocaciones, desde nuestro marco teórico no las trataríamos todas como tal sino como otro tipo de unidades fraseológicas o como sintagmas libres.

El trabajo que sigue se articula de la siguiente manera: empezamos por describir la metodología de la construcción de la LCE; a continuación describimos un test diseñado a partir de la LCE; y, por último, apuntamos las conclusiones y esbozamos las tareas futuras de este trabajo.

2. METODOLOGÍA DE LA CONSTRUCCIÓN DE LA LCE

La lingüística de corpus permitió crear listas de palabras frecuentes con cierta facilidad. Sin embargo, la creación de listas de colocaciones conlleva una dificultad añadida: la identificación de colocaciones en corpus. Por lo tanto, para la LCE seguimos un método complejo que desarrollamos en dos fases y según unos criterios previamente definidos. En la primera fase hacemos una selección de las bases que vamos a utilizar; en la segunda, escogemos los colocativos que forman colocaciones con nuestras bases previamente seleccionadas. Posteriormente, asignamos una frecuencia a cada una de las colocaciones. Además, hemos restringido nuestra lista a las colocaciones que tienen como base un sustantivo, incluyendo así cuatro tipos de colocaciones según su esquema sintáctico: nombre más adjetivo, verbo más nombre, nombre más verbo (en el que el sustantivo funciona como sujeto) y nombre de nombre (el segundo nombre es la base de la colocación).

El objetivo de la primera fase de la construcción de la LCE es la extracción de cincuenta sustantivos que nos sirvan como bases para formar las colocaciones de nuestra lista. Hemos realizado la selección a partir de *A frequency dictionary of Spanish. Core vocabulary for learners* (Davies, 2006), diccionario que recoge las cinco mil palabras más frecuentes del español. Dividimos estas cinco mil palabras en cinco grupos de mil que denominamos "bandas", de cada una de las cuales hemos escogido diez sustantivos de forma aleatoria. Una vez seleccionadas las cincuenta bases pasamos a la segunda fase la selección de cuatro colocativos (uno por cada tipo de colocación) para cada base.

En lugar de seleccionar los colocativos a partir de medidas estadísticas en un corpus (Martinez y Schmitt 2012; Shin and Nation 2008), nos hemos decidido por su selección a partir de la entradas de

un diccionario de igual modo que hacen Gyllstad (2007) y Nizonkiza (2012) en sus test de competencia colocacional. Nosotros hemos utilizado un diccionario combinatorio, el *Diccionario combinatorio práctico del español actual* (Bosque, 2006). Hemos sustituido algunas de las bases que formaban la primera versión de la lista de bases por dos razones: porque no aparecían en el *Práctico* o porque no se formaba un mínimo de tres colocaciones por base.

Para la selección de los colocativos en la entrada del diccionario combinatorio hemos seguido unos criterios que pasamos a enumerar:

a) Si el colocativo pertenece a otro sentido de los que establece el *Práctico*, lo descartamos. Por ejemplo, en la entrada de *leche* encontramos palabras combinables con el sentido 'bebida' (que vamos a incluir) o 'crema', solo seleccionaríamos los que combinan con el primero.
b) Con la intención de evitar colocaciones o muy poco frecuentes, todos los colocativos deben estar entre las cinco mil palabras más frecuentes del diccionario de frecuencias (Davies, 2006).
c) En el caso de que varias palabas cumplan todos los criterios anteriores tenemos en cuenta la transparencia y carácter fraseológico de la combinación. En el caso de *negocio redondo* y *negocio bueno*, aunque las dos son colocaciones elegimos la primera para formar parte de nuestra lista.

Una vez seleccionadas las colocaciones de la lista, buscamos su frecuencia en la lengua en un corpus de referencia. Al igual que la frecuencia se considera un indicador de la utilidad de las palabras de la lengua (Milton, 2007), lo mismo debe ocurrir con las colocaciones: su frecuencia puede ser un indicador de su utilidad o pertinencia en el aprendizaje. La LCE no es una lista de las colocaciones más frecuentes, sino una lista de colocaciones a partir de bases y colocativos frecuentes que incluye la información de frecuencia.

Para conocer la frecuencia de las colocaciones de nuestra lista hemos seguido parcialmente la metodología utilizada para asignar frecuencia a las colocaciones del *Diccionario de colocaciones del español* (Vincze y Alonso Ramos 2013). Debemos tener en cuenta que la frecuencia elevada de una palabra no es sinónimo de que la frecuencia de la colocación sea alta y viceversa. Por ejemplo, una colocación como *pedir disculpas* es una de las colocaciones más

frecuentes de nuestra lista y su base lematizada, *disculpa*, no es una de nuestras bases más frecuentes. Estas autoras establecen dos pasos para la asignación de frecuencia de la colocación: en primer lugar, se busca la frecuencia de la base en un corpus y en segundo lugar, se asigna la frecuencia a la colocación a partir de unas gramáticas diseñadas para las búsquedas en el corpus *esTenTen* (Kilgarriff, Rychly, Smirz, & Tugwell, 2004). La frecuencia de la colocación es una frecuencia ponderada en la que se tiene en cuenta tanto la frecuencia de la base como de la colocación como unidad.

Una vez terminado el proceso de elaboración de la lista de colocaciones vemos que de las 200 previstas, hemos obtenido 164 colocaciones.

3. DISEÑO DE UN TEST A PARTIR DE LA LCE

Con la primera versión de la LCE hemos decidido crear un test piloto. Seleccionamos diez colocaciones que siguen el esquema sintáctico nombre más adjetivo, dos de cada una de las bandas de frecuencia de las que hemos extraído las bases de las colocaciones a partir del *Diccionario de frecuencias del español* (Davies, 2006). Las colocaciones seleccionadas han sido las siguientes: *barba cerrada*, *lógica aplastante*, *zapato plano*, *leche entera*, *negocio redondo*, *posibilidad lejana*, *colonia fresca*, *queso fresco*, *tacto áspero* e *infección grave*. Decidimos medir tanto la comprensión como la producción en dos formatos diferentes que administramos con una diferencia de quince días a cada participante. Para la producción elegimos un formato abierto en el que se debía escribir el colocativo de una colocación presente en contexto con una glosa entre paréntesis que explica el sentido que le da el colocativo a la base; es controlado porque incluimos la primera letra del colocativo. Para los test de comprensión se usó un test cerrado con la colocación descontextualizada en el que se debía elegir entre varias glosas la que mejor definiera el sentido de la colocación.

(1) Ítem del test de producción
Se puso ***zapato*** (sin tacón) ***p*** ...

(2) Ítem del test de comprensión

Zapato plano

1. Con tacón
2. Sin tacón
3. Pequeño
4. Liso, sin adornos

El test se pasó a un grupo reducido de diez personas: cuatro nativos y seis no nativos, de diferentes lenguas maternas (japonés, alemán o italiano). En esta ocasión nuestro principal objetivo no era evaluar la competencia colocacional de quienes realizaron el test sino el test en sí mismo, tanto los formatos como los ítems seleccionados. Comprobamos la dificultad del ítem dividiendo el número de respuestas correctas entre el número de personas que realizaron el test. Tal y como cabía esperar, la dificultad del ítem es mayor en el test de producción que en el de comprensión. De todas formas, el aspecto más destacable es el desajuste de frecuencias y aciertos. Pese a la hipótesis inicial de que las colocaciones más frecuentes son más sencillas y por lo tanto serían las que más aciertos acumularían hemos comprobado cómo esta hipótesis no se cumple en nuestro test piloto. Aunque la colocación menos frecuente, *barba cerrada*, ocasionó bastantes problemas, la segunda colocación más frecuente de las diez seleccionadas, *negocio redondo*, es también la colocación que más problemas ocasiona tanto en producción como en comprensión. La relación entre las colocaciones más y menos problemáticas parece estar también relacionado con las que designan referencialmente un objeto.

4. CONCLUSIONES Y LÍNEAS FUTURAS

La metodología escogida para la LCE nos permite contar con una lista de colocaciones variada, tanto por la pluralidad de esquemas sintácticos (verbo + nombre, nombre + verbo, nombre + adjetivo y nombre de nombre) como por la amplitud de la distribución de las frecuencias de las colocaciones de la lista. Estas características convierten a la LCE en la herramienta adecuada para la selección de ítems de un test que mida la competencia colocacional.

Sin embargo, la baja frecuencia de alguna de las colocaciones de la LCE podría ser un criterio para su exclusión, podríamos establecer un número mínimo de ocurrencias en el corpus de referencia para aceptar una colocación. Además, la LCE puede no ser lo suficientemente amplia si nuestro objetivo es también pedagógico y no solo evaluativo; por lo que entre las líneas futuras está la creación de un banco de colocaciones en el que las colocaciones se organicen temáticamente y se les asigne un nivel buscando una utilidad inmediata en las aulas.

REFERENCIAS BIBLIOGRÁFICAS

Alonso Ramos, M. 2010. "No importa si la llamas o no colocación, descríbela". En Carmen Mellando *et al*., *La fraseografía del S. XXI: Nuevas propuestas para el español y el alemán,* págs. 55–80. Berlin: Frank & Timme

Bosque, I. 2006. *Diccionario combinatorio práctico del español contemporáneo: las palabras en su contexto*. Madrid: SM

Cobb, T. 2007. Why and how to use frequency lists to learn words. *LEXTUTOR*. En http://www.lextutor.ca/research/

Coxhead, A. 2000. "A New Academic Word List". *TESOL Quarterly*, 34(2), 213–238.

Daller, H., Milton, J. y Treffers-Daller, J. 2007. *Modelling and Assessing Vocabulary Knowledge*. Cambridge: Cambridge University Press.

Davies, M. 2006. *A Frequency Dictionary of Spanish*. New York, London: Routledge.

Gardner, D. y Davies, M. 2013. "A New Academic Vocabulary List". *Applied Linguistics*, págs. 1–24.

Gyllstad, H. 2007. *Testing English Collocations. Developing Receptive Test for Use with Advanced Swedish Learners*. Lund University.

Hausmann, F. J. 1998. "O diccionario de colocacións. Criterios de organización". *Actas do I Coloquio Galego de Fraseoloxía*, págs. 63–83. Santiago de Compostela: Centro Ramón Piñeiro, Xunta de Galicia.

Herrera Soler, H., Martínez Arias, R. y Amengual Pizarro, M. 2011. *Estadística aplicada a la investigacion lingüística*. Madrid: Editorial EOS

Higueras, M. 2006. *Las colocaciones y su enseñanza en clase de ELE*. Madrid: Arco Libros.

Kilgarriff, A., Rychly, P., Smirz, P. y Tugwell, D. 2004. "The Sketch Engine". *Proceedings Euralex 2004*, págs.105–116.

Laufer, B., & Goldstein, Z. 2004. "Testing Vocabulary Knowledge: Size , Strength , and Computer Adaptiveness". *Language Learnin*, 54: 399–436.

Lewis, M. 1993. *Teaching collocation. Further Developments in the Lexical Approach*. London: Teacher Training.

Martinez, R., 2011. *The development of a corpus-informed list of formulaic sequences for language pedagogy*. Nottingham University. Tesis doctoral

Martinez, R. y Schmitt, N. 2012. "A Phrasal Expressions List". *Applied Linguistics*, 33: 299–320.

Mel'čuk, I., Clas, A. y Polguère, A. 1995. *Introduction à la lexicologie explicative et combinatoire*. Louvain-la-Neuve: Duculot.

Milton, J. 2007. "Lexical profiles, learning styles and the construct validity of lexical size test". En Daller, H.; J. Milton y J. Treffers-Daller *Modelling and Assessing Vocabulary Knowledge*, *págs.* 47–59. Cambridge: Cambridge University Press.

Moreno Jaén, M. 2009. *Recopilación, desarrollo pedagógico y evaluación de un banco de colocaciones frecuentes de la lengua inglesa a traves de la lingüística de corpus y computacional*. Universidad de Granada. Tesis doctoral.

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.

Read, J. 2011. *Assessing Vocabulary*. Cambridge: Cambridge University Press.

Shin, D. y Nation, P. 2008. "Beyond single words: the most frequent collocations in spoken English". *ELT Journal*, 62: 339–348.

Sinclair, J. M. 1991. *Corpus, Concordance, Collocation*. Oxford: Oxford University Press.

Vincze, O. y Alonso Ramos, M. 2013. "Incorporating frequency information in a collocation dictionary: Establishing a methodology". *Procedia - Social and Behavioral Sciences*, 96: 241-248.