

Estudos Papers

As novas tecnoloxías no ensino/aprendizaxe das colocacións en inglés¹

Using the new technologies to teach and learn collocations in English

Mario Cal Varela

Universidade de Santiago de Compostela

Mario.cal@usc.es

Francisco Javier Fernández Polo

Universidade de Santiago de Compostela

Xabier.fernandez@usc.es

Resumo: Neste artigo preséntanse unha serie de actividades para o ensino das “colocacións”, ou combinacións frecuentes de palabras, na aula de inglés como lingua estranxeira, nas que se empregan ferramentas da Lingüística de corpus e materiais dispoñibles na Rede. A primeira parte do traballo contén unha definición do concepto de colocación e unha breve panorámica das tendencias actuais no ensino de linguas estranxeiras e do papel que desempeñan as novas tecnoloxías. No artigo deféndese que vivimos un momento propicio para a introdución de prácticas de aprendizaxe baseadas en datos, dada a ampla variedade de dicionarios e corpus lingüísticos, a máis de ferramentas para a súa explotación dispoñibles na Rede, a miúdo de forma gratuíta. A segunda parte do artigo contén unha serie de actividades prácticas para o ensino das colocacións léxicas en inglés. En sintonía coa crecente popularidade das pedagogías centradas no discente no contexto particular do ensino de linguas estranxeiras, as catro actividades presentadas e ilustradas nesta sección poñen a énfase no papel activo do estudante, o verdadeiro protagonista dunha aprendizaxe na que o profesor desempeña un papel de guía e de facilitador. As catro actividades están organizadas de menor a maior complexidade e, xa que logo, están previstas para estudantes con distintos niveis de madurez e distintos graos de independencia ou capacidade para asumir a responsabilidade da súa propia aprendizaxe.

Palabras clave: Aprendizaxe baseada en datos, colocacións, corpus, ensino do inglés.

Abstract: In this paper, we propose a range of practical activities that may be implemented in the ESL classroom to learn English collocations using the affordances and tools of corpus linguistics

¹ Data de recepción: 07.06.2019. Data de aceptación: 05.09.2019.

and the Internet. The first part of the study provides a definition of the key concept of collocation as well as an overview of current tendencies in the teaching of foreign languages and the role of the new technologies. The time is ripe for the introduction of data-driven learning practices in the ESL classroom given the wide variety of free-access dictionaries and corpora available on the web, as well as the tools for their exploitation. In the second, central part of the study, we present a number of practical activities for the teaching of the collocational patterns of English. In line with the current popularity of learner-centered pedagogies in foreign-language teaching, the four activities fully presented and illustrated in this section put a great emphasis on the active role of the students as the true protagonists of their own learning, with the teacher becoming a guide and a facilitator. The four activities are organized in such a way that they show an increasing degree of complexity, therefore, of student maturity and learner independence.

Key words: Collocations, corpus, data-driven learning, ESL learning.

1. Introducción

As combinacións pouco naturais de palabras son un fenómeno de interlingua, un dos síntomas máis evidentes da falta de dominio dunha lingua estranxeira, que é, a máis, susceptible de xerar problemas na comunicación. Non obstante, a fraseoloxía, e a combinatoria léxica en xeral, foron durante moito tempo áreas relativamente esquecidas no ensino das linguas estranxeiras, que tradicionalmente deu prioridade á gramática e ó vocabulario entendido como elementos léxicos individuais, sen dúbida porque as teorías lingüísticas prestaban especial atención a estes niveis de descrición da lingua. Porén, a partir dos anos oitenta e, sobre todo, a partir dos traballos de John Sinclair (1991) no mundo anglosaxón, aumentou o interese polo estudo da combinatoria léxica. Estes traballos foron quen de amosar que as combinacións fixas de palabras son un fenómeno ubicuo e constitúen un principio organizativo importante nas linguas que desempeña, ademais, un papel central na adquisición dunha lingua: en esencia, ao aprender unha lingua memorizamos combinacións de palabras, fixas e recorrentes.

Estes avances no terreo da Lingüística téñense traducido nun maior interese por estas cuestións tamén no ensino das linguas estranxeiras (Nattinger & Carrico 1992; Meunier & Granger 2008). Programas e manuais dedican cada vez máis atención á aprendizaxe da fraseoloxía e, nos manuais de inglés como lingua estranxeira, en particular, son cada vez máis habituais as seccións específicas para o estudo das denominadas “colocacións”. Na lingüística anglosaxoa adoita empregarse o termo *collocation* (Sinclair 1991) para referirse a aquelas combinacións recorrentes de dous elementos léxicos, especialmente cando pertencen a algún tipo de patrón sintáctico, por exemplo, un substantivo e o adxectivo que o modifica nunha frase nominal (*the plain truth*) ou un verbo e o seu suxeito ou complemento (*tell the truth*). Foi, sobre todo, o desenvolvemento da Lingüística de corpus e das ferramentas de análise lingüística asociadas o que permitiu realizar avances significativos na descrición destas combinacións. A información así obtida tense aplicado, por exemplo, na mellora dos dicionarios destinados a aprendices de linguas (pero non só), cuxas entradas conteñen con frecuencia información detallada tanto sobre a variedade de significados das palabras coma tamén sobre as súas posibilidades combinatorias.

Eses mesmos corpus e ferramentas posúen un gran potencial para a o ensino e a aprendizaxe das colocacións na aula de linguas estranxeiras. De feito, a aprendizaxe de linguas baseada en datos (os extraídos dos corpus, por exemplo) ten experimentado un importante desenvolvemento nas dúas últimas décadas (Guilquin & Granger 2010). Trátase dunha aprendizaxe máis activa na que, empregando corpus de textos e ferramentas de análise computerizada, os aprendices enfróntanse aos datos lingüísticos reais en busca de respostas ás súas preguntas. É unha forma de aprendizaxe centrada no aprendiz, máis activa e, xa que logo, posiblemente tamén máis efectiva (Koosha & Jafarpour 2006), aínda que non exenta de problemas e dificultades para a súa aplicación nas aulas.

No presente artigo, ofrecemos unha breve panorámica destas cuestións e, sobre todo, presentamos unha serie de propostas de actividades de ensino/aprendizaxe das colocacións en inglés baseada en datos e co apoio das novas tecnoloxías. Na primeira parte do traballo, definimos o concepto de colocación, sobre todo, dende o punto de vista da súa relevancia para a aprendizaxe de linguas, así como os principios, as vantaxes e os problemas da aprendizaxe baseada en datos, particularmente para a aprendizaxe das unidades fraseolóxicas. Na segunda parte, descríbense en detalle catro propostas de actividades para o ensino das colocacións en inglés que teñen no punto de mira catro perfís distintos de aprendizaxe, en función, principalmente, do nivel de competencia lingüística e do grao de independencia ou autonomía dos discentes.

2. As colocacións dentro da fraseoloxía

As colocacións son combinacións de palabras frecuentes, recorrentes e semifixas, por exemplo, *manter unha reunión*, *promulgar unha lei*, *idea brillante*, *amarga derrota* etc., asociadas na mente dos falantes. Na lingüística anglosaxoa, o concepto remóntase a J. R. Firth, nos anos 50, para quen as colocacións dunha palabra, a maneira en que adoita aparecer de forma repetida a carón doutras, son unha parte integral do seu significado (Carter & McCarthy 1988: 32). Esta idea tería importantes repercusións no desenvolvemento dos dicionarios de inglés e doutras linguas a partires dos anos oitenta (Corpas Pastor 2001), tales como os derivados do proxecto COBUILD. Un dos seus principais responsables, John Sinclair (1991) sostiña que as combinacións repetidas de palabras son un fenómeno ubicuo e, xa que logo, un importante principio organizativo das linguas, o que el chama *the idiom principle*.

O termo “colocación” posúe distintos valores en Lingüística (Boers & Webb 2018). Na Lingüística de corpus, por exemplo, refírese á relación que vencella palabras que teñen algún grao de probabilidade de coaparecer, superior ao que marcaría o azar. Neste sentido, é preciso distinguir as colocacións doutras combinacións máis ou menos frecuentes de palabras reveladas polos estudos de corpus como son os “feixes léxicos” (ingl. *lexical bundles*, Biber & al. 1999), pero nos que os compoñentes individuais non están ligados por ningún tipo de relación sintáctica nin semántica.

En Fraseoloxía, segundo Copas Pastor (1996: 66), trataríase de “unidades fraseolóxicas formadas por dúas unidades léxicas en relación sintáctica, que no constitúen por si

mismas actos de habla ni enunciados”. Son combinacións estables e usuais, fixadas na norma, entre cuxos compoñentes existe algún tipo de “restrición combinatoria” de natureza semántica². O carácter recorrente e restritivo da relación é o que define as colocacións e as distingue doutras combinacións posibles pero “libres” de palabras. Así, nos seguintes exemplos, o primeiro elemento do par sería unha colocación, pero non o segundo: *revogar/abolir unha lei* fronte a *revisar/ler unha lei*; *un rotundo fracaso* fronte a *un predicible fracaso*; ou *levantar a veda* fronte a *falar da veda*.

As colocacións están formadas por dous colocativos, unha base e un colocado, onde a primeira determina a natureza do(s) segundo(s), de aí o carácter idiomático da combinación. Por exemplo, as colocacións da palabra *lei*, que tomamos como base, terían como posibles colocativos verbos coma *promulgar*, *aprobar*, *elaborar*, *sacar*, *votar*, *derrogar* etc. A diferenza doutras expresións semellantes, coma as locucións idiomáticas (*idioms* en inglés), por exemplo *ser (unha persoa) de lei*, malia existir unha relación forte entre os compoñentes dunha colocación, cada palabra individual segue a manter o seu valor intrínseco no seo da combinación, amosando unha certa “transparencia semántica” (Boers & Webb 2018: 77). Con todo, nalgúns colocacións obsérvase unha especialización semántica (Corpas Pastor 2001: 93) dalgún dos seus colocativos, dando como resultado unha metaforización xeral da expresión, por exemplo en *levantar unha prohibición*, o que fai ás veces difícil distinguir entre colocacións e locucións idiomáticas, como ocorre, por exemplo, co carácter polisémico da expresión *meter un gol*, que menciona a propia Corpas Pastor.

Na bibliografía especializada, é habitual distinguir entre dous tipos de colocacións, gramaticais e léxicas (Benson & al. 1986). As primeiras están formadas por unha palabra léxica, fundamentalmente un verbo, un substantivo ou un adxectivo e a preposición ou tipo de cláusula que lle segue. Así, un verbo coma *falar* “rexer” e, xa que logo, ten como posibles colocativos as preposicións *de* (algo) ou *con* (alguén). Pola súa banda, as colocacións léxicas están formadas por dúas palabras léxicas e inclúiría os seguintes tipos, principalmente: verbo + substantivo (*acadar un obxectivo*), substantivo + adxectivo (*inmenso pracer*), substantivo + substantivo (*banco de peixe*), adverbio + adxectivo (*profundamente durmido*) ou verbo + adverbio (*falar seriamente*).

3. As colocacións no ensino do inglés como lingua estranxeira

Malia que o tema das colocacións non suscitou gran interese na aprendizaxe das linguas estranxeiras ata finais do século pasado, nas dúas últimas décadas ten habido unha morea de estudos sobre a cuestión (Boers & Webb 2018). Con todo, este interese non se ten traducido nunha toma de conciencia xeralizada nas aulas. Como apunta Römer (2011: 211), “information and exercises on typical collocations are as yet largely missing from language-teaching course books (or they are inadequate)”.

² Corpas Pastor distingue entre dous enfoques básicos do concepto de colocación: o da escola británica e a Lingüística de corpus, no que o elemento determinante da colocación é a frecuencia de coaparición das palabras, e o da corrente semántica, onde a esencia da colocación é a existencia dalgún tipo de restrición semántica entre a base e os seus colocados.

As colocacións son centrais na aprendizaxe dunha lingua estranxeira e o seu uso correcto unha boa medida do grao de dominio da lingua adquirido polo aprendiz. O principal problema na súa aprendizaxe ten que ver coa arbitrariedade destas combinacións e, xa que logo, coa súa impredecibilidade: por exemplo, en galego dicimos *prestar atención* pero non, malia que podería ser perfectamente posible, **dar atención a algo*. A máis, varían dunha lingua a outra (e.g., gal. *prestar atención*, ingl. *pay attention*, fr. *faire attention* ou gal. *aprobar un exame*, ingl. *pass an exam*, fr. *réussir un examen*), o que motiva calcos innecesarios e, como resultado, combinacións léxicas inaceptables para un falante nativo. Xa que logo, fronte a colocacións non problemáticas para o aprendiz, por coincidiren nas dúas linguas (gal. *correr un risco*, ingl. *run a risk*), outras adoitan dar lugar a erros habituais (gal. *cometer un erro*, ingl. **commit a mistake*, no canto de *make a mistake*; gal. *celebrar unha reunión*, ingl. **celebrate a meeting*, no canto de *hold a meeting*; ou gal. *prestar atención*, ingl. **lend attention*, por *pay attention*).

Segundo Bahns (1993), o ensino das colocacións xera fundamentalmente tres interrogantes: 1) deben ensinársese ou adquirense ao tempo que se traballa coa lingua?; 2) cales, das moitas colocacións posibles dunha lingua, deben priorizarse?; e 3) que tipo de actividades e métodos de ensino son os máis adecuados para o seu ensino?

Canto á primeira das preguntas, para Lewis (2000), o nivel de competencia no uso das colocacións non se deriva automaticamente do dominio do vocabulario ou da gramática da lingua estranxeira e deberían, xa que logo, constituír un obxecto de aprendizaxe autónomo. Para este autor, moitos estudantes de nivel intermedio estancáanse na aprendizaxe, porque “they have not been trained to notice which words go with which”. O problema radica en que nun nivel intermedio de aprendizaxe, o alumno ten a impresión de coñecer a maioría do vocabulario de calquera texto que lle presentamos, pero non percibe ou subestima a importancia da súa combinatoria. Entende as colocacións dun texto, pero non é quen de usalas activamente. Segundo Lewis, a solución pasa por cambiar o foco da aprendizaxe: no canto de introducir estruturas sintácticas máis complexas e vocabulario máis esotérico, o que o aprendiz de nivel intermedio ou avanzado realmente precisa é tomar conciencia da existencia dos patróns combinatorios típicos das palabras que xa coñece e aprender a usalos activamente. En definitiva, o ensino das colocacións merece, para este autor, un tratamento específico e complementario doutros contidos.

Canto ó segundo dos interrogantes, como apunta Bahns (1993), dado o inmenso número de colocacións dunha lingua (por exemplo, o dicionario Oxford de colocacións en liña contén máis de 150.000 para a lingua inglesa), cómpre decidir en cales se ha centrar a aprendizaxe. A máis da relativa frecuencia de uso das distintas colocacións, dous son os factores relevantes á hora de decidir as prioridades do ensino neste ámbito: o nivel de lingua do aprendiz e a súa lingua materna. En relación co primeiro, segundo apunta Bahns, as colocacións que han tratarse nun curso de nivel básico non poden ser as mesmas cás que cómpre ensinar nun curso avanzado. Canto á segunda cuestión, dado que os problemas dos aprendices a miúdo son o resultado de interferencias da súa propia lingua materna (Nesselhauf 2003), cómpre centrar a atención naquelas colocacións que máis difiren entre a súa lingua e a lingua estranxeira que está a aprender (Bahns 1993), o que significa, por exemplo, que o

catálogo de colocacións que é relevante para un aprendiz galego de inglés non podería ser o mesmo có que se empregue para un estudante alemán ou polonés.

Canto á metodoloxía, Boers e Lindstromberg (2008) (citados en Vasiljevic 2014) falan de tres fases no ensino-aprendizaxe das colocacións: 1) concienciación da súa existencia e importancia, 2) memorización e 3) repaso ou uso activo. A abordaxe adecuada do ensino precisa traballar todas elas na aula a través de actividades específicas, con tarefas como as que describe, por exemplo, a propia Vasiljevic.

De maneira máis xeral, dado que a miúdo as dificultades na aprendizaxe das colocacións están en que lle pasan desapercibidas ao aprendiz, quen non ve nelas un problema, as claves do seu ensino pasan por espertar a súa consciencia sobre a existencia e a importancia destas combinacións, ao tempo que estimulamos a súa curiosidade. Isto é o que en ensino do inglés os especialistas chaman *language awareness* ou *language consciousness raising*, que se aplica para o ensino da gramática dunha lingua (cf. Rutherford 1987, Rutherford & Sharwood Smith 1988) e que se pode extrapolar para o ensino do vocabulario.

Neste mesmo capítulo da metodoloxía de ensino para a aprendizaxe das colocacións nunha lingua estranxeira, segundo unha corrente de opinión moi estendida actualmente na aprendizaxe de linguas, esta debe centrarse fundamentalmente no discente, quen debe desenvolver a súa autonomía para decidir o que e mailo como da experiencia de aprendizaxe. Esta independencia vese favorecida na actualidade pola existencia de abundantes recursos de fácil acceso que pode empregar para explorar os temas que lle interesen en cada momento e segundo o método que máis lle conveña. A dispoñibilidade de recursos gratuítos como dicionarios de todo tipo e unha cantidade ilimitada de datos lingüísticos na Rede, a máis de programas informáticos para a súa análise, ofrece grandes posibilidades para o estudo das colocacións na aula de linguas estranxeiras, como veremos nas seccións que seguen.

4. A aprendizaxe baseada en datos

De entre os profundos cambios experimentados pola descrición e a teoría lingüísticas desde mediados do século pasado, cabe salientar o extraordinario interese espertado pola variabilidade estruturada característica da fala e o desenvolvemento de ferramentas e métodos cada vez máis refinados para o seu estudo. A recollida sistemática de mostras de lingua auténticas en forma de coleccións, ou corpus, representativos de variedades lingüísticas específicas —xeográficas, sociais ou situacionais— forma hoxe parte habitual da investigación lingüística desde moi distintas perspectivas. Como consecuencia lóxica, o modo en que se ensinan e aprenden as linguas ten mudado tamén profundamente en liña coas novas achegas.

O impacto máis inmediato e visible foi seguramente a transformación dos materiais empregados no ensino de linguas estranxeiras. Os froitos da análise de corpus abríronse camiño ata os libros de texto e condicionaron tanto os contidos como o modo de ensinalos. Pola súa banda, as gramáticas de referencia e os dicionarios enriqueceron as súas explicacións de estruturas e léxico con información moito máis detallada sobre frecuencia, rexistro, combinacións máis habituais etc. O proxecto COBUILD, posto en

marcha en 1980 por John Sinclair desde a Universidade de Birmingham, constitúe un fito fundamental neste proceso. Hoxe en día mesmo as versións dispoñibles gratuitamente en liña dos dicionarios de inglés máis coñecidos incorporan as achegas máis recentes da Lingüística de corpus. Así, por exemplo, o *Longman Dictionary of Contemporary English Online*³ inclúe como parte das entradas seccións sobre colocacións e abundantes “exemplos extraídos do corpus”. O *Macmillan Dictionary*⁴ en liña, pola súa banda, destaca en letra vermella as 7.500 palabras máis comúns en inglés e clasifícaa en tres niveis de frecuencia (1-3 estrelas), de acordo cos datos extraídos dun corpus.

A introdución directa do traballo con corpus na aula de lingua tardou algo máis en chegar. Para que ese tipo de propostas pedagóxicas resultasen practicables, debían darse certas condicións. A primeira e máis importante xa chegara en realidade nos anos 70, da man da aprendizaxe comunicativa, que redefiniu os papeis de profesor e alumno para cederlle a este último espazo, autonomía e responsabilidade na súa aprendizaxe. A segunda, tamén esencial, implicaba unha dispoñibilidade razoable de ferramentas apropiadas, tanto para a exploración de corpus como para a propia compilación das necesarias coleccións de textos. A eclosión de Internet a partir dos anos 90 e os rápidos avances desde entón no que fai a almacenamento, acceso e procesamento de textos permitiron salvar estes obstáculos. O traballo de entusiastas pioneiros (Johns 1994; Tribble & Jones 1997) indicou o camiño e as enormes posibilidades desta metodoloxía e, desde hai xa trinta anos, a denominada aprendizaxe baseada en datos —*data-driven learning* é o termo máis frecuentemente empregado en inglés— ocupa un lugar fixo entre os enfoques didácticos máis comentados. Desde 1994 lévanse celebrado xa trece edicións do congreso bienal TaLC, *Teaching and Language Corpora*, no que especialistas de todo o mundo comparten as súas experiencias co uso de corpus lingüísticos con fins pedagóxicas. Non faltan tampouco obras dirixidas ao profesorado de linguas en xeral (Aijmer 2009; Bennett 2010; O’Keeffe, McCarthy, & Carter 2007; Reppen 2010; Wichmann 1997) que abordan as dificultades técnicas e ofrecen orientacións prácticas, xogando así un papel divulgativo esencial neste ámbito.

Polo que respecta ás particularidades definitorias deste enfoque metodolóxico, cómpre salientar a súa adaptabilidade a distintos contextos e finalidades didácticas. Boulton (2011) enténdeo como un concepto de carácter prototípico, sen un conxunto de trazos necesarios e suficientes. Podemos condensar a súa proposta facendo referencia a catro dimensións ou aspectos centrais. O primeiro é que a aprendizaxe baseada en datos implica o traballo directo con ferramentas de corpus na aula e cos resultados máis habituais que proporcionan, en particular as liñas de concordancia que exemplificaremos máis adiante. Con estudantes principiantes ou menos afeitos ao traballo con estes instrumentos, pódese, non obstante, partir de versións simplificadas dos resultados da análise presentadas en forma de documento de traballo. Este método implica, en segundo lugar, traballar con datos extraídos dun corpus, aínda que a noción de corpus debe entenderse aquí nun sentido moi amplo, que abarque tanto o acceso ás grandes compilacións dispoñibles hoxe como a coleccións máis modestas reunidas *ad hoc* ou mesmo contemple actividades para as

³ <https://www.ldoceonline.com/>.

⁴ <https://www.macmillandictionary.com/>.

que se emprega un único texto extenso. En terceiro lugar, as aplicacións máis típicas implican unha aproximación fundamentalmente indutiva aos datos. A exploración directa de datos auténticos na procura de regularidades por parte do estudante é, sen dúbida, un trazo central. Define un contexto máis natural de adquisición da lingua na que o discente asume o papel máis estimulante de investigador e abre mesmo a posibilidade de aprendizaxes imprevistas (Bernardini 2000 e 2002). Con todo, alternativas con maior intervención do docente nas que se propón a comprobación dalgunha hipótese relacionada con algunha regra non están tampouco excluídas, sempre e cando se manteña unha esencia construtivista e o estudante conserve a iniciativa (Flowerdew 2015). Por último, a aprendizaxe baseada en datos tende a asociarse co estudo da segunda lingua en niveis avanzados, dado que son os estudantes cunha base consolidada os que en principio poden tirar máis partido destas propostas. Hai, porén, experiencias positivas con estudantes de niveis máis baixos (Boulton 2010: 139) e coa exploración da lingua materna en ámbitos como, por exemplo, a tradución (Beeby et al. 2009).

Sen dúbida, a incorporación dos corpus lingüísticos á práctica pedagóxica constitúe un considerable desafío para profesores e estudantes. Esixe algúns coñecementos previos e uns medios técnicos mínimos, ademais dun certo investimento de tempo tanto na preparación como na posta en práctica das actividades. Son factores nada desprezables que seguramente explican unha difusión menor da esperable dada a xa non curta traxectoria desta metodoloxía no ensino de linguas e os datos en xeral alentadores sobre o seu impacto positivo na aprendizaxe (Römer 2011; Tribble 2015). Á hora de contemplar a posibilidade de integrala na propia práctica docente, cómpre non perder de vista as vantaxes que a experiencia acumulada pon de manifesto.

A máis obvia das achegas positivas desta metodoloxía á aprendizaxe é que expón ao discente a mostras de lingua auténtica, coa súa variabilidade intacta, favorecendo así a familiarización co uso lingüístico en situacións comunicativas reais. Para o docente non nativo os corpus constitúen unha fonte de enorme credibilidade da que obter datos obxectivos sobre os que apoiar as súas recomendacións e explicacións sobre as formas apropiadas para un contexto concreto. Son, xa que logo, un recurso “empoderador” (Mair 2002: 121). O traballo con múltiples exemplos dunha estrutura ou un fenómeno lingüístico concreto desenvolve a capacidade de recoñecer e prestar atención consciente a elementos lingüísticos salientables, o que pode xogar un papel clave na súa adquisición (Flowerdew 2015; Schmidt 2001). Pola outra banda, esta exposición intensiva a un elemento léxico determinado en múltiples contextos contribúe non só a incrementar a “amplitude” do vocabulario dispoñible, senón tamén a “profundidade” dese coñecemento (Cobb 1999). O carácter exploratorio e indutivo do emprego de corpus estimula a curiosidade e mellora a capacidade de formular preguntas proveitosas e procurar información adecuada para darlles resposta, ademais de facilitar a adquisición de aspectos non previstos inicialmente, destrezas todas esenciais para desenvolver verdadeira autonomía na aprendizaxe. Xa para rematar este rápido reconto de motivos para dar o paso, cabe engadir finalmente que esta metodoloxía fomenta tamén a actitude crítica cara ás ferramentas e recursos de aprendizaxe e mellora a comprensión da información que estes proporcionan, ao

“desvelar” e traer a primeiro plano un aspecto chave do proceso moderno de elaboración de dicionarios e gramáticas.

A introdución das ferramentas de corpus na aula enriquece o catálogo de métodos e técnicas a dispor do docente para abordar certos aspectos da aprendizaxe dunha lingua estranxeira. A sutil fronteira entre o léxico e a gramática na que se sitúa a fraseoloxía é unha área idónea para demostrar a súa eficacia. Cómpre adaptar o uso concreto que se fai da técnica ás condicións específicas de cada situación de ensino, tendo en consideración os obxectivos particulares, o nivel e as preferencias de aprendizaxe dos alumnos e a dispoñibilidade de medios. Así mesmo, é necesario axustar as expectativas sobre os seus efectos positivos ao medio e longo prazo, dado que a súa eficacia depende en boa medida de cambios na cultura de aprendizaxe que se producen amodo.

Nas seccións seguintes facemos algunhas propostas, a modo de ilustración, en grao crecente de complexidade para distintos perfís de alumno e para distintos obxectivos do curso. As experiencias das que damos conta refírense exclusivamente á aprendizaxe de inglés como lingua estranxeira, pero os modos de proceder serían igualmente aplicables para outras linguas. En varios dos casos, os exemplos xorden de aplicacións reais resultantes da experimentación persoal dos autores con estudantes de perfil moi diverso, en concreto, estudantes de inglés de nivel intermedio, estudantes de tradución e alumnos de máster dun curso sobre inglés para fins académicos. Comezaremos por usos indirectos da análise de corpus nos que a aproximación é máis ben híbrida, indutiva-dedutiva e nos que a mediación do profesor é, xa que logo, máis explícita.

5. Os dicionarios en liña como recurso para aprender colocacións na aula de inglés

Como xa vimos anteriormente, para a lingüística firthiana e neo-firthiana, un compoñente central do significado dunha palabra é o que se deriva das súas colocacións habituais. Na actualidade, a meirande parte dos dicionarios xerais de inglés parten deste principio na súa confección, de aí que conteñan abundante información sobre as colocacións habituais das palabras, tanto na sección central das entradas coma en seccións especiais aparte. Xa de xeito máis específico, tamén contamos con dicionarios de colocacións, a miúdo elaborados polo mesmo equipo responsable dos dicionarios xerais, que conteñen información máis completa e mellor organizada sobre a cuestión. Outra ferramenta en liña que ofrece información útil sobre os patróns combinatorios das palabras é SKELL⁵, unha aplicación derivada de Sketch Engine especialmente concibida para o ensino e a aprendizaxe de linguas, unha de cuxas funcións, Word Sketch, permite obter información básica sobre as combinacións máis frecuentes dunha palabra. Coma SKELL, a meirande parte dos dicionarios modernos de inglés, incluídos algúns dicionarios de colocacións, dispoñen de versións en liña. Pola súa gran accesibilidade, a través da Rede e sen rexistro previo, e por seren de balde, estas ferramentas e outras semellantes resultan moi útiles para o ensino e a aprendizaxe das colocacións na aula de inglés, como se amosa a continuación.

⁵ <https://skell.sketchengine.co.uk/run.cgi/skell>.

Como apunta Lewis (2000), unha situación onde se percibe claramente a importancia das colocacións é cando temos que explicarlle aos alumnos as diferenzas entre pares de palabras sinónimas ou cuasisinónimas, como serían *illness* e *disease* ou *powerful* e *strong*. A diferenza entre elas non está na “definición” das mesmas senón nas palabras coas que se adoitan combinar, nos seus contextos de uso: en inglés podemos dicir *a strong tea* ou *a strong argument*, pero non **a strong car*, do mesmo xeito que podemos dicir *a powerful car* ou *a powerful argument* pero non **a powerful tea*. Xa que logo, unha posible maneira que ten o aprendiz de “aproximarse” ás diferenzas entre estas dúas palabras e mediante a comparación das súas colocacións habituais.

Para esta tarefa, centrarémonos no par *illness* e *disease*. Pedirémolles ao aprendiz que consulte a información das entradas respectivas das dúas palabras (Figura 1) no *Online Oxford Collocation Dictionary* en liña⁶.

disease	<i>noun</i>
ADI common obscure , rare dangerous , serious mild chronic acute degenerative deadly , fatal , incurable , killer , terminal <i>fears of a new killer disease</i> curable preventable communicable , contagious , infectious non-communicable congenital , hereditary , inherited childhood <i>childhood diseases such as mumps and chicken pox</i> tropical insect-borne , water-borne occupational bowel , (coronary) heart , liver , etc. mental sexually transmitted , social , venereal <i>patients suffering from venereal disease</i> circulatory , respiratory Alzheimer's , Crohn's , etc.	
illness	<i>noun</i>
ADI fatal , incurable , terminal debilitating , life-threatening , major , serious , severe minor lingering , long , long-standing , long-term , prolonged brief , short final , last acute , chronic infectious painful depressive , mental , psychiatric , psychotic psychosomatic physical respiratory childhood	

Figura 1. Entradas dos termos “disease” e “illness” no *Online Oxford Collocation Dictionary*

O primeiro que cómpre facer é amosarlle como está organizada a información na entrada, como as palabras tenden a agruparse por similitudes semánticas, o que facilita a aprendizaxe, por exemplo, de sinónimos e antónimos; así, *fatal*, *incurable* e *terminal* aparecen ao carón unha da outra para indicar que son sinónimos, pero tamén *common* e *rare* para amosar que son contrarios. Agora que “ven” como está organizada a información, deberían ser quen de clasificar os adxectivos das dúas entradas en función do seu significado xeral, por exemplo, adxectivos que indican perigo de morte, gravidade, duración, aparato (*respiratory*, *circulatory*) ou órgano afectado (*heart*, *liver*, *blood*) etc. O seguinte paso é identificar coincidencias e diferenzas entre as dúas palabras, ver que grupos de adxectivos ou modificadores comparten e cales tenden a ser máis habituais con cada unha delas. Para facilitar a comparación poderían utilizar un diagrama coma o da Figura 2, formado por dous círculos que se superpoñen parcialmente e, xa que logo, teñen unha zona en común.

⁶ <http://www.freecollocation.com/>

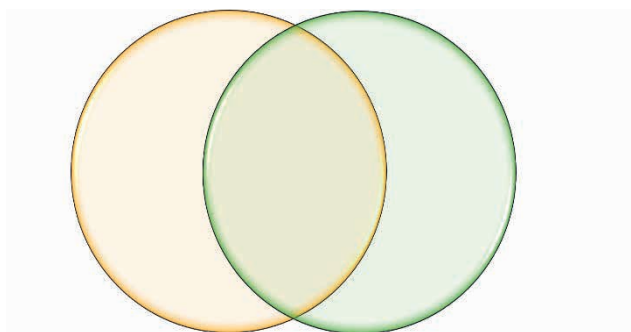


Figura 2. Diagrama baleiro para un exercicio sobre colocacións

A comparación permite ver que os dous substantivos se combinan de forma igual de natural con algúns dos grupos semánticos que formaron: falamos de *serious illness* ou de *serious disease*, de *fatal illness* ou *disease*, *childhood illness* ou *disease* etc. Con todo, tamén se observan algunhas “especializacións”: *disease*, pero non *illness*, combínase con modificadores que indican o tipo concreto de doenza (*Alzheimer’s/Crohn’s disease*, pero non **Alzheimer’s illness*); *disease* tamén parece combinarse de forma máis natural con enfermidades de carácter físico (*bowel, liver, heart*), mentras que *illness* amosa unha tendencia a “especializarse” en enfermidades mentais (*psychiatric illness, depressive illness* etc., malia que falamos de *mental illness* ou *mental disease*). *Disease* tamén se combina máis naturalmente con adxectivos como *water-born, hereditary, tropical* etc., que permiten clasificar as enfermidades en función da súa orixe, o lugar onde se adoita dar etc. Por contra, en inglés fálase de *final* ou *last illness* pero non de **final disease* para referirse a unha enfermidade que dá como resultado a morte do afectado. Algunhas destas diferenzas remiten ó feito de que só *illness*, pero non *disease*, posúe o significado de condición de estar enfermo, a máis de poder designar a doenza propiamente dita. Podemos tamén conciencialos das limitacións do dicionario e animalos a que completen a información; por exemplo, mesmo se no dicionario só se recollen as expresións *childhood illness* ou *childhood disease*, ámbolos dous substantivos poden recibir outros modificadores referidos ás franxas de idade nas que adoita darse a doenza, e podemos invitalos a que exploren en internet o uso de expresións como *old-age* ou *middle-age illness/disease, adult* ou *pre-natal illness/disease* etc. Podemos igualmente aproveitar para identificar e aclarar o valor dalgunhas expresións de significado “escuro”; por exemplo, ver con eles que significa realmente a expresión *acute disease*, que en realidade é unha enfermidade repentina e de curta duración. Podemos, por último, chamar a súa atención sobre falsos amigos que se cadra ignoran, coma que o adxectivo *communicable* en inglés, referido a unha enfermidade, ten o valor de “contaxiosa” e non de “comunicable”, que de feito carecería de sentido.

Obviamente, como complemento desta actividade de descubrimento e “concienciación” das diferenzas nos patróns colocacionais das dúas palabras, cumprirá realizar algún tipo de actividade práctica complementaria na que o discente teña que empregar o aprendido en contextos concretos. Neste sentido, é importante lembrar que o ensino do vocabulario

dunha lingua estranxeira débese realizar sempre dentro dun contexto determinado, onde o aprendiz teña que facer un uso activo do aprendido ou, como apuntaba Vasiljevic (2014) (*vid. supra*) a propósito especificamente das colocacións na aprendizaxe dunha lingua estranxeira, a concienciación e a memorización destas expresións deben completarse co repaso e co uso activo das mesmas.

6. WebCorp como paso intermedio cara ás ferramentas de corpus

O obxectivo dos creadores de Webcorp era poder extraer información lingüística dos textos da web de forma semellante a como se fai cun corpus convencional (Renouf & al. 2007). Xa que logo, permite buscar palabras, frases ou secuencias descontinuas (mediante comodíns) en textos dispoñibles na Rede. Entre outras funcionalidades especialmente útiles está a posibilidade de aplicar filtros nas procuras e así restrinxir a busca a linguas concretas ou por dominio de busca (.es, www.elmundo.es etc.), o que supón, en definitiva, poder explorar distintas variedades de lingua (xeográficas, rexistros, estilos máis ou menos formais etc.).

Orixinalmente foi concibido especialmente para o estudo da lingua que se emprega nos múltiples novos ámbitos da comunicación vía Internet (chats, blogs etc.) e, en particular, para o estudo de novas creacións fraseolóxicas, neoloxismos léxicos e novas modas lingüísticas en xeral, que non se poden investigar adecuadamente cos corpus tradicionais. Con todo, a ferramenta pode facer as veces dun corpus “normal” para explorar calquera aspecto da lingua. Malia dispor de funcionalidades de busca moi limitadas en comparación con moitos dos corpus máis coñecidos, dende o punto de vista pedagóxico, ten a vantaxe de ser de acceso gratuito en liña e de posuír un interface relativamente sinxelo de usar para calquera usuario non experto.

3) <http://www.telegraph.co.uk/science/2017/11/14/air-pollution-bigger-risk-people-with-a-b-blood-types-study/>
Text, Wordlist, text/html, UTF8 (Content-type), 2017-11-14 (Meta tag)

6: Telegraph Science Air pollution is a bigger risk for people with A, B, or AB blood types, study
7: with type A, B, or AB blood, age at greater risk Credit: Rayart Graphics / Alamy Stock Photo
8: consider staying indoors to minimize their risk if they had underlying heart conditions, such as
9: a heart attack but it is the first time that the risk has also been linked to blood type. A study of
10: Heart Institute in Utah, US, found that the risk of a heart attack or chest pain doubled for
11: levels. London's polluted air in contrast the risk rose only rose by 40 per cent for those with
12: for every additional 10 micrograms over 20, the risk to people with type A, B, or AB blood increased
13: B, or AB and they are thought to be at greater risk of heart problems because their blood contains
14: Premium 06 Jan 2018, 3:13pm Families put at risk because NHS is refusing to test for deadly

10) <http://www.independent.co.uk/news/health/intelligence-mental-illness-iq-study-findings-depression-a8005801.html>
Text, Wordlist, text/html, UTF8 (Content-type), 2017-10-17 (Server header)

71: us: News Health Intelligent people more at risk of mental illness, study finds Study surveys
72: found that people with higher IQs are more at risk of developing mental illness. A team of US
73: healthy obese people are still at a higher risk of heart disease or a stroke than those with a
74: Babies born to fathers aged under 25 have higher risk of autism New research has found that babies
75: under the age of 25 or over 51 are at higher risk of developing autism and other social disorders.
76: es/iStockphoto 13/37 Cycling to work could halve risk of cancer and heart disease' Commuters who swap
77: their car or bus pass for a bike could cut their risk of developing heart disease and cancer by almost
78: cyclists. Cycling to work is linked to a lower risk of developing cancer by 45 per cent and
79: Millions of people are putting themselves at risk by cooking their rice incorrectly, scientists
80: to death. A diet of Japanese rice was linked to a lower risk of death, known as *kareoshi*, as they
81: is well known that high blood pressure is a risk factor for dementia, so the results of a new
82: drug, venlafaxine, was shown to increase the risk users engaging in suicidal thoughts and attempts
83: antidepressants in pregnancy could double the risk of autism in children taking antidepressants
84: during pregnancy could almost double the risk of a child being diagnosed with autism in the
85: aracetamol to treat mild fevers. As a result, the risk of developing diabetes or will also kidney, heart
86: brain. PA 35/37 Drinking alcohol could reduce risk of diabetes A new study shows that drinking
87: three to four days a week could reduce the risk of diabetes. Wine was found to be most effective
88: was found to be most effective in reducing the risk due to the chemical compounds that balance blood

11) <http://www.telegraph.co.uk/news/worldnews/islamic-state/11629161/Mapped-The-countries-facing-the-greatest-threat-from>
Text, Wordlist, text/html, UTF8 (Content-type), 2015-06-26 (Meta tag)

89: is the biggest threat, according to global risk management company, Aon Interactive: Terror
90: Russian aggression was the cause of its higher risk rating. (The UK has a risk rating of 2, the same
91: cause of its higher risk rating. (The UK has a risk rating of 2, the same as last year - readers can
92: well as the threat from Islamic terrorism, Aon Risk Solutions, the global risk management business
93: terrorism, Aon Risk Solutions, the global risk management business who produce the risk report:
94: global risk management business who produce the risk report for their customers, said it was 'no
95: Violence Map produced together with The Risk Advisory Group, the report found there had been

Figura 3. Regularidades observadas nas liñas de concordancia do substantivo “risk” (WebCorp)

WebCorp permite procurar un termo chave (*keyword*) na Rede, pero ademáis, en comparación cun buscador tradicional, ten a vantaxe pedagóxica de que presenta os resultados en liñas de concordancias, que facilitan notablemente a identificación por parte do aprendiz dos patróns colocacionais dunha palabra (Guilquin & Granger 2010: 362). Nunha liña de concordancias, como se pode ver na Figura 3, a palabra clave aparece destacada e aliñada no centro, cun pouco de contexto a cada lado, o que nos permite extraer información sobre os seus significados, a súa gramática e, o máis importante para nós, as súas colocacións habituais.

Unha actividade simple que se pode realizar na aula é buscar e identificar as colocacións dunha determinada palabra aplicando algún dos filtros de WebCorp por defecto, por exemplo, o que permite restrinxir a busca á prensa británica “seria” (*UK Broadsheet Newspapers*). Alternativamente, como se ve no exemplo da Figura 3 co substantivo *risk*, é o profesor o que pode levar a cabo a procura e seleccionar unha ou varias pequenas secuencias dentro da liña de concordancias que conteñan información especialmente rica e variada dende o punto de vista colocacional. Xeraríamos entón un documento PDF coas concordancias seleccionadas, que logo compartiríamos cos aprendices. A continuación, podemos pedirllas que abran o ficheiro no seu ordenador ou na súa tableta e que empreguen unha ferramenta de anotación para, como suxire Sripicharn (2010: 381), arrodear con distintas cores nos seus dispositivos, sucesivamente, as preposicións que adoitan ter *risk* como complemento (*at* na Figura 3), os adxectivos que o premodifican (*high*, *low*, *big*, *great*), os substantivos que pode modificar (*risk management*, *risk factor*, *risk rating*, *risk report*), os verbos dos que *risk* é complemento (*minimize*, *cut*, *halve*,

double, reduce), os verbos dos que pode ser suxeito (*rise*), as preposicións que rexe cando vai seguido dun complemento preposicional (*of, to, for*) etc.

A decisión de deixarlle ao aprendiz que realice a busca e saque as súas propias conclusións ou, en cambio, de seleccionar para eles aquelas secuencias da liña de concordancias máis ricas e orientar así a súa análise dependerá do grao de madurez e/ou do nivel de lingua do discente (Sripicharn 2010; Guilquin & Granger 2010): un adulto cun nivel avanzado terá sen dúbida unha marxe de tolerancia máis alta ó “ruído” ou datos non relevantes que inevitablemente xera a busca e non precisará indicacións detalladas do profesor.

WordCorp dispón doutras funcionalidades particularmente útiles para a exploración das colocacións. Ao final dos resultados da busca ofrece varias posibilidades de afinamento (*Post Search Options*, Figura 4). Por exemplo, é posible organizar os resultados alfabeticamente a partir da primeira, da segunda palabra etc. á esquerda ou á dereita da *keyword*, o que nos permite identificar de forma inmediata aquelas formas que máis tenden a aparecer nesa posición en relación coa palabra clave, xa que logo, as colocacións máis frecuentes. Tamén é posible xerar un informe estatístico básico do perfil colocacional (función *collocates*) da palabra clave, que lle permite ó estudante identificar aquelas formas que tenden a aparecer con máis frecuencia no entorno da palabra clave e a posición na que adoitan situarse con respecto a ela: unha, dúas, tres ou catro posicións á esquerda ou á dereita (Figura 5). O principal valor pedagóxico desta ferramenta radica en que nos permite identificar dunha ollada as combinacións máis frecuentes, xa que logo, máis interesantes dende o punto de vista da aprendizaxe, para logo poder estudalas nos exemplos que obtemos na lista de concordancias. A táboa mesmo permite anticipar a relación desas palabras coa palabra clave: por exemplo, se a palabra clave é un verbo, calquera substantivo que tenda a repetirse nas posicións 1 e 2 á súa dereita, será posiblemente un complemento; se a *keyword* é un substantivo, calquera adxectivo recorrente na posición 1 á súa esquerda será un premodificador típico etc.

Post Search Options

Show URLs: ☒ Span: 50 characters

Show Collocates: ☒ Exclude Stopwords: ☐

Sort by: Words to the right Position: 1

Make all above options case insensitive: ☐

Filter by Date: No filter Range: YYYY-MM-DD - YYYY-MM-DD

Restablecer Submit

Figura 4. Panel de opcións de WebCorp para filtrar os resultados dunha busca

A título de exemplo, na Figura 5 amósanse os resultados dunha busca na que se aplicaron como filtros o dominio *bbc.co.uk/news* (só noticias da BBC) e a palabra filtro *health* para limitar os resultados ás noticias sobre saúde. Nela observamos, xa máis claramente do que podíamos ver na secuencia de liñas de concordancias (Figura 3), a forte tendencia do substantivo *risk* a “seleccionar” complementos coas preposicións *of* e *to*, pero tamén o elevado número de aparicións da combinación *at risk*. Podemos ver igualmente que *risk*, neste contexto particular das noticias sobre saúde, adoita formar frases con substantivos coma *health*, *cancer*, *problems* ou *disease*, funcionando non só como premodificadores (e.g., *health risk*, *cancer risk* etc.), senón tamén dentro de frases que complementan a *risk* (*risk of heart disease*) e a combinarse con adxectivos como *high* ou *increased* e máis aínda a funcionar como complemento do propio verbo *increase* (*increase the risk of* etc.). Unha vez identificados estes importantes patróns colocacionais, o estudante pode logo empregar a lista de concordancias para ver exemplos concretos de todas estas combinacións.

Collocates

Word	L4	L3	L2	L1	R1	R2	R3	R4	Total
of	5	7	3	4	166	2	5	11	203
the	6	7	17	50	1	8	7	8	104
a	5	15	40	10	1	10	8	3	92
at	1	5	18	29	4	0	1	1	59
to	9	15	5	0	9	6	6	7	57
health	0	1	3	7	0	21	18	3	53
cancer	1	2	2	15	0	15	15	3	53
and	10	3	7	0	4	0	7	21	52
heart	3	0	4	7	2	22	2	4	44
increased	0	1	4	30	0	0	0	1	36
increase	1	3	27	0	0	1	0	0	32
problems	0	0	0	0	0	0	19	13	32
in	3	4	2	3	2	3	3	10	30
disease	0	1	0	2	0	2	22	1	28
your	1	0	0	24	0	0	0	1	26
have	3	20	1	1	0	0	0	1	26
is	6	3	5	0	3	0	3	5	25
higher	0	1	1	19	0	0	1	1	23

Figura 5. Palabras máis frecuentes á esquerda e dereita do substantivo “risk” (WebCorp)

7. Os grandes corpus

WebCorp constitúe un paso intermedio moi útil entre os resultados obtidos cos motores de busca cos que os estudantes están ben familiarizados e os formatos de resultados máis propios dos corpus académicos. No pasado, estes últimos eran ferramentas imponentes,

de uso complexo e acceso xeralmente moi restrinxido. Hoxe están en moitos casos a dispor de calquera usuario gratuitamente a través de interfaces de fácil uso. É o caso, por exemplo, do repositorio que propoñemos aquí como fonte de datos para este seguinte nivel. Á hora de introducir estes recursos na aula de lingua, malia a facilidade apuntada, cómpre, en todo caso, contemplar algunhas sesións de formación básica sobre a ferramenta para garantir unha explotación adecuada deste recurso por parte dos estudantes.

O banco de datos compilado por Mark Davies, profesor de Lingüística da Universidade Brigham Young (EE.UU.), e o seu equipo hai xa máis dunha década contén máis dunha ducia de corpus de enorme tamaño e continúa a incrementarse con novos textos e coleccións de forma regular. Na súa plataforma en liña⁷ atópanse integrados, entre outros, os 100 millóns de palabras do British National Corpus ou creacións do equipo do profesor Davies, como o GloWbE, Global Web-based English, que contén 1.900 millóns de palabras representativas das variedades de inglés de vinte países, ou o COCA, o corpus de inglés americano contemporáneo, que contén 560 millóns de palabras correspondentes ao período 1990-2017 e que continúa a medrar en 20 millóns adicionais cada ano. Todos estes recursos son accesibles ao usuario xeral a través dun rexistro gratuito, que permite 50 buscas diarias, ou dunha licencia ampliada cun custo duns 30 dólares anuais. De acordo coa información ofrecida polo propio portal, un total de 130.000 usuarios empregan esta ferramenta cada mes, para fins de investigación, para a creación de materiais didácticos ou como fonte de información para a tradución de textos.

Os distintos corpus aloxados no sitio comparten unha mesma arquitectura e unha ferramenta de busca que funciona cunha sintaxe sinxela, accesible a calquera estudante de linguas de nivel intermedio. Todas as coleccións están lematizadas e etiquetadas por clase de palabras, o que permite buscas moi específicas de estruturas recorrentes. Calquera observación sobre unha combinación habitual de palabras feita na clase ou sinalada nos materiais do curso pode servir de punto de partida para unha viaxe exploratoria por este inmenso banco de datos. Así, por exemplo, coa opción “List” [lista] do buscador, en apenas uns pasos podemos obter unha cantidade inxente de datos en resposta a unha cuestión simple como a de que adxectivos se empregan con máis frecuencia para describir a cor, a lonxitude ou o tipo de pelo (“ADJ hair”: *black hair*, *blonde hair* e *dark hair* son os tres máis frecuentes) ou en que orde tenden a colocarse cando se combinan dúas ou máis clases de adxectivos (“ADJ ADJ hair”, Figura 6) e someter así a proba a regra recollida na gramática ou proposta como hipótese polo docente.

⁷ <https://corpus.byu.edu/corpora.asp>.



Figura 6. Lista dos resultados máis frecuentes en COCA da busca “ADJ ADJ hair”

Unha busca máis aberta da combinación de substantivos cunha preposición determinada (por exemplo, “NOUN in”) pode constituír o primeiro paso dunha actividade na que o estudante deba identificar cales dos termos extraídos do corpus son predicados nominalizados que requiren un complemento preposicional específico (por exemplo, *increase in*, *change in*, *interest in*, *difference in*) e completar a relación con outras preposicións frecuentes.

Como vimos ao describir WebCorp, a busca dunha palabra no seu contexto, opción “KWIC (Keyword in Context)”, devolve casos do termo chave xunto cunha porción de contexto previo e posterior ordenados por algún dos termos que o seguen ou o preceden, segundo a preferencia establecida polo investigador. Ademais da disposición ordenada dos resultados, no caso dos corpus de Mark Davies e o seu equipo, a utilización de cores diferentes para cada clase de palabras é unha característica adicional da aplicación que facilita a exploración. Malia esta axuda, a riqueza da información que proporciona o corpus co formato KWIC pode resultar excesiva e pouco doada de procesar por parte daqueles estudantes menos avanzados, polo que cómpre limitar a atención ás estruturas realmente frecuentes. A propia ferramenta ofrece, porén, un modo máis doado de sistematizar esa información a través da opción de colocacións (“Collocates”), coa que se obtén un listado dos termos cos que adoita combinarse a nosa palabra chave, nun intervalo de palabras á dereita ou á esquerda fixado polo investigador.



Figura 7. Vista parcial dos resultados da busca de combinacións da palabra “reason”, no intervalo de dúas palabras no contexto seguinte (COCA)

A modo de exemplo, imaxinemos unha situación na que, como parte da preparación para un debate, un estudante necesite ampliar o seu repertorio de expresións para argumentar a favor ou en contra dunha determinada postura. A súa procura incluíría seguramente termos tales como *reason* [razón, motivo], *proof* ou *evidence* [proba(s)] e o seu obxectivo sería familiarizarse cos contextos e as expresións máis habituais das que forman parte estes termos. A súa exploración podería comezar, por exemplo, por buscar as palabras que seguen con máis frecuencia a “reason(s)”. Coa opción “Collocates”, a busca das posibles combinacións do lema “REASON_N” [razón, motivo_substantivo] no intervalo de dúas palabras no contexto seguinte, tal e como amosa a parte superior esquerda da Figura 7, dá como resultado unha relación de palabras ordenadas por frecuencia que recolle parcialmente a parte inferior esquerda da figura. Cada elemento desa lista é, de feito, un hiperenlace que dá acceso á concordancia correspondente á combinación en cuestión, tal e como se mostra na parte dereita da figura, resultado de seguir o enlace ao cuarto elemento da lista (“doubt”). Deste xeito, as estruturas que lle poden resultar útiles ao estudante preséntanselle de forma máis sistematizada en forma de subconxuntos de exemplos similares.

8. O aprendiz como compilador de corpus

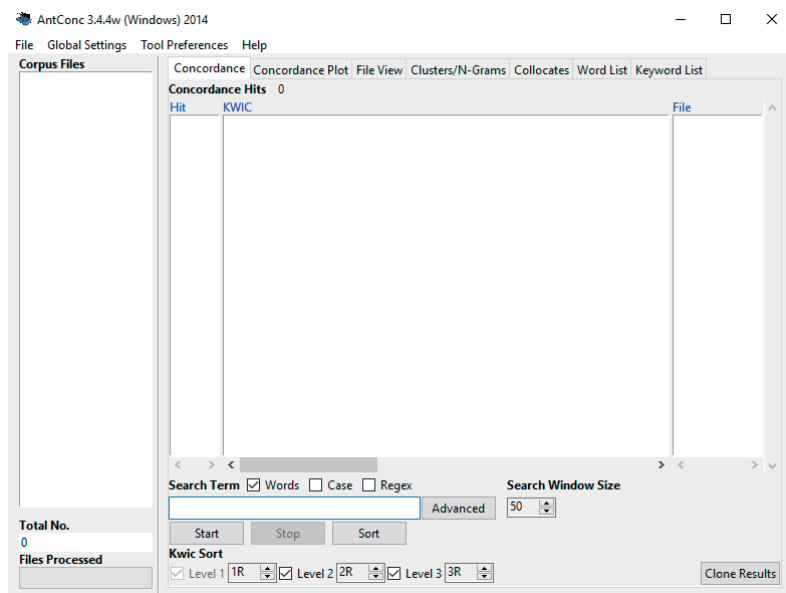


Figura 8. Interfaz do programa de concordancia AntConc, de Laurence Anthony

Unha vez que o estudante está familiarizado coa noción de corpus e cos formatos de resultados que se empregan de forma máis habitual, está listo para o seguinte paso, que non é outro que realizar a súa propia compilación de textos. Esa colección pode estar simplemente constituída polo conxunto de textos empregados ao longo dun curso, de xeito que a exploración pode converterse nunha actividade de repaso. Con todo, é seguramente máis estimulante concibila como unha actividade cooperativa na que se pode involucrar o conxunto da clase. Os criterios de inclusión de posibles textos poden ser obxecto de discusión e dar pé a interesantes reflexións sobre o concepto de rexistro, a calidade das fontes, o grao de especificidade temática etc. A participación activa dos estudantes na escolla de materiais conferiralles a estes unha familiaridade da que carecen os textos máis “anónimos” dos corpus tradicionais (Aston 2002: 11), o cal constitúe unha certa vantaxe de cara á explotación posterior do corpus, ao facer, por exemplo, os contextos das liñas de concordancia máis recoñecibles para o estudante. Compilar un corpus *ad hoc* deste tipo pode ser de especial utilidade cando o discente ten que prepararse para empregar un rexistro específico, por exemplo, para realizar unha presentación sobre un tema de especialidade no contexto dunha experiencia de aprendizaxe integrada de contidos curriculares e lingua estranxeira (AICLE).

Xunto cos textos, é necesario contar con algún programa informático que permita a extracción de información sen complicacións técnicas innecesarias. AntConc é unha ferramenta que cumpre perfectamente ese propósito. Foi desenvolvida por Laurence

Anthony, profesor de ciencia e enxeñería da Universidade de Waseda (Xapón), e está dispoñible gratuitamente na súa páxina web⁸. O seu sobrio interface de usuario (Figura 8) inclúe todas as funcións básicas, tales como a extracción de liñas de concordancia (*Concordance*), as colocacións (*Collocates*) ou a lista de palabras ordenadas por frecuencia de aparición (*Word List*). Non permite a inclusión da categoría gramatical nas buscas, ou o emprego de lemas, pero a utilización do asterisco * como comodín posibilita a busca de variantes dunha palabra. A secuencia “reason*”, por exemplo, identificaría calquera das seguintes formas: *reason, reasons, reasoning, reasoned* ou *reasonable*.

As compilacións de textos reunidas para explorar con esta ferramenta na aula serán loxicamente de tamaño reducido, da orde dunhas decenas de miles de palabras. Con todo, ao estar recollidas cuns obxectivos moi específicos e presentar, xa que logo, considerable homoxeneidade temática e de rexistro e tipo textual, delas pódese extraer un número comparativamente elevado de expresións, combinacións léxicas etc. características do ámbito representado. En comparación, os corpus estándar, malia o seu tamaño, poden conter, en realidade, apenas un feixe de textos sobre o tema específico de interese para o estudante.

Levar á práctica esta actividade na aula non supón excesivas complicacións. Unha vez discutidos os obxectivos específicos, fíxanse uns criterios para a selección de textos válidos: o tema pode ser moi específico (por exemplo, as vacinas) ou máis ben xeral (por exemplo, a saúde), o rexistro pode ser máis académico ou máis xeral ou divulgativo etc., en función da necesidade específica do discente. Unha vez acordados os criterios, distribúese entre os membros da clase a tarefa de achegar textos coa axuda de calquera dos motores de busca máis populares. Para dar unha idea, para xerar o exemplo presentado na Figura 9, recolléronse só 13 textos, que en conxunto suman algo menos de 17.000 palabras, de diversos tipos de fontes: apuntamentos de bioloxía, blogs sobre saúde, textos periodísticos divulgativos, entre outros. Unha vez gravados os documentos en formato TXT e cargados na aplicación AntConc⁹, o primeiro paso da exploración pode consistir en extraer unha lista de palabras ordenadas por frecuencia de aparición. Fóra das palabras gramaticais que adoitan ocupar os primeiros postos (*the, to, and, a, of* etc.), podemos centrar a atención sucesivamente nos termos específicos que figuran nos postos altos desa relación e tomar nota das estruturas e combinacións nas que se integran. No exemplo da Figura 9, pódese observar que o termo *symptoms* [síntomas], no posto 39 da lista con 58 casos, pode combinarse con verbos como *have, experience, show* para expresar que alguén “ten, experimenta ou amosa” síntomas dunha doenza, que son posibles *cold symptoms* e *(any) symptoms of a cold* [síntomas do arrefriado], *flu symptoms* [síntomas da gripe] e *(further) symptoms of (swine) flu* [(outros) síntomas da gripe (porcina)] etc.

⁸ <http://www.laurenceanthony.net/software.html>

⁹ Previamente débese seleccionar na configuración xeral (*Global settings*) o sistema de caracteres de Europa Occidental (*Western Europe “Latin1” (iso-8859-1)*).

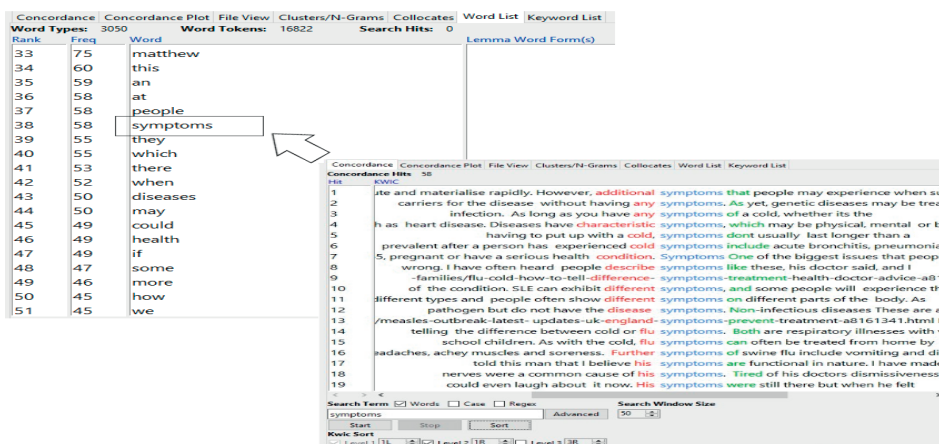


Figura 9. Paneis de resultados de AntConc: Lista de palabras (ordenadas por frecuencia) na parte superior esquerda da imaxe e Concordancia (KWIK) na parte inferior dereita

En definitiva, cunha pequena colección de textos podemos obter información abundante e valiosa sobre pautas fraseolóxicas e combinatorias, e, se cadra, formular novas preguntas. Podemos, por exemplo, preguntármolos se a presenza de certos elementos adicionais na frase favorece o uso de estruturas do tipo *cold symptoms* fronte a *symptoms of a cold*, e acudir a fontes como COCA para buscar exemplos de “*NOUN symptoms*” fronte a “*symptoms of*”. Calquera observación concreta pode, deste xeito, converterse en punto de partida para unha nova tarefa de exploración que expón aos estudantes a múltiples exemplos de combinacións frecuentes de palabras.

9. Reflexións finais

Na aprendizaxe dunha lingua estranxeira, a familiaridade coas combinacións fixas ou habituais de palabras é un aspecto importante do nivel de fluidez acadado. Non obstante, nos métodos máis tradicionais de ensino este aspecto estaba esencialmente desatendido, coa excepción das frases idiomáticas ou, no caso do inglés, das combinacións de verbos con partículas adverbiais, os temidos “*phrasal verbs*”, que, por ser expresións de significado pouco transparente, entrañaban especial dificultade. Os avances na descrición lingüística na segunda metade do século XX e, sobre todo, o rápido desenvolvemento da Lingüística de corpus trouxeron a primeiro plano esta cuestión das combinacións frecuentes e das “preferencias” combinatorias das palabras. Desde hai xa tres décadas, as denominadas colocacións e os fenómenos fraseolóxicos entendidos nun sentido amplo veñen tendo unha presenza significativa nos materiais e recursos didácticos.

A nosa intención con este traballo non era outra que presentar este terreo fronteirizo entre o léxico e a gramática como un ámbito especialmente apto para estimular a curiosidade dos estudantes e fomentar a súa iniciativa e autonomía na aprendizaxe. A través de Internet, docentes e estudantes teñen hoxe ao seu dispor unha fonte inesgotable de “materia prima”

que cómpre manexar con espírito crítico e método para poder explotar as súas inmensas posibilidades. O percorrido que presentamos aquí por distintas ferramentas e materiais constitúe unha proposta para achegarse de xeito gradual a ese proceso de adquisición das competencias necesarias.

Non cabe dúbida de que a aprendizaxe baseada en datos e o emprego de corpus en concreto presentan certas dificultades técnicas e condicionantes materiais que explican o escepticismo co que a miúdo se reciben propostas nesta liña por parte do profesorado de linguas. Confiamos en que as actividades e recursos presentados aquí demostren que ningunha desas dificultades supón un obstáculo insuperable que deba impedir a integración nalgunha medida deste enfoque na práctica didáctica das aulas de linguas. Os efectos positivos a medio e longo prazo serían mesmo maiores cós inmediatos.

10. Referencias bibliográficas

- AIJMER, Karin (ed.) (2009): *Corpora and language teaching*. Ámsterdam/Filadelfia: John Benjamins.
- ASTON, Guy (2002): “The learner as corpus designer”, en KETTERMANN, Bernhard e MARKO, Georg (eds.), *Teaching and learning by doing corpus analysis: Proceedings of the Fourth International Conference on Teaching and Language Corpora, Graz 19-24 July, 2000*, 9–25. Ámsterdam/Nova York: Rodopi.
- BAHNS, Jens (1993): “Lexical collocations: A contrastive view”, en *ELT Journal*, 47(1), 56–63.
- BEEBY, Allison, RODRÍGUEZ-INÉS, Patricia e SÁNCHEZ-GIJÓN, Pilar (eds.) (2009): *Corpus use and translating: Corpus use for learning to translate and learning corpus use to translate* (Vol. 82). Ámsterdam/Filadelfia: John Benjamins Publishing.
- BENNETT, Gena R. (2010): *Using corpora in the language learning classroom: Corpus linguistics for teachers*. Ann Arbor: University of Michigan Press.
- BENSON, Morton, BENSON, Evelyn e ILSON, Robert (1986): *The BBI combinatory dictionary of English. A guide to word combinations*. Ámsterdam/Filadelfia: John Benjamins.
- BERNARDINI, Silvia (2000): “Systematising serendipity: Proposals for concordancing large corpora with language learners”, en BURNARD, Lou e MCENERY, Tony (eds.), *Rethinking language pedagogy from a corpus perspective*, 225–234. Fráncfort: Peter Lang.
- (2002): “Exploring new directions for discovery learning”, en KETTERMANN, Bernhard e MARKO, Georg (eds.), *Teaching and learning by doing corpus analysis: Proceedings of the Fourth International Conference on Teaching and Language Corpora, Graz 19-24 July, 2000*, 165–182. Ámsterdam/Nova York: Rodopi.
- BIBER, Douglas, JOHANSSON, Stig, LEECH, Geoffrey, CONRAD, Susan e FINNEGAN, Edward (1999): *Longman grammar of spoken and written English*. Londres: Pearson.

- BOERS, Frank e LINDSTROMBERG, Seth (2008): *Teaching chunks of language*. Helbling Languages.
- BOERS, Frank e WEBB, Stuart (2018): “Teaching and learning collocation in adult second and foreign language learning”, en *Language Teaching*, 51(1), 77–89.
- BOULTON, Alex (2010): “Learning outcomes from corpus consultation”, en MORENO JAÉN, M.; SERRANO VALVERDE, F. e CALZADA PÉREZ, M. (Eds.), *Exploring new paths in language pedagogy: Lexis and corpus-based language teaching*, 129–144. Equinox.
- (2011): “Data-driven learning: The perpetual enigma”, en GOŹDŹ-ROSZKOWSKI, Stanislaw (ed.) *Explorations across languages and corpora*, 563–580. Fráncfort: Peter Lang.
- CARTER, Ronald e MCCARTHY, Michael (1988): *Vocabulary and language teaching*. Londres: Pearson.
- COBB, Tom (1999): “Breadth and depth of lexical acquisition with hands-on concordancing”, en *Computer Assisted Language Learning*, 12(4), 345–360.
- CORPAS PASTOR, Gloria (1996): *Manual de fraseología española*. Madrid: Gredos.
- (2001): “En torno al concepto de colocación”, en *Euskera*, 46 (1), 89–108.
- FLOWERDEW, Lynne (2015): “Data-driven learning and language learning theories: Whither the twain shall meet”, en LENKO-SZYMAŃSKA, Agnieszka e BOULTON, Alex (eds.), *Multiple affordances of language corpora for data-driven learning*, 15–36. Ámsterdam/Filadelfia: John Benjamins.
- GUILQUIN, Gaëtanelle e GRANGER, Sylviane (2010): “How can data-driven learning be used in language teaching?” en O’KEEFFE, Anne e MCCARTHY, Michael (eds.), *The Routledge handbook of corpus linguistics*, 359–370. Londres: Routledge.
- JOHNS, Tim F. (1994): “From printout to handout: Grammar and vocabulary teaching in the context of data-driven learning”, en ODLIN, Terence (ed.), *Perspectives on pedagogical grammar*, 27–45. Cambridge: Cambridge University Press.
- KOOSHA, Mansour e JAFARPOUR, Ali Akbar (2006): “Data-driven learning and teaching collocation of prepositions: The case of Iranian EFL adult learners”, en *Asian EFL Journal Quarterly*, 8(4), 192–209.
- LEWIS, Michael (2000): “There is nothing as practical as a good theory”, en LEWIS, Michael (ed.), *Teaching collocation. Further developments in the lexical approach*, 10–27. Hove: Language Teaching Publications.
- MAIR, Christian (2002): “Empowering non-native speakers: The hidden surplus value of corpora in continental English departments”, en KETTERMANN, Bernhard e MARKO, Georg (eds.), *Teaching and learning by doing corpus analysis: Proceedings of the Fourth International Conference on Teaching and Language Corpora, Graz 19-24 July, 2000*, 119–130. Ámsterdam/Nova York: Rodopi.
- MEUNIER, Fanny e GRANGER, Sylviane (eds.) (2008): *Phraseology in foreign language learning and teaching*. Ámsterdam/Filadelfia: John Benjamins.
- NATTINGER, James R. e CARRICO, Jeanette S. de (1992): *Lexical phrases and language teaching*. Oxford: Oxford University Press.

- NESSELHAUF, Nadja (2003): “The use of collocations by advanced learners of English and some implications for teaching”, en *Applied Linguistics*, 24(2), 223–242.
- O’KEEFFE, Anne, MCCARTHY, Michael e CARTER, Ronald (2007): *From corpus to classroom: language use and language teaching*. Cambridge: Cambridge University Press.
- RENOUF, Antoinette; KEHOE, Andrew e BANERJEE, Jay (2007): “WebCorp: an integrated system for web text search”, en HUNDT, Marianne; NESSELHAUF, Nadja e BIEWER, Carolin (eds.), *Corpus linguistics and the Web*, 47–67. Ámsterdam: Brill-Rodopi.
- REPPEN, Randi (2010): *Using corpora in the language classroom*. Cambridge: Cambridge University Press.
- RÖMER, Ute (2011): “Corpus research applications in second language teaching”, en *Annual Review of Applied Linguistics*, 31, 205–225.
- RUTHERFORD, William (1987): *Second language grammar: Learning and teaching*. Nova York: Longman.
- RUTHERFORD, William e SHARWOOD SMITH, Michael (1988): *Grammar and second language teaching: A book of readings*. Nova York: Newbury House.
- SCHMIDT, Richard (2001): “Attention”, en ROBINSON, Peter (ed.), *Cognition and second language instruction*, 3–32. Cambridge: Cambridge University Press.
- SINCLAIR, John (1991): *Corpus, concordance, collocation*. Oxford: Oxford University Press.
- SRIPICHARN, Passapong (2010): “How can we prepare learners for using language corpora”, en O’KEEFFE, Anne e MCCARTHY, Michael (eds.), *The Routledge handbook of corpus linguistics*, 371–384. Londres: Routledge.
- TRIBBLE, Christopher (2015): “Teaching and language corpora: Perspectives from a personal journey”, en LEŃKO-SZYMAŃSKA, Agnieszka e BOULTON, Alex (eds.), *Multiple affordances of language corpora for data-driven learning*, 37–62. Ámsterdam/Filadelfia: John Benjamins.
- TRIBBLE, Christopher e JONES, Glyn (1997): *Concordances in the classroom: A resource guide for teachers*. Houston: Athelstan Publications.
- VASILJEVIC, Zorana (2014): “Teaching collocations in a second language: Why, what and how?”, en *ELTA Journal*, 2(2), 48–73. Consultado en: www.eltajournal.org.rs.
- WICHMANN, Anne (ed.) (1997): *Teaching and language corpora*. Londres: Longman.