

Estudo das colocacións a través da análise de corpus

Un estudo das colocacións co verbo galego *caer* e o correspondente inglés *to fall*

M^a Carmen Paz Roca¹

University of Wolverhampton

O obxectivo deste traballo é facer unha análise comparativa das colocacións co verbo galego *caer* e o correspondente inglés *to fall*. Tal estudio está baseado na análise de frecuencias de coaparición deses verbos en dous córpora: o CORGA (Corpus de Referencia do Galego Actual) para o galego e o BNC (British National Corpus) para o inglés. A intención é máis a de ser un experimento sobre que nos pode aportar un corpus á hora de analizar aspectos fraseolóxicos, que un estudo cuantitativo exacto. O traballo en si consta de dúas partes: a primeira é unha reflexión sobre o que un corpus nos pode aportar á hora de facer este tipo de estudos e a segunda é a presentación e análise comparativa dos resultados.

The aim of this study is making a comparative analysis of collocations with the Galician verb *caer* and the corresponding English one *to fall*. Such study is based on the analysis of co-occurrence frequencies of these verbs in two corpora: CORGA (Corpus de Referencia do Galego Actual) for Galician and BNC (British National Corpus) for English. The purpose is to make an experiment of how corpora can contribute to the analysis of phraseology, rather than an accurate quantitative study. The work has two main parts: the first one is a reflection upon what a corpus can contribute to this type of study and the second one is the presentation and comparative analysis of the results.

¹ O meu agradecemento ó Research Group in Computational Linguistics da University of Wolverhampton e en especial ó profesor Ruslan Mitkov e a Constantín Orarsan, por poñeren á miña disposición tódolos medios e apoio necesarios para a elaboración deste traballo.

1. Por que usar c rpora?

O uso de c rpora na investigaci n lexicogr fica non   un descubrimento do s culo XX, pero a forma de procesalos para obter informaci n sobre o uso real da lingua   algo relacionado co desenvolvemento da inform tica.   principio, a maior a dos traballos de lexicograf a computacional centr base nas palabras e frecuencias, mais a fraseolox a e o estudo da fixaci n a trav s de c rpora foi adquirindo m is relevancia nas  ltimas d cadas.

Lexic grafos como Samuel Johnson, ou mesmo os centos de persoas que traballaron para compila-lo *Oxford English Dictionary*, xa usaban un corpus. Con todo, a s a idea de como empregalo para sabe-lo significado dunha palabra base base en recoller l xico lendo textos e facendo fichas con citas que conti an cada lema e mailos contextos correspondentes. Podemos imaxina-lo tama o dun proxecto destas caracter sticas hai douscentos anos, cando todo se fac a sobre papel e non hab a ordenador para organizar ou levar conta do traballo. Hoxe en d a son tam n moitos os dicionarios que se ve en compilando desde o enfoque da simple base de datos, s  que agora resulta moito m is r pido coa axuda dos ordenadores. As  e todo, cada vez   m is frecuente que os dicionarios modernos estean baseados nun corpus. O obxectivo deste novo enfoque non   usa-los ordenadores para face-lo mesmo traballo que fac an Samuel Johnson e os seus contempor neos, sen n para face-lo traballo dunha forma que s  podemos facer grazas   inform tica.

Biber describe a diferenza entre un m todo e mailo outro como segue:

The primary difference is that citation slips represent only those contexts that a human reader happens to notice (in some cases representing only the more unusual uses). In contrast, a corpus-based listing provides an exhaustive listing of all occurrences and contexts (Biber et alii 1998: 26) .

Merc    uso dos ordenadores, ademais de ter unha lista, podemos ter tam n automaticamente as frecuencias de uso en distintos rexistros e variedades de lingua; e isto vale tanto para a lexicograf a en xeral coma para o caso concreto da investigaci n fraseol xica.

En resumo, moitas das listas fraseol xicas coas que contamos compil ronse a base de ler textos e reconece-la fixaci n e os contextos de uso de xeito manual ou ben a partir de traballo de campo ou da memoria do propio autor. As  e todo, ter un corpus facilita a an lise de frecuencias e mais a elaboraci n das listas de palabras en contexto.

2. CORGA, BNC e a fraseolox a

Tendo en conta o exposto anteriormente, o obxectivo deste traballo   presentar un estudo pr ctico que leve   reflexi n sobre como pode contribuí-la an lise de corpus   estudo das colocaci ns e das expresi ns fixas. Para iso, escoll ronse as colocaci ns co verbo galego *caer* e o correspondente ingl s *to fall* e os c rpora empregados foron o CORGA (Corpus de Referencia do Galego Actual) e o BNC (British National Corpus).

A elección do enfoque fundaméntase no crecente interese pola elaboración de córpora galegos (CORGA, TILGA, CLUVI) nos últimos anos que, malia ofrecer un campo de traballo moi amplo, aínda están sen explotar. Non é así no caso do inglés, onde a análise de corpus é xa unha tradición. Por iso pode ser interesante ver qué se pode facer co BNC e qué se podería facer cos nosos córpora galegos tamén.

Neste punto é importante ter en conta as diferenzas entre ámbolos dous córpora. Por un lado, o CORGA, que foi elaborado por un equipo do Centro Ramón Piñeiro para a Investigación en Humanidades, é aínda un proxecto aberto. Contén na actualidade 17.646.907 palabras de fontes diversas desde 1975 ata a actualidade, sobre todo de estilo periodístico. Por outro lado, o BNC é un proxecto xa completo, levado a cabo por un equipo dirixido pola Oxford University Press. Contén 100 millóns de palabras e está considerada a maior colección de textos de inglés británico, que ademais se compilou co obxectivo primordial de evitar sobrerrepresentar textos idiosincráticos.²

A análise demostrou que no CORGA hai moitos segmentos que se repiten debido á similitude das fontes e do tema. De aí que non tódolos elementos ou series de elementos que presentan unha frecuencia alta se poidan considerar como fixados na lingua. Por exemplo, o feito de incluír unha gran cantidade de textos periodísticos leva a que a cobertura dun mesmo acontecemento durante varias semanas devolva frecuencias moi altas para combinacións de palabras que non teñen tal representatividade na lingua. Por exemplo, *caer un proxectil* aparece 5 veces no CORGA en noticias relacionadas coa guerra de Bosnia; e esa é a mesma frecuencia que *caer enfermo*. Sen embargo, a asociación entre *proxectil* e *caer* non parece tan frecuente no contexto global da lingua galega.

Outra diferenza importante entre o BNC e mailo CORGA é o motor de busca. O do BNC (Sara98) permite a busca directamente por colocacións e frecuencias. Ademais, dános o *z-score*³ da relación entre dous elementos que aparecen xuntos con certa frecuencia. Incluso é posible determina-lo número de palabras cara á dereita ou cara á esquerda que queremos considerar para calcula-la frecuencia de coaparición, así como a categoría gramatical da palabra ou palabras que buscamos ou unha cadea de categorías gramaticais determinada.

O CORGA é un corpus máis pequeno e o motor de busca non permite facer buscas tan complexas. Polo tanto, as frecuencias de coaparición hai que calculalas sobre unha base de datos elaborada a partir dos resultados que o corpus nos proporciona. Os resultados que obtemos son moi semellantes ós que se poden obter a partir dunha base de datos ou mesmo dun documento de texto. Por exemplo, non é posible especifica-la categoría gramatical ou a cadea de categorías gramaticais buscada. Así pois, podemos limita-la busca para ve-los casos en que *caer* vai seguido de *desde* ou de *de*, ou de *vítima* ou

²“avoids over-representing idiosyncratic texts” (<http://www.natcorp.ox.ac.uk/>).

³ O *z-score* é unha medida estatística. Neste caso, expresa cal é a probabilidade de que dous elementos non teñan relación entre si. Por exemplo, se no corpus aparece a colocación *caer na tentación*, o *z-score* mide a probabilidade de que *caer* e *tentación* teñan correlación ou simplemente aparezan con certa frecuencia, sen haber relación entre eles. Polo tanto, canto máis alto sexa o *z-score*, máis probable é que dúas palabras que aparecen xuntas sexan unha colocación.

derrotado; pero o que non nos permite é analizar automaticamente a frecuencia de aparición de *caer* seguido de preposición, substantivo, adxectivo, etc.

Dado que os dous córpora son diferentes, a información resultante e mailo proceso de análise tamén son diferentes. Disto resulta que os datos expostos a continuación teñen máis un valor metodolóxico ca estrictamente cuantitativo.

3. Procedemento

Esta análise foi levada a cabo en tres pasos: análise bibliográfica, exploración e explotación dos dous córpora e, finalmente, análise comparativa dos resultados nas dúas linguas.

Para empezar, o manual *Corpus Linguistics* de Douglas Biber e *Fixed Expressions and Idioms in English: A Corpus-Based Approach* de Rosamund Moon, foron decisivos á hora de defini-lo punto de partida. O primeiro ofrece unha explicación clara e precisa do uso de córpora en lingüística, incluíndo exemplos de cada unha das posibilidades. E o de Moon é unha análise completa e minuciosa da fraseoloxía inglesa baseándose na análise de corpus. De aí a decisión de combina-la análise de córpora en dúas linguas co campo da fraseoloxía.

Unha vez exploradas as posibilidades de investigación, fíxose a elección do tema e do procedemento. A primeira vista a comparación entre o inglés e o galego parecía difícil debido á diferenza estrutural das dúas linguas: a primeira moi analítica e a segunda máis sintética. Esas diferenzas entre o inglés e o galego nótanse sobre todo na variación morfolóxica. Tal é o caso do verbo *caer*. A desinencia verbal cambia segundo a persoa gramatical (*eu caio, ti caes...*), o tempo (*eu caio, eu caía...*) e o modo (*eu caio, eu caía...*). E algúns verbos irregulares presentan moita máis variación aínda. Véxase, por exemplo, o verbo *ir* (*eu fun, eu irei, eu vaía...*) ou *ser* (*eu son, ti es, nós sexamos, eles foron...*).

Nun intento de face-la busca eficiente e ó mesmo tempo minuciosa, seguiu-se este procedemento: dividi-lo verbo entre elementos constantes e elementos variables. O elemento constante era a raíz *ca-* e os variables *-e-* e máis *-ia-*. A combinación do elemento constante, máis os dous variables e máis calquera outro segmento, dános tódalas formas posibles de *caer*. Pero o inconveniente deste proceder é que formas que non pertencen ó paradigma verbal de *caer* pero que teñen esa combinación de caracteres poden aparecer como resultados da busca. Por exemplo, cando buscamos *cae*^{*4}, aparecen palabras como *Caetano, Caeiro, CAE* (Comités Abertos de Estudantes), etc. e cando se busca *caía*^{*5} obtéñense resultados como *caiado*. Por iso, un dos pasos á hora de puli-los resultados é determinar se tódolos termos que se están tendo en conta pertencen ó paradigma que buscamos ou non.

⁴ O asterisco significa que a busca é para calquera forma que coincida co que nós escribimos no motor de busca, seguido de calquera outro carácter ou serie de caracteres.

⁵ Hai unha opción de busca que nos permite non ter en conta os tiles. Polo tanto, pedir resultados que conteñan *caía*^{*} inclúe tanto formas como *caía* coma outras como *caíamos*.

Neste sentido, as dificultades que presenta un verbo irregular inglés non teñen comparación coas do galego. Por exemplo, o verbo irregular *to fall* só ten cinco formas en todo o paradigma: *fall* para tódalas persoas, tempos e modos de presente e futuro, excepto a terceira persoa de singular de presente que é *falls*, *fell* (pasado simple), *fallen* (participio de pasado) e *falling* (xerundio). Estas son as únicas formas que se buscaron para a análise de colocacións co verbo *to fall* en inglés. Sobra dicir que, de tratarse dun verbo regular, a busca tería sido aínda máis simple.

4. Análise dos resultados

Xa queda dito que os resultados obtidos nos dous casos foron moi diferentes. No caso do galego, a busca de colocacións co verbo *caer* deu lugar a unha lista de contextos de aparición do verbo. Despois transformouse esta lista nunha base de datos para atopar as coaparicións e determinar con que elementos colocaba, e ese foi o listado que permitiu conta-la frecuencia de aparición dun determinado segmento. É importante ter en conta neste punto que se entendeu por colocación para facer este estudo.

Por unha banda, John Sinclair, referente imprescindible na análise de corpus en inglés, considera que “collocation is the occurrence of two or more words within a short space of each other in a text” (Sinclair 1991: 170). Esta definición non resultou de todo axeitada para este traballo porque se temos en conta só a simple coaparición e frecuencia, segmentos como *has fallen* serían colocacións en vez de tempos compostos do verbo.

Por outra banda, F. J. Hausmann limita moito máis ca Sinclair o concepto de colocación. Para Hausmann “a combinación dunha palabra autosemántica cunha sinsemántica axeitada é unha colocación” (Hausmann 1998: 66), entendéndose por palabras autosemánticas aquelas que forman parte do núcleo do vocabulario, son semanticamente autónomas e poden definirse sen ter en conta o contexto; e por palabras sinsemánticas aquelas que “non podemos definir sen referencia a un contexto preciso porque son semanticamente dependentes” (*Ibidem*: 65). Así, unha colocación é unha combinación de palabras que “para a codificación dun texto está dispoñible en bloque” (*Ibidem*:66). Esta definición de Hausmann implica unha relación semántica entre os compoñentes da colocación que vai alén da pura coaparición. É dicir, no caso das colocacións, os distintos elementos coaparecen porque a colocación está fixada en bloque e a aparición do elemento sinsemántico está supeditada á aparición dun determinado elemento autosemántico.

Ademais de colocacións, a definición que se adoptou para este traballo inclúe tamén o que Hausmann considera locucións: aquelas que “non empregan as palabras nin como autosemánticas, nin como sinsemánticas, senón como extrasemánticas” (Hausmann 1998: 66). Por exemplo, se dicimos que *Xoán caeu nas redes da mafia*, *caer* e *redes* teñen un significado figurado que depende do contexto (non se trata das mesmas redes que cando un peixe cae na rede dun pescador). O mesmo podería dicirse do inglés *fall in love* [namorarse] ou *fall foul* [poñerse a mal / chocar].

Tendo en conta as devanditas posturas sobre o que é e o que non é unha colocación, para o presente traballo consideráronse colocacións aqueles segmentos que presentan unha

fixación alta dentro dos respectivos cónpora, pero tamén se tivo en conta a relación semántica entre eles. Non se fixo distinción do tipo de relación semántica e por iso entran na nosa clasificación de colocación o que Hausmann incluía dentro das locucións. Deste xeito, aínda que o concepto pode parecer demasiado amplo, permitiu descartar combinacións coma *has fallen* ou *is falling* para o inglés, e outras como *teñen caído* ou *ía caer* para o galego, porque a relación que existe entre os elementos desas unidades é gramatical e non semántica. Ó mesmo tempo, o concepto de colocación do que se parte é suficientemente amplo para incluír combinacións difíciles de clasificar. O que nos interesa neste caso é con que palabras combinan os verbos *caer* e *to fall* máis ca que significado extrasemántico adquiere ese bloque (que, sen dúbida, podería ser tema doutro estudo).

Unha vez feito o cribado segundo o criterio que vimos de explicar, vexamos como se analizaron as frecuencias. Para o galego, nun corpus que proporcionou 516 contextos diferentes de aparición do verbo *caer*, consideráronse colocacións aqueles segmentos con frecuencia de aparición superior a 2. Con todo, ó non te-lo *z-score* e ó seren textos moi parecidos, non se pode asegurar que os elementos de frecuencia 1 non sexan colocación nin que tódolos que teñen frecuencia superior a 2 si o son.

Pola contra, no caso do BNC, Sara98 permite buscar automaticamente as frecuencias e consideráronse colocacións os segmentos con *z-score* superior a 10 e con frecuencia de coaparición superior a 6 por cada 1000 casos. O *z-score* dá idea da relación semántica entre palabras en tanto que mide a probabilidade de que haxa relación entre dous elementos máis alá da pura frecuencia de coaparición. Por exemplo, en *caer chuzos* o *z-score* mide a probabilidade de que estas dúas palabras teñan ou non relación semántica. Isto faise analizando con que outros elementos combinan cada unha das palabras dentro do corpus. Se *chuzos* combina con *caer* en máis casos que con calquera outra palabra, o *z-score* será máis alto.

A táboa que segue dá conta das palabras ou cadeas de palabras que coaparecen con *caer* en galego e a frecuencia de aparición no CORGA en números absolutos cando fora maior de 2. En ámbalas dúas táboas, a do galego e máis a do inglés, na terceira columna aparece a correspondencia na outra lingua. Esta marcouse con letra negra e cursiva cando na outra lingua o corpus non proporcionou esa colocación, ou ben cando na outra lingua iso non é unha colocación co verbo *caer* ou *to fall*. Cando simplemente se fixo unha aliñación cos resultados devoltos polos dous cónpora, a correspondencia aparece só en letra negra.

Ademais, o número que aparece á esquerda entre parénteses corresponde á clasificación semántica das distintas colocacións segundo a lenda que aparece a continuación: *caer* asociado a (1) fenómenos meteorolóxicos (2) situacións ou estados negativos, (3) categorizacións, (4) índices, (5) lugar onde / de onde, (6) partes do corpo ou (7) outros (sen clasificar).

CAER COLOCA CON	frec.	corresp. inglés
(5) (2) (3) en (incluíndo as contraccións <i>no, na, nos, nas</i>)	240	into (+mar, desgracia, trampa, desuso, tentación, nas mans de) on (+superficie plana: no chan, día da semana: caer en luns) in (+lugar) to lapse into (caer nun erro, en malos hábitos, vicios, nun tópico)
(5) sobre	39	on / over (+persoa, lugar)
(5) de (incluíndo as contraccións <i>do, da, dos, das</i>)	37	from (+moto, andamio) off (+cabalo)
(5 / 7?) conta	31	to realise (caer na conta)
(6) mans	30	hands (caer nas súas mans: to fall into his hands)
(5) por	24	over (caer por un barranco: to fall over a cliff) drop in on (deixouse caer por Santiago: He dropped in on Santiago)
(4) (un) %	23	% (prezos, índices coma o do paro, bolsa, cotización)
(5) chan	22	floor
(5) mar	19	sea (ó mar, no mar: into the sea)
(2) trampa	16	trap (nunha trampa: into a trap)
(1) neve	15	snow
(1) chuvia	12	rain
(2) derrotado	11	to be defeated
(4) bolsa	11	the Stock Exchange dropped (caeu a bolsa)
(2) ante	10	before
(2) tentación	10	temptation (caer na tentación: to fall into temptation, <i>to give in to temptation</i>)
(5) (2) enriba / encima	10	on/over
(7) en sorte	10	to one's lot
(5) á auga	9	into the water
(5) desde	9	from (desde un segundo piso)
(7) bombas	9	bombs (tamén <i>bombs drop</i>)
(3) ben	8	I like him (cáeme ben)
(5) a	8	into, on, onto
(7) deixa	7	to drop in (deixou caer algo ou deixouse caer, é dicir, pola casa de alguén: he dropped in someone's house) V. por . let fall (deixou cae-la carteira: he let his purse fall)
(2) baixo	6	below, under

(2) desgracia	6	disgrace (caer en desgracia: to fall into disgrace)	(6) de xeonllos	4	on one's knees
(2) equipo	6	team	(7) árbores	4	trees
(7) coma / como	6	to come as (caer coma un tiro: to come as a blow)	(7) muro	4	the wall falls down (cae a casa / caer: fall down)
(1) a noite	5	nightfall	(7) no sono	4	asleep
(1) raio	5	to strike (caeu un raio: lightning struck)	(7) proxectil	4	projectiles / missiles drop (caen proxectís)
(2) enfermo	5	ill	(1) a tarde	3	evening
(4) prezos	5	prices	(1) chuzos	3	to pour / to rain cats and dogs
(5) (1) auga	5	water (caer á auga: to fall into water, pero “caer auga = chover”: to rain, to pour)	(2) asasinado	3	murdered / assassinated
(7) caía quen caía	5	no matter who may fall	(2) erro	3	to make a mistake, to lapse into an error (caer nun erro)
(2) avión	4	to crash, to come down (o avión caeu: the plane crashed / the plane came down)	(2) lesionado	3	injured
(3) da lista	4	to drop from the list	(2) redes	3	trap (caer nas súas redes: to fall into his trap) network (caeu unha rede de traficantes de droga: a drug-dealers' network fell)
(4) consumo	4	consumption	(3) mal	3	not to suit (o martes cáeme mal: Tuesday doesn't suit me) not to like / to get on badly with / not to get on well with (non me cae ben María: I don't like María / I get on badly with María / I don't get on well with María)
(4) mercados	4	market			
(5) (2) diante de	4	before, in the face of			
(5) onde	4	where			
(6) (5) cara (subst. e prep.)	4	to die of shame (cae-la cara de vergoña: to die of shame) towards (caeu cara ó interior: it fell towards the inside)			

(3) tópico	3	<i>to lapse into a cliché</i> (non caían nos tópicos da Galicia de charanga e pandeireta: they wouldn't lapse into the clichés of the brass band and tambourine Galicia)	(2) lazo	2	<i>trap</i> (caer no lazo: to fall into the trap)
(5) escaleira	3	<i>to fall down the stairs</i> (caer polas escaleiras)	(2) lei	2	<i>law</i> (que caia todo o peso da lei sobre os terroristas: let all the weight of the law fall on terrorists)
(5) ó	3	<i>into</i> (cando implica “dentro de”, p.ex. caer ó mar: to fall into the sea / caer ó río: to fall into the river) <i>on</i> (+ superficie plana. Caer ó chan: to fall on the floor)	(2) morta	2	<i>to lie dead</i> (caeu morta: she lay dead)
(6) brazos	3	<i>arms</i> (caer nos seus brazos: to fall into his arms)	(2) preso de	2	<i>prey of stricken</i> (desde que caera preso dun extraño autismo: since he had fallen stricken with a strange autism) <i>stricken</i> (presa do pánico: panic-stricken)
(6) de cabeza	3	<i>headfirst / on one's head</i>	(2) redondo	2	<i>in a heap / to collapse</i>
(6) de cu	3	<i>on one's backside</i>	(2) rendido	2	<i>to drop with tiredness</i> (caín rendido: I dropped with tiredness)
(7) folla	3	<i>leaf</i>	(2) todo o peso	2	<i>all the weight</i>
(1) trebón	2	<i>heavy rain</i>	(2) vicio	2	<i>to lapse into a bad habit</i> (caer nun vicio)
(2) abatido	2	<i>shot</i> (caeu abatido pola policía: he was shot down by the police)	(2) vítima	2	<i>victim</i>
(2) eliminado	2	<i>to be eliminated</i>	(3 / 7?) día	2	<i>on</i> (cae en martes: it falls on a Tuesday)
(2) esquecemento	2	<i>oblivion</i> (caer no esquecemento: to fall / to sink into oblivion)	(3) en gracia	2	<i>to make a hit with s.o.</i> (caeulle en gracia: she made a hit with him)
			(4) cotizacións	2	<i>prices</i>

(4) emprego	2	employment	(6) cabelo	2	hair (o cabelo caíalle polos ombreiros: her hair fell down her shoulders)
(4) índice	2	fall (off)	(6) de fociños	2	to fall flat on one's face
(4) produción	2	production	(6) debruzado	2	flat on one's face / headlong
(4) puntos	2	points	(7) bágoas	2	to drop tears (caían bágoas: tears would drop)
(4) rendibilidade	2	profitability	(7) casa	2	house (caeu unha casa: a house fell down)
(5) (3) dentro	2	into	(7) granada	2	grenade / shell (tamén <i>to drop</i>)
(5) a/á terra	2	to land (os satélites caeron a terra: satellites landed) to come down / to crash (o avión caeu á terra: the plane came down / the plane crashed)	(7) pano	2	curtain
(5) río	2	river (caer ó / no río: to fall into the river)	(7) polo seu propio peso	2	on its own (o sistema cae polo seu propio peso: the system falls on its own)

E a seguinte táboa é a correspondente ó inglés:

TO FALL COLOCA CON	frec.	Z-SCORE	CORRESP. GALEGO
(5) (2) into	405	21.6	en (+ mar, desgracia, trampa, desuso, tentación, nas mans de, <i>desembocar</i>) a / ó / á... (ó mar, ó río) dentro
(4) per	160	10.6	%
(7) love	145	10.6	namorarse (to fall in love)

(4) cent	144	12.8	(per cent) %
(4) by	115	10.3	un (+porcentaxe)
(7) asleep	110	26.3	no sono (non son exactamente equivalentes. "to fall asleep" é "dormirse")
(7 / 2) short	90	13.8	quedarse curto , <i>non alcanzar, non ser suficiente, non cumpli-las expectativas</i>

(4) below	65	10.5	por debaixo de (p.ex. índices. “it fell below 3%: <i>caeu por debaixo do 3%</i>)	(7) wicket	28	49.0	pau (expresión relacionada co mundo do golf)
(2) down	62	18.8	caer, fracasar	(2) victim	27	11.5	víctima (to fall victim of: <i>caer vítima de</i>)
(3) within	62	14.8	dentro (dun grupo, categoría)	(7) tree	25	12.8	árbore
(4) prices	60	12.8	prezos	(2) foul	24	13.1	poñerse a mal / chocar (con algo ou alguén) (to fall foul of sb.: <i>poñerse a mal con alguén</i>)
(3) category	55	23.3	categoría (to fall within a category: <i>entrar dentro dunha categoría.</i>)	(7) let	24	14.5	deixar
(7) over	49	10.9	devecer (por) (to fall over oneself to do sth: <i>devecer por facer algo</i>)	(2) apart	23	11.9	desfacerse, desmoronarse, fracasar
(1) snow	45	13.8	neve	(4) rising	22	32.9	crecente, á alza (the stocks rise and fall: <i>as accións suben e baixan</i>)
(1) rain	44	30.1	chuvia	(4) steadily	22	20.7	sen parar, ininterrumpidamente
(4) %	39	21.8	%	(4) rise	21	20.8	V. “rising”
(4) sharply	34	10.5	bruscamente, en picado (p.ex. un índice)	(2 / 7) behind	20	12.7	rezagarse, quedarse atrás
(7) since (*non parece unha colocación a pesar da frecuencia)	34	14.2	desde (temporal. Production fell since 1999: a produción caeu / baixou desde 1999)	(2) disuse	18	29.9	desuso
(4) profits	33	13.2	beneficios, ganancias	(2) trap	18	15.3	trampa
(4) rates	32	16.7	proporción, %	(4) index	18	27.3	índice
				(7) leaves	17	21.8	follas
				(2) dramatically	16	10.5	dramaticamente, de maneira espectacular
				(4) inflation	14	21.5	inflación

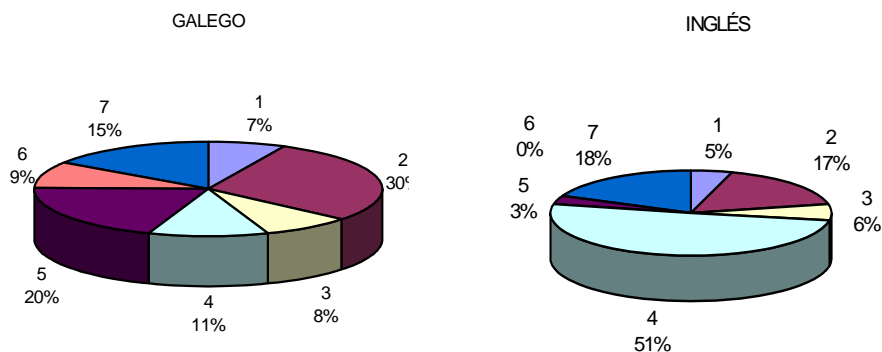
inflation				shadow			in shadow.
(4) values	14	16.2	valores	(2) danger	10	13.1	perigo
(2) prey	13	10.2	víctima	(4) p.c. (per cent)	10	20.3	V. “%”
(4) consumption	13	23.5	consumo	(4) pound	10	12.1	libra (cotización)
(4) points	13	10.4	puntos	(4) rises	10	24.5	V. “rising”
(4) rose	13	12.4	subiron. Fell and rose: baixaron e subiron (accións, índices)	(7) bombs	10	18.1	bombas
(4) stop	13	10.3	parar de (to stop falling: parar de baixar)	(1) dusk	9	17.5	crepúsculo, a tarde
(4) industrial	12	10.9	industrial (asociado coa produción industrial)	(4) billion	8	11.7	mil millóns
(4) output	12	15.7	producción	(4) dollar	8	18.5	V. “pound” (cotización)
(4) temperature	12	18.7	temperatura (caer ou baixar)	(4) GDP	8	27.3	P.I.B.
(4) demand	11	10.2	demanda	(7) silent	8	13.8	quedar en silencio / quedar calado (h e fell silent: quedou en silencio / calado)
(4) investment	11	10.2	inversión	(2) ill	7	10.1	enfermo
(5) floor	11	10.2	chan (to fall on the floor: caer ó/no chan)	(3) frame	7	12.4	marco (to fall within a frame (of): caer dentro do marco (de))
(5) onto	11	14.2	en. “to fall on” = “to fall onto” (+superficie plana)	(4) stands	7	11.8	posicións
(7) shadow	11	11.2	quedarse en sombra: to fall	(3) scope	6	10.3	ámbito
				(4) pre-tax	6	22.5	preimposto / pretaxa (índices)
				(4) retail	6	13.4	venda
				(7) ash	6	17.7	cinza

5. Conclusións

Unha vez vistas as correspondencias e as frecuencias de aparición nas dúas linguas, debemos facer dous tipos de valoración. En primeiro lugar, está a cuestión da pura análise de resultados e, en segundo lugar, a metodoloxía e a utilidade dun corpus para este tipo de análise.

Canto ós resultados, cabe destacar que os dous verbos, *caer* e *to fall*, non teñen sempre o mesmo comportamento nas dúas linguas, aínda que moitas das palabras que colocan con un en galego colocan tamén co outro en inglés. O punto de diverxencia máis importante entre os dous ten que ver co feito de que *to fall* é un verbo frasal en inglés, é dicir, que unha vez combinado cunha partícula (con forma igual ás preposicións pero función diferente) pasan a significar algo que non ten que ver co significado orixinal do verbo. Véxase por exemplo o caso de *fall over* 'devecer'. Neste caso, a análise dos dous córpora non nos axuda a atopalo paralelismo a menos que ámbolos dous conteñan a versión dos mesmos textos en linguas diferentes. Ademais, as correspondencias non son sempre directas. Por exemplo, aínda que *into* en inglés sexa equivalente a *dentro*, non sempre que temos *fall into* en inglés teremos *caer dentro* en galego. Por iso, a aliñación de correspondencias debe ter en conta os contextos que o corpus nos proporciona e non só a lista de palabras que colocan con cada verbo.

Por outro lado, e tamén dentro da análise dos resultados, este é o resultado da tentativa de clasificación semántica:



(1) fenómenos meteorolóxicos, (2) situacións ou estados negativos, (3) categorizacións, (4) índices, (5) lugar onde / de onde, (6) partes do corpo, (7) outros

Os gráficos mostran a distribución porcentual das colocacións segundo o campo semántico ó que se asociaron. Os números que aparecen enriba das porcentaxes corresponden ós números asignados ós distintos grupos semánticos nas táboas de correspondencias.

A primeira vista apréciase unha diferenza notable entre os campos semánticos cos que se asocia *caer* e *to fall*. Podemos pensar que o tipo de textos que inclúen os dous córpora é diferente e, polo tanto, cabería esperar resultados diferentes. Pero casos coma o da ausencia total de colocacións relacionadas coas partes do corpo en inglés, ou a case ausencia das relacionadas co lugar, poñen de manifesto as diferenzas entre estes verbos que inicialmente parecían equivalentes.

No caso das colocacións relacionadas con partes do corpo, no corpus galego hai algunhas do tipo *cae-la cara de vergoña*, outras do tipo *caer de xeonllos*, ou do tipo *caer nos brazos de alguén*. No inglés non aparece ningunha dentro da mostra seleccionada, aínda que si se pode dicir *to fall face down* ‘caer de fociños’ ou *to fall on one’s knees* ‘caer de xeonllos’. O que revela a análise dos dous córpora neste caso pode ser simplemente a diferente frecuencia de uso destas expresións nas dúas linguas.

Por outro lado, a abundancia de colocacións relacionadas con índices no corpus británico podería ter que ver con que *to fall* se usa moito máis en inglés cando se describen gráficos ou a evolución dun determinado índice do que se pode usar en galego *caer*. *Caer* ten unha connotación moito máis negativa e “drástica” ca *to fall*. Por exemplo, non é o mesmo *baixaron os prezos ca caeron os prezos*. No primeiro caso, a baixada non ten por que ser grande, mentres que no segundo a baixada é moi significativa. En inglés, aínda habendo unha lixeira diferenza entre *prizes lowered* e *prizes fell*, poderíamos usar *to fall* para un contexto moito máis amplo.

Non menos vistosa é a diferenza no que atinxe ós fenómenos meteorolóxicos. Neste caso, a diverxencia podería deberse incluso á diferenza estrutural das dúas linguas. En inglés, cando se fala de fenómenos meteorolóxicos, emprégase directamente o verbo (*it rains, it pours, it snows...*) e en galego o uso de expresións do estilo *caeu unha nevada* ou *caeu unha treboada* está máis restrinxido á expresión literaria.

Canto á metodoloxía e, a pesar de tódalas deficiencias que poida te-lo traballo, o uso de córpora para a análise fraseolóxica e en particular das colocacións é produtivo. Un dos valores importantes que nos aporta é a posibilidade de establecer frecuencias de uso. Se ben eses valores son máis significativos canto máis grande e variado sexa o corpus, non deixa de ser unha orientación á hora de determinar que é o que realmente ten fixación e que é o que non. Ademais, a frecuencia vai acompañada de contextos reais de uso da lingua, que en caso de que o corpus estea anotado pode levarnos a determinar se un fraseoloxismo ou colocación dados son máis propios dunha variedade de lingua ou doutra (análise que non se fixo neste caso). Cando se teñen córpora paralelos a utilidade é moito maior, xa que a aliñación é moito máis fácil, exacta e útil para os estudos ou traballos que se fagan a partir dela.

Por último, e non menos importante, un valor indiscutible deste tipo de análise é o que manifestaban os lingüistas Jenny Thomas e Mick Short: “although things that are discovered from corpus analysis are ‘obvious’, they only become ‘obvious’ once the corpus has revealed them to us, i.e. they are not reliably recovered from the lexicographer’s innate knowledge and understanding of the language” (Thomas e Short 1996: 263).

6. Bibliografía

6.1 Referencias de internet

<http://www.natcorp.ox.ac.uk/> para o *BNC (British National Corpus)*

<http://corpus.cirp.es/corga/> para o *CORGA (Corpus de Referencia do Galego Actual)*

6.2 Outras referencias

BIBER, D. et alii (1998): *Corpus Linguistics: Investigating Language Structure and Use*. C.U.P, Cambridge.

CARBALLEIRA ANLLO, X.M. (coord.) et alii (2000): *Gran Diccionario Xerais da Lingua*. Edicións Xerais, Vigo.

CORPAS PASTOR, G. (1996): *Manual de Fraseología Española*. Gredos, Madrid.

HAUSMANN, F. J. (1998): “O diccionario de colocacións. Criterios de organización” en FERRO RUIBAL, X. (coord.): *Actas do I coloquio galego de fraseoloxía*. Centro Ramón Piñeiro para a Investigación en Humanidades-Xunta de Galicia, Santiago de Compostela, 63-81.

MOON, R. (1998): *Fixed Expressions and Idioms in English: A Corpus-Based Approach*. Clarendon Press, Oxford.

SINCLAIR, J. et alii (ed.) (1987): *Collins Cobuild English Dictionary*. Collins, London.

(1995): *Collins Cobuild Dictionary of Idioms*. Harper Collins, London.

SINCLAIR, J. (1991): *Corpus, Concordance, Collocation*. O.U.P, Oxford.

THOMAS, J. e SHORT, M. (ed.) (1996): *Using Corpora for Language Research*. Longman, London / New York.