

---

# Especialización y Prototipicidad en Binomios N y N

Ignacio Rodríguez Sánchez  
Universidad Autónoma de Querétaro  
igrodsan@uaq.mx

## Abstract

El trabajo que aquí se presenta es una investigación sobre binomios de estructura *N y N* (dos sustantivos unidos por la conjunción *y*). Esta investigación tiene un carácter exploratorio y se inscribe en la corriente neofirthiana de la lingüística de corpus. A nivel teórico y metodológico partimos de un enfoque guiado por datos, basándonos en los siguientes corpus: Corpus del Español, CORDE, Googlebooks, y EsTenTen11 (Sketchengine).

Aquí se abordan asuntos relacionados con la naturaleza de este tipo de colocaciones: su frecuencia, la estadística de la información mutua, la dispersión y el grado de reversibilidad para identificarlas; las relaciones que se establecen entre ambas partes; la especialización del nodo (como primera o segunda parte de los binomios) y, finalmente, la prototipicidad de algunos elementos. Se concluye que el concepto de binomio tal como se entendía a partir de su definición clásica, se diluye, y se propone una visión integrada en un sistema dinámico de interacciones complejas e impredecibles.

**Keywords:** sustantivo y sustantivo; binomio; colocación; lingüística de corpus

## 1 Antecedentes

La definición de Sinclair del principio de idiomaticidad representa para algunos el momento de un cambio de paradigma en la lingüística: “*The principle of idiom is that a language user has available to him or her a large number of semi-preconstructed phrases that constitute single choices, even though they might appear to be analysable into segments*” (Sinclair, 1991: 110). Emparentadas con este trabajo, cristalizan visiones del lenguaje como la gramática de patrones (Hunston & Francis, 2000), la teoría de activación léxica (Hoey, 2005), los trabajos de Biber sobre grupos léxicos (1999; Biber & Barbieri, 2007), el análisis coloconstruccional, y otros trabajos como Wray (2002, 2008) y Corrigan et al. (2009a, 2009b).

La definición clásica de binomio corresponde a Malkiel (1959), que en su estudio comparativo de binomios irreversibles en varias lenguas los define como “[...] *the sequence of two words pertaining to the same form-class, placed on an identical level of syntactic hierarchy, and ordinarily connected by some kind of lexical link*”. Según García-Page (2008: 347), un “binomio fraseológico comprende, básicamente, las construcciones simétricas compuestas por dos sintagmas coordinados y los esquemas prepositivos, y, marginalmente ciertas construcciones asindéticas o yuxtapuestas”. Para muchos autores (Almela Pérez, 2006; García-Page, 1998, 2008; Malkiel, 1959) un binomio digno de estudio es básicamente irreversible, mien-

tras que para otros (Moon, 1998) puede no serlo. Un binomio irreversible sería el que no permitiera revertir el orden de sus dos componentes, como por ejemplo *a tontas y a locas*, *coser y cantar*, *cal y canto*. A diferencia de otras expresiones fraseológicas, el significado del binomio a veces sí se puede deducir de la suma de sus partes.

En este trabajo esperamos mostrar que es productivo estudiar los binomios desde una perspectiva que los incluya como colocaciones y no exclusivamente como expresiones idiomáticas. En la tradición fraseológica, sin embargo, hay posturas encontradas sobre este asunto. García-Page (2008: 12) argumenta que la colocación no es una estructura fija y que por tanto no debe ser objeto de estudio de la fraseología (concepción estrecha de la fraseología). Sin embargo, él mismo reconoce la dificultad de establecer definiciones claras: “¿Es *mesa redonda* o *dinero negro* una locución (...), una colocación o un compuesto?” (p.13). Por otra parte, otros estudiosos coinciden con nuestra visión: Corpas Pastor (1996: 52) propone una división de la fraseología en la que las colocaciones tienen perfecta cabida y que coincide plenamente con la de los fraseólogos anglosajones (concepción ancha de la fraseología).

Evert (2009: 1212) señala que el concepto de colocación es uno de los más controvertidos de la lingüística. Las diferencias entre lo que los neofirthianos y los fraseólogos entienden por ese mismo término ha creado una gran confusión en todos los campos. Desde nuestro punto de vista, coincidimos con Stubbs (1996: 172) sobre el hecho de que las intuiciones de los hablantes nativos sobre las colocaciones son muy imprecisas y no pueden de ninguna manera documentar con detalle dichas colocaciones. A una conclusión similar llega Alderson (2007) en el estudio en que compara datos de corpus y apreciaciones de lingüistas sobre la frecuencia de ciertas palabras.

Desde la psicolingüística, la teoría de activación léxica (*Lexical Priming Theory*) de Hoey (2005) supone un marco teórico adecuado para estudiar el fenómeno que nos proponemos abordar. Esta teoría considera que los hablantes hacemos de modo subconsciente complicadas asociaciones léxicas (semánticas, pragmáticas, de colocaciones y de coligaciones) con un género, estilo y situación social. Finalmente, Hoey sostiene que, también de manera subconsciente, percibimos la posición que ocupa una palabra en un texto, la cohesión que esta produce o deja de producir y las relaciones textuales que contribuye a formar. La teoría de activación léxica se apoya en principios psicolingüísticos como que las palabras de mayor frecuencia se activan antes y más en la mente de los hablantes, lo cual favorece un acceso rápido y fácil al lexicón. Esta activación favorece (y a la vez limita y restringe) la combinación entre las palabras. Si, como dice Giammarresi (2010: 262), almacenamos en el lexicón secuencias formulaicas enteras para ahorrar esfuerzo en el procesamiento, entonces es más lógico esperar que, dada una elección entre dos formas de transmitir un mensaje, una formulaica y otra no formulaica, sea la primera opción la que se genere antes.

En español, además del texto de García Page (2008) hay dos excelentes trabajos sobre binomios: el ya mencionado de Almela Pérez (2006) sobre binomios irreversibles y otro del propio García-Page (1998) sobre binomios antitéticos. Ambos trabajos tienen un carácter descriptivo basado en (no guiado por) corpus. Estos trabajos contienen listados de binomios que, aparentemente, se han ido recogiendo de

manera intuitiva. Para nuestra investigación y por motivos de espacio nos referiremos exclusivamente a algunos datos que nos proporciona Almela Pérez (2006).

Los criterios que usa Almela Pérez (2006, págs. 141-146) para definir qué es un binomio son:

- (1) Los binomios constan de dos lexemas.
- (2) Son una secuencia infratextual.
- (3) Tienen una estructura paralelística.
- (4) Sus formas -léxicas y funcionales- son inmutables.
- (5) Forman una secuencia indescomponible.
- (6) Los miembros son inseparables.
- (7) Tienen un significado composicional o idiomático.

## 2 Preguntas de Investigación

Nuestra intención es señalar que, aparte de los criterios mencionados arriba, hay explicaciones de tipo psicolingüístico (basadas en datos cuantitativos) que contribuyen a explicar las formas que adoptan los binomios:

- (1) ¿La frecuencia con que ocurre el binomio y los índices de relación (en concreto la Información mutua -IM en adelante) pueden servir para identificar los binomios? Es decir, si una colocación binomial tiene 20 casos en un corpus y una información mutua de 8 ¿habremos identificado a un binomio irreversible?
- (2) ¿Qué relaciones se establecen entre  $N_1$  y  $N_2$ ? ¿La frecuencia de uso de cada una de las palabras que constituyen el binomio va asociada a su posición en el binomio? Tomando el ejemplo de *aventuras* y *desventuras*, el hecho de que *aventuras* sea más frecuente que *desventuras* ¿nos está diciendo algo sobre el orden en que ese binomio se lexicalizó?
- (3) ¿Se especializan algunas palabras en ser  $N_1$  o de un binomio (siendo  $N_1$  la primera parte del binomio, y  $N_2$  la segunda)? Es decir, si sabemos que la palabra *señor* siempre aparece como  $N_2$  cuando se combina con ciertos sustantivos de un mismo campo semántico (*amigo y señor, amo y señor, dueño y señor, esposo y señor, marido y señor, padre y señor, primo y señor, tío y señor* que aparecen como  $N_1$ ), ¿no es acaso lógico que el binomio *rey y señor* también haya lexicalizado en el mismo orden que sus cohipónimos?
- (4) Aparte de las relaciones semánticas, ¿qué otro tipo relaciones se establecen entre los diferentes colocativos de un binomio?

### 3 Extracción de datos a partir de un corpus

Los datos para este estudio proceden en primera instancia de Corpus del Español ([Davies, 2002] en adelante, CDE). En este corpus se realizó la búsqueda [NN\*] Y [NN\*] de los siglos XIX y XX, con un límite de 5000 casos. Del resultado obtenido se eliminaron los casos con frecuencia menor a cuatro y quedaron 2482 colocaciones binomiales. Por razones prácticas, se eliminaron posteriormente los casos de binomios con sustantivo repetido (por ejemplo, *años y años*). Nuestra base de datos original tampoco incluyó los binomios en los que se intercalan artículos, posesivos o preposiciones entre los dos sustantivos.

En una investigación previa (Rodríguez, 2013: 291) se mencionó la conveniencia de contrastar esos datos con los de corpus más grandes, que es lo que se empieza a hacer en esta investigación, complementando los datos del CDE con los del CREA, Googlebooks (interface de Mark Davies), y EsTenTen11 (Sketchengine).

## 4 Resultados

### 4.1 Información Mutua y Frecuencia

En respuesta a la primera pregunta de investigación, se calculó un índice de relación (IM) entre los dos términos del binomio para cada caso, siguiendo la fórmula basada en Oakes (1998, págs. 63-65). Según Evert (2009: 1229), la IM se debe combinar siempre con una frecuencia mínima (en nuestro caso 4) para equilibrar el sesgo hacia palabras de alta frecuencia.

Tal y como se esperaba, esta estadística mostró que la inmensa mayoría binomios tenían una IM significativa. Solo 140 binomios de los 2482 tenían una IM inferior a 3, que es, según Hunston (2002: 71), la cifra a partir de la cual se suele considerar que una colocación es significativa. Ejemplos de colocaciones y binomios cuya IM mutua es menor a 3 pero con cierto grado de fijación serían: *tiempo y forma*, *fondo y forma*, *padre y señor*, *tierra y libertad*, *forma y manera*, *cuerpo y sangre*, *vida y muerte* (5% de los casos). Esto apunta la imposibilidad de distinguir exclusivamente por métodos estadísticos una colocación de una no colocación (tal y como mencionan Evert [2009: 1242] y Cantos & Sánchez [2002]).

La frecuencia y la IM, pues, pueden ayudar a descubrir formulaicidad de un binomio pero puede que no sean los únicos criterios que los identifiquen. Junto a estos dos criterios, el de la dispersión es otro elemento útil, que contribuye a eliminar elementos estilísticos propios de los textos incluidos en el corpus. Por ejemplo, “Silvicultura y pesca” es un binomio que se repite 67 veces en el corpus pero solo aparece en un texto, que es una enciclopedia.

Sobre la segunda pregunta de investigación, es decir el orden distributivo de los dos términos de los binomios (las relaciones de  $N_1$  con  $N_2$ ) encontramos dos tipos de explicaciones. Por un lado, García-Paige (2008: 347) señala que las posibles explicaciones del orden de los binomios son tanto de tipo semán-

tico (“los principios del “egocentrismo” (...), de jerarquía social (...), de ordenación cronológica u espacial (...), de disposición de contrarios”) como “fonéticos, morfológicos y léxicos”. Estamos de acuerdo con él cuando comenta que este es un tipo de trabajo pendiente en el español, pero creemos que la semántica cognitiva (la tesis de la cognición corporeizada que desarrollaron casi simultáneamente Talmy [1988] y Johnson [1990] parece ser el referente de “los principios del egocentrismo”) tendría algo que añadir en este sentido, en especial la integración (*blending*) conceptual que sugieren Fauconnier & Turner (2003).

Por otro lado, García-Page (2008: 348) hace referencia a que “el orden distributivo también se ha querido ver a veces en el esquema que dispone el constituyente silábicamente más corto en el primer lugar del binomio y el más largo en el segundo (...) una prueba del valor icónico de los binomios”. De ser cierta esta apreciación -no se aportan datos empíricos en español- nosotros optaríamos por ofrecer para este fenómeno una explicación de carácter estadístico y psicolingüístico: simplemente que la primera parte del binomio es más frecuente que la segunda, como mencionamos antes al hablar de la teoría de activación léxica de Hoey. Este orden basado en la frecuencia permitiría al hablante un acceso más rápido al lexicón y facilitaría, por tanto, la fluidez en la comunicación. Lo corto de las palabras más frecuentes (y en primer término) vendría explicado por el corolario a la ley de Zipf (1949) que se refiere a que existe una relación directa entre la longitud de una palabra y su frecuencia (Davies, 2006: 164).

Así, en el binomio *tiempo y forma*,  $N_1$  (*tiempo*) tiende a ser más frecuente que  $N_2$  (*forma*) y en el binomio *oferta y demanda*,  $N_1$  (*oferta*) tiende a ser más frecuente que  $N_2$  (*demanda*). Esta tendencia viene confirmada por los datos de la Tabla 1.

Casos en que $N_1$ es más frecuente que $N_2$	Casos en que $N_2$ es más frecuente que $N_1$	Casos en que $N_1$ es tan frecuente como $N_2$	<b>Total</b>
784	462	17	<b>1263</b>

**Tabla 1: Comparación de la frecuencia de uso de  $N_1$  y  $N_2$  de los binomios.**

## 4.2 Dispersión

El valor de la estadística de información mutua combinado con una medida simple de dispersión contribuye a una mayor precisión en la identificación de binomios. Del total de 2483 binomios identificados inicialmente, 320 aparecen en el CDE de Mark Davies en solo un texto (independientemente de la frecuencia), lo cual equivale al 12.9% del total de binomios. Asimismo, se procedió a buscar en el CORDE el número de documentos en los que aparecía cada uno de los 2483 binomios. En este caso se obtuvo que 446 (18%) aparecían como máximo en un solo documento (de estos, 277 no aparecen ni una sola vez).

El número de binomios que reunían ambas condiciones (solo un documento del CDE de Mark Davies y uno o ninguno en el CORDE) es de 198 (es decir, el 8%).

### 4.3 Reversibilidad

En cuanto a la irreversibilidad del binomio, ya se ha mencionado que esta es la característica esencial del binomio (Almela Pérez, 2006: 155; Malkiel, 1959: 113; García-Page, 2008: 329). García-Page es el único que reconoce que hay excepciones a la irreversibilidad. Según nuestra investigación, la reversibilidad de los binomios es un fenómeno que ocurre con mayor asiduidad de lo que suele considerarse.

En primer lugar, la en cuanto a la irreversibilidad, hay 346 binomios revertidos (del total de 2483 binomios), lo cual supone un 28% del total. Quince ejemplos y sus frecuencias aparecen en la Tabla 2

En la lista de Almela Pérez (2006) hay aproximadamente unos 90 binomios que coinciden con la estructura de los binomios que analizamos nosotros en nuestra base de datos. De estos 90 analizamos una muestra de 30 binomios que encajan con la estructura que se examina en este trabajo. Así, en la lista de binomios irreversibles que da Almela aparecen binomios que en realidad son reversibles, como *día y noche* y *noche y día* (Almela Pérez, 2006: 148-9); y otros casos como *pan y agua*, *cuerpo y alma*, *besos y abrazos*, *calidad y cantidad*, *cielo y tierra*, *uñas y dientes*, y *pies y manos*). Según nuestras estimaciones, un 30% de los binomios que Almela Pérez lista como irreversibles, no lo son en realidad.

<b>Binomio</b>	<b>FREC.</b>	<b>Binomio revertido</b>	<b>FREC.</b>
hombres y mujeres	426	hombres y mujeres	53
oro y plata	213	plata y oro	54
día y noche	171	noche y día	107
cuerpo y alma	136	alma y cuerpo	22
puertas y ventanas	136	ventanas y puertas	15
blanco y negro	136	negro y blanco	4
mujeres y niños	120	niños y mujeres	17
flora y fauna	116	fauna y flora	20
calles y plazas	100	plazas y calles	24
pies y manos	85	manos y pies	22
usos y costumbres	82	costumbres y usos	9
radio y televisión	80	televisión y radio	6
sangre y fuego	73	fuego y sangre	14
petróleo y gas	65	gas y petróleo	7
flor y nata	46	nata y flor	5

**Tabla 2: Ejemplos de pares de binomios revertidos en el CDE con sus frecuencias.**

La Tabla 2 muestra que en la mayoría de los casos, hay una gran diferencia entre las frecuencias de los dos binomios. Hay varias posibles explicaciones para esta situación. Por un lado podemos estar ante casos en que los procesos de lexicalización de la colocación no han terminado todavía pues el binomio supuestamente irreversible todavía no ha desplazado completamente al binomio revertido. Otra explicación, tal vez complementaria, es que un binomio, puede usarse solamente en determinados contextos.

Para nosotros resulta, pues, importante señalar que más que hablar de binomios irreversibles tal vez sea más fértil hablar de un cierto tipo de colocaciones, en  $N_1$  y  $N_2$ , con un alto grado de lexicalización.

#### 4.4 Especialización como $N_1$ o $N_2$

Con la muestra de 30 binomios  $N_1$  y  $N_2$  de Almela Pérez realizamos un análisis más pormenorizado.

Binomio	$N_1$ y N	Nº de tipos	N y $N_1$	Nº de tipos	$N_2$ y N	Nº de tipos
acoso y derribo	acoso y N	5	N y acoso	2	derribo y N	0
agua y ajo	agua y N	107	N y agua	95	ajo y N	6
ajos y cebollas	ajos y N	7	N y ajos	2	cebollas y N	0
alfa y omega	alfa y N	3	N y alfa	0	omega y N	0
armas y bagajes	armas y N	100	N y armas	70	bagajes y N	4
besos y abrazos	besos y N	19	N y besos	13	abrazos y N	12
bombo y platillo	bombo y N	3	N y bombo	1	platillo y N	0
bromas y veras	bromas y N	16	N y bromas	12	veras y N	3
cal y canto	cal y N	12	N y cal	12	canto y N	0
calidad y cantidad	calidad y N	80	N y calidad	75	cantidad y N	22
capa y espada	capa y N	11	N y capa	12	espada y N	19
cara y cruz	cara y N	33	N y cara	45	cruz y N	18
carne y hueso	carne y N	46	N y carne	41	hueso y N	14
carretera y manta	carretera y N	5	N y carretera	3	manta y N	2
causas y efectos	causas y N	24	N y causas	16	efectos y N	21
cielo y tierra	cielo y N	25	N y cielo	11	tierra y N	90
ciencia y conciencia	ciencia y N	61	N y ciencia	32	conciencia y N	38
cruz y raya	cruz y N	18	N y cruz	16	raya y N	1
cuenta y riesgo	cuenta y N	23	N y cuenta	37	riesgo y N	16
cuerpo y alma	cuerpo y N	59	N y cuerpo	33	alma y N	43
día y noche	día y N	34	N y día	11	noche y N	24
garbo y salero	garbo y N	15	N y garbo	7	salero y N	2
golpe y porrazo	golpe y N	10	N y golpe	4	porrazo y N	0
ida y vuelta	ida y N	0	N y ida	0	vuelta y N	7

Binomio	N <sub>1</sub> y N	Nº de tipos	N y N <sub>1</sub>	Nº de tipos	N <sub>2</sub> y N	Nº de tipos
moco y baba	moco y N	3	N y moco	0	baba y N	1
pan y agua	pan y N	54	N y pan	30	agua y N	107
pecho y espada	pecho y N	36	N y pecho	15	espalda y N	13
sangre y fuego	sangre y N	103	N y sangre	84	fuego y N	34
uñas y dientes	uñas y N	8	N y uñas	8	dientes y N	25
viento y marea	viento y N	19	N y viento	15	marea y N	2

**Tabla 3: Análisis de 30 binomios según el número de combinaciones con otros sustantivos (CDE de Mark Davies).**

La Tabla 3 muestra en cuántos tipos de colocaciones aparecen como nodos N<sub>1</sub> y N<sub>2</sub>. Por ejemplo en cuanto al primer binomio de la tabla, *acoso y derribo*, encontramos 5 combinaciones del N<sub>1</sub>, *acoso*, con otro sustantivo (*acoso y abuso, acoso y vigilancia, acoso y protección, acoso y persecución y acoso y derribo*); dos casos en que *acoso* aparece como segunda parte del binomio (*violación y acoso y persecución y acoso*). No se registran casos en que el N<sub>2</sub>, *derribo*, sea primera parte de otro binomio. Es decir que, según nuestros datos, parece que *acoso* se especializa como N<sub>1</sub> pues aparece predominantemente en esa posición. Si hacemos una comparación entre las columnas tres y cinco, tenemos que, de los treinta ejemplos tomados, hay solo 3 en que N y N<sub>1</sub> tiene más tipos de binomios que N<sub>1</sub> y N (se trata de los casos *capa y espada, cara y cruz y cuenta y riesgo*). En estos casos, *espada, cruz y riesgo* son nodos que se combinan más veces como N<sub>2</sub>. Hay también tres casos en los que la colocación N<sub>1</sub> y N, tiene los mismos tipos que N y N<sub>1</sub> (*cal y canto, ida y vuelta y uñas y dientes*). Excluidos estos 6 casos, el 80% de los casos el nodo más productivo es N<sub>1</sub>.

Si comparamos del mismo modo el número de tipo de colocaciones N<sub>1</sub> y N con N<sub>2</sub> y N (columnas 3 y 7), tenemos resultados muy similares: 83% de predominio de N<sub>1</sub> y solo 17% de predominio de N<sub>2</sub> (*capa y espada, cielo y tierra, ida y vuelta, pan y agua, y uñas y dientes*).

¿Se puede deducir pues que las palabras que están en la primera parte de un binomio tienen una potencia combinatoria mayor y esta ayuda a mantener en esa posición? La respuesta es sí, pero con salvaguardas: los casos en que no hay especialización clara (por ejemplo *valor*) y los de binomios que se especializan como N<sub>2</sub> como *confianza, director y libertad*.

## 4.5 Prototipicidad

En cuanto a las relaciones entre colocativos, un análisis somero realizada con los corpus más grandes arrojan resultados comparables a las relaciones de N<sub>1</sub> y N<sub>2</sub>.

Aparte de las relaciones semánticas esperables (hiponimia, sinonimia, hiperonimia, etc.) y las relaciones morfológicas lógicas (misma categoría gramatical), resulta llamativo hallar relaciones sobre la estructura léxica relacionadas en el sonido (*clang association*): con el número de sílabas, la posición de la sílaba tónica, similitudes fonéticas y el uso de afijos. Aquí nos referimos a relaciones entre colocativos, no entre elementos de un mismo binomio como en *troche y moche y tomo y lomo*.



Por ejemplo en la Tabla 4 vemos que de las 10 primeras colocaciones de *sumisión* ordenadas por su IM, hay 5 que son palabras llanas de cuatro sílabas 4 de las cuales tres terminan en *-ismo*, y hay dos parejas que son parecidas en su inicio y su final: *obsecuencia* y *obediencia*, y *abyección* y *abnegación*.

	<b>Binomio</b>	<b>Frecuencia</b>	<b>IM</b>
1	servilismo	42	8.24
2	obediencia	190	7.74
3	obsecuencia	17	7.15
4	vasallaje	12	6.9
5	entreguismo	10	6.64
6	docilidad	11	6.61
7	conformismo	15	6.33
8	pasividad	31	6.13
9	abyección	6	6.13
10	abnegación	10	5.97

**Tabla 4: Colocativos en binomios de *sumisión* y (Fuente: [www.SketchEngine.co.uk](http://www.SketchEngine.co.uk), EsTenTen11, American, TreeTagger).**

No es infrecuente hallar este tipo de situación. La Tabla 5 muestra que entre las diez colocaciones más frecuentes de *privaciones*, tiene gran parecido varias parejas: *austeridades* y *penalidades*, *escaseces* y *estrecheces*, *pobrezas* y *durezas*.

	<b>Palabra</b>	<b>Frecuencia</b>	<b>IM</b>
1	escaseces	161	9.98
2	fatigas	1338	9.8
3	abstinencias	49	9.72
4	austeridades	45	9.55
5	penalidades	641	9.51
6	sufrimientos	1379	9.44
7	estrecheces	172	9.37
8	pobrezas	56	9.15
9	miserias	979	9.07
10	penurias	282	8.94
11	durezas	43	8.91

**Tabla 5: Colocativos en binomios de *privaciones* y (Fuente: Googlebooks Spanish).**

## 5 Conclusiones

En primer lugar conviene recordar que los trabajos sobre binomios, y el nuestro no es una excepción, simplifican el valor del binomio pues en realidad esta estructura nunca deja de ser parte de una secuencia mayor con una función específica.

Hecha esta acotación, creemos que vale la pena explorar los binomios partiendo de principios básicos de la psicolingüística, tomando en cuenta la frecuencia de las palabras que aparecen como  $N_1$  y como  $N_2$ . Aunque haya motivos de toda índole, parece ser que la frecuencia de uso del  $N_1$  tiende a ser más alta que la del  $N_2$ . También hemos mostrado que hay nodos que se especializan en  $N_1$  o en  $N_2$ , y que normalmente los nodos que aparecen como  $N_1$  forman más binomios que los  $N_2$ .

El tamaño y la configuración del corpus son factores capitales en estudios cuantitativos porque no solo determinan las frecuencias, las estadísticas de asociación; las medidas de dispersión sino que nos van a señalar la variabilidad de relaciones conceptuales (secuencia de tiempo, de espacio, causa-efecto, jerarquía, etc). Asimismo, podemos observar fenómenos insospechados como la coincidencia en algunas colocaciones en el inicio y/o el final de la palabra (efecto tina/bañera) y la similitud del número de sílabas, del patrón consonántico, la posición de la sílaba acentuada, coincidencias fonéticas.

En definitiva, creemos que estos datos exploratorios respaldan la tesis de Hoey de que los hablantes registramos el uso de palabras, su contexto, su posición y más aspectos de los que no parecemos estar conscientes. Queda pues diluido el concepto de binomio tal como lo conciben Malkiel (1959) y Almela Pérez (2006), pasando a formar parte de un sistema dinámico de interacciones complejas e impredecibles en que se influyen mutuamente el uso, el procesamiento, el aprendizaje y la estructura de la lengua (Ellis & Frey 2009).

Este trabajo señala que los métodos semiautomáticos generados en una investigación guiada por datos sobre los binomios arrojan resultados ricos y complejos que obligan a replantearnos formas básicas generadas por la introspección.

## 6 Referencias

- Alderson, J. C. (2007). Judging the Frequency of English Words. *Applied Linguistics*, 28(3), 383-409 doi:10.1093/applin/amm024
- Almela Pérez, R. (2006). Binomios (irreversibles) en español. *LEA: Lingüística Española Actual*, 28(2), 135-160.
- Biber, D. (1999). Longman grammar of spoken and written English. Londres: Longman.
- Biber, D., & Barbieri, F. (2007). Lexical bundles in university spoken and written registers. *English for Specific Purposes*, 26(3), 263-286. doi:10.1016/j.esp.2006.08.003
- Cantos, P., & Sánchez, A. (2002). Lexical Constellations: What Collocates Fail to Tell. *International Journal of Corpus Linguistics*, 6(2), 199-228. doi:10.1075/ijcl.6.2.02can
- Corpas Pastor, G. (1996). *Manual de fraseología española*. Madrid: Gredos.
- Corrigan, R., Moravcsik, E. A., Ouali, H., & Wheatley (Eds.). (2009a). *Formulaic language* (Vol. 1). Amsterdam/ New York: John Benjamins Publishing Company.

- Corrigan, R., Moravcsik, E. A., Ouali, H., & Wheatley (Eds.). (2009b). *Formulaic language. (Vol. 2 Acquisition, loss, psychological reality, and functional explanations)*. Amsterdam/New York: John Benjamins Publishing Company.
- Cowie, A. P. (2006). Phraseology. *Encyclopedia of language & linguistics*. Amsterdam: Elsevier.
- Davies, M. (2002). *Corpus del Español: 100 million words, 1200s-1900s*. <http://www.corpusdelespanol.org> [12/032013]
- Davies, M. (2006). A frequency dictionary of Spanish: core vocabulary for learners. Abingdon: Routledge.
- Ellis, N. C. (1998). Emergentism, Connectionism and Language Learning. *Language Learning*, 48(4), 631-664. doi:10.1111/0023-8333.00063
- Ellis, N. C. (2008). The Dynamics of Second Language Emergence: Cycles of Language Use, Language Change, and Language Acquisition. *The Modern Language Journal*, 92(2), 232-249. doi:10.1111/j.1540-4781.2008.00716.x
- Ellis, N. C., & Frey, E. (2009). The psycholinguistic reality of collocation and semantic prosody (2). En R. Corrigan, E. A. Moravcsik, H. Ouali, & Wheatley (eds.), *Formulaic language (Vol. 2 Acquisition, loss, psychological reality, and functional explanations)*. Amsterdam/New York: John Benjamins Publishing Company.
- Evert, S. (2009). Corpora and collocations. En A. Lüdeling & M. Kytö (Eds.), *Corpus linguistics: an international handbook* (Vols. 1-2, Vol. 2, págs. 1212-1248). W. de Gruyter.
- Fauconnier, G., & Turner, M. (2003). *The Way We Think: Conceptual Blending And The Mind's Hidden Complexities*. Basic Books.
- García-Page, M. (1998). Binomios fraseológicos antitéticos. En G. Wotjak (Ed.), *Estudios de fraseología y fraseografía del español actual* (págs. 195-202). Madrid/Frankfurt am Main: Iberoamericana/Vervuert.
- García-Page, M. (2008). Introducción a la fraseología española. Estudio de las locuciones. Barcelona: An-thropos.
- van Geert, P. (2008). The Dynamic Systems Approach in the Study of L1 and L2 Acquisition: An Introduction. *The Modern Language Journal*, 92(2), 179-199. doi:10.1111/j.1540-4781.2008.00713.x
- Gries, S. T., & Stefanowitsch, A. (2007). *Corpora in Cognitive Linguistics: Corpus-Based Approaches to Syntax and Lexis*. Berlin/New York: Walter de Gruyter.
- Hoey, M. (2005). *Lexical priming: a new theory of words and language*. Abingdon: Routledge.
- Hunston, S. (2002). *Corpora in applied linguistics*. Cambridge: Cambridge University Press.
- Hunston, S., & Francis, G. (2000). *Pattern grammar: a corpus-driven approach to the lexical grammar of English*. Amsterdam/New York: John Benjamins Publishing Company.
- Johnson, M. (1987). *The body in the mind: the bodily basis of meaning, imagination, and reason*. Chicago: University of Chicago Press.
- Kilgarriff, Adam, Pavel Rychly, Pavel Smrz, David Tugwell. *The Sketch Engine. Proc EURALEX 2004*, Lorient, France; Pp 105-116, <http://www.sketchengine.co.uk> [12/032013]
- Malkiel, Y. (1959). Studies in irreversible binomials. *Lingua*, 21, 142-155.
- Moon, R. (1998). *Fixed expressions and idioms in English: a corpus-based approach*. Oxford / New York: Clarendon Press.
- Oakes, M. P. (1998). *Statistics for corpus linguistics*. Edinburgh: Edinburgh University Press.
- Real Academia Española: Banco de datos (CORDE) [en línea]. *Corpus diacrónico del español actual*. <<http://www.rae.es>> [12/03/2013]
- Rodríguez Sánchez, I. (2013). Frequency and Specialization in Spanish Binomials N y N. *Procedia - Social and Behavioral Sciences*, 95, 284-292. doi:10.1016/j.sbspro.2013.10.649
- Sinclair, J. M. (1991). *Corpus, concordance, collocation*. Oxford / New York: Oxford University Press.
- Stubbs, M. (1996). *Text and corpus analysis*. Blackwell Publishers.
- Talmy, L. (1988). Force dynamics in language and cognition. *Cognitive Science: A Multidisciplinary Journal*, 12(1), 49-100.
- Tognini-Bonelli, E. (2001). *Corpus linguistics at work*. Amsterdam/New York: John Benjamins Publishing Company.

---

Wray, A. (2002). *Formulaic language and the lexicon*. Cambridge University Press.

Wray, A. (2008). *Formulaic language: pushing the boundaries*. Oxford / New York: Oxford University Press.

Zipf, G. K. (1949). *Human behavior and the principle of least effort: an introduction to human ecology*. Cambridge Mass.: Addison-Wesley Press