

# **LINGÜÍSTICA DE CORPUS Y GRAMÁTICA INDUCTIVA: UNA APROXIMACIÓN**

## **Inmaculada Martínez Martínez**

(Centro Internacional de Estudios Superiores del Español. CIESE-Comillas, España)

[martinezi@fundacioncomillas.es](mailto:martinezi@fundacioncomillas.es)

Fecha de recepción: 9-7-2017 / Fecha de aceptación: 21-12-2017

### **RESUMEN:**

La presente investigación muestra cómo incorporar los corpus lingüísticos a la clase de español de los negocios: mediante la gramática inductiva que estimula en los aprendientes su consciencia lingüística. Ponemos el foco, así, en la conexión entre la concepción del aprendizaje como comprobación de hipótesis y las ideas más recientes conectadas con los corpus y el estudiante como investigador (Johansson, 2009). Si bien es cierto que la Lingüística de corpus goza de un momento de esplendor, no se ha debatido aún la relevancia de cuestiones como la atención y la consciencia, o la formación de hipótesis, que son centrales en el uso de los corpus. El objetivo se alcanza después de apuntalar en el adecuado marco teórico la investigación, para pasar más tarde a la presentación estructurada y ejemplificada de la metodología con la que han de llevarse al aula las muestras de corpus. Al hacerlo, los beneficios didácticos que se alcanzan parecen cruciales: entre otros, la potenciación de destrezas cognitivas, clave para el aprendizaje de la L2; el nuevo rol que adopta el estudiante como investigador de la lengua y la capacidad del docente de ser más autónomo y menos dependiente, por tanto, de los materiales diseñados por otros. El siguiente paso será abordar estudios experimentales que midan la efectividad de los corpus en la enseñanza de la lengua.

**Palabras clave:** Aprendizaje guiado por datos; Lingüística de corpus; Gramática inductiva; Realce del input; Consciencia lingüística.

### **ABSTRACT:**

This paper tries to show how to use linguistics corpus in a Business Spanish course by means of the inductive grammar that encourages language awareness in the learners.

We focus on the connection between the concept of learning as hypothesis testing and recent ideas related with the corpus and the learner as an investigator (Johansson, 2009). Although it is true that Corpus Linguistics enjoy great prestige nowadays, there are still questions –like attention, awareness or hypothesis building– to discuss that are essential for the use of corpora. The goal is achieved by placing our research in the appropriate theoretical framework followed by an introduction of a structured methodology accompanied by examples that show how to make use of a corpus in the classroom. In doing so, the didactic benefits that are reached seem crucial: among others, the strengthening of cognitive skills, key for the second language learning; the new role for the student who is seen now as a language investigator and the teacher's ability of being more autonomous and, therefore, less dependent from the activities designed by others. Experimental studies will be needed hereafter in order to measure the effectiveness of corpora in the teaching of Spanish language.

**Key Words:** Data-driven learning; Corpus linguistics; Inductive grammar; Input enhancement; Language awareness.

## **INTRODUCCIÓN**

Resulta obvio que la lengua es tan vasta y variada que nunca podremos proporcionarles a nuestros aprendientes una descripción abarcable de esta como un todo. Sí podremos, sin embargo, suministrarles pautas o directrices que los guíen en su camino de aprendizaje (Willis & Willis, 1996). En este sentido, una manera útil de acercarse a la lengua puede ser mediante actividades que los impulsen a reflexionar sobre muestras de lengua para trazar sus propias conclusiones sobre cómo funciona el sistema lingüístico. Este tipo de actividades se conocen como actividades de despertar la consciencia (*consciousness raising activities*) y activan en la mente del alumno importantes destrezas cognitivas, clave para su aprendizaje en la segunda lengua (Schmidt, 1990; Fotos, 1994; Autor, 2004).

Por otro lado, usos de la lengua que se pasan por alto o resultan implícitos en un solo texto pueden destacarse y ser más explícitos si manejamos la información intertextual que nos proporcionan los corpus lingüísticos y su evidencia (Willis & Willis, 1996; Hunston, 2002; Tsui, 2005). Así, la técnica básica de análisis de un corpus, las líneas de concordancia, se convierte en una valiosa herramienta que permite trabajar con un gran número de ejemplos de un ítem, presentes en muchos textos y que juntos estos ejemplos aparezcan en su contexto original (Evison, 2010).

La unión de ambos marcos teóricos –la Lingüística de corpus y el Cognitivismo– parece contribuir, por tanto, a una mejora en la calidad de la enseñanza y el aprendizaje de la lengua, en concreto, del discurso profesional del español de los negocios. Ello supone despertar la consciencia lingüística mediante la herramienta de los corpus textuales.

Los beneficios que pueden surgir a partir del trabajo en clase con ambos enfoques parecen indudables: al estimular en los aprendientes la observación y el análisis lingüístico, estamos reforzando su tendencia natural a explorar el sentido de la lengua y a sistematizarla (Willis & Willis, 1996). Los estamos conduciendo, además, a un aprendizaje autónomo a partir de muestras reales de lengua, muestras a las que ellos podrán acudir siempre que lo necesiten para confirmar o refutar sus hipótesis lingüísticas.

La relevancia de cuestiones de índole cognitiva como la atención y la consciencia, el input o la formación de hipótesis está fuera de toda duda. También lo está el hecho de que los citados aspectos resultan centrales en el uso de los corpus para la enseñanza de una lengua extranjera. Sin embargo, llegados a este punto, lo que parece evidente es el hecho de que dichos conceptos aún no se han abordado en profundidad (Johansson, 2009). Tampoco en el caso del español como lengua extranjera o del español de los negocios, el ámbito que aquí se aborda.

El presente artículo aspira a ser una sencilla propuesta para la inclusión en las clases de español de los negocios de la gramática inductiva a través de la herramienta de los corpus lingüísticos; en nuestro caso concreto, a través del Corpus Comillas del Español de los Negocios (en adelante, CORPEN Comillas)<sup>1</sup>, que se encuentra en sus fases iniciales de elaboración.

Se comenzará por sustentar la investigación en el correspondiente marco teórico que una las disciplinas mencionadas: la Lingüística de corpus y la Gramática inductiva. Se mostrará cómo fomentar mediante el uso de corpus, destrezas cognitivas tan importantes como el advertir, el despertar la consciencia o la focalización en la forma, dado que todas ellas suponen un beneficio a medio y largo plazo en la adquisición de la lengua por parte de los aprendientes. Todo ello encontrará su nexo inevitable en el enfoque conocido como 'aprendizaje guiado por datos' (*Data-Driven Learning*), de plena actualidad en el ámbito de la Lingüística

<sup>1</sup> CORPEN Comillas es un corpus específico que tiene como fin servir para obtener las características del español de los negocios en España y América en el periodo comprendido entre los años 2010 y 2020. Contará con 12 millones de formas y lo está llevando a cabo un equipo de investigación del Centro Internacional de Estudios Superiores del Español (CIESE-Comillas).

aplicada. Finalmente, se desglosarán las implicaciones didácticas que todo ello conlleva para la clase de español de los negocios y se apuntarán ciertos comentarios finales y algunas perspectivas de investigación.

## **MARCO TEÓRICO**

### ***Lingüística de corpus***

#### Fundamentos

Recordamos aquí que un corpus es un conjunto amplio de textos de los más diversos tipos, representativos del estado de una lengua, en formato electrónico, codificados de modo que sea posible obtener de ellos los casos de una determinada forma o expresión en la totalidad del conjunto (Sinclair, 2004).

Es la compilación electrónica, estandarizada conforme a unos parámetros definitorios, la que permite la recuperación selectiva de la información. Esta compilación debe ir acompañada de las correspondientes herramientas de aplicación informática que permitan la recuperación posterior de la información para todo investigador que se acerque al estudio de la lengua (Sinclair, 1996).

El acceso a los corpus nos permite, a través de las líneas de frecuencia que más adelante desentrañaremos, la búsqueda de modelos gramaticales y de significados de las palabras en contexto. Aquí subyacen dos principios teóricos que debemos fijar antes de seguir avanzando:

- Si la búsqueda de una forma en el corpus nos arroja sus numerosas ocurrencias junto a otras formas, y gracias a estas múltiples co-ocurrencias, podemos deducir su comportamiento y significado, podremos concluir que hay una indistinción entre el modelo y su significado, ya que distinguir entre significados es distinguir entre modelos de uso (Sinclair, 1991; Hunston, 2002).
- Si el significado de una forma en el corpus puede deducirse de su co-aparición con otras formas en los textos, podremos concluir que el significado de las palabras se deduce por el contexto (O'Keeffe, McCarthy & Carter, 2007). Es, precisamente, este énfasis en el contexto

lo que provoca el cambio terminológico de 'lenguajes con fines específicos' a 'discursos profesionales', desde la aproximación pragmática y sociocognitiva (López Ferrero, 2002). De hecho, los corpus pueden promover una mejor comprensión de la relación entre los modelos gramaticales y sus contextos de uso (O'Keeffe et al., 2007; Parodi, 2008).

### Aplicaciones

Son numerosos los ámbitos en los que los investigadores de la Lingüística aplicada a la enseñanza de segundas lenguas, pueden volcar sus investigaciones basadas en las evidencias que obtienen de su trabajo con los corpus: los diccionarios, las gramáticas, la evaluación, la investigación, el diseño de syllabus, las actividades de clase, los libros de texto o el material de enseñanza. De ellos, al menos las actividades de clase, los libros de texto y los materiales de enseñanza, entroncarán directamente con el propósito de esta investigación.

Si el profesor –no solo el investigador- puede usar los corpus para crear sus propios materiales, estamos apuntando a una dirección distinta a la que consideraba al profesor como un destinatario pasivo o un implementador de los syllabus y los métodos de otros. El docente de lenguas es, bajo la perspectiva que aquí se sostiene, un creador activo –y, por lo tanto, autónomo- de sus propios materiales y actividades de clase (Nunan, 2004; Svalberg, 2016). Estos materiales los elaborará a partir de la evidencia que le muestran los corpus y no a partir de sus propias intuiciones o de las intuiciones de otros (O'Keeffe et al., 2007).

### Recursos

Entre las herramientas de los corpus disponibles para trabajar la gramática y el léxico de manera inductiva en clase, nos centraremos en las listas de frecuencia y en las líneas de concordancia. Es importante señalar que, mediante este acercamiento, la lengua se convertirá no solo en objetivo sino también en medio de aprendizaje (Svalberg, 2016).

- Listas de frecuencia

Constituyen una relación de todas las palabras del corpus en orden descendente con respecto a su frecuencia de aparición (Gilquin & Granger, 2010). Mostramos en la Figura 1 la apariencia que tiene la lista de frecuencia del CORPEN Comillas en el estado actual en el que se encuentra:

N	Word	Freq.	%	Texts	%ispersion
1	#	320	9,85	15	100,00
2	DE	184	5,66	15	100,00
3	EL	155	4,77	15	100,00
4	EN	126	3,88	15	100,00
5	LOS	107	3,29	15	100,00
6	LA	105	3,23	15	100,00
7	UN	98	3,02	15	100,00
8	Y	78	2,40	15	100,00
9	LAS	66	2,03	15	100,00
10	DEL	60	1,85	15	100,00
11	A	58	1,79	15	100,00
12	Á	51	1,57	15	100,00
13	IBEX	46	1,42	15	100,00
14	POR	41	1,26	15	100,00
15	SE	39	1,20	15	100,00
16	QUE	37	1,14	15	100,00
17	MERCADOS	30	0,92	15	100,00
18	CON	29	0,89	15	100,00
19	MAYORES	28	0,86	15	100,00
20	REGISTRAN	28	0,86	15	100,00
21	PUNTOS	26	0,80	15	100,00
22	LO	23	0,71	15	100,00
23	BOLSA	20	0,62	15	100,00
24	CIERRE	20	0,62	15	100,00
25	SUBIDAS	20	0,62	15	100,00

**FIGURA 1. Lista de las 25 palabras más frecuentes en CORPEN Comillas.**  
**Fuente: elaboración propia**

Resultan interesantes estas listas, pues se convierten en una herramienta más para aprender sobre la lengua. O'Keeffe et al. (2007) destaca la importancia de este recurso al señalar que una lista de este tipo será un importante índice del léxico que debería incluir un syllabus básico, un manual o una gramática. Para los profesores parece esencial, en definitiva, conocer qué palabras o expresiones se repiten más, de manera que estén seguros de que enfatizan en sus clases las formas frecuentes (Chambers, 2010).

Entre las diversas aplicaciones didácticas de esta herramienta, resulta útil comparar, por ejemplo, dos listas de frecuencia, la de un corpus general y la de otro específico, con el objetivo de sensibilizar al aprendiente sobre la especificidad léxica de los discursos profesionales. Comparar, así mismo, pequeños corpus especializados de, al menos, dos lenguas, puede hacer que el estudiante tenga una mirada más profunda al uso intercultural de la lengua en contextos específicos (O'Keeffe et al.,

2007). Por último, pueden usarse estas formas como punto de partida para los análisis de concordancias (Gilquin & Granger, 2010) que trataremos a continuación.

- Líneas de concordancias

Es la técnica analítica básica de un corpus y la más conocida. Es valiosa porque permite, por un lado, que se visualice un gran número de ejemplos de una forma y, por otro lado, que tales ejemplos aparezcan juntos en su contexto original (Evison, 2010). Cada forma es indexada y, algo muy importante, se incluye la referencia al lugar de ocurrencia en el texto correspondiente (Sinclair, 1991).

La apariencia visual es la que se muestra en la Figura 2, con la palabra objeto de búsqueda destacada en el centro de la línea, normalmente realizada en negrita por el software de la aplicación<sup>2</sup>:

N	Concordance	Set	Tag	Word #	Sen	Sen	Par	Par	Hea	Hea	Sec	Sec	File
1	, respectivamente. Por otro lado, <b>Bolsas</b> y Mercados (-0,37%) registra			137	8	16%	0	84%			0	84%	PE2014_0003.x
2	de estrÃ©s a la banca europea, <b>las bolsas</b> cierran con signo mixto pero			25	2	69%	0	10%			0	10%	PE2014_0009.x
3	las alzas registradas en su inicio, <b>las bolsas</b> europeas cerraban en negativo.			82	4	73%	0	40%			0	40%	PE2014_0010.x
4	de cierre de mercados Cautela en <b>las bolsas</b>			11	2	53%	0	5%			0	5%	PE2014_0008.x
5	Pese a la escalada registrada por <b>las bolsas</b> europeas el pasado viernes, al			28	3	28%	0	12%			0	12%	PE2014_0008.x
6	centran la atenciÃ³n en <b>las bolsas</b> europeas. Los mercados			30	2	97%	0	12%			0	12%	PE2014_0012.x
7	crecimiento en EE.UU. impulsa a <b>las bolsas</b>			15	3	17%	0	9%			0	9%	PE2014_0013.x
8	en EE.UU. daban un giro a <b>las bolsas</b> que se recuperaban de las caÃ			52	5	15%	0	31%			0	31%	PE2014_0013.x
9	de mercados Cierre mixto en <b>las bolsas</b>			12	2	37%	0	5%			0	5%	PE2014_0012.x
10	<b>Las bolsas</b> europeas cerraban al alza tras			14	2	33%	0	7%			0	7%	PE2014_0002.x
11	de cierre de mercados Subidas en <b>las bolsas</b>			11	2	37%	0	7%			0	7%	PE2014_0003.x
12	favorables desde Asia y <b>las bolsas</b> europeas no dudaban en			29	2	52%	0	13%			0	13%	PE2014_0001.x
13	ContinÃ³an las subidas en <b>las bolsas</b>			13	2	31%	0	6%			0	6%	PE2014_0002.x
14	cierre de mercados Ascensos en <b>las bolsas</b>			11	2	48%	0	7%			0	7%	PE2014_0006.x
15	35 encabezaba las subidas en <b>las bolsas</b> gracias a la banca y a			31	3	34%	0	21%			0	21%	PE2014_0006.x
16	cierre de mercados Descensos en <b>las bolsas</b>			11	2	43%	0	6%			0	6%	PE2014_0004.x
17	del Ibex 35, las compaÃ±Ã­as IAG y <b>Bolsas</b> y Mercados registran las			161	8	63%	0	78%			0	78%	PE2014_0014.x
18	. Por otro lado, FCC (-4,93%) y <b>Bolsas</b> y Mercados (-4,11%) registran			151	8	23%	0	85%			0	85%	PE2014_0004.x

**FIGURA 2. Líneas de concordancias para la búsqueda de la forma “bolsas”.**

**Fuente: elaboración propia**

Nos acercamos a continuación a la Gramática inductiva y a las operaciones cognitivas que la conforman para descubrir posteriormente cómo podremos abordar dichos procesos en clase desde el enfoque conocido como Aprendizaje guiado por datos.

<sup>2</sup> En nuestro caso el software empleado para el trabajo con corpus en clase es *WordSmith Tools* (<http://www.lexically.net/wordsmith/>)

## ***Gramática inductiva***

Nos recuerda O'Keeffe et al.(2007) que en el ámbito de los negocios para el cual se crea CORPEN Comillas, tanto como aprender a usar una lengua de manera apropiada, deberían ser primordiales la observación detallada y profunda, así como el despertar la consciencia de las formas lingüísticas. Los corpus nos capacitan para llevar a cabo tal análisis en el contexto reflexivo de la clase.

Analizaremos, en primer lugar, la concienciación formal, proceso que da nombre al enfoque. Abordaremos más tarde el realce del input que el docente realiza para desarrollar en el aprendiente las operaciones cognitivas del advertir, la consciencia y la focalización en la forma. Todos estos procesos serán posibles a través de las muestras de corpus.

### Concienciación formal<sup>3</sup> (*Consciousness Raising*)

La concienciación formal parece convertirse en un proceso esencial para la adquisición de la lengua, pues, al desarrollar en el aprendiente conocimiento sobre un rasgo lingüístico, tiende a ser más consciente de dicho rasgo en el input comunicativo que reciba (Fotos, 1994). Se trataría, en suma, de despertar la consciencia del estudiante sobre los elementos lingüísticos problemáticos en la L2 y hacerlo a través de la interacción (Svalberg, 2016), es decir, a partir de actividades que permitan el intercambio comunicativo dentro del aula.

Estamos, pues, ante la teoría según la cual manipular o hacer resaltar parte del input de una forma artificial, afectará de forma positiva al conocimiento del aprendiente y, por lo tanto, a su comportamiento lingüístico (Sharwood Smith, 1993). La lógica de la argumentación es que, al exponer al alumno a la L2 de forma realizada, estamos abordando una batería completa de diferentes mecanismos de procesamiento lingüístico que es necesario considerar.

La concienciación formal desarrolla esa fase intermedia entre el conocimiento explícito –suscitado por la instrucción formal- y el implícito –o eventual adquisición de la forma lingüística-, paso previo a su producción posterior en la comunicación (Ellis, 1995). Parece claro, por tanto, que todas las operaciones que realice el aprendiente,

<sup>3</sup> Varios son los términos que se emplean para designar este enfoque en la bibliografía española: “sensibilización lingüística” (Martín Peris, 1998: 24), “atención a la forma” (Martín Peris, 1998: 27) o “instrucción formal” (Coronado, 1998: 83). “Concienciación formal” parece reflejar con mayor transparencia los dos aspectos clave: la consciencia y la forma lingüística sobre la que se ejercita dicha consciencia.



cuando lo exponamos a muestras reales de corpus y elaboremos actividades sobre ellas, pasarán necesariamente por la concienciación formal.

### Realce del input (*Input Enhancement*)

Es preciso señalar que la consciencia sobre fragmentos del caudal lingüístico es solo posible porque previamente el profesor ha resaltado esas formas concretas que quiere que sean advertidas, focalizadas y, finalmente, adquiridas por el aprendiente. Esta operación llevada a cabo por el docente es denominada realce del input, o más recientemente, "realce visible del input" (*visual input enhancement*) y "realce textual del input" (*textual enhancement of input*) (Svalberg, 2016: 403).

El profesor puede, a partir de este material ya destacado, preparar, agrupar, dar color o subrayar aquellas formas sobre las que desea focalizar la atención de sus estudiantes (Sripicharn, 2010).

### Advertir (*Noticing*)

Schmidt (1990) sostiene que el aprendizaje subliminal –el llevado a cabo sin ninguna colaboración de la consciencia- es absolutamente imposible y que advertir es condición necesaria para convertir el input en *intake*<sup>4</sup>. Así, el aprendizaje incidental – en el que sí incide o interviene la consciencia- es tanto posible como efectivo, cuando las necesidades de una tarea centran la atención en lo que va a ser aprendido por parte del aprendiente de la L2.

Ahora bien, solo el advertir no parece suficiente. Para que el procesamiento de la forma sea lo bastante firme y asegure la retención, deben darse otras condiciones que quizá hayan de pasar por la consciencia y la atención a la forma.

### Consciencia (*Awareness*)

El concepto de consciencia parece situarse en un punto intermedio entre el input advertido y el input comprendido (Gass, 1988). Estamos ante un concepto de plena vigencia<sup>5</sup> que integra, entre otras dimensiones, el conocimiento sobre la

<sup>4</sup> Con input nos referimos a la información percibida, mientras que intake es el término empleado para designar la información que es advertida por el aprendiente.

<sup>5</sup> La consciencia lingüística (*Language Awareness*), abarca, no solo la estructura de la lengua y el vocabulario, sino también, cómo la lengua funciona social y culturalmente. Su actualidad se muestra a través de una Asociación para la

lengua y no solo la habilidad para usarla (Svalberg, 2016). De ahí que también se emplee con frecuencia el término 'consciencia lingüística'. En este componente cognitivo de la consciencia, que incluye la percepción consciente y la sensibilidad lingüística, nos posicionamos a la hora de defender el trabajo con corpus en el aula.

### Focalización en la forma (*Focus on Form*)

Schmidt (1990) desarrolla la tesis de que la noción de consciencia es útil porque se vincula a otros conceptos relacionados como el de la atención y la atención selectiva y esta última es, precisamente, la que posibilita la focalización en la forma.

Este proceso de atención selectiva es uno de los más importantes de la concienciación formal y es fácil deducir su sintonía con las muestras de corpus, pues a partir de ellas el docente puede dirigir la observación de los alumnos sobre determinadas formas y sobre determinadas conexiones forma-significado. Y lo más importante: al hacerlo, les provee de herramientas que les ayudarán a reconocer esas estructuras lingüísticas en su interlengua y a darse cuenta de que necesitan modificación.

Se pretende capitalizar, así, las propias tendencias analógicas del aprendiz y forzar con ello un cambio en su competencia. Esta táctica que se adopta asume el modelo de comprobación de hipótesis y conecta de forma estrecha con las ideas más recientes en torno a los corpus y al estudiante como investigador (Johansson, 2009).

Por último, técnicas como el señalamiento (*saliency*), la redundancia (*redundancy*) y la frecuencia de presentación de cláusulas marcadas (*markedness*), son técnicas presentes en el trabajo con corpus y actividades todas ellas definitorias de la concienciación formal.

## **APRENDIZAJE GUIADO POR DATOS**

### ***Definición y objetivos***

Todo el marco teórico arriba expuesto encuentra su concreción en la clase de español de los negocios a través del Aprendizaje guiado por datos (*Data-Driven*

---

Consciencia Lingüística (Association for Language Awareness, ALA), nacida en 1994 y que tuvo entre sus fundadores a Leo Van Lier. Sostiene una página web (<http://www.languageawareness.org>) y una publicación (ALA Journal).

*Learning, DDL*) y de la herramienta principal para su puesta en práctica, el Corpus del Español de los Negocios.

El Aprendizaje guiado por datos es una aplicación del Aprendizaje de la Lengua Asistido por Ordenador (ALAO), que incluye el uso de los corpus lingüísticos para la adquisición de la lengua extranjera. En dicho aprendizaje los aprendientes experimentan el uso de un corpus a través de tareas guiadas o a través de materiales basados en evidencias del corpus, como las líneas de concordancias (Cruz Piñol, 2012).

En definitiva, consiste en usar las herramientas y técnicas de los corpus lingüísticos con objetivos pedagógicos (Gilquin & Granger, 2010). El objetivo de este enfoque no es otro que el de poner a disposición de los aprendientes muestras de lengua auténtica para el aprendizaje de la gramática de manera inductiva, sin reglas y a través de la reflexión (Larsen-Freeman & Long, 1994).

Se trata de un proceso que enfrenta al aprendiente, tan directamente como sea posible, con los datos reales de la lengua y hace de él un investigador lingüístico (Johns, 2002), "un segundo Sherlock Holmes", en palabras de Johns (1997:101). O un "viajero", en palabras de Sripicharn, en el sentido de que los aprendientes actúan como viajeros que exploran un corpus para encontrar posibles respuestas a sus búsquedas lingüísticas (Sripicharn, 2010: 372).

Por todo ello, debemos señalar que la inclusión –y explotación– de los corpus en la clase de ENE, no solo realza el papel de la gramática en la adquisición de una segunda lengua. También nos permite acercarnos a su comprensión y uso a través del importante mecanismo cognitivo de la inducción por parte del estudiante. Para ello debemos enfocarles antes la lupa sobre la gramática que queramos trabajar en clase. Y lo haremos, como ya se ha señalado, al realzar el input colocando banderas sobre él<sup>7</sup> (Sharwood Smith, 1993).

Este realce lo llevaremos a cabo sobre los componentes que la Lingüística de corpus nos haya revelado previamente como fundamentales para nuestra didáctica específica. En nuestro caso concreto, una buena pedagogía de la lengua de los negocios debe incluir áreas como la deixis personal, la modalidad, la protección o la vaguedad (O'Keeffe et al., 2007).

Este acercamiento inductivo, en el cual el estudiante juega un papel activo en el proceso de aprendizaje, es la esencia del Aprendizaje guiado por datos. Y tiene mucho

<sup>7</sup> Traducción literal y sumamente visual de la expresión acuñada por Sharwood Smith (1993) "putting flags in the input".

que ver con las corrientes actuales de la pedagogía, que hacen a los estudiantes más activos, reflexivos y autónomos en su aprendizaje (Chambers, 2010).

### **Tipología de ejercicios**

Las actividades en las que se concreta lo visto líneas más arriba, pueden localizarse en un rango que va desde las más dirigidas por el profesor hasta las llevadas a cabo por el propio aprendiente. Las primeras serán adaptaciones a partir del corpus, por lo tanto acceso indirecto a este, y las segundas serán ya consultas personales al corpus, por lo tanto acceso directo (McCullough, 2001; Chambers, 2010)<sup>8</sup>. Los docentes pueden usar un proceso de andamiaje como el que aquí se muestra, en el que los estudiantes vayan progresivamente familiarizándose con las herramientas y la metodología de este enfoque (Granath, 2009):

Como punto de partida, puede ofrecérseles a los estudiantes las líneas de concordancia pero editadas de forma que contengan frases completas. En ellas, la palabra buscada o nodo aparecerá resaltada en negrita o en un color diferente, lo cual les ayudará a detectar el comportamiento de dicha palabra en contexto, guiados en todo momento por el profesor.

Otra manera de visualizar la información puede ser idéntica a la que aparece tras la búsqueda en el programa de concordancias, pero pidiendo al estudiante que focalice de manera consciente su investigación. En esta segunda fase se le puede pedir que indague sobre las combinaciones que se repiten con mayor insistencia en torno a la forma realzada. Así, por ejemplo, en las líneas de concordancia mostradas en la Figura 3, el aprendiente descubrirá la frecuencia de aparición de la combinación "ascensos/descensos en las bolsas".

<sup>8</sup> Hay autores que prefieren referirse a esta distinción como acercamiento débil (*soft*) y fuerte (*hard*) (Gabrielatos, 2005) y otros prefieren llamarlo acercamiento pasivo (*hands-off*) o activo (*hands-on*) (Bolton, 2010). En Frankenberg-García (2016: 391).

N	Concordance	Set	Tag	Word #	Sen	Sen	Par	Par	Hea	Hea	Sec	Sec	File
1	, respectivamente. Por otro lado, Bolsas y Mercados (-0,37%) registra			137	8	16%	0	84%			0	84%	PE2014_0003.x
2	de estr�s a la banca europea, las bolsas cierran con signo mixto pero			25	2	69%	0	10%			0	10%	PE2014_0009.x
3	las alzas registradas en su inicio, las bolsas europeas cerraban en negativo.			82	4	73%	0	40%			0	40%	PE2014_0010.x
4	de cierre de mercados Cautela en las bolsas			11	2	53%	0	5%			0	5%	PE2014_0008.x
5	Pese a la escalada registrada por las bolsas europeas el pasado viernes, al			28	3	28%	0	12%			0	12%	PE2014_0008.x
6	centran la atenci�n en las bolsas europeas. Los mercados			30	2	97%	0	12%			0	12%	PE2014_0012.x
7	crecimiento en EE.UU. impulsa a las bolsas			15	3	17%	0	9%			0	9%	PE2014_0013.x
8	en EE.UU. daban un giro a las bolsas que se recuperaban de las ca�			52	5	15%	0	31%			0	31%	PE2014_0013.x
9	de mercados Cierre mixto en las bolsas			12	2	37%	0	5%			0	5%	PE2014_0012.x
10	Las bolsas europeas cerraban al alza tras			14	2	33%	0	7%			0	7%	PE2014_0002.x
11	de cierre de mercados Subidas en las bolsas			11	2	37%	0	7%			0	7%	PE2014_0003.x
12	favorables desde Asia y las bolsas europeas no dudaban en			29	2	52%	0	13%			0	13%	PE2014_0001.x
13	Contin�an las subidas en las bolsas			13	2	31%	0	6%			0	6%	PE2014_0002.x
14	cierre de mercados Ascensos en las bolsas			11	2	48%	0	7%			0	7%	PE2014_0006.x
15	35 encabezaba las subidas en las bolsas gracias a la banca y a			31	3	34%	0	21%			0	21%	PE2014_0006.x
16	cierre de mercados Descensos en las bolsas			11	2	43%	0	6%			0	6%	PE2014_0004.x

**FIGURA 3. Co-apariciones del t rmino "bolsas".**

**Fuente: elaboraci n propia**

Tambi n puede ped rsele al estudiante que en la misma b squeda observe con atenci n las formas m s alejadas de la palabra clave para llegar, as , inductivamente, a descubrir la combinatoria de los verbos que acompa an frecuentemente al t rmino "bolsas": "registrar, cerrar, recuperar, etc."

El profesor puede tambi n seleccionar inicialmente estas l neas de concordancia y agregarles un conjunto de preguntas para guiar al alumno a la hora de advertir la informaci n relevante y despertar as  su consciencia (Hunston, 2002; Sripicharn, 2010). Las cuestiones como las que se muestran en la Figura 4 facilitar n en el estudiante su proceso de familiarizaci n con la metodolog a y estimular n en  l la generaci n y comprobaci n de hip tesis.

Modelos de preguntas:

1.  Cu l es la forma que queremos analizar?
2.  A qu  n mero pertenece, singular o plural?
3.  Qu  sustantivos aparecen antes de esta forma?  Qu  otros tipos de palabras aparecen antes de "bolsas"?
4.  Qu  formas verbales se construyen con esta forma? Re nelas todas y busca similitudes
5.  Qu  expresiones m s frecuentes forma la palabra "bolsas"?

**FIGURA 4. Ejemplo de consciencia guiada por el profesor.**

**Fuente: elaboraci n propia**

A continuaci n, con otro modelo de ejercicio como el que aparece en la Figura 5, se puede emplear la evidencia negativa –o errores del propio estudiante– para

focalizar sobre ellos la búsqueda en el corpus y, por tanto, la consciencia sobre la forma localizada.

Tarea: las palabras o frases subrayadas contienen errores o usos inapropiados. Busca algunas líneas de concordancia y usa la información obtenida para sugerir una versión revisada de cada frase. Te proponemos las palabras de búsqueda:

Versión original	Palabras de búsqueda	Versión revisada
“La <u>mejor</u> cantidad de ventas se produce en el primer trimestre”	“ <b>mayor</b> ”	
“El volumen de ventas ha subido y por eso ahora los datos son <u>mayores</u> ”	“ <b>mejor</b> ”	
“La <u>calidad</u> es la principal característica de nuestros productos”	“ <b>calidad</b> ”	
“Nuestros principales <u>calidades</u> son la puntualidad y los buenos precios”	“ <b>calidad</b> ”	

**FIGURA 5. Consciencia del estudiante sobre sus propios errores.**

**Fuente: elaboración propia**

Por último, y en un estadio superior de aprendizaje, se le conduce al estudiante a que formule por sí mismo cuestiones lingüísticas y que intente darles respuesta mediante la consulta directa a muestras de corpus. Así puede indagar sobre el uso de la determinación que advierte en las búsquedas anteriores: el porqué en determinados contextos decimos “las bolsas”, “los mercados”, “las subidas”... y no \*unas bolsas, \*unas subidas, \*unos mercados.

Como hemos podido comprobar, los corpus son una manera de realzar y focalizar el input para el estudiante (Johansson, 2009), de forma que este lleve a cabo toda clase de procesos cognitivos que estimulan su aprendizaje.

## **IMPLICACIONES DIDÁCTICAS**

Los beneficios didácticos que se derivarían de la implementación de los corpus con un tratamiento inductivo en la clase de español han recorrido transversalmente este artículo. Procede ahora su sistematización y la incorporación de otros que aún no se han mencionado, en aras de una mayor claridad conceptual:

### ***La perspectiva del aprendiente***

El alumno adopta bajo este enfoque un papel más activo e independiente, más autónomo (O'Keeffe et al. 2007), al dirigir sus búsquedas en el corpus hacia las carencias encontradas en su propia interlengua y controlar, así, su propio aprendizaje. En este sentido, las ideas más interesantes sobre el uso de los corpus en la enseñanza de la lengua tienen que ver con la noción del estudiante como investigador (Johns, 2002). El aprendizaje efectivo de la lengua es, en sí mismo, una forma de investigación lingüística: las concordancias facilitan dicho proceso mediante estrategias de aprendizaje inductivas. En particular, las estrategias para percibir similitudes y diferencias y las de formación y comprobación de hipótesis (Johansson, 2009).

### ***La perspectiva del docente***

El profesor se convierte aquí en un activo diseñador de sus propios materiales que adapta en todo momento al considerar el perfil y el ritmo de aprendizaje de sus propios estudiantes. Será también él, por tanto, más autónomo (Svalberg, 2016). Además, su papel cambia, pues ya no es él solo la única fuente de conocimiento en la lengua objeto: es un facilitador que ayuda a interpretar los datos y que da consejos sobre cómo buscar mejor en el corpus. Todo ello obligará al docente a adoptar un acercamiento descriptivo, más que prescriptivo, a la lengua, todo un reto para él (Chambers, 2010).

### ***La perspectiva del aprendizaje***

El aprendizaje así concebido resulta más motivador, al incluir un elemento de descubrimiento, lo cual redundará en una mayor adquisición (Johansson, 2009). Algunos autores como Svalberg (2016: 403) hablan en este sentido de "compromiso con el aprendizaje" (*Engagement With Learning, EWL*). Con él destacan el acuerdo tácito en el que el estudiante debe involucrarse: se asume en este sentido que la interacción con otros –compañeros y profesores– facilitará esta construcción consciente del conocimiento lingüístico.

### ***La perspectiva de la metodología***

Los materiales que se emplean representan muestras de lengua auténtica y no lengua artificial o inventada, con el estímulo que ello supone para los estudiantes (Chambers, 2010). Puede desencadenar, además, el que los creadores de materiales para la enseñanza del español de los negocios comiencen a incorporar los datos de los corpus lingüísticos, como ya están llevando a cabo en el caso del inglés en contextos específicos (O'Keeffe et al. 2007).

Por otro lado, las ventajas que entraña la técnica se conforman a medio-largo plazo. Más allá de la adquisición de un ítem, puede tener efectos a largo plazo sobre el desarrollo de destrezas cognitivas reutilizables en su proceso de aprendizaje (Gilquin & Granger, 2010).

Junto a ello, todas las directrices del enfoque están orientadas a partir de un acercamiento de abajo hacia arriba (*bottom-up*)<sup>9</sup>, es decir, extraído de la propia realidad de la clase para llegar a las conclusiones pedagógicas pertinentes: entre ellas, que el lenguaje observado es un objeto de estudio más válido que la intuición; que es muy importante la frecuencia de aparición de un ítem lingüístico y que un corpus puede producir resultados que pueden extrapolarse a categorías más generales de la lengua (Hunston, 2002).

Por último, las búsquedas en el corpus estimulan la reflexión y encajan bien con la concienciación formal y, por tanto, con el entrenamiento en la inferencia: los aprendientes pueden adquirir un número de destrezas cruciales como el predecir, el observar y el advertir, entre otras.

Bajo esta perspectiva, la creatividad se introduce también en la clase de español mediante el uso de los corpus lingüísticos. El objetivo es desarrollar en los aprendientes la consciencia de las propiedades y funciones de los patrones lingüísticos que trabajan creativamente en la comunicación diaria (O'Keeffe et al. 2007).

## **CONCLUSIONES**

La Lingüística de corpus es una metodología que está en plena ebullición (Parodi, 2008) y que reporta importantes ventajas como las arriba analizadas. Dichas

<sup>9</sup> Se trata de los dos enfoques para la descripción de los géneros a partir de corpus que propone Biber (Biber *et al.* 2007). Uno de estos enfoques opera de arriba abajo (*top-down*) y el otro, en la dirección opuesta (*bottom-up*). En este último caso utiliza un *software* específico para después analizar los textos y extraer sus características principales, las que servirán de base y de evidencia para las especificaciones curriculares.



ventajas nos incitan a incluirlo como un complemento en la clase del español de los negocios.

Es cierto también que muchos detractores invocan obstáculos importantes, entre otros el tiempo que consume para el profesor preparar los materiales y para el estudiante familiarizarse con la técnica. Así mismo el hecho de que los profesores pueden no tener la formación suficiente para manejar los programas de búsqueda (Frankenberg-Garcia, 2016), o el hecho de el estudiante pueda realizar inferencias equivocadas (Gilquin & Granger, 2010). Pese a ello, estamos ante un enfoque prometedor que no debería abandonarse (Chambers, 2010); entre otras razones, porque lleva a la clase una lengua real, porque motiva a los estudiantes por el elemento de descubrimiento que encierra, y porque desarrolla destrezas cognitivas esenciales.

Sin embargo, es preciso admitir que existe todavía una brecha entre la Lingüística de corpus y la Didáctica de la lengua (McCarthy & O'Keeffe, 2010) y que se necesitan estudios empíricos sobre la efectividad de los corpus en la enseñanza de la lengua (Johansson, 2009). No debemos dejar de reclamar prudencia en este sentido. También en el hecho de que los corpus con finalidad didáctica son un apoyo, pues ni sustituyen a la comunicación ni pueden reemplazar al profesor (Johansson, 2009; Svalberg, 2016).

Su inclusión en la clase de español de los negocios, sin embargo, fortalecerá los procesos naturales de adquisición de la lengua en la medida en que hará más sencillo a los aprendientes advertir y experimentar secuencias repetidas de uso (Johansson, 2009). No debemos olvidar, por último, que las mejores condiciones para el aprendizaje de una lengua tienen lugar cuando se produce un acercamiento combinado, que incluya lo implícito y lo explícito, lo inductivo y lo deductivo en el propio estudiante.

## **REFERENCIAS BIBLIOGRÁFICAS**

Chambers, A. (2010). What is data-driven learning? En A. O'Keeffe & M. McCarthy (Eds.), *The Routledge Handbook of Corpus Linguistics* (pp. 345-358). Nueva York: Routledge.

Cruz Piñol, M. (2012). *Lingüística de corpus y enseñanza del español como 2/L*. Madrid: Arco Libros.

Ellis, R. (1995). *The Study of Second Language Acquisition*. Oxford: Oxford University Press.

Evison, J. (2010). What are the basis of analysing a corpus? En A. O'Keeffe & M. McCarthy (Eds.), *The Routledge Handbook of Corpus Linguistics* (pp. 345-358). Nueva York: Routledge.

Fotos, S. (1994). Integrating Grammar Instruction and Communicative Language Use Through Grammar Consciousness-Raising Tasks. *TESOL Quarterly* 28 (2), 323-351.

Frankenberg-Garcia, A. (2016). Corpora en ELT. En G. Hall (Ed.) *The Routledge Handbook of English Language Teaching* (pp. 383-398). Nueva York: Routledge.

Gass, S. (1988). Integrating research areas: a framework for second language studies. *Applied Linguistics* 9. 198-217.

Gilquin, G. & Granger, S. (2010). How can data-driven learning be used in language teaching? En A. O'Keeffe & M. McCarthy (Eds.), *The Routledge Handbook of Corpus Linguistics* (pp. 359-370). Nueva York: Routledge.

Granath, S. (2009). Who benefits from learning how to use corpora? En K. Aijmer (Ed.), *Corpora and Language Teaching* (pp.47-65). Amsterdam: John Benjamins Publishing Company.

Hunston, S. (2002), *Corpora in Applied Linguistics*. Cambridge: Cambridge University Press.

Johansson, S. (2009). Some thoughts on corpora and second language acquisition. En K. Aijmer (Ed.), *Corpora and Language Teaching* (pp.33-44). Amsterdam: John Benjamins Publishing Company.

Johns, T. (1997). Contexts: The background, development and trialling of a Concordance-based CALL Program. En A. Wichmann, S. Fligelstone, T. McEnery, & G. Knowles, (Eds.), *Teaching and language corpora* (pp. 100-115). London: Longman.

Johns, T. (2002). Data-driven learning: the perpetual challenge. En B. Kettemann & G. Marko (Eds.), *Teaching and Learning by Doing Corpus Linguistics* (pp.107-117). Amsterdam: Rodopi.

Larsen-Freeman, D. & Long, M. H. (1994). *Introducción al Estudio de la Adquisición de Segundas Lenguas*. Madrid: Gredos.

López Ferrero, C. (2002). Aproximación al análisis de los discursos profesionales. *Revista Signos*, 35(51-52), 195-215. Recuperado el 6 de abril, 2016, de <https://dx.doi.org/10.4067/S0718-09342002005100013>

Martínez Martínez, I. (2004). *Nuevas perspectivas en la Enseñanza-Aprendizaje de ELE para japoneses: la concienciación formal*. Tesis doctoral [en línea]. Disponible en: <http://eprints.ucm.es/4338/>. Madrid: Universidad Complutense de Madrid, Servicio de Publicaciones.

McCarthy, M. & O'Keeffe, A. (2010). Historical perspective: what are corpora and how have they evolved? En A. O'Keeffe & M. McCarthy (Eds.), *The Routledge Handbook of Corpus Linguistics* (pp. 359-370). Nueva York: Routledge.

McCullough, J. L. (2001). Los usos de los corpórea de textos en la enseñanza de lenguas. En M. Trenchs Parera (Ed.), *Nuevas tecnologías para el autoaprendizaje y la didáctica de lenguas* (pp. 125-140). Barcelona: Milenio.

Nunan, D. (2004). *Task-Based Language Teaching*. Cambridge: Cambridge University Press

O'Keeffe, A., McCarthy, M., & Carter, R. (2007). *From Corpus to Classroom: Language use and Language teaching*. Cambridge: Cambridge University Press.

Parodi, G. (2008). "Lingüística de corpus: una introducción al ámbito". *RLA. Revista de lingüística teórica y aplicada*, 46 (1), 93-119. Recuperado el 16 de octubre, 2016, de <https://dx.doi.org/10.4067/S0718-48832008000100006>

Schmidt, R. (1990). The Role of Consciousness in Second Language Learning. *Applied Linguistics*, 11 (2), 127-158.

Sharwood Smith, M. (1993). Input Enhancement in Instructed SLA. Theoretical Bases. *Studies in Second Language Acquisition* 15, 165-179.

Sinclair, J. (1991). *Corpus, concordance, collocation*. Oxford: Oxford University Press.

Sinclair, J. (1996). The search for units of meaning. *Textus* 9 (1), 75-106.

Sinclair, J. (2004). *Trust the Text: Language, Corpus and Discourse*. London: Routledge.

Sripicharn, P. (2010). How can we prepare learners for using language corpora? En A. O´Keeffe & M. McCarthy (Eds.), *The Routledge Handbook of Corpus Linguistics* (pp. 371-384). Nueva York: Routledge.

Svalberg, A. M-L (2016). Language Awareness. En G. Hall (Ed.), *The Routledge Handbook of English Language Teaching* (pp. 399-412) Nueva York: Routledge.

Tsui, A.B.M. (2005). ESL Teacher´s Questions and Corpus Evidence. *International Journal of Corpus Linguistics*, 10 (3), 335-356.

Willis, J. & Willis, D. (1996). Consciousness-raising activities in the classroom. En J. Willis & D. Willis (Eds.), *Challenge and change in Language Teaching*. London: Heinemann.