



La utilización de métodos de aprendizaje *data-driven* en la enseñanza del inglés jurídico

Autoras: Dra. María José Marín Pérez; Dra. Camino Rea Rizzo

Institución u Organismo al que pertenecen: Universidad de Murcia. Departamento de Filología Inglesa.

Indique uno o varios de los seis temas de Interés: (Marque con una {x})

{ } Enseñanza bilingüe e internacionalización

{ } Movilidad, equipos colaborativos y sistemas de coordinación

{X} Experiencias de innovación apoyadas en el uso de TIC. Nuevos escenarios tecnológicos para la enseñanza y el aprendizaje.

{ } Nuevos modelos de enseñanza y metodologías innovadoras. Experiencias de aprendizaje flexible. Acción tutorial.

{ } Organización escolar. Atención a la diversidad.

{ } Políticas educativas y reformas en enseñanza superior. Sistemas de evaluación. Calidad y docencia.

Idioma en el que se va a realizar la defensa: (Marque con una {x})

{X} Español { } Inglés

Resumen.

Esta comunicación presenta la propuesta de dos actividades para la enseñanza de la terminología legal aplicando una metodología DDL (*data-driven learning*). El número de experimentos pedagógicos realizados en este área del inglés para fines específicos es muy reducido (Boulton, 2010). Así se deduce de la revisión de la literatura que se ofrece en las secciones 1 y 2 donde también se reflejan los argumentos a favor y en contra del uso de los corpora en la enseñanza de lenguas. Así pues, la sección 4 de este estudio sugiere dos actividades basadas en un corpus legal, *BLaRC*, diseñado y compilado por Marín (2013) dada la escasez de corpora legales disponibles. Asimismo, en esta sección se ofrece un modelo de experimentación pedagógico para su futura implementación. Estas actividades se conciben, no como un sustituto de otros métodos de enseñanza, sino como complemento a materiales tales como los libros de texto

Palabras Clave: Inglés jurídico; corpus legal; metodología DDL; Inglés para Fines Específicos (IFE)

Abstract

This paper presents the proposal of two activities aimed at fostering the acquisition of legal terminology using a DDL (*data-driven learning*) methodology. The number of pedagogical experiments carried out in this ESP branch is very reduced (Boulton, 2010). This is why the gap in the area must be filled by designing and implementing activities using legal corpora as the source to obtain the information from. The activities presented in section 4 of this study are based on *BLaRC*, an 8.85 million-word corpus of British judicial decisions designed and compiled by Marín (2013). A review of the literature is also offered in sections 1 and 2, which show the pros and cons of the use of this type of teaching methods, according to the authors cited, and the methodological void existing in the area. This study ends with the proposal of an experimental research design for the implementation of the activities following it, which are conceived to act as support to already existing materials such as textbooks.

Keywords: legal corpora; specialised terminology; DDL methodology; English for Specific Purposes (ESP)

Texto

LA UTILIZACIÓN DE MÉTODOS DE APRENDIZAJE *DATA-DRIVEN* EN LA ENSEÑANZA DEL INGLÉS JURÍDICO

1. Introducción

El uso de corpora lingüísticos se ha extendido en los últimos veinticinco años en la enseñanza de idiomas gracias a los primeros pasos dados por Johns (1986; 1991), que acuña el término *data-driven learning -DDL-* (aprendizaje basado en datos). Los métodos didácticos DDL promueven el estudio de la lengua basado en la observación de concordancias, es decir, ejemplos del uso real de palabras clave en su contexto (o *keywords in context -KWC-*) que se extraen de corpora lingüísticos utilizando programas de software diseñados con ese fin como *Wordsmith 5.0* (Scott, 2008).

Sinclair (1991), uno de los padres de lingüística del corpus, define los corpora como una colección de textos que reflejan el uso natural de la lengua seleccionados para caracterizar el estado o una variedad de esta. Más tarde redefiniría el término introduciendo aspectos como el formato electrónico de los textos, los criterios de selección de estos, o su representatividad (Sinclair, 2005). Él sería uno de los impulsores y continuadores del trabajo de Johns en el área del aprendizaje *data-driven* (Sinclair, 2003), desarrollando la enseñanza de lenguas a través del uso de corpora y siendo considerado como una referencia fundamental en este campo en los últimos veinticinco años (McEney y Xiao, 2010).

Algunos autores (Flowerdew, 2009; Swales, 1990; Widdowson, 1998) muestran ciertas objeciones respecto del uso de los corpora en la enseñanza de lenguas argumentando que emplean muestras descontextualizadas y truncadas que se examinan de manera aislada, fomentando un modelo de aprendizaje inductivo que no siempre es recomendable. Sin embargo, son muchos los que señalan las virtudes de los métodos DDL por diversos motivos. Por una parte, los consideran motivadores y, sobre todo, destacan su capacidad de desarrollar en los alumnos

estrategias de aprendizaje y una mayor autonomía (Boulton, 2011), convirtiéndolos, en palabras de Johns, en “detectives de la lengua” (Johns, 1997:101).

Sea como fuere, son muchos más los autores que apoyan el uso de los corpora en la enseñanza de lenguas que aquellos que no lo recomiendan, aunque, como subraya Boulton (2010), aún queda un largo camino por recorrer en la experimentación utilizando métodos de enseñanza DDL.

Dada la escasez de experimentos que apliquen la metodología DDL en el área del inglés jurídico (Boulton, 2010), esta comunicación presenta una propuesta para la aplicación de este tipo de métodos a la enseñanza de esta variedad del inglés para fines específicos (IFE). Con ese fin, la sección 2 se dedica a la descripción del corpus que se emplea para la planificación de las actividades propuestas en la sección 4. A continuación, en la sección 3, se trata la relación entre los listados de vocabulario y la enseñanza de lenguas. Asimismo, se describe el proceso de extracción de términos jurídicos de *BLaRC*, el corpus que se emplea como fuente para la obtención de datos. Para terminar, la sección 4 presenta un método de implementación de las actividades que se proponen así como las actividades mismas seguidas por la conclusión, en la sección 5.

2. Aplicaciones de los corpora a la enseñanza de lenguas: características de BLaRC, *British Law Report Corpus*

A pesar de los apoyos y el interés que despierta el uso de los corpora en la enseñanza de lenguas, como se señala en la sección anterior, el número y la disponibilidad de estos en inglés jurídico es muy reducido. Así lo afirman Marín y Rea (2012) en un revisión de los corpora legales existentes online entre los que destaca *BoLC*, *the Bononia Legal Corpus* (<http://corpora.dslo.unibo.it/BoLC/BOLCCorpQueryENG.html>)

En consecuencia, se diseña *BLaRC*, un corpus de textos jurídicos británicos de 8,85 millones de palabras, que se procesa para la obtención de la información necesaria para la elaboración de las actividades que se proponen en la sección 4 de esta comunicación. Marín (2013) ofrece una descripción detallada del proceso de compilación y los criterios empleados para su diseño.

BLaRC está formado por 1.228 sentencias dictadas por tribunales del Reino Unido pertenecientes a las distintas jurisdicciones en las que se organiza el sistema legal británico. El corpus se basa en un único género legal, el de las sentencias judiciales, dada la importancia de este dentro de los sistemas legales *common law*. El Reino Unido se considera un país perteneciente al *common law* o derecho anglosajón donde el derecho emana principalmente de la jurisprudencia.

La estructura del corpus responde, por un lado, a criterios geográficos, esto es, el Reino Unido se divide en tres áreas con sus propios sistemas judiciales independientes: Inglaterra y Gales, Escocia e Irlanda del Norte. En estas áreas, la terminología jurídica a menudo varía debido a la independencia de los tres sistemas entre sí, por ejemplo, en Escocia el demandante sería el *pursuer*, mientras que en Inglaterra o Gales se denominaría *claimant*. Por otro lado, se encuentran aquellos países de la *Commonwealth* británica que aún aceptan al *Privy Council* como corte de último recurso dentro de sus sistemas legales, por lo que también se establece una categoría aparte para los textos provenientes de estos países. Por último, se asigna también una categoría propia a los tribunales cuya jurisdicción se extiende a

todo el Reino Unido tales como el Tribunal Supremo o la red de tribunales administrativos que opera en toda Gran Bretaña. En resumen, *BLaRC* está organizado en cinco categorías principales que responden a criterios geográfico-jurisdiccionales dada la influencia de estos sobre la posible variación terminológica. Debido a la escasez de corpora disponibles y su utilidad para la enseñanza del inglés jurídico, *BLaRC* está disponible en la web *Lextutor* de Tom Cobb (http://www.lex Tutor.ca/concordancers/concord_e.html) y pronto lo estará también en *Sketch Engine* (<http://www.sketchengine.co.uk/>), de Adam Kilgarriff, para su consulta.

3. Vocabulario general vs. vocabulario especializado: le extracción de listados de términos para su uso en el aula

Como afirma Sánchez (2000), los intentos de producir listas de vocabulario generales para su utilización en el aula se remontan a siglos atrás. Sánchez (2000) señala a Thorndike y Lorge (1944) como los primeros autores que estudian los distintos significados de las palabras en sus listados con el fin de organizarlos.

Basándose en el trabajo de estos autores, West (1953) ofrece un listado que incluye las 2.000 familias de palabras más frecuentes del inglés, la *General Service list of English Words (GSL)*, especificando los distintos sentidos de las palabras incluidas en ella.

La existencia de estos inventarios de vocabulario supone una fuente de información básica para la enseñanza de idiomas pues permite recurrir a ellos para la secuenciación de actividades de léxico. Nation y Beglar (2007) afirman que, conociendo las 8.000 familias de palabras más frecuentes del inglés, un estudiante de lengua extranjera puede llegar a entender hasta un 98% de varios tipos de textos orales y escritos como novelas o noticias de periódico.

En lo que respecta al inglés jurídico, no existen listados estándar a los que el profesor de IFE pueda recurrir para la elaboración de actividades o tests de vocabulario, por lo que se hace necesario la obtención de estos listados a través del uso de métodos de extracción automática de términos o métodos ATR (Automatic Term Recognition), sobre todo en el caso de utilizar un corpus de gran tamaño como *BLaRC*.

En su tesis doctoral, Marín (2013) evalúa los diez métodos ATR que se citan en la tabla 1 tras su aplicación a un corpus piloto de 2,6 millones de palabras extraído de *BLaRC*. Esta evaluación se lleva a cabo midiendo los niveles de precisión alcanzados por cada uno de ellos en función de las coincidencias encontradas entre los listados de candidatos a término obtenidos con cada uno de los métodos y un glosario de 10.088 términos jurídicos también compilado *ad hoc*. La tabla 1 ilustra los resultados obtenidos tras realizar esta evaluación. Los porcentajes reflejan el número de coincidencias encontradas entre las listas obtenidas con los métodos ATR y los términos recogidos en nuestro glosario legal. Si un candidato a término coincide con un término del glosario, automáticamente se considera término real y por lo tanto aumenta el porcentaje de precisión alcanzado por cada método ATR. En la tabla también se indica el procedimiento de implementación de los métodos, esto es, online (se ofrece la dirección web donde se puede encontrar el software), manual (aplicando las fórmulas que los autores proporcionan en la literatura), a través de

una herramienta java diseñada por Zang et al. (2008), *JATE tools*, o utilizando el software diseñado por los autores.

Método ATR	Procedimiento de implementación	Precisión media
Chung, 2003	Manual	42,25%
Church y Gale, 1995	JATE tools (Zang et al., 2008)	48%
Drouin, 2003	Online: http://termostat.ling.umontreal.ca/index.php?lang=en_CA	73,45%
Frantzi y Ananiadou, 1999	JATE tools	52,43%
Ki y Liu, 2008	Manual	64,75%
Nazar y Cabré, 2012	Online: http://terminus.upf.edu/cgi-bin/terminus2.0/terminus.pl	71,5%
Park et al., 2002	JATE tools	52,71%
Sclano y Velardi, 2007	Online: http://lcl.uniroma1.it/sso/index.jsp?returnURL=%2Ftermextractor%2F	48,29%
Scott, 2008	<i>Wordsmith 5.0</i> (Scott, 2008)	62%
Sparck Jones, 1972	Manual	57,3%

Tabla 1. Métodos ATR evaluados; procedimiento de implementación y porcentaje de precisión

Tras su evaluación, los métodos ATR diseñados por Drouin (2003) y Nazar y Cabré (2012) resultan ser los más precisos en la identificación de términos jurídicos (su porcentaje de precisión media se indica en rojo), por lo que se aplican al corpus de 8,85 millones de palabras, obteniendo un total de 2.848 términos jurídicos compuestos por una o más de una palabra, que se ordenan en función del nivel de especialización calculado por cada método.

4. Propuesta de actividades

Como señalan Harwood (2005) y Boulton (2012), los corpora complementan adecuadamente otros materiales para la enseñanza del IFE como los libros de texto. Por este motivo, utilizando *BLaRC* como la fuente de información para la obtención de datos, se diseñan dos actividades que podrían utilizarse para complementar libros de texto de inglés jurídico como *Professional English in Use: Law* (Brown y Rice 2007), que está indicado para alumnos cuyo nivel de inglés se corresponde con el nivel B2 del Marco Común Europeo de Referencia para las Lenguas¹ (MCERL). Las actividades propuestas estarían enfocadas a alumnos de primer curso del Grado en Derecho de la Universidad de Murcia en cuyo currículum se incluye el inglés jurídico como asignatura obligatoria. Para la implementación de las actividades y la medición de los posibles resultados de estas, se llevaría a cabo una selección aleatoria de alumnos de este nivel. Por este motivo, sería necesaria la realización de un test inicial de evaluación para descartar a aquellos de los seleccionados que no tuvieran el nivel mínimo requerido para el curso, esto es, el nivel B1, como se recoge

¹ Documento completo del MCERL: http://www.coe.int/t/dg4/linguistic/source/framework_en.pdf

en la guía de la asignatura. Este test inicial también serviría para dividir el grupo de control en varios subgrupos dependiendo del nivel de competencia reflejado en estas pruebas iniciales. De esta manera, sería posible observar los resultados de la implementación de las actividades por subgrupos y comparar su progresión.

Para medir los resultados de las actividades, se administraría otro pre-test a los alumnos (el mismo para todos los subgrupos) donde se incluirían los contenidos de las actividades sugeridas a continuación con el fin de obtener información sobre sus conocimientos previos y poder comparar los resultados con los post-tests, que deberían ser iguales a los anteriores para poder medir con exactitud el efecto de las actividades basadas en corpus (Hamp-Lyons, 1985).

La propuesta de actividades que se desarrolla a continuación es una propuesta piloto que debería ampliarse en el futuro incluyendo un mayor número de tareas basadas en corpora, que se planificarían a lo largo de todo el curso de inglés jurídico, de esta manera podría observarse el efecto de la metodología *data-driven* a más largo plazo. También sería deseable llevar a cabo un estudio en el que se compararan resultados entre un grupo experimental y uno de control en los que se aplicaran distintas metodologías, utilizando actividades DDL sólo en el primero de ellos.

En lo que respecta a los experimentos pedagógicos llevados a cabo en los últimos veinticinco años (Boulton, 2010), un número muy reducido de ellos se centra en la utilización de métodos DDL para la enseñanza del inglés jurídico, por lo que la implementación de actividades de este tipo y la observación de los resultados en el aprendizaje deben ser explorados en mayor profundidad.

Para la realización de estas actividades, los alumnos recibirían una sesión de formación para el manejo de los corpora a través del uso de herramientas como *Wordsmith* (Scott, 2008). Si se recurre a este tipo de actividades de manera regular, una sesión bastaría para que los alumnos adquiriesen las habilidades necesarias para que el método de trabajo no supusiera una dificultad añadida. En lo concerniente a la integración de las actividades en el currículum del curso, por tratarse de tareas que complementarían a un libro de texto, habría que relacionarlas con los bloques de contenidos del libro en cuestión, bien lingüísticos, bien temáticos, según conviniese. La actividad 1, podría llevarse a cabo dentro de la unidad 1 del libro de texto de Brown y Rice (2007) titulada *Legal Systems* donde se recogen tanto los términos *proceed* como *regulate*. La actividad 2, por estar más enfocada al contenido que a la forma de los términos, debería programarse para mediados del curso, cuando los alumnos ya estuvieran algo más familiarizados con la terminología jurídica británica habiendo estudiado unidades como *Sources of law: common law* (unidad 3); *Criminal justice and criminal proceedings* (unidad 5) o *European Union Law* (unidad 8).

Actividad 1

El uso de los corpora en la clase de IFE permite explorar distintos niveles de la lengua tales como el nivel morfológico, sintáctico, semántico o discursivo. Las actividades que se sugieren a continuación se centran en los tres primeros.

La primera actividad consistiría en pedir a los alumnos que formaran las familias de palabras (Bauer y Nation, 1993) de una serie de términos seleccionados entre los más representativos de los extraídos de *BLaRC*. Deberían por tanto hacer una búsqueda en el corpus de posibles derivados de los términos sugeridos por medio

de la adición de prefijos y sufijos para confirmar sus intuiciones. Una de las ventajas de una actividad de estas características es la de hacer reflexionar a los alumnos de manera consciente sobre los mecanismos de formación de palabras, facilitando la posterior aplicación de las reglas inferidas a otros casos similares. Este tipo de actividades, muy frecuentes en los libros de IFE, a menudo carecen de un contexto que haga más significativos los términos pertenecientes a cada una de las familias. Es precisamente en ese punto en el que las actividades basadas en corpus suponen un valor añadido puesto que los alumnos, al realizar las búsquedas de los derivados en el corpus, obtienen una serie concordancias, es decir, tienen que examinar los contextos de uso de los términos derivados.

Una vez identificados los términos derivados, los alumnos deberían realizar una actividad “fill in the gaps” en la que, utilizando líneas de concordancias incompletas seleccionadas y filtradas por el profesor, deberían completar los huecos de las frases con los términos adecuados en cada caso. Esta actividad actuaría a modo de feedback confirmando lo aprendido de manera inductiva observando los ejemplos de uso real de la lengua obtenidos del corpus por ellos mismos.

La figuras 1 y 2 ilustran algunas de las concordancias obtenidas para los términos derivados del verbo *proceed*.

N	Concordance
1	the Landlord of a fair trial. It is for the LVT to make its own procedural decisions, in the exercise of its discretion, in the
2	Bansil gave evidence on it. The Tribunal is satisfied that no procedural irregularities occurred which affected the
3	unsupported by any other documentation. 11. There is a procedural reason why the absence of a document of this
4	fairly and, in particular, may- (a) apply any provision or procedural rules which applied to the proceedings before
5	appeal as time-barred but for the introduction of the new procedural rules consequent upon the creation of the new
6	authority was entitled to bring appeal proceedings, and the procedural challenges were in my view arguable. Costs: 87.

Figura 1. Concordancias asociadas a *procedural*

N	Concordance
1	for acceptance of an offer made under Part 36 of the Civil Procedure Rules, namely 21 days, was less than
2	23.2). Court procedure was now controlled by the Civil Procedure Rules (CPR). Part 44 of the CPR dealt with the
3	the High Court and county courts (paragraph 23.2). Court procedure was now controlled by the Civil Procedure Rules
4	should be dealt with under the written representations procedure. Witness statements were lodged from Louise
5	a letter accompanying this decision gives details for the procedure to be followed. P R Francis FRICS.
6	and not before. The accompanying letter sets out the procedure for making costs submissions in writing. 1 ©

Figura 2. Concordancias asociadas a *procedure*

Actividad 2

La segunda actividad exploraría el nivel sintáctico y semántico de la lengua pues versaría sobre la identificación de los colocados más frecuentes de *law*. El concepto de colocado que se emplea en este caso se corresponde con la definición del término de la escuela de Birmingham. John Sinclair define colocado como “the occurrence of two or more words within a short space of each other in a text” (Sinclair, 1991: 170). *Wordsmith*, el software utilizado para el procesamiento de *BLaRC*, nuestro corpus de inglés jurídico, permite configurar los ajustes para la obtención de los colocados de un término. Se sugeriría a los alumnos que emplearan la medida *Mutual Information* y seleccionaran aquellos colocados inmediatos a la izquierda de *law* (ordenados tal y como aparecen en la figura 3) cuya frecuencia fuera mayor, se limitaría el número de concordancias para evitar que la actividad resultara excesivamente costosa.

Una vez ordenados y seleccionados, los alumnos tendrían que identificar aquellos que se refieren a ramas del derecho como *domestic law*, *community law* o *criminal law* y emparejarlos con sus correspondientes definiciones, obtenidas de un diccionario especializado (Saint-Dahl, 1999), que el profesor habría preparado con anterioridad. Los alumnos podrían también consultar el contexto de uso de los colocados (pulsando en la pestaña “source text”, como se aprecia bajo la tabla de la figura 3) de manera que les resultara más sencillo realizar esta actividad, sobre todo en aquellos casos en los que desconociera el significado del término en cuestión. Para confirmar sus intuiciones sobre el uso de los términos identificados en el ejercicio, también tendrían que corregir una serie de ejemplos extraídos del corpus que el profesor previamente habría manipulado para que los colocados referidos a las ramas del derecho no estuvieran en la frase correcta. Finalmente, el profesor elaboraría un *cloze*, o texto incompleto, que los alumnos tendrían que ir completando con los términos extraídos.

File Edit View Compute Settings Windows Help																
N	Word	With	lation	Texts	Total	al Left	Right	L5	L4	L3	L2	L1	entre	R1	R2	
1	THE	law	0.000	913	0,145	6,587	3,558	924	870	522	1,576	2,695	0	252	1,128	
2	OF	law	0.000	810	6,890	4,501	2,389	423	366	768	1,287	1,657	0	936	262	
3	COMMON	law	0.000	223	1,164	1,133	31	6	6	5	1	1,115	0	1	3	
4	IN	law	0.000	733	3,976	2,421	1,555	291	444	343	548	795	0	570	241	
5	DOMESTIC	law	0.000	119	510	503	7	2	8	7	21	465	0	0	3	
6	COMMUNITY	law	0.000	56	463	453	10	3	4	6	3	437	0	0	2	
7	ENGLISH	law	0.000	150	523	499	24	8	10	7	51	423	0	0	5	
8	INTERNATIONAL	law	0.000	54	401	362	39	5	7	9	16	325	0	1	16	
9	CASE	law	0.000	256	515	423	92	34	38	19	10	322	0	9	7	
10	CRIMINAL	law	0.000	84	353	309	44	8	10	8	2	281	0	0	8	
11	PUBLIC	law	0.000	121	331	264	67	12	6	3	20	223	0	3	8	
12	BY	law	0.000	419	1,219	772	447	86	70	149	295	172	0	132	81	
13	NATIONAL	law	0.000	62	198	176	22	8	6	10	11	141	0	0	5	
14	REFORM	law	0.000	30	208	146	62	1	3	11	6	125	0	57	1	
15	PRIVATE	law	0.000	48	158	143	15	17	2	10	19	95	0	0	2	

oncordance collocates plot patterns clusters filenames follow up source text notes

Figura 3. Listado de colocados inmediatos a la izquierda de *law* ordenados según su frecuencia

5. Conclusión

Este estudio ha ofrecido una propuesta para la aplicación de la metodología *data-driven*, basada en corpus lingüísticos, a la enseñanza del inglés jurídico. Debido a la escasez y poca disponibilidad de los corpus legales existentes, se recurre a *BlaRC* un corpus de sentencias judiciales británicas compilado por Marín (2013) como fuente para la obtención de datos.

Como consecuencia de la inexistencia de listados especializados de términos jurídicos, la sección 3 ha presentado el proceso y los resultados de la evaluación de 10 métodos ATR diferentes para la identificación de estos términos en *BLaRC*.

Tras ofrecer una revisión de la literatura dedicada al uso de los corpora para la enseñanza de lenguas, se llega a la conclusión de que aún queda un largo camino por andar dentro de área del inglés jurídico por lo que se diseñan dos actividades centradas en los niveles morfológico, sintáctico y semántico de la lengua basadas en *BlaRC*, nuestro corpus legal en la sección 4. Además, se plantea un modelo de experimental para la medición de los posibles resultados derivados de su implementación.

Bibliografía y Referencias

- Bauer, L., Nation, I.S.P. (1993). Word Families. *International Journal of Lexicography*, 6 (4), 253-279.
- Boulton, A. (2010). Data-driven learning: Taking the computer out of the equation. *Language Learning*, 60, 534-572.
- Boulton, A. (2011). Data Driven Learning: the Perpetual Enigma. In Roszkowski, S. and Lewandowska-Tomaszczyk, B. (eds.) *Explorations across Languages and Corpora* (563-580). Frankfurt: Peter Lang.
- Boulton, A. (2012). Corpus consultation for ESP. A review of empirical research. In Boulton, A., Carter-Thomas, S., Rowley-Jolivet, E. (eds.) *Corpus-Informed Research and Learning in ESP. Issues and Applications* (261-292). John Benjamins Publishing Company.
- Brown, G.D., Rice, S. (2007). *Professional English in Use: Law*. Cambridge: Cambridge University Press.
- Chung, T. M. (2003). A corpus comparison approach for terminology extraction. *Terminology*, 9 (2), 221-246.
- Church, K.W., Gale, W. (1995). Inverse Document Frequency (IDF): A measure of Deviations from Poisson. *Proceedings of the Third Workshop on Very Large Corpora*. Cambridge: Massachusetts Institute of Technology.
- Drouin, P. (2003). Term extraction using non-technical corpora as a point of leverage. *Terminology*, 9 (1), 99-117.
- Flowerdew, L. (2009). Applying corpus linguistics to pedagogy: A critical evaluation. *International Journal of Corpus Linguistics*, 14 (3), 393-417.
- Frantzi, K.T., Ananiadou, S. (1999). The c/nc value domain independent method for multi-word term extraction. *Journal of Natural Language Processing*, 3 (2), 115-127.
- Hamp-Lyons, L. (1985). Two Approaches to Teaching Reading: A Classroom-Based Study. *Reading in a Foreign Language*, 3 (1), 363-73.
- Harwood, N. (2005). What do we want EAP teaching materials for?. *Journal of English for Academic Purposes*, 4, 149-161.
- Johns, T. (1986). Microconcord: A language-learner's research tool. *System*, 14 (2), 151-162.
- Johns, T. (1991). Should you be persuaded: Two examples of data-driven learning. *English Language Research Journal*, 4, 1-16.
- Johns, T. (1997). Contexts: The background, development and trialling of a concordance-based CALL program. In Wichmann, A., Figelston, S., McEnery, T. and Knowles, G. (eds.) *Teaching and Language Corpora* (100-115). London: Longman.
- Kit, C., Liu, X. (2008). Measuring mono-word termhood by rank difference via corpus comparison. *Terminology*, 14 (2), 204-229.
- Marín, M.J. and Rea, C. (2012). Structure and design of the BLRC: a legal corpus of judicial decisions from the UK. *Journal of English Studies*, 10. La Rioja: Servicio de Publicaciones de la Universidad de La Rioja. <http://www.unirioja.es/servicios/sp/ei/jes/jes10.shtml>
- Marín, M.J. (2013) *Identification and analysis of the specialised vocabulary of british law reports : a corpus-driven study of this legal genre at the core of Common Law legal systems*. Tesis Doctoral. Universidad de Murcia.

- McEnery, T., Xiao, R. (2010). What corpora can offer in language teaching and learning. In Hinkel, E. (ed.) *Handbook of Research in Second Language Teaching and Learning* (364-380). London: Routledge.
- Nation, I.S.P., Beglar, D. (2007). A vocabulary size test. *The Language Teacher*, 31 (7), 9-13.
- Nazar, R., Cabré, M.T. (2012). Supervised Learning Algorithms Applied to Terminology Extraction. In Aguado de Cea, G., Suárez-Figueroa, M.C., García-Castro, R. and Montiel-Ponsoda, E. (eds.) *Proceedings of the 10th Terminology and Knowledge Engineering Conference (TKE 2012)*. Madrid: Ontology Engineering Group, Association for Terminology and Knowledge Transfer.
- Oxford English Dictionary*. (2002) Version 3.0. Electronic Resource. Oxford: Oxford University Press.
- Park, Y., Byrd, R.J., Boguraev, B. (2002). Automatic Glossary Extraction: Beyond Terminology Association. In *Proceedings of COLING '02 19th International Conference on Computational Linguistics*. Taipei, Taiwan.
- Saint-Dahl, H. (1999). *Dahl's Law Dictionary*. New York: William S. Hein and Co. Inc.
- Sánchez, A. (2000). Language Teaching before and after 'Digitized Corpora'. Three Main Issues. *Cuadernos de Filología Inglesa*, 9.1. Murcia: Servicio de Publicaciones, Universidad de Murcia.
- Sclano, F., Velardi, P., (2007). A Web Application to Learn the Common Terminology of Interest Groups and Research Communities. *Proceedings of the Conference TIA-2007*. Sophia Antipolis, France.
- Scott, M. (2008). *WordSmith Tools version 5* [computer software]. Liverpool: Lexical Analysis Software.
- Sinclair, J. (1991). *Corpus, Concordance and Collocation*. Oxford: Oxford University Press.
- Sinclair, J. (2003). *Reading Concordances: An Introduction*. London: Longman.
- Sinclair, J. (2005). Corpus and Text: Basic Principles. In Wynne, M. (ed.). *Developing Linguistic Corpora: a Guide to Good Practice*. AHDS Literature, Languages and Linguistics: University of Oxford.
- Sparck Jones, K. (1972). A statistical interpretation of term specificity and its application in retrieval. *Journal of Documentation*, 28, 11-21.
- Swales, J. (1990). *Genre Analysis. English in Academic and Research Settings*. Cambridge: Cambridge University Press.
- Thorndike, E.L., Lorge, I. (1944). *The teacher's Word Book of 30,000 Words*. Teachers' College: Columbia University.
- West, M. (1953). *A General Service List of English Words*. London: Longman.
- Widdowson, H. (1998). Communication and community: The pragmatics of ESP. *English for Specific Purposes*, 17 (1), 3-14.
- Zhang, Z., Iria, J., Brewster, C., Ciravegna, F. (2008). A Comparative Evaluation of Term Recognition Algorithms. In *Proceedings of the sixth International Conference on Language Resources and Evaluation, (LREC 2008)*. Marrakech, Morocco.