

TRABAJO FINAL DE GRADO EN TRADUCCIÓN E INTERPRETACIÓN
Departamento de Traducción y Comunicación

TÍTULO

**EL CORPUS: UN BUEN COMPLEMENTO PARA
LA TAREA DE TRADUCIR**

Autora: Andrea Fachina Rivera

Tutora: María Calzada Pérez

Fecha de lectura: junio 2015



Resumen:

Con el presente Trabajo de Fin de Grado se aspira a evidenciar la importancia de los corpus en la traducción y, en concreto, se pretende demostrar que son un buen complemento de otros recursos lingüísticos de consulta habitual durante la práctica traductora. Para ello, se compila un archivo electrónico con artículos sobre baloncesto *Sportsnews* de 364.562 palabras compuesto por dos corpus: *Sportsnews_en* (en inglés y de 190.417 palabras) y *Sportsnews_es* (en español y de 174.145 palabras). Estos artículos pertenecen a cuatro periódicos importantes de los ámbitos angloparlante (*The New York Post* y *The New York Times*) e hispanoparlante (*El País* y *El Mundo*). Con estos textos de uso real, se realizan tres tipos de análisis lingüísticos diferentes. En el primer tipo se dan una serie de pinceladas descriptivas del lenguaje, en inglés y en español, de los corpus estudiados; en el segundo, se identifican algunas diferencias entre los periódicos de una misma lengua; finalmente, en el tercer tipo de análisis se examina qué aporta la metodología de corpus frente a una importante fuente de información para los traductores: un glosario especializado, en nuestro caso, del ámbito del baloncesto. Para llevar a cabo estos análisis, se estudian las trescientas palabras clave de *Sportsnews_en* y de *Sportsnews_es*. En definitiva, este trabajo permite indagar en la mejora de la práctica traductora a través de la descripción y el análisis lingüísticos de textos de uso real. Además, conoceremos qué información extra aportan los corpus al traductor y, en conjunto, cómo se trabaja con ellos.

Palabras clave:

corpus, prensa deportiva, baloncesto, descripción, traducción

Índice

1. Motivación y justificación	4
1.1. Contextualización	6
1.1.1. El lenguaje periodístico del deporte y del baloncesto en español y en inglés.....	7
1.1.2. Corpus y glosarios: recursos de interés para el traductor.....	11
1.1.2.1. Corpus.....	11
1.1.2.2. Glosarios	12
2. Metodología	13
2.1. Fuentes de análisis	13
2.2. Método de compilación de los corpus	15
2.3. Método de análisis	17
3. Resultados de los análisis	22
3.1. Descripción del lenguaje baloncestístico de <i>Sportsnews_en</i> y <i>Sportnews_es</i>	22
3.1.1.1. Datos estadísticos básicos.....	23
3.1.1.2. Sustantivos.....	24
3.1.1.3. Pronombres.....	25
3.1.1.4. Verbos.....	25
3.1.1.5. Adjetivos	29
3.2. Comparación de periódicos dentro de un mismo corpus	29
3.3. Comparación de corpus y glosario	32
4. Conclusiones	34
5. Bibliografía	37
6. Anexos	40

1. Motivación y justificación

Inicialmente este trabajo aspira a demostrar que la metodología de corpus es útil para un traductor como complemento informativo y terminológico de otras herramientas, como por ejemplo, un glosario.

El uso de la metodología de corpus en traducción e interpretación es un ámbito en el que los estudiantes no tenemos la oportunidad de profundizar, y que está totalmente relacionado con las estrategias que se adoptan durante la práctica de trasvase lingüístico. Según Olohan, los corpus “cover numerous topics within linguistics: lexical studies, grammar, semantics, pragmatics and discourse analysis, sociolinguistics, stylistics and text linguistics, historical linguistics, dialectology and variation studies and psycholinguistics” (Olohan, s.f., p. 419). Por ello, parecen, *a priori*, aportar información útil y variada para profesionales, como los traductores e intérpretes que lidian todos los días con este tipo de matices lingüísticos. Es decir, de acuerdo con esta autora, los corpus no solo sirven para arrojar equivalencias de aplicación inmediata, lo cual es cierto especialmente en el caso de los corpus paralelos, sino que son una fuente de información lingüística que refuerza los conocimientos relacionados con el trasvase.

La decisión de elegir como Trabajo de Fin de Grado el uso de corpus surgió, por un lado, de mi interés por las herramientas electrónicas, presente a lo largo de toda la carrera. La curiosidad sobre este tema me llevó a sondear qué es lo que podría aprender de esta metodología que me fuera útil especialmente con vistas a un futuro inmediato. Dentro de unos meses, realizaré mis prácticas en el servicio lingüístico de la empresa Xerox, en donde tendré que enfrentarme a nuevas tecnologías de tratamiento y explotación de textos de campos diversos (médico, técnico, audiovisual, etc.). Este trabajo me ofrecía la oportunidad de iniciarme en el uso de herramientas que solo conocía superficialmente, trabajar con ellas (creando incluso listados de palabras especializadas cuando fuera necesario), aprender de ellas y adaptarme a sus métodos. Asimismo, siempre he tenido interés por aprender más sobre el deporte del baloncesto y pensé que este trabajo contribuiría también a satisfacer mi curiosidad al tiempo que me permitiría experimentar en un ámbito en el que carezco de especialización. Así pues, este trabajo no solo aborda el estudio de qué información arrojan los corpus y cómo pueden complementar a otros recursos, como los glosarios, sino que también es una especie de prueba de lo que podría hacer ante futuros encargos laborales.

En conjunto, el presente Trabajo de Fin de Grado se inicia con una parte más teórica que sirve de contextualización del estudio. En ella, se definirá brevemente el lenguaje que se emplea en los textos periodísticos, concretamente en los del ámbito del baloncesto, y se presentarán los recursos empleados para acercarnos al objeto de estudio. A continuación, se presenta una parte metodológica en la que se acomete la compilación y tratamiento previo de un archivo electrónico con artículos sobre baloncesto *Sportsnews* compuesto por dos corpus: *Sportsnews_en* (en inglés) y *Sportsnews_es* (en español). A este apartado, le sigue una parte analítica tripartita con la que se pretende confirmar o refutar las ideas que se defienden en la parte teórica. Por último, este trabajo se cierra con un apartado de conclusiones finales en las que explicaré las conclusiones que derivan de los tres análisis y la relación del trabajo con la carrera.

1.1. Contextualización

A lo largo de los años, la traducción de los textos especializados en el deporte ha ido ganando importancia, dada la creciente popularización de esta actividad de competición y ocio. A día de hoy, desempeña una función importante, por ejemplo, en la expansión internacional de empresas de material deportivo y facilita la actividad de empresas de *marketing*. Hay momentos en los que la comunicación multilingüe es frenética; se traducen e interpretan muchos comunicados y kits de prensa de entidades y clubes. Cuando hay olimpiadas, por ejemplo, las agencias deportivas, los organismos olímpicos y los medios necesitan traducir una gran cantidad de documentos. Así pues, podría decirse que, para un traductor y para un intérprete, es un ámbito que merece la pena explorar.

Al empezar un estudio sobre la traducción del deporte la primera característica que parece evidente es que, muchas veces para llevar a cabo este tipo de traducciones, es necesario tener un alto nivel de especialización, puesto que cada deporte tiene su propia jerga. Esta impresión parece venir refrendada por la opinión de traductores e intérpretes profesionales como Song Dongsheng, quien durante las olimpiadas de Beijing afirmó en la revista *China Today* que “los diccionarios se quedan atrás del desarrollo de las distintas disciplinas deportivas, por lo que es difícil encontrar la terminología actual en muchos deportes” (Yohong, L., s.f.). En este tipo de textos, como en tantos otros, el traductor no solo tiene que trasladar la información hacia su lengua materna sino que además tiene que ser experto en el área de trabajo para entender ciertas expresiones y fraseología. Así pues, no solo se ha de tener en cuenta el sentido de los textos sino que se debe estar atento a la forma estética en la que estos se presentan en los dos idiomas, puesto que varían de una cultura a otra. Todo ello pasa por dominar de forma activa el lenguaje deportivo.

1.1.1. El lenguaje periodístico del deporte y del baloncesto en español y en inglés

En 1977, Lázaro Carreter, en su libro *Lenguaje en periodismo escrito*, defendía que el lenguaje periodístico en español presentaba un estilo estándar culto. En aquella época, los periodistas mezclaban registros coloquiales, administrativos e incluso literarios (Lázaro, 1977). En nuestro tiempo, según Hernando Cuadrado (2002), se apuesta principalmente por un lenguaje directo, sencillo y cercano, a veces hasta coloquial sin llegar a la vulgaridad.

Desde el punto de vista gramatical, Hernando Cuadrado (2002) recuerda que la sencillez estilística se busca mediante los verbos en pretérito perfecto simple y en presente con valor histórico. Como botón de muestra recuperamos este ejemplo: *Fallece el americano...*

A través de la sencillez, se pretende que cualquier aficionado pueda entender el texto sin demasiado esfuerzo y, por ello, se promueve el empleo de un vocabulario inteligible, cercano al lector medio, recurriendo en grado mínimo a los tecnicismos.

Dentro del periodismo deportivo en español, Hernández Alonso (2012) argumenta que la inteligibilidad no está reñida con el uso de figuras retóricas. Y así, sobre todo en las victorias internacionales, se produce un derroche de metáforas, como son las palabras: *eternidad* y *San Iker* (Hernández, 2012, pp. 75-78). Otra característica idiosincrásica del lenguaje deportivo, según este mismo autor es la abundancia de polisemia, que facilita la creación de neologismos y permite que el periodista exprese las ideas de manera más concisa. A través de la derivación se inventan nuevas palabras para trasladar la idea de forma más clara y concisa; como por ejemplo el término *clasicazo* (Hernández, 2012, p. 96). Por otro lado, también destaca el empleo de la elipsis. El titular *70-68, once horas, set y partido*, con el cual el periódico *ABC* resume que el partido ha durado más de lo habitual, es un claro ejemplo de elipsis. A través de su uso, el periodista evita información innecesaria y consigue encontrar expresiones de otras lenguas distintas a las habituales. Por consiguiente, de este *léxico ampliado*, al cual hace referencia Hernández Alonso, nos llama la atención el uso de los llamados compuestos sintagmáticos, los cuales han ido ganando importancia gracias a la repetición de estructuras propias de este tipo de lenguaje; como por ejemplo, el uso de estos dos vocablos juntos: *estilo Barça* (Hernández, 2012, p. 88).

Como colofón al repaso, necesariamente breve por motivos de espacio, que hacemos aquí sobre el estilo de los textos en español vinculados con el deporte, cabe resaltar que, en ellos, abunda la incorrección de las formas de algunos tiempos verbales o de derivados léxicos debido a la rapidez en la que se escriben los artículos; por ejemplo, el uso del presente perfecto *han sido detenidos*, que debe estar en pretérito indefinido *fueron detenidos* (Hernando, 2002, pp. 271-272).

Por lo que se refiere al tema más concreto del lenguaje del baloncesto en español, la repercusión que tiene la NBA por todo el mundo y el éxito de los equipos y las selecciones en campeonatos internacionales en las últimas décadas ha dado lugar a que la terminología de este deporte se haya universalizado de una forma natural. Bernárdez García en su libro *El deporte correctamente hablado* escribe: “La calle impone sus normas y a la gente más joven le gusta calzar zapatillas de *basket* o llevar unas *baskets*” (Bernárdez, 1998, p. 26). De hecho, según Rojas Torrijos (2014) en su entrada *Extranjerismos necesarios y superfluos en las crónicas de baloncesto*, uno de los deportes en el que más abunda el uso de extranjerismos es el baloncesto.

Muchas veces el uso de vocablos foráneos está bastante justificado, cuando la adaptación al castellano no define su concepto con precisión o pierde matices (Rojas, 2014). Por lo tanto, podría decirse que conforme a este autor, hay dos tipos de voces extranjeras: las que, por la difusión de su uso o por su dificultad para ser traducidas o entendidas, resultan más necesarias y aquellas otras que son más prescindibles, ya que no definen una realidad nueva, diferente de la que ya existe en nuestro idioma. Un ejemplo de estos últimos es el término *captura* (calco de *capture*) que en español es un *rebote*. Por otro lado, se podría decir que el caso del término *alley-oop* es diferente. Se trata de un extranjerismo necesario, ya que exige en español una equivalencia mucho más larga y compleja para definir la jugada con exactitud y sin perder ningún matiz. Así pues, su traducción, *pase alto hacia el aro*, no dice suficiente del significado de esta acción (Rojas, 2014). La enciclopedia *Merriam Webster* define el término *alley-oop* como “A basketball play in which a leaping player catches a pass above the basket and immediately dunks the ball”.

En cuanto al lenguaje periodístico en inglés, o *journalese*, se caracteriza por estar muy estereotipado. De acuerdo con Dickson y Skole “journalese is the particular code in which journalists report a story. It is a pattern of language —a jargon— that never appears in normal conversation”. Los periodistas suelen emplear y repetir expresiones, a veces grandilocuentes, lo cual nos lleva a pensar en el lenguaje literario (Dickson & Skole, 2012). Por ejemplo, se utilizan palabras como *magnitude* en lugar de *size* o *possesses* por *has*. Asimismo, en cuanto a la oración, por ejemplo en los titulares, el inglés tiende a utilizar en mayor medida tiempos verbales en pasado que el español (Lara, 2011). Además, cabe destacar que el inglés tiende a utilizar más el estilo nominal que el español, que prefiere el estilo verbal.

Por lo demás, de acuerdo con Steve Rushin en su entrada *The Specialized Language of Sports* (2010) el lenguaje deportivo en inglés hace uso de una jerga muy técnica, al contrario de lo que ocurre en español; es decir, abundan las palabras especializadas, como *to podium* o *slurve* (compuesta por *slider* y *curve*). Se trata de un lenguaje muy descriptivo y muchas veces, figurativo que, como en español, a veces se intensifica con metáforas del campo de batalla. Un ejemplo de ello es este titular que podemos encontrar en la sección de baloncesto del periódico *The New York Times: In a Battle of the Lowly, the Knicks End Up a Bit Lower*.

El lenguaje que se utiliza en inglés sobre el deporte del baloncesto en concreto está muy extendido. La lengua evoluciona junto a los deportes. Puede que en un momento dado el entrenador de un equipo o el comentarista de una televisión invente una nueva palabra y que, a partir de ese instante, ese término pase a convertirse en uno más de la lengua en cuestión. Muchas veces, sobre todo en baloncesto, hay varios términos para referirse a un mismo concepto, lo cual puede llevar a confusiones. Por ejemplo, podemos llamar *free throw line* o *charity stripe* a la línea que no se debe sobrepasar en el lanzamiento de un tiro libre.

En conclusión, el baloncesto se practica en todas las partes del mundo y contiene un gran listado de terminología general, pero también hay términos más especializados. Para organizar todos esos conceptos, podemos encontrar en Internet algunos recursos, como por ejemplo los glosarios, que pueden ayudar al traductor en las consultas de traducción, aunque hay que asegurarse de su fiabilidad y adecuación.

1.1.2. Corpus y glosarios: recursos de interés para el traductor

Si bien son muchos los recursos que se encuentran al alcance de los traductores e intérpretes para recabar información acerca de los textos sobre los que trabajan, en este apartado nos centraremos, por motivos de espacio, solo en los corpus y los glosarios. Así, sentaremos las bases para extraer conclusiones en cuanto a la comparación de ambos tipos de recursos con vistas a identificar la información más característica que arrojan los corpus.

1.1.2.1. Corpus

En la actualidad, podemos considerar que estamos ante una gran revolución de las nuevas tecnologías. La introducción de nuevos instrumentos de análisis ha supuesto grandes cambios en la disciplina de la lexicografía (Rojo, 2009).

Uno de los instrumentos más novedosos son los corpus, de los que Tognini Bonelli afirma “Given that a corpus is a collection of texts, the aim of corpus linguistics has rightly been seen as the analysis and description of language use, as realised in text(s)” (Tognini, 2010, pp. 18-19). Los corpus, por tanto, pueden ser muy útiles para descubrir aspectos del uso del lenguaje que pueden pasar desapercibidos al ojo humano (Reppen, 2010).

Existen diversos tipos de corpus, dependiendo de los objetivos para los que se desee su uso. En general, según Villayandre (2010), los principales parámetros para clasificar los corpus son: el número de lenguas de los textos, el tamaño del corpus según la cantidad de palabras, el grado de especialización de los textos, etc.

Según su número de lenguas, los corpus pueden ser monolingües o multilingües/bilingües (Villayandre, 2010). Dentro de estos últimos, se pueden distinguir los corpus comparables, los paralelos y los alineados. Los primeros están compuestos por textos parecidos en una, dos o más lenguas; los paralelos son “original, source language-texts in language A and their translated versions in language B” (Baker, 1995, p. 230); y, por último, los corpus alineados son “corpus paralelos en los que los textos están situados unos al lado de otros por párrafos o frases, de tal forma que sea más fácil extraer las equivalencias de traducción” (Villayandre, 2010, p. 29).

En cuanto al tamaño, basándonos en la tipología de corpus según Vargas (2006), “si el ámbito que queremos estudiar es muy delimitado y circunscrito bastará con contar con un corpus de tamaño pequeño a mediano (de 80.000 a un millón de palabras)” (Vargas, 2006, p. 11), aunque bien es cierto que, para que un corpus sea verdaderamente representativo tiene que tener millones de palabras (Sánchez & Cantos, 1998). Por tanto, en la selección de los textos, hay que asegurarse de que unas de las características de nuestro corpus sean el equilibrio y la representatividad por parte del contenido lingüístico (Vargas, 2006) y uno de los parámetros cruciales para que un corpus sea representativo de una lengua es la cantidad de palabras.

1.1.2.2. Glosarios

De acuerdo con Cabré en su artículo *El traductor y la terminología: necesidad y compromiso* (2000), de entre los diversos recursos que pueden consultar los traductores, destacan los glosarios, piezas esenciales para llevar a cabo una traducción rápida y eficaz. Según el *Diccionario de la Real Academia Española*, los glosarios son “catálogos de palabras de una misma disciplina, de un mismo campo de estudio, etc., definidas o comentadas”. Son, en consecuencia, instrumentos que permiten agrupar tecnicismos de acuerdo con cada rama y facilitan la localización de las palabras.

Conforme a la entrada *Terminologías y glosarios de traducción: Cuál es su importancia*, publicada en la página web *JR Language Translation Services, INC.* (2012), los glosarios pueden contener la explicación de un término, como en los diccionarios monolingües, y la equivalencia de este en otro idioma o ambas cosas. Algunas de las ventajas de la utilización de glosarios especializados son las siguientes: el proyecto de traducción final es más uniforme, se reduce el tiempo del proceso de traducción y se resuelven posibles problemas de ambigüedad entre términos (JR Language Translation Services, INC., 2012).

En definitiva, un glosario de términos resulta muy práctico para los proyectos de traducción de las grandes empresas, ya que todas las personas involucradas en ese mismo proceso pueden partir de un lenguaje común.

2. Metodología

En esta sección nos centraremos en presentar las fuentes de análisis del presente trabajo así como las fases de elaboración del mismo. Recordamos aquí que nuestro TFG parte con el objetivo de sondear el potencial informativo más característico de la metodología de corpus.

2.1. Fuentes de análisis

Para la consecución del objetivo mencionado arriba he utilizado, por un lado, dos corpus comparables compilados por separado, a los que hemos dado los nombres de *Sportsnews_en* y *Sportsnews_es* y, por otro, un glosario especializado de baloncesto.

Por lo que se refiere a los corpus, *Sportsnews_en* está compuesto por artículos sobre baloncesto publicados en dos periódicos norteamericanos, *The New York Times* (99.466 palabras) y *The New York Post* (90.951 palabras), que suman 190.417 palabras. Por su parte, el segundo de los corpus, *Sportsnews_es*, se compone de artículos equiparables publicados en dos periódicos españoles, *El País* (83.078 palabras) y *El Mundo* (91.067 palabras), que en total contiene 174.145 palabras.

Como vemos, la suma de los dos corpus es de más de trescientas cincuenta mil palabras. Se ha intentado aproximar el número de palabras de ambos corpus, a pesar de que no se ha conseguido una paridad total. Cabe explicar aquí que se han elegido dos periódicos de cada lengua para que hubiera cierto nivel de variedad lingüística, pero por motivos de espacio y tiempo, no se han incorporado más diarios, lo cual parecería conveniente para trabajos futuros.

De igual manera, cabe destacar que, debido a las limitaciones de un Trabajo de Fin de Grado, durante el análisis se han tenido que obviar muchos de los datos obtenidos del corpus. Si bien, los corpus proporcionan un amplio caudal de información de las diversas categorías lingüísticas presentes en los textos, aquí nos centraremos en sus palabras clave o términos idiosincrásicos, que consideramos son una de las piezas principales de todo proyecto traductor.

Por otro lado, la tercera fuente de información del trabajo es un glosario especializado de baloncesto bilingüe (inglés-español), elaborado por un experto en el ámbito, el entrenador de baloncesto Juan Ángel Lozano Piedehierro¹. Tras haber buscado en Internet varios glosarios especializados, me llamó la atención este en concreto por su extensión y número de término. Además, su autor es un entrenador de la Federación de Baloncesto Española (FEB), por lo que también es fiable. Se trata de uno de los pocos glosarios con original y traducción que hay disponibles en línea y me pareció que sería interesante utilizarlo en este estudio para extraer conclusiones.

¹ Glosario especializado de baloncesto. Recuperado de http://www.desdeelbanquillo.es/articulos/basket_ingles.pdf

2.2. Método de compilación de los corpus

La selección de los artículos que componen los corpus *Sportsnews_en* y *Sportsnews_es* ha sido muy precisa. Todos ellos pertenecen a la sección de deportes de cuatro periódicos: *The New York Post*, *The New York Times*, *El País* y *El Mundo*. He decidido extraer los textos de estos cuatro periódicos porque son publicaciones generalistas que no se dedican solamente al deporte, sino que dedican a este tema una sola sección. De esta manera, los textos serán más comparables, puesto que se extraen de periódicos con las características similares. Además, por lo que respecta a los periódicos en inglés, me he decantado por escoger dos de los Estados Unidos, dado que la popularidad del baloncesto es mayor en este país que en el Reino Unido.

La fase de creación del corpus es fundamental para llevar a cabo los análisis correspondientes partiendo de este tipo de recursos. Así pues, es necesario seguir un preciso método para la compilación del corpus.

En primer lugar, me descargué los artículos, por separado, desde LexisNexis Academic, una potente base de datos que proporciona información actual nacional e internacional y contiene un gran número de artículos de prensa, revistas y otras fuentes.

Dentro de esta base de datos, me dirigí a la barra de navegación de la izquierda e hice clic en *Power Search*. En el cuadro de parámetros que aparece, en *Search Terms* escribí los términos *baloncesto/basketball* para extraer los textos de los cuatro periódicos. En *Specify Date* seleccioné la opción *Previous year* del desplegable y dentro de *Add Section Search*, en *Section* seleccioné *Subject* y, por último, en *Term(s)* escribí de nuevo *baloncesto/basketball* e hice clic en *Add*. Por último, mediante el botón rojo *Search* he accedido a otro cuadro de diálogo para seleccionar el periódico deseado y una vez han aparecido los artículos correspondientes, he ido seleccionando minuciosamente los que eran más indicados (véase Figura 1).

LexisNexis® Academic Home English

General Searching

- » Easy Search™
- » Power Search
- » **Tip:** Click the headings below to view links to specialized search forms and other useful features.

Power Search Help Clear

Use of this service is subject to Terms and Conditions

Search Type: Terms & Connectors Natural Language

Search Terms: Search

Specify Date: Date is before .

Add Index Terms:

Select Source:

By Type:

By Name:
Start typing a title like New York Times

» Try also Find Sources Or Browse Sources

Add Section Search:

Add search term(s) within a specific document section

Connector: And Or

Section:

Term(s): Add to Search

News

US Legal

Legal

International Companies

Subject Areas

Sources

Help & Instructions

About LexisNexis | Terms and Conditions | Privacy & Cookies Policy (Updated)
Copyright © 2015 LexisNexis, a division of Reed Elsevier Inc. All rights reserved.

Figura 1: Búsqueda de artículos en LexisNexis Academic

Una vez seleccionados y descargados los textos, dimos paso al proceso de tratamiento y limpieza de los mismos. Tras la eliminación de espacios en blanco y la conversión de símbolos de codificación ilegibles, discriminamos (con etiquetas propias de XML y la ayuda del Editor JEdit 5.2.0) la información útil para nuestro análisis (el texto de los artículos propiamente dichos, que ahora aparece entre las etiquetas <text> y </text>) de información despreciable para nuestro trabajo (principalmente los metadatos encerrados entre las etiquetas <header></header>, <title></title>, <autor> </autor> <section></section> y <length></length>) (véase Figura 2).

```

NYT_3-partdh: Bloc de notas
Archivo Edición Formato Ver Ayuda
<?xml version="1.0" encoding="UTF-8" standalone="no"?>
<!DOCTYPE sportsnews SYSTEM "sportsnews.dtd">
<sportsnews>
<header>
<doc_number>197 of 200 DOCUMENTS</doc_number>
<journal>The New York Post</journal>
<date>February 28, 2011 Monday</date>
<edition>NA</edition>
</header>
<body>
<title>SHOT IN THE ARM - BILLUPS MAKES 'BIG' DIFFERENCE FOR KNICKS</title>
<author>MIKE VACCARO</author>
<section>Late City Final; Pg. 48</section>
<length>417 words</length>
<text>
MIAMI - If this felt familiar to you, Knicks fan, the Knicks and the Heat
battling to the final seconds of a tight, taught, close basketball game, it
should. If you have a long enough memory, you know this used to be
standard-operating procedure for these two teams.

Chauncey Billups? He doesn't have to go back that far. It wasn't that long ago
that Billups was the heart, soul, motor and engine of a Detroit Pistons team
that won one NBA title (in which he was voted Finals MVP), went to game 7 of
another, and qualified for six straight Eastern Conference Finals.

So this felt like the greatest possible flashback for him, too.

"It really felt like a playoff game," Billups said last night, in the wake of
the Knicks' thrilling 91-86 victory over the Heat at American Airlines Arena.
"It felt like my old Detroit-versus-Miami days, coming down here in May, trying
to advance to the Finals. It really was high intensity out here. It was a good
feeling."

Billups' numbers may look pedestrian: 16 points, two assists, five turnovers,

```

Figura 2: Documento etiquetado con J-Edit

2.3. Método de análisis

Con el corpus preparado, procedimos a analizar los datos obtenidos, y para ello, utilizamos dos corpus monolingües de control, el *Bank of English* para el inglés y el *Corpus de Referencia del Español Actual* para el español. Con el fin de analizar los resultados del corpus utilizamos WordSmith (versión 6.0). Se trata de un generador de concordancias muy completo para llevar a cabo análisis lingüísticos completos. Mediante este tipo de programas, se pueden obtener, entre muchas otras cosas, estadísticas de las características básicas de los corpus, listas de las palabras más frecuentes y clave, resultados de concordancia entre palabras, etc.

En concreto, para este trabajo me he centrado en los resultados de las palabras más frecuentes en *Sportsnews_en* y *Sportsnews_es*, en las palabras clave de ambos corpus y en las colocaciones y grupos de vocablos más habituales.

Para generar datos que permitan los análisis, en primer lugar, introduce en WordSmith los textos de cada corpus y así generé las listas de Wordlists y Keywords. Para seleccionar los textos deseados, hice clic en *WordList* y, después, *File, New* y *Choose texts now* (véase Figura 3).

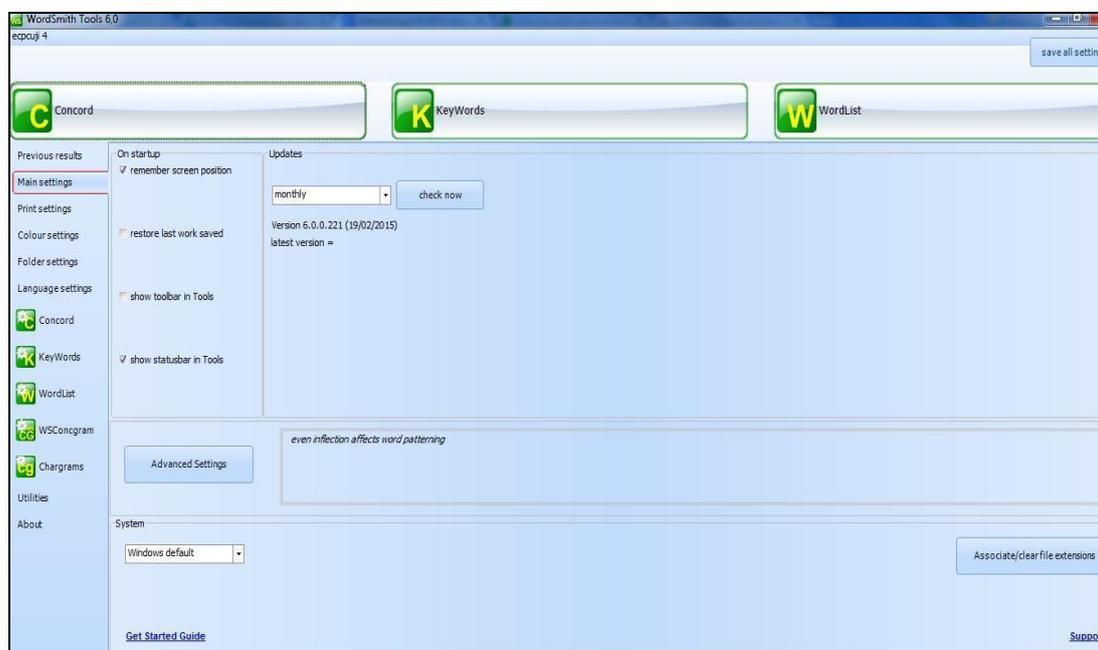


Figura 3: Análisis lingüístico con WordSmith

Antes de continuar con el siguiente paso, es importante asegurarse de que en la configuración del programa, dentro de *Language Settings*, esté seleccionada la lengua correcta, en nuestro caso, el inglés o el español, según el corpus que se analiza.

Por otro lado, en *Main Settings*, y dentro, en *Advanced Settings*, en *Tags and Markup* hicimos clic en *Only Part of File* para seleccionar la parte del texto que queremos analizar. Esto podemos hacerlo por nuestro trabajo previo de etiquetado en XML, que nos permite discriminar entre el texto que queremos analizar y el que deseamos despreciar. Como ya quedó claro arriba, en nuestro estudio, analizamos el texto en sí (entre las etiquetas `<text>` y `</text>`), dejando fuera el título y las otras partes del documento, como son por ejemplo, la fecha, el *copyright* o el número de palabras. Para ello, en el cuadro de parámetros seleccionamos la pestaña *Sections to keep* y en el primer desplegable escribí `<text> to </text>`. De esta manera, no se analizan las palabras que no forman parte del texto (véase Figura 4).

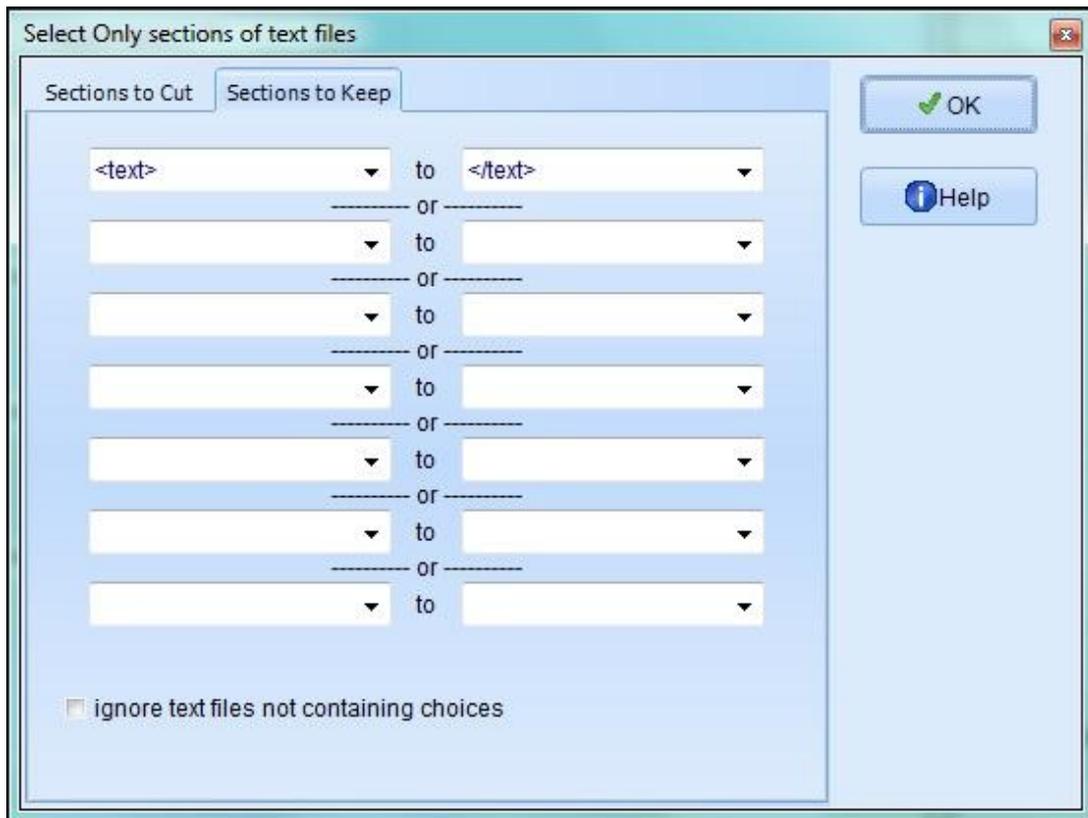


Figura 4: Análisis de las palabras más frecuentes

En resumen, como primer paso, introduje todos los textos de los cuatro periódicos por separado y creé una Wordlist para cada uno de ellos. Por otro lado, con el propósito de obtener las palabras clave o Keywords de *Sportnews_en* y *Sportnews_es*, generé dos Wordlists para ambos corpus. Para la creación de las Wordlists, entré en *WordList, Files, New* y en la pestaña *Main* hice clic en *Make a wordlist now*. Una vez apareció la lista, la guardé seleccionando en el menú *File* y dentro de *Save as, Wordlist* (véase Figura 5).

N	Word	Freq.	%	Texts	% Lemmas	Set
1	DE	9.084	5,22	552	98,40	
2	EL	6.946	3,99	547	97,50	
3	#	5.969	3,43	523	93,23	
4	EN	5.903	3,39	530	94,47	
5	LA	5.781	3,32	538	95,90	
6	Y	4.787	2,75	501	89,30	
7	QUE	4.537	2,61	459	81,82	
8	A	3.189	1,83	456	81,28	
9	LOS	3.058	1,76	461	82,17	
10	UN	2.320	1,33	396	70,59	
11	CON	2.166	1,24	439	78,25	
12	DEL	1.931	1,11	441	78,61	
13	SE	1.833	1,05	423	75,40	
14	SU	1.668	0,96	400	71,30	
15	POR	1.617	0,93	416	74,15	
16	NO	1.451	0,83	328	58,47	
17	UNA	1.245	0,71	365	65,06	
18	LAS	1.229	0,71	334	59,54	
19	AL	1.222	0,70	405	72,19	
20	PARA	1.201	0,69	345	61,50	
21	ES	1.026	0,59	270	48,13	
22	MÁS	940	0,54	293	52,23	
23	COMO	885	0,51	278	49,55	
24	LO	779	0,45	280	49,91	
25	PERO	758	0,44	271	48,31	
26	EQUIPO	624	0,36	284	50,62	
27	HA	609	0,35	227	40,46	

frequency | alphabetical | statistics | filenames | notes

15.907 entries | Row 1 | Spanish | DE

Figura 5: Lista de las palabras más frecuentes en español

Para obtener las listas de Keywords de *Sportsnews_en* y *Sportsnews_es*, se debía proceder de manera diferente. En primer lugar, mi tutora me facilitó las listas de las palabras clave del inglés y del español a partir de los corpus monolingües de control, mencionados arriba, en formato lst. para poder abrirlas en WordSmith.

Con estas listas preparadas, ya podía crear las Keywords correspondientes haciendo clic en el botón *KeyWords*, *File* y *New*. En el cuadro de parámetros, en la pestaña *KeyWords*, debajo de *Wordlists*, por un lado, subí las Wordlists de *Sportnews_en* y *Sportnews_es* creadas en su momento y, por otro, las listas preparadas de los corpus de control con las palabras clave del inglés y del español (véase Figura 6).

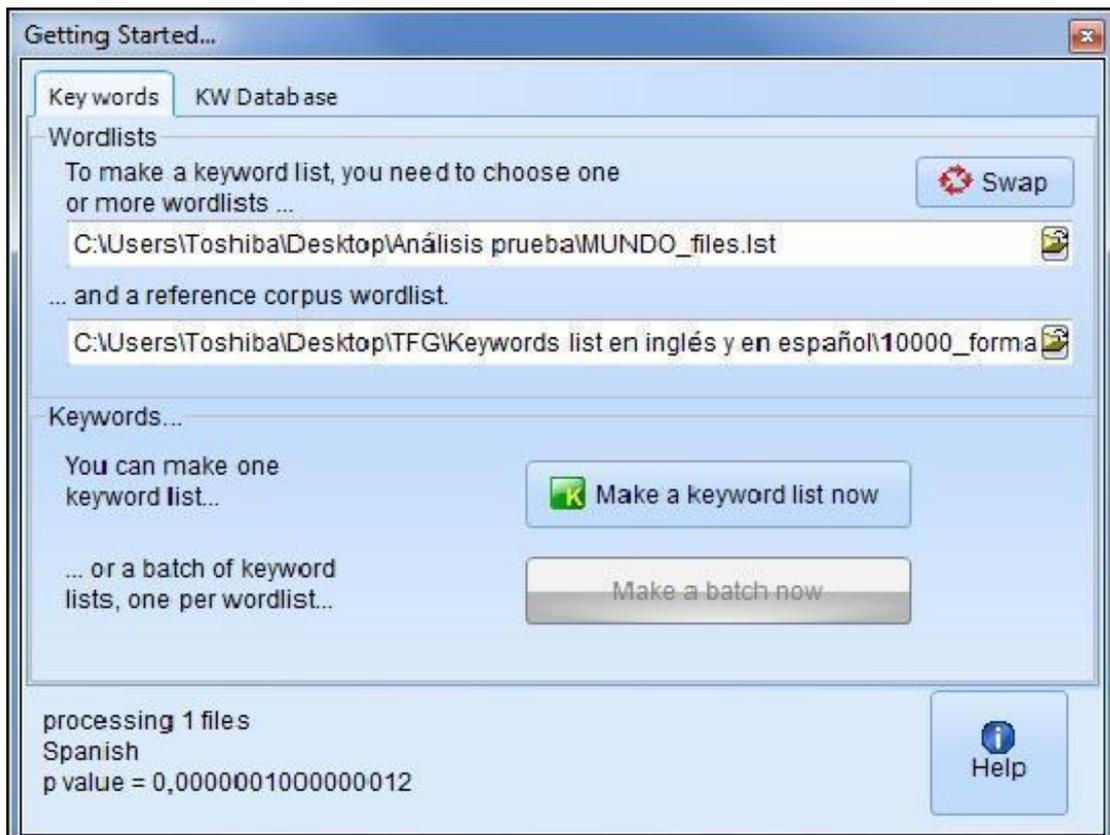


Figura 6: Análisis de las palabras clave

Finalmente, clasifiqué las primeras trescientas palabras clave de cada corpus por categorías en inglés (véase Anexo 1) y en español (véase Anexo 2).

Teniendo en cuenta el número de palabras clave del corpus, decidí examinar las primeras trescientas palabras clave obtenidas mediante las listas de Keywords.

3. Resultados de los análisis

En este estudio nos vamos a centrar en tres análisis. En primer lugar, describiremos el lenguaje baloncestístico de *Sportsnews_en* y *Sportsnews_es*, sobre todo en cuanto a estadísticas, sustantivos, pronombres, adjetivos y verbos. En el segundo análisis, se compararán los periódicos dentro de un mismo corpus, en cuanto a estadísticas y a sus diferencias más significativas. Cabe destacar que es interesante describir estas características y contrastarlas con la teoría presentada en la contextualización, puesto que consideramos que, para el traductor, puede ser relevante conocer los estilos de redacción de cada periódico. Por último, contrastaremos el corpus con el glosario centrándonos sobre todo en los términos, ya que, en teoría, en el lenguaje de los textos de baloncesto se emplea terminología especializada.

3.1. Descripción del lenguaje baloncestístico de *Sportsnews_en* y *Sportnews_es*

Para este estudio, utilizamos principalmente las listas de palabras clave de *Sportnews_en* y *Sportnews_es*, que contiene los vocablos más idiosincrásicos de ambos corpus.

En este apartado nos centraremos en el análisis de datos estadísticos y algunas características de los sustantivos, pronombres, verbos y adjetivos de los corpus estudiados. Llegado a este punto, cabe destacar que en este análisis no se aspira a afirmar que los rasgos identificados en este Trabajo de Fin de Grado son las características definitorias de la lengua baloncestística, sino que han de aplicarse exclusivamente a estos corpus y refrendarse con estudios posteriores.

Asimismo, conviene advertir de que, en nuestro estudio, descartamos los nombres geográficos por considerar que emiten una información ambigua, es decir, puede ser un metadato sobre el lugar en el que se publica el texto o ubicación de la noticia.

3.1.1.1. Datos estadísticos básicos

El *standardised TTR* (*types tokens ratio*), que encontramos en el cuadro de estadísticas de las listas de frecuencia de los textos de baloncesto de los cuatro periódicos (véase Anexo 3) y se obtiene dependiendo del número total de palabras de cada corpus, hace referencia a la cantidad de palabras diferentes que tiene cada corpus y, por lo tanto, el grado de su riqueza léxica; a mayor variedad, mayor riqueza del lenguaje.

Para empezar, exponemos a continuación los datos estadísticos relativos al *standardised TTR* de *Sportsnews_en* y *Sportsnews_es* (véase Tabla 1).

Tabla 1: Datos *standardised TTR* de *Sportsnews_en* y *Sportsnews_es*

Standardised TTR de Sportsnews_en		Standardised TTR de Sportsnews_es	
The New York Times	The New York Post	El País	El Mundo
43,59	47,25	44,40	45,80

Los resultados de las estadísticas muestran que los periódicos en español son más homogéneos mientras que los periódicos en inglés son más dispares. Además, teniendo en cuenta las cifras de *standardised TTR*, podríamos decir que, por ejemplo, *The New York Post* tiene mucha más riqueza léxica que *El País* o *El Mundo*. La valoración de estos datos, más allá de lo expuesto hasta ahora, no es posible si no existen trabajos futuros que corroboren nuestros resultados. En cualquier caso, esta información es interesante porque las estrategias traductorales pueden variar dependiendo de si el texto fuente u origen presentan una mayor o menor variedad léxica.

3.1.1.2. Sustantivos

Los sustantivos comunes, que no términos, que se utilizan en los corpus, tanto en inglés (véase Anexo 1) como en español (véase Anexo 2), son muy generales y de un registro no excesivamente formal. Se confirma, por lo tanto, la información del apartado de contextualización, sobre todo para el español, donde se explicaba que los artículos deportivos prefieren un lenguaje directo, sencillo y claro.

Por un lado, en inglés, aparecen muchos nombres comunes que hacen referencia al participante de la acción, como por ejemplo: *guys*, *men*, *girls* y *women*. A este respecto, es interesante destacar que, en esta lengua son más frecuentes los sustantivos que hacen referencia al hombre que los que aluden a la mujer, reflejando el dato sociológico de que el deporte masculino atrae mucho más la atención de la prensa que el femenino.

Además, los corpus nos aportan además una información que no contemplaban los estudios teóricos. Tanto el inglés como el español coinciden en un *sobreuso* significativo (recordemos que las palabras clave tienen valor estadístico) de nombres que tienen la función de ubicar el tiempo. De hecho, los tres primeros sustantivos clave del baloncesto en inglés son: *time*, *years* y *minutes*; y en español: *años*, *minutos* y *tiempo*.

Por otro lado, como constatan las estadísticas (véase Anexo 3), en general, los textos en inglés tienen una mayor riqueza léxica; sin embargo, como ya se ha mencionado anteriormente, los resultados del análisis de los sustantivos muestran que los sustantivos en español son más variados; por ejemplo, el español distingue *hombres* y *compañeros*.

3.1.1.3. Pronombres

En cuanto al estudio de los pronombres resultantes de los corpus en inglés (véase Anexo 1) y en español (Anexo 2), en *Sportnews_en* destacamos el hecho de que casi todos los pronombres son personales. Entre estos, podríamos destacar *they* y *we*; lo cual nos viene a confirmar la tendencia que se tiene, en el ámbito del deporte, a hablar del equipo, en conjunto, o a vincularse afectivamente con él.

En cambio, nuestros datos reflejan el hecho de que, en *Sportsnews_es*, apenas se utilizan los pronombres personales; por ello no sorprende que la mayoría de pronombres clave sean demostrativos: *eso, ese, esa, esos, esas*. Lo que resulta más interesante es que se eligen estos demostrativos frente a *este, esta, estos estas o aquel, aquella, aquellos, aquellas*. Este dato podría ser importante para decantar la traducción de vocablos tan normales y a veces complicados de trasvasar, como por ejemplo: *this/these, that/those*.

3.1.1.4. Verbos

El verbo está considerado como una de las categorías lingüísticas más relevantes. Por este motivo, deben tenerse en cuenta una serie de aspectos a la hora de traducir este tipo de palabras. Como sabemos, dentro del texto, los tiempos verbales se suelen utilizar de maneras distintas en inglés y en español.

Por un lado, los resultados del corpus en español confirman la información teórica (véase sección 1.1.1.) acerca de la tendencia que tiene el texto periodístico en español a utilizar verbos en presente perfecto, puesto que la primera forma verbal de la lista de las trescientas palabras clave es *han* (véase Tabla 7).

EP_EM_keywordlist.kws										
File Edit View Compute Settings Windows Help										
N	Key word	Freq.	%	Texts	RC. Freq.	RC. %	Keyness	P	Lemmas	Set
1	#	5.969	3,43	523	0		79.45	0,000		
2	QUE	4.537	2,61	459	0		60.35	0,000		
3	CON	2.166	1,24	439	0		28.78	0,000		
4	UNA	1.245	0,71	365	0		16.53	0,000		
5	MÁS	940	0,54	293	0		12.48	0,000		
6	COMO	885	0,51	278	0		11.75	0,000		
7	MADRID	607	0,35	204	0		8.06	0,000		
8	SUS	559	0,32	249	0		7.42	0,000		
9	DOS	521	0,30	248	0		6.91	0,000		
10	PUNTOS	475	0,27	251	0		6.30	0,000		
11	AÑOS	437	0,25	199	0		5.80	0,000		
12	BALONCESTO	358	0,21	223	0		4.75	0,000		
13	DESDE	351	0,20	177	0		4.66	0,000		
14	NBA	336	0,19	152	0		4.46	0,000		
15	TAMBIÉN	311	0,18	194	0		4.12	0,000		
16	CUANDO	295	0,17	169	0		3.91	0,000		
17	GASOL	292	0,17	123	0		3.87	0,000		
18	TEMPORADA	268	0,15	179	0		3.55	0,000		
19	PAU	260	0,15	106	0		3.45	0,000		
20	MUY	246	0,14	139	0		3.26	0,000		
21	ESPAÑA	243	0,14	117	0		3.22	0,000		
22	AUNQUE	230	0,13	136	0		3.05	0,000		
23	HAN	221	0,13	124	0		2.93	0,000		
24	MINUTOS	207	0,12	118	0		2.74	0,000		
25	PORQUE	204	0,12	118	0		2.70	0,000		

Figura 7: Lista de las primeras 25 palabras clave en español

Por otro lado, tras observar los *patterns* de *has*, descubrimos que en inglés también predomina de forma significativa el uso del presente perfecto (véase figura 8), lo cual demuestra que, como se ha mencionado en la parte teórica, puede que el lenguaje periodístico en inglés recurra más a los verbos en pasado, pero en el lenguaje baloncestístico de *Sportsnews_en* se utiliza más el presente perfecto.

N	L5	L4	L3	L2	L1	Centre	R1	R2	R3	R4	R5
1	THE	THE	THE	THE	HE	HAS	BEEN	THE	OF	THE	THE
2	IN	IN	OF	AND	WHO		HAD	TO	TO	OF	TO
3	TO	TO	IN	THAT	THAT		TO	BEEN	THE	AND	AND
4	AND	OF	THIS	OF	IT		WON	HIS	ON	TO	IN
5	OF	AND	AND	HE	AND		THE	IN	IN	IN	OF
6	ON	AS	THAT	SEASON	TEAM		NOT	MORE	WITH	FOR	THIS
7	HAS	SEASON	SAID	TEAM	WHICH		MADE	ONE	AND	THAT	WITH
8	WITH	FOR	NEW	THIS	SHE		BECOME	IT	HIS	AS	HE
9	HE	ON	YEAR	IN	STILL		LED	AND	GAMES	HIS	FOR
10	IT	IS	TO	BUT	ALSO		NEVER	AN	ALL	SEASON	IS
11	BUT	WITH	ON	SAID	STATE		DONE	CHANCE	AS	WITH	ON
12	IS	AT	LEAGUE	GAME	BUT		PLAYED	ON	FOR	ITS	HAS
13	FROM	THIS	BUT	THOUGH	KIDD		TAKEN	WITH	THAT	THIS	SEASON
14	THAT	BE	AS	KNICKS	STOUDE		MISSED	BE	AN	SINCE	THAT
15	THIS	ONE	WITH	HIS	JACKSON		LOST	THAT	LOT	FROM	AT
16	HIS	NOT	IS	AS	THERE		GONE	FOR	IS	ON	AS
17	LAST	13	COACH	STOUDE	DEFENS		SAID	THEM	POINTS	GAMES	HIS
18	NOW	THAT	BOB	COACH	WILLIAM		HELPED	LOT	STRAIGH	MORE	FROM
19	ALL	KNICKS	NO	NO	KRZYZEV		CHANGE	AS	IF	TEAM	GAMES
20	GAME	ARE	GUARD	STATE	BUTLER		IMPROVE	PLAYING	BY	HE	TEAM
21	AN	HIS	AT	SOMETH	FISHER		TWO	NEW	THIS	SAID	MOST
22	WOULD	GUARD	TIME	WHAT	ALWAYS		MORE	GOOD	AT	GAME	AN
23	FOR	HIGH	WHAT	MY	BASKET		ALWAYS	INTO	TEAM	TEAMS	PLAYER
24	SEASON	MY	FOR	JASON	WHAT		HIS	ITS	INTO	WHAT	SAID

Figura 8: *Patterns de has*

Además, del corpus en inglés, podemos destacar la variedad y abundancia de los verbos; de las trescientas palabras clave, más de un tercio son verbos, y sesenta y ocho de ellos son diferentes (véase Anexo 4). Así pues, se podría confirmar la variedad léxica que, como demuestran las estadísticas anteriormente expuestas, son características de estos textos en inglés. En cambio, de *Sportsnews_es*, de las trescientas palabras, sólo treinta y dos son verbos y diecinueve de ellos son diferentes (véase Anexo 5).

Según el corpus, la abundancia de los verbos no es precisamente característica de los textos en español. En otras palabras, se nos plantea la hipótesis de que en los textos de este ámbito se tiende a utilizar más verbos distintos en inglés que en español. No obstante, es interesante destacar que, en teoría, como observamos en el apartado de contextualización (sección 1.1.1.), el inglés tiende más a nominalizar mientras que el español prefiere el estilo verbal.

Seguidamente, etiquetamos los verbos clave en inglés y en español obtenidos de las listas siguiendo el criterio de clasificación de verbos de Halliday (2014) (véase Tabla 2).

Tabla 2: Análisis de verbos (Halliday)

Tipos de verbos	Verbos clave en español y en inglés del ámbito del baloncesto
Materiales	<p>EN: go, score, play, do make, win, come, include, miss, finish, start, take, call, look, give, throw, held, see, end, shoot, bring, coach, use, turn, visit, begin, leave, return, name, lead, try, rank, average</p> <p>ES: hacer, ganar, acabar, quedar, jugar, quedar, pasar, imponer, anotar, perder, poder</p>
Mentales	<p>EN: want, expect, know, feel</p> <p>ES: deber, esperar, saber</p>
Verbales	<p>EN: say, ask</p> <p>ES: ninguno</p>
Relacionales	<p>EN: be, seem, become, got, have</p> <p>ES: ser, estar, parecer, tener</p>
Existenciales	<p>EN: (there) be</p> <p>ES: haber</p>

Tras clasificar los resultados, descartamos el verbo *to be*, puesto que puede ser ambiguo (existencial, relacional o la base para otras construcciones verbales como *be able to*).

Basándonos en la tabla anterior, observamos que, tanto en inglés como en español, predominan los verbos materiales, lo cual parece lógico. El baloncesto es un deporte muy activo y en artículos sobre él se narra lo que sucede o se hace, por ejemplo, durante un partido. Podría decirse que esa actividad se ve reflejada en el lenguaje. En consecuencia, parece importante que el traductor esté especialmente preparado para describir acciones, más que para valorar procesos mentales o relacionar participantes con sus cualidades o posesiones.

3.1.1.5. Adjetivos

A través del análisis de los adjetivos confirmamos de nuevo que, en concreto, en este ámbito se utiliza de forma significativa un lenguaje muy general, puesto que los adjetivos, tanto en inglés (véase Anexo 1) como en español (véase Anexo 2), son los más básicos. Por ejemplo, los que más se utilizan en inglés son los superlativos: *best*, *biggest*, *worst* y *greatest*. En este sentido, en español hay mucha más variedad de adjetivos, aunque los más frecuentes son también los más básicos, como por ejemplo: *mejores*, *mayor* y *máximo*. No obstante, que haya más variedad en español parece poder explicarse por el hecho de que, en inglés, los sustantivos pueden utilizarse en posición adjetiva, rasgo poco común en español. Por otro lado, es interesante observar que muchos de los adjetivos analizados son superlativos, puesto que mediante su uso se demuestra la intensidad con la que se describen las acciones.

3.2. Comparación de periódicos dentro de un mismo corpus

Confrontamos aquí, los datos correspondientes a los diarios de cada uno de los corpus. De esta forma que lograremos así señalar características de divergencia entre los periódicos de cada corpus.

Si comparamos ahora las diferencias lingüísticas entre *The New York Times* y *The New York Post*, podemos apreciar rasgos diferenciales de todo tipo (desde el formato a la ideología), cuyo conocimiento por parte del traductor puede influir en su trabajo. Por ejemplo, se observa que el primer periódico escribe con puntos las siglas *N.C.A.A.* (*National Collegiate Athletic Association*) (véase figura 8), mientras que el segundo las usa sin puntos (véase Figura 9).

N	Concordance	Set	Tag
1	Next Challenge: The Atlantic 10 · N.C.A.A. Tournament: Best of the		
2	Next Challenge: The Atlantic 10 · N.C.A.A. Tournament: Best of the		
3	victory. Golden State became the 10th N.B.A. team to win at least 35 games		
4	led the Lakers and the Bulls to 11 N.B.A. titles. (PHOTOGRAPH BY		
5	Los Angeles, basking in a record 11 N.B.A. championships, flirting with		
6	31 points. Carlos Boozer, in his 13th N.B.A. season, and Jordan Clarkson,		
7	Connecticut (33-4) makes its 13th N.C.A.A. Final Four appearance since		
8	averaged 24.3 points per game in 15 N.B.A. seasons and was inducted into		
9	by Phog Allen that won the 1952 N.C.A.A. championship, and he		
10	Desk; Pg. 2 475 words The 1968 N.B.A. All-Star Game, at Madison		
11	Doherty, a forward for Smith's 1982 N.C.A.A. champions and later the		
12	than that, fire away. According to 1986 N.C.A.A. Division I statistics, the year		
13	Dean Smith after North Carolina's 1993 N.C.A.A. title. (PHOTOGRAPH BY		
14	Jackson after the Bulls won the 1998 N.B.A. title. The triangle offense fit		
15	words The indelible star of the 2006 N.C.A.A. tournament knows precisely		
16	Pg. 1 1019 words After the 2006 N.C.A.A. men's basketball tournament		
17	20-year scoring decline from the 2011 N.C.A.A. statistics. The typical college		
18	this analysis in March when the 2012 N.C.A.A. tournament field is		
19	. tournaments. This year marks the 27th N.C.A.A. tournament appearance for		
20	beyond the arc won 32 of the first 48 N.C.A.A. men's basketball tournament		
21	and the Utah Jazz opened the 1999-04 N.B.A. season in Tulsa, the league		

Figura 8: N.C.A.A. en The New York Post

Concord			
File Edit View Compute Settings Windows Help			
N	Concordance	Set	Tag
1	to ignore the fact Butler won 10 NCAA games in 2010 and 2011.		
2	" Hardy said. Keady took Purdue to 18 NCAA Tournaments, after assisting		
3	champion Louisville in the 1981 NCAA Tournament. Post columnist		
4	explaining to viewers watching the 1981 NCAA basketball tournament that they		
5	world by upsetting Houston in the 1983 NCAA championship game, the birth		
6	Rhode Island team in contention for an NCAA Tournament spot this season.		
7	No. 14 seed Davidson poised for an NCAA upset. Getty Images		
8	. The trial is scheduled for June and an NCAA loss there could force a		
9	Arena over South Florida, an NCAA Tournament success story last		
10	and eight of nine to emerge as an NCAA Tournament team. St. John's		
11	. It sets up another epic UConn-Duke NCAA Tournament game. In the		
12	(16-5, 8-2) on pace to reach its first NCAA Tournament since 1999, led by		
13	. He beat Northeastern in his first NCAA Tournament game as the		
14	, leading the Jaspers to their first NCAA Tournament appearance in a		
15	play it, anyway. For all the four-flushing NCAA Tournament expert TV analysis		
16	LANGUAGE: ENGLISH GRAPHIC: : NCAA jerseys: Lasalle 55 Ramon		
17	THE BEST One man's Top 10 greatest NCAA Tournament shots in the last 35		
18	one of the most improbable upsets in NCAA Tournament history. Wisconsin		
19	as the unheralded Blackbirds will be in NCAA Tournament next week. AP;		
20	and No. 3 in front of Marquette on its NCAA Tournament bracket line. It sees		
21	Thursday NA Syracuse gives itself NCAA ban Zach Braziller Late City		
22	Post March 21, 2013 Thursday NA <title>NCAA Final Four AN		

Figura 9: NCAA en The New York Post

Por lo que respecta *El País* y *El Mundo*, existen muy pocas diferencias entre ambos. Sin embargo, es interesante destacar que el primero utiliza más palabras que hacen referencia a la mujer; como por ejemplo: *la*, *femenino* o *esa* (véase Figura 10).

N	Key word	Freq.	%	Texts	RC. Freq.	RC. %	Keyness	P	Lemmas	S
1	SOLO	105	0,13	70	16	0,02	81,69	0,0000		
2	HIGHLIGHT	55	0,07	55	0		81,43	0,0000		
3	LA	3.034	3,65	348	2.747	3,02	54,60	0,0000		
4	ANTE	233	0,28	137	134	0,15	36,98	0,0000		
5	DE	4.602	5,54	358	4.482	4,92	33,49	0,0000		
6	ESPAÑOLA	77	0,09	58	25	0,03	32,81	0,0000		
7	FEMENINO	58	0,07	28	15	0,02	31,16	0,0000		
8	EQUIPO	365	0,44	160	259	0,28	29,25	0,0000		
9	MENOS	34	0,04	27	104	0,11	-31,10	0,0000		
10	ESA	29	0,03	25	102	0,11	-36,70	0,0000		
11	TODO	88	0,11	55	203	0,22	-36,82	0,0000		
12	NO	570	0,69	155	881	0,97	-42,03	0,0000		
13	#	2.602	3,13	339	3.367	3,70	-42,09	0,0000		
14	SÓLO	39	0,05	23	159	0,17	-67,47	0,0000		
15	MADRID	186	0,22	91	421	0,46	-73,35	0,0000		
16	AUNQUE	46	0,06	36	184	0,20	-76,57	0,0000		

Figura 10: Diferencias entre *El País* y *El Mundo*

En el siguiente apartado, abandonamos el estudio de diarios independientes para retomar el análisis de *Sportnews_en* y *Sportnews_es*, con el fin de identificar la información más característica que arrojan los corpus frente a las de una de las fuentes de documentación estrella del traductor: un glosario especializado. Nos centraremos para hacerlo en el contraste de términos baloncestísticos propiamente dichos.

3.3. Comparación de corpus y glosario

Como ya hemos avanzado, en este apartado centramos nuestro estudio en el análisis de algunos términos muy ilustrativos. Más en concreto, y de nuevo por motivos de espacio, nos centramos exclusivamente en *Sportnews_en* y comparamos los datos de este corpus con el glosario especializado para distinguir la información añadida que aportan los corpus.

Para empezar, destacamos el término *guard*. Según el glosario, los *guards* son el jugador base y el escolta; no obstante, en nuestro corpus encontramos *guard* en singular. El conflicto se presenta cuando en el glosario aparece este término dos veces en singular acompañado por dos adjetivos *point guard* (base) y *shooting guard* (escolta). Por lo tanto, en este caso, si en el texto original no se entiende a qué jugador concreto se refiere, el glosario no es suficiente para saber cómo traducir este término, puesto que se necesita información contextual para hacerlo (véase Figura 11).

N	Concordance	Set	Tag	Word #
1	Holman at City College. A 5-foot-11 guard, he played with Red Holzman,			535
2	, 6-foot guard. David Stockton, 5-11 guard. Steven Gray, 6-foot-5 guard.			516
3	in question was Jonte Hall, a 5-foot-2 guard known as Too Tall. And then			790
4	, 5-11 guard. Steven Gray, 6-foot-5 guard. Mike Hart, six-foot-six guard.			520
5	prospect in Delon Wright, a 6-foot-5 guard/forward who can do a little bit of			551
6	to Dayton, Ohio, just as Skip Dillard, a guard for the No. 1 team in the nation,			469
7	he would kick the ball back out to a guard, re-establish his position, and			394
8	be "aggressive" in looking at free agent guard J.R. Smith. Smith, a former			60
9	up memories of Hakim Warrick and guard Juan Fernandez is as steady,			532
10	said. In this game, Gallagher and guard Alexa Dietrich epitomized			293
11	Craft will get a huge test from Arizona guard Mark Lyons, who has had 50			912
12	Devils with another talented scorer at guard, and one that can create his			412
13	power St. Rita and Kentucky-bound guard Charles Matthews. Alkins had			288
14	Butler winning, 71-70. The former Butler guard Shelvin Mack committed a foul			322
15	Cavs led 62-52 on two free throws by guard Malcolm Brogdon with 4:21			215
16	STARS OF THE DAY n Cincinnati guard JaQuon Parker posted a			611
17	It is tough. Sawy. Cerebral. Clever. Guard-driven. And year-round. Those			46
18	The N.B.A. fined Los Angeles Clippers guard Chris Paul \$25,000 for criticizing			413
19	, it was announced that Clippers guard Chauncey Billups had a torn			98
20	, just for ourselves," Connecticut guard Renee Montgomery said. "The			141
21	the way in the defense on Holy Cross guard Keith Simmons, the Patriot			538

Figura 11: Lista de concordancias del término "guard"

Por último, en este corpus distinguimos una palabra clave, *starters*, que no se encuentra en el glosario. Se trata de un término que se utiliza en muchos contextos del deporte en general y que, por ejemplo, en fútbol sería equivalente al *once inicial*, pero en baloncesto hace referencia al *quinteto inicial*. Sin embargo, para saber con certeza la equivalencia de este término, un corpus sería de gran ayuda, puesto que nos muestra el uso de estos términos en el lenguaje real. Un traductor no experto en este ámbito no tiene por qué saber que es *quinteto* y no, por ejemplo, *cinco inicial*, como se deduciría a partir del término que se emplea en fútbol: *once*.

4. Conclusiones

Este apartado de conclusiones lo dividiré en dos secciones. Por un lado, presentaré una serie de conclusiones acerca de la información lingüística que resulta del estudio y, por otro, haré hincapié en la relación de los corpus con el ámbito de la traducción.

En este trabajo teníamos el objetivo de demostrar la importancia de los corpus para la traducción y, en particular, analizar qué tipo de información pueden aportar para complementar a otros recursos lingüísticos. Para ello, hemos compilado *Sportsnews* y hemos realizado tres tipos de análisis.

Del primer análisis, la primera conclusión que resulta evidente es que los corpus son fuentes de terminología muy prolíficas. Durante la práctica traductora, permiten reducir el tiempo de documentación, puesto que están en formato digital y pueden explotarse mediante herramientas de software, y además, examinar las diferencias entre las equivalencias de términos o construcciones y realizar el estudio de la terminología especializada de un ámbito. Asimismo, este tipo de recursos pueden mostrar al traductor características de un lenguaje determinado en una lengua en concreto de manera inmediata. Por último, podría decirse que, a través de los corpus, se tiene la certeza de que estos resultados son el reflejo de un lenguaje de uso real de textos actuales.

Los resultados de este primer análisis también revelan que, en general, el tipo de textos examinados aquí, tanto en inglés como en español, utiliza una terminología precisa, pero no muy especializada. No obstante, se han obtenido datos muy interesantes que confirman o refutan las ideas de los expertos de este lenguaje. Por ejemplo, el corpus corrobora muchos de los datos que proponen los expertos, por lo tanto, a través de este recurso el traductor gana seguridad. En cambio, al contrario de lo que defienden los expertos respecto al estilo del lenguaje periodístico, este análisis nos sugiere que en este ámbito en concreto, el español prefiere el estilo nominal al estilo verbal.

El segundo análisis nos sugiere que, por un lado, a través del corpus podemos obtener información relevante para redactar de una manera u otra dependiendo del cliente de la traducción. Sin ir más lejos, existen diferencias, por ejemplo, de carácter tipográfico entre periódicos. Por otro lado, se han obtenido resultados que apuntan a conclusiones ideológicas más que lingüísticas, pero un traductor siempre se beneficia de un conocimiento certero de la ideología del medio en el que se ha publicado el texto meta, que sin duda le influye en su toma de decisiones en todos los niveles.

Por último, del tercer análisis se deriva la conclusión de que, los corpus aportan al glosario ejemplos y el contexto de los términos. Muchas veces, el traductor tiene que elegir de entre muchos términos el más apropiado y algunos glosarios, como el que se ha utilizado en este trabajo, solamente nos proporcionan equivalencias. A través de las listas de concordancias, se puede observar, por ejemplo, cómo se emplean los términos en la oración, con qué colocan o con qué frecuencia. Por lo tanto, el traductor se beneficia de información que, como hemos visto, el glosario no suele ofrecer.

En líneas generales, el glosario contiene casi todos los términos que propone *Sportsnews*. Sin embargo, en el glosario se obvian los términos más generales que precisamente, como vienen a confirmar los corpus, son los que más se utilizan en la traducción periodística y en concreto, en los textos de baloncesto.

En definitiva, tras observar las conclusiones de los tres análisis, podría decirse que los resultados son suficientes para demostrar que un archivo de corpus es un complemento ideal para cualquier otro recurso, puesto que el traductor se beneficia de muchas enseñanzas, desde diferentes puntos de vista lingüísticos, y, a la vez, los corpus le acercan la posibilidad de ponerse en la piel de otros expertos y conocer muchos de los aspectos característicos de otros ámbitos.

En cuanto a la relación del trabajo con la traducción, podría decirse que, muchos lingüistas sostienen que, para traducir o interpretar, es muy importante reunir una serie de competencias textuales y comunicativas. Desde luego, es indudable que, a través los tres análisis anteriores, observamos implícita o explícitamente que de los corpus se puede extraer, por un lado, información especializada para traductores que no tienen por qué conocer todos los ámbitos a fondo y, por otro lado, información lingüística que no resulta tan evidente. En este sentido no cabe duda de que la metodología y la lingüística de corpus es un recurso muy competente para el mundo de la traducción.

Para terminar, cabe destacar que me he sentido muy cómoda trabajando este ámbito y en un futuro me gustaría seguir analizando varios aspectos lingüísticos que, por motivos de espacio, me he dejado por estudiar. Por ejemplo, me gustaría indagar más sobre la riqueza léxica de los textos, puesto que considero que es una información lingüística muy relevante en la traducción. De igual manera, me gustaría tener la oportunidad de aplicar los conocimientos adquiridos sobre la metodología de corpus, puesto que considero que es muy importante que el traductor pueda beneficiarse de unas fuentes terminológicas tan grandes. Este trabajo me ha permitido completar la excelente formación que he recibido en el Departamento de Traducción y Comunicación.

5. Bibliografía

- Baker, M. (1995). Corpora in Translation Studies. An Overview and Suggestions for Future Research. *Target* 7(2). Recuperado de <http://wenku.baidu.com/view/505899fcf61fb7360b4c65ac.html> el 2 de febrero de 2015
- Baker, M. (1996). *Corpus-based translation studies: the challenges that lie ahead*. En H. Somers (Ed.).
- Bernárdez, J. (1998). *El deporte correctamente hablado*. Universidad de Oviedo. Vicerrectorado de Extensión Universitaria. Servicio de Publicaciones.
- Bernardos, M^a S. (s.f.). Calíope: herramienta para gestionar un corpus y un glosario de términos informáticos. *Archivo Digital UPM*. Recuperado de http://oa.upm.es/6506/1/Cal%C3%ADope,_herramienta.pdf el 5 de marzo de 2015
- Blasco, E.J. (25 de junio de 2010). 70-68, once horas, set y partido. *ABC.es*. Recuperado de <http://hemeroteca.abc.es/nav/Navigate.exe/hemeroteca/madrid/abc/2010/06/25/093.html> el 1 de abril de 2015
- Boyle, R. (2006). *Sports journalism context and issues*. Londres: SAGE. Recuperado de http://books.google.es/books/about/Sports_Journalism.html?id=ffE31oI21wQC&redir_esc=y el 2 de enero de 2015
- Cabré, M^a T. El traductor y la terminología: necesidad y compromiso. *Revista de traducción Medtrad*. Recuperado de http://www.medtrad.org/panacea/IndiceGeneral/n2_EditorialCabre.pdf el 22 de marzo de 2015
- Cacciola, S. (19 de marzo de 2015) In a Battle of the Lowly the Knicks End Up a Bit Lower. *The New York Times*. Recuperado de <http://www.nytimes.com/2015/03/20/sports/basketball/nba-roundup.html> el 27 de marzo de 2015
- Castañón, J. (s.f.) El lenguaje deportivo del nuevo siglo. *Idioma y deporte*. Recuperado de <http://www.idiomaydeporte.com/articulos/el-lenguaje-deportivo-del-nuevo-siglo.php> el 6 de enero de 2015
- Castañón, J. (2012). El lenguaje periodístico del deporte en el idioma español del siglo XXI. Historia y Comunicación Social. *Revistas científicas complutenses*. Recuperado de: <http://revistas.ucm.es/index.php/HICS/article/view/40613> el 6 de enero de 2015
- De Arcos, A. (2011). *Tratamiento de la información deportiva en la prensa: la crónica como género prevalente. El caso de los encuentros de fútbol entre Real Madrid y F.C Barcelona*. [Tesis inédita de doctorado, Universidad de Málaga].

Recuperado de
http://riuma.uma.es/xmlui/bitstream/handle/10630/4848/TDR_NARANJO_ARC OS.pdf?sequence=6 el 4 de enero de 2015

Desde el banquillo.es. Recuperado de <http://www.desdeelbanquillo.es/> el 5 de febrero de 2015

Dickson, P., & Skole, R. (2012). *Journalese: A Dictionary for Deciphering the News*. Marion Street Press.

Halliday, M., & Matthiessen, C. (2014). *Halliday's introduction to functional grammar (Fourth Ed.)*.

Hernández Guerrero, M^aJ. (2008). La traducción periodística en los diarios españoles de información general. *Asociación Ibérica de Estudios de Traducción e Interpretación*. Recuperado de http://www.aieti.eu/pubs/actas/III/AIETI_3_MJHG_Traduccion.pdf el 22 de febrero de 2015

Hernández, N. (2012). *Tendencias en el lenguaje deportivo actual*. Madrid: Vision Libros.

Hernando, L.A. (2002). Sobre la configuración lingüística del mensaje periodístico. *Estudios sobre el mensaje periodístico. Universidad Complutense de Madrid*. Recuperado de http://pendientedemigracion.ucm.es/info/emp/Numer_08/Art/4-09-1.pdf el 10 de febrero de 2015

Kantzow, M., & Stubbs, S. (2000). *Targeting Media*. Glebe, N.S.W.: Blake Education. Recuperado de http://books.google.es/books/about/Targeting_Media.html?id=qZ4PFpQpWdEC&redir_esc=y el 19 de diciembre de 2014

Lara, D. (2011). *Estudio morfosintáctico contrastivo de titulares de periódicos en inglés y en español*. [Tesis presentada a la Escuela de Lenguas y Traducción de la Universidad Católica de Temuco]. Recuperado de http://www.academia.edu/4045889/Estudio_morfosint%C3%A1ctico_contrastivo_de_titulares_de_peri%C3%B3dicos_en_Ingl%C3%A9s_y_en_Espa%C3%B1ol el 15 de febrero de 2015

Lázaro, F. (1977). *Lenguaje en periodismo escrito*. Madrid: Fundación Juan March.

LexisNexis. Recuperado de <http://academic.lexisnexis.nl/> el 19 de diciembre de 2014

Martínez Moreno, R. (2013). Breve análisis del lenguaje periodístico deportivo en España. *Papel de periódico*. Recuperado de <http://papeldeperiodico.com/2013/09/25/breve-analisis-del-lenguaje-periodistico-deportivo-en-espana/> el 11 de enero de 2015

Merriam Webster. Recuperado de <http://www.merriam-webster.com/> el 10 de febrero de 2015

Oliva, C. Fundamentos lingüísticos del lenguaje periodístico y jurídico-administrativo (2011). *Index.comunicación: Revista científica en el ámbito de la*

Comunicación Aplicada. Recuperado de <http://dialnet.unirioja.es/servlet/articulo?codigo=3689640> el 21 de enero de 2015

- Olohan, M. Corpus Linguistics and Translation Studies: Interaction and Reaction. *Linguistica Antverpiensia, New Series – Themes in Translation Studies (LANS – TTS)*. Recuperado de <https://lans-tts.ua.ac.be/index.php/LANS-TTS/index> el 2 de enero de 2015
- Ospina, D. (s.f.) El Glosario como estrategia didáctica. *Aprende en Línea. Programa de Integración de Tecnologías a la Docencia - Universidad de Antioquia*. Recuperado de <http://aprendeenlinea.udea.edu.co/boa/contenidos.php/6d62732f9c16ddb566a8a61062cf478f/19/estilo/aHR0cDovL2FwcmVuZGVlbmxbmVhLnVhZWEuZWwR1LmNvL2VzdGl3b3MvYXp1bF9jb3Jwb3JhdGl2by5jc3M=/1/contenido/> el 20 de febrero de 2015
- Real Academia Española. Recuperado de <http://www.rae.es/> el 18 de marzo de 2015
- Reppen, R. (2010). Building a corpus. What are the basics? En O' Keeffe, A & Michael McCarthy (Eds.), *The Routledge handbook of corpus linguistics*. Londres: Routledge.
- Rojas, J.L. (2013). Periodismo deportivo. Hacia la innovación y emprendimiento en la red. En Sobrados, M. *Presente y futuro en el periodismo especializado*. Madrid: Fragua.
- Rojas, J.L. (2014). Extranjerismos necesarios y superfluos en las crónicas de baloncesto. *Periodismo deportivo de calidad*. Recuperado de <http://periodismodeportivodecalidad.blogspot.com.es/2014/08/extranjerismos-necesarios-y-superfluos.html> el 10 de febrero de 2015
- Rojo, G. (2009). Sobre la construcción de diccionarios basados en corpus. *Revista tradumàtica. Universitat Autònoma de Barcelona*. Recuperado de <http://webs2002.uab.es/tradumatica/revista/num7/articles/02/02.pdf> el 13 de marzo de 2015.
- Rushin, S. (2010). The specialized language of sports. *SI*. Recuperado de <http://www.si.com/more-sports/2010/08/11/word-play> el 25 de febrero de 2015
- Sánchez, A & Cantos, P. (1998). El ritmo incremental de palabras nuevas en los repertorios de textos. Estudio experimental y comparativo basado en dos corpus lingüísticos equivalentes de cuatro millones de palabras, de las lenguas inglesa, y española y en cinco autores de ambas lenguas. *Departamento de Filología Inglesa. Universidad de Murcia*. Recuperado de <http://www.um.es/lacell/miembros/asp/textos/Atlantis.htm> el 5 de abril de 2015
- Terminologías y glosarios de traducción: Cuál es su importancia. (2012). *JR Language Translation Services, inc*. Recuperado de http://es.jrlanguage.com/blog-de-traducccion/index.php/terminologias-y-glosarios-de-traducccion-cual-es-su-importancia_00368/ el 2 de febrero de 2015

Tognini, E. (2010). Theoretical overview of the evolution of corpus linguistics. En O'Keefe, A & Michael McCarthy (Eds.), *The Routledge handbook of corpus linguistics*. Londres: Routledge.

Vargas, C. (2006). Diseño de un corpus especializado con fines terminográficos: el Corpus de la Piedra Natural. En *Universidad de Alicante*. Recuperado de http://www.ua.es/personal/chelo.vargas/Documentos/chelo_vargas_sierra.pdf el 1 de abril de 2015

Villayandre, M. Lingüística Computacional II. Curso monográfico sobre Lingüística de corpus. *Departamento de Filología Hispánica y Clásica. Universidad de León*. Recuperado de <http://fhyc.unileon.es/Milka/LCII/Corpus4.pdf> el 22 de marzo de 2015

Yohong, L. (s.f.). Intérpretes deportivos acarician su oportunidad. *China Today*. Recuperado de <http://www.chinatoday.com.cn/hoy/2007n/s2007n11/p80.htm> el 12 de enero de 2015

6. Anexos

Anexo 1

Sustantivos	<ol style="list-style-type: none"> 1. TIME 2. YEARS 3. MINUTES 4. GUYS 5. TIMES 6. SECONDS 7. THINGS 8. MEN'S 9. SCHOOL 10.STATE 11.GIRLS 12.DAYS 13.WOMEN'S 14.YEAR'S 15.PEOPLE 16.MONTHS 17.PROGRAMS 18.SCHOOLS 19.KIDS 20.HANDS
Pronombres	<ol style="list-style-type: none"> 1. IT'S 2. I

	<ol style="list-style-type: none"> 3. THAT 4. HE'S 5. I'M 6. WE'RE 7. THEY'RE 8. YOU'RE 9. I'VE 10. WE'VE 11. WE'LL 12. YOU'VE 13. THAT'S 14. WHAT'S 15. THIS 16. THEY'VE
<p>Verbos</p>	<ol style="list-style-type: none"> 1. SAID 2. BEEN 3. GOING 4. SCORED 5. PLAYED 6. MADE 7. GOT 8. WON 9. PLAYING 10. WINNING 11. LED 12. ADDED 13. TOOK 14. LOST 15. WENT 16. CAME 17. TOLD 18. SEEDED 19. KNOW 20. PLAY 21. LOSING 22. GETTING 23. HAVING 24. THERE'S 25. LEADING 26. COMING 27. INCLUDING 28. MAKING 29. MISSED 30. FINISHED 31. STARTED 32. SCORING 33. ASKED 34. TRYING 35. WIN

	36. DONE 37. USED 38. DOING 39. COACHING 40. WANTED 41. TAKE 42. RANKED 43. CALLED 44. BEING 45. LOOKING 46. GAVE 47. STARTING 48. GO 49. MAKES 50. HELD 51. LOOKED 52. SEEN 53. FELT 54. ENDED 55. SHOOTING 56. HELPED 57. BECAME 58. COMES 59. TURNED 60. VISITING 61. BEGAN 62. EXPECTED 63. SEEMED 64. KNOWN 65. REMAINING 66. SEEMS 67. KNOWS 68. GIVEN 69. KNEW 70. SAW 71. AVERAGING 72. BROUGHT 73. RETURNED 74. TAKING 75. FELL 76. GOES 77. GETS 78. TALKING 79. WATCHING 80. NEEDED 81. COACHED 82. SUPPOSED 83. RECEIVED 84. SHOWED
--	---

	85. SEED 86. BEAT 87. WATCHED 88. COMBINED 89. SENT 90. REBOUNDED 91. LIKE 92. INJURED 93. APPEARED 94. SAY 95. AVERAGED 96. LOOKS 97. SPENT 98. ENTERING 99. WORKED 100. ANNOUNCED 101. MISSING 102. BLOCKED 103. DEFENDING
Verbos auxiliares y modales	1. WAS 2. IS 3. HAD 4. HAS 5. WERE 6. ARE 7. DID 8. DON'T 9. DIDN'T 10. DOES 11. CAN'T 12. WASN'T 13. DOESN'T 14. ISN'T 15. WON'T 16. COULDN'T
Adverbios	1. NO
Adjetivos	1. LAST 2. FOUR 3. FIRST 4. BEST 5. BIGGEST 6. ONE 7. NEW 8. WELL 9. WORST 10. NATIONAL 11. GREATEST

Preposiciones o locuciones preposicionales	1. ON 2. OUT 3. TO 4. UP 5. AS 6. AFTER 7. OFF 8. BECAUSE 9. ACCORDING 10. BEFORE
Conjunciones o locuciones conjuntivas	1. SO
Términos del baloncesto y del deporte en general	1. POINTS 2. PLAYERS 3. TEAM 4. GAMES 5. TEAMS 6. BASKETBALL 7. REBOUNDS 8. GAME 9. COACH 10. SHOTS 11. FANS 12. TOURNAMENT 13. ASSISTS 14. SEASON 15. FINALS 16. SPORTS 17. PLAYOFF 18. OFFENSE 19. COACHES 20. TEAM'S 21. PLAYOFFS 22. SEASONS 23. PLAYS 24. TEAMMATES 25. GUARD 26. POINTER 27. ROUND 28. TITLE 29. FINAL 30. THROWS (free) 31. OVERTIME 32. POINTERS 33. TURNOVERS 34. WINS 35. FORWARD 36. SEMIFINALS

	<p>37.FREE 38.SHOT 39.LOSSES 40.CHAMPIONSHIPS 41.STARTERS 42.TOP 43.LEAGUE’S 44.DEFENSE 45.DUNK 46.OPPONENTS 47.MATCHUP 48.LINEUP 49.POSTSEASON 50.TITLES 51.LOCKER 52.FRESHMAN 53.STREAK 54.VICTORIES 55.HALFTIME 56.GUARDS 57.CHAMPIONSHIP 58.SOPHOMORE 59.SENIORS</p>
<p>Nombres propios</p>	<p>1. KNICKS 2. NBA 3. N.C.A.A 4. NETS 5. JAMES 6. STOUDEMIRE 7. WILLIAMS 8. CELTICS 9. SPURS 10.LAKERS 11.NCAA 12.HAWKS 13.LEBRON 14.JOHN’S 15.CHRIS 16.STAR 17.WARRIORS 18.CLIPPERS 19.UCONN 20.SYRACUSE 21.CAVALIERS 22.GARNETT 23.NOTRE 24.BULLS 25.MAVERICKS 26.ESPN</p>

	27.CARMELO 28.ANTHONY 29.STARS 30.RAPTORS 31.THOMAS 32.AMAR'E 33.PISTONS 34.RUTGERS 35.PELICANS 36.DAVIS 37.FISHER 38.DUKE 39.PACERS 40.ROOKIE 41.ROSTER 42.HEAT 43.A.B.A. 44.JACKSON 45.MILWAUKEE 46.BLAZERS 47.KIDD 48.KOBE 49.BARCLAYS 50.DURANT
Nombres geográficos	1. DALLAS 2. KANSAS 3. LOUSVILLE 4. MEMPHIS 5. LOS 6. ANGELES 7. ORLEANS 8. TEXAS 9. EAST 10.ORLANDO 11.BROOKLYN 12.NORTH

Anexo 2

Sustantivos	1. AÑOS 2. MINUTOS 3. AÑO 4. TIEMPO 5. MUNDO 6. VECES 7. MOMENTO
--------------------	--

	8. COSAS 9. CABEZA 10. CUENTA 11. CAMINO 12. NOCHE 13. CASA 14. MANERA 15. HOMBRE 16. PROBLEMAS 17. MESES 18. MOMENTOS 19. ESTADO 20. SEGUNDOS 21. ESTADOS 22. SEMANA 23. MES 24. COMPAÑEROS 25. HOMBRES 26. FRANQUICIA 27. EJEMPLO 28. DUDA 29. PASO 30. CAPAZ 31. METROS 32. MANO 33. SEPTIEMBRE 34. CHOQUE 35. DUDAS 36. EUROS
Pronombres	1. QUE 2. SUS 3. ESO 4. ME 5. ESE 6. NOS 7. ESA 8. ESOS 9. ESAS
Verbos	1. HACE 2. HAY 3. SON 4. HECHO 5. PARECE 6. GANÓ 7. SEA 8. TENEMOS 9. ACABÓ 10. QUEDA 11. JUGANDO

	<ul style="list-style-type: none"> 12. QUEDÓ 13. ESPERA 14. PASÓ 15. GANADO 16. DADO 17. SABE 18. IMPUSO 19. JUGAR 20. JUGÓ 21. JUEGA 22. PERDIÓ
Verbos auxiliares y modales	<ul style="list-style-type: none"> 1. HAN 2. ESTABA 3. HEMOS 4. PUDO 5. ESTAMOS 6. HUBO 7. DEBE 8. HAYA 9. PUEDEN
Adverbios	<ul style="list-style-type: none"> 1. MÁS 2. COMO 3. HOY 4. TAMBIÉN 5. CUANDO 6. MUY 7. DESPUÉS 8. SIEMPRE 9. MUCHO 10. DONDE 11. MENOS 12. NUNCA 13. ADEMÁS 14. NADA 15. POCO 16. AÚN 17. QUÉ 18. ENTONCES 19. APENAS 20. MUCHOS 21. CÓMO 22. TAMPOCO 23. MAÑANA 24. DEMASIADO 25. ESPECIALMENTE 26. AYER
Adjetivos	<ul style="list-style-type: none"> 1. DOS 2. MUNDIAL 3. ESPAÑOL

	<ol style="list-style-type: none"> 4. UNO 5. PASADO 6. UNA 7. CADA 8. SEGUNDO 9. UNOS 10. MISMO 11. ESPAÑOLA 12. SEGUNDA 13. OCHO 14. GRANDES 15. MEJORES 16. EUROPEO 17. NUEVO 18. BLANCOS 19. FEMENINO 20. MAYOR 21. MEDIA 22. AMBOS 23. PRÓXIMO 24. NUEVA 25. MISMA 26. NUEVE 27. PASADA 28. ALGUNOS 29. MAL 30. MÁXIMO 31. ESPAÑOLES 32. BUEN 33. MEDIO 34. ESTADOUNIDENSE 35. MUCHAS 36. BUENA 37. MÁXIMA 38. AZULGRANA 39. GRANDE 40. ALGUNAS 41. CINCO 42. CANARIA 43. MEJOR 44. CUATRO 45. TERCER 46. SIETE
<p>Preposiciones o locuciones preposicionales</p>	<ol style="list-style-type: none"> 1. CON 2. DESDE 3. PORQUE 4. SEGÚN 5. TRAS 6. ANTE

Conjunciones o locuciones conjuntivas	<ol style="list-style-type: none"> 1. PESE 2. PUES 3. EMBARGO
Términos del baloncesto y del deporte en general	<ol style="list-style-type: none"> 1. PUNTOS 2. BALONCESTO 3. TEMPORADA 4. REBOTES 5. EQUIPO 6. EQUIPOS 7. DEFENSA 8. TRIPLES 9. PÍVOT 10. CANCHA 11. COPA 12. LIGA 13. ATAQUE 14. BASE 15. ENCUENTRO 16. ASISTENCIAS 17. PUESTO 18. CAMPEÓN 19. COMPETICIÓN 20. DESCANSO 21. FINAL 22. DERROTAS 23. BASKET 24. PARTIDO 25. JUGADOR 26. PÍVOTS 27. CAMPEONATO 28. SELECCIONADOR 29. JUGADORES 30. BANQUILLO 31. TEMPORADAS 32. ALERO 33. REBOTE 34. CANASTA 35. PRÓRROGA 36. MARCADOR 37. SELECCIÓN 38. ENTRENADOR 39. QUINTETO 40. CAMBIO 41. MVP 42. CONCENTRACIÓN 43. TOP 44. JUEGO 45. ZONA 46. CUARTOS

	<ol style="list-style-type: none"> 47. RIVAL 48. DERROTA 49. VICTORIAS 50. TRIPLE 51. LESIÓN 52. BALÓN 53. TIROS 54. PISTA 55. TÉCNICO 56. INICIAL 57. TIRO 58. DUELO 59. RIVALES 60. TORNEO 61. ESTRELLA 62. ACIERTO 63. FALTAS 64. JORNADA 65. CLUB 66. ESCOLTA 67. DEPORTES 68. PLANTILLA 69. JUEGOS 70. VENTAJA 71. AFICIONADOS
<p>Nombres propios</p>	<ol style="list-style-type: none"> 1. NBA 2. GASOL 3. PAU 4. MARC 5. EUROLIGA 6. LASO 7. UNICAJA 8. RUDY 9. LAKERS 10. BULLS 11. MIROTIC 12. LLULL 13. ACB 14. KUTXA 15. BARÇA 16. ORENGA 17. ENDESA 18. SAN 19. MACCABI 20. ABRINES 21. BASKONIA 22. RICKY 23. TOMIC 24. PASCUAL

	<ol style="list-style-type: none"> 25. BARCELONA 26. CSKA 27. LEBRON 28. OLYMPIACOS 29. GRIZZLIES 30. JAMES 31. CASO 32. CAI 33. KOBE 34. BRYANT 35. CARROLL 36. XAVI 37. JACKSON 38. JOVENTUT 39. OLESON 40. PANATHINAIKOS 41. PALAU 42. SLAUGHTER 43. BOUROUSIS 44. NAVARRO 45. NOCIONI 46. EUROBASKET 47. SPURS 48. SERGIO 49. PLAYOFFS 50. GALATASARAY 51. MEJRI 52. DOELLMAN 53. NACHBAR 54. DEPORTE 55. PARTIDOS 56. REAL 57. REYES 58. RODRÍGUEZ 59. LABORAL 60. FELIPE 61. CALDERÓN
<p>Nombres geográficos</p>	<ol style="list-style-type: none"> 1. MADRID 2. ESPAÑA 3. EUROPA 4. CHICAGO 5. MEMPHIS 6. ÁNGELES 7. VALENCIA 8. TURQUÍA 9. BILBAO

Anexo 3

NYT_files.lst								
File Edit View Compute Settings Windows Help								
N	text file	file size	tokens (running words) in	tokens used for word list	sum of entries	types (distinct words)	type/toke ratio	standardised TTR
1	Overall	1.400.892	99.466	95.487		8.763	9,18	43,59
2	NYT_2-partaa.xml	4.218	243	237		150	63,29	
3	NYT_2-partab.xml	6.882	500	497		243	48,89	
4	NYT_2-partac.xml	10.156	808	800		367	45,88	
5	NYT_2-partad.xml	9.436	695	671		367	54,69	
6	NYT_2-partae.xml	5.414	344	326		185	56,75	
7	NYT_2-partag.xml	7.492	482	445		200	44,94	
8	NYT_2-partah.xml	3.614	178	172		93	54,07	
9	NYT_2-partai.xml	4.496	244	244		143	58,61	
10	NYT_2-partaj.xml	8.386	543	528		259	49,05	
11	NYT_2-partak.xml	9.872	695	686		353	51,46	
12	NYT_2-partal.xml	9.208	684	662		335	50,60	
13	NYT_2-partam.xml	7.712	577	556		265	47,66	
14	NYT_2-partan.xml	5.600	378	342		171	50,00	
15	NYT_2-partao.xml	7.062	498	489		261	53,37	
16	NYT_2-partap.xml	5.510	375	374		196	52,41	

NYP_files.lst								
File Edit View Compute Settings Windows Help								
N	text file	file size	tokens (running words) in	tokens used for word list	sum of entries	types (distinct words)	type/toke ratio	standardised TTR
1	Overall	1.338.146	90.951	88.202		8.728	9,90	47,25
2	NYP_2-partaa.xml	8.174	564	559		277	49,55	
3	NYP_2-partab.xml	9.548	703	682		357	52,35	
4	NYP_2-partac.xml	4.074	228	222		116	52,25	
5	NYP_2-partad.xml	9.290	649	642		282	43,93	
6	NYP_2-partae.xml	11.462	863	802		383	47,76	
7	NYP_2-partaf.xml	9.214	661	643		278	43,23	
8	NYP_2-partag.xml	8.980	660	637		281	44,11	
9	NYP_2-partah.xml	7.446	501	485		253	52,16	
10	NYP_2-partai.xml	5.480	337	330		201	60,91	
11	NYP_2-partaj.xml	9.794	696	688		352	51,16	
12	NYP_2-partak.xml	7.752	500	492		252	51,22	
13	NYP_2-partal.xml	5.068	295	288		177	61,46	
14	NYP_2-partam.xml	4.516	260	259		175	67,57	
15	NYP_2-partan.xml	7.644	493	485		284	58,56	
16	NYP_2-partao.xml	10.132	758	731		344	47,06	
17	NYP_2-partap.xml	10.350	740	734		333	45,37	
18	NYP_2-partaq.xml	10.122	748	702		350	49,86	
19	NYP_2-partar.xml	4.524	287	259		167	64,48	
20	NYP_2-partas.xml	2.278	83	83		70	84,34	

EP_files.lst									
File Edit View Compute Settings Windows Help									
N	text file	file size	tokens (running words) in	tokens used for word list	sum of entries	types (distinct words)	type/token ratio	standard	TTR
1	Overall	1.505.91	83.078	80.476		9.714	12,07	44,40	
2	EP_2_1-partaa.xml	1.944	43	40		35	87,50		
3	EP_2_1-partab.xml	2.426	88	82		58	70,73		
4	EP_2_1-partac.xml	6.942	417	374		224	59,89		
5	EP_2_1-partad.xml	2.198	66	65		47	72,31		
6	EP_2_1-partae.xml	2.542	91	86		60	69,77		
7	EP_2_1-partaf.xml	2.280	69	69		52	75,36		
8	EP_2_1-partag.xml	6.890	430	418		209	50,00		
9	EP_2_1-partah.xml	2.136	62	58		43	74,14		
10	EP_2_1-partai.xml	6.020	351	344		186	54,07		
11	EP_2_1-partaj.xml	6.500	398	385		200	51,95		
12	EP_2_1-partak.xml	4.924	298	284		160	56,34		
13	EP_2_1-partal.xml	1.958	44	40		34	85,00		
14	EP_2_1-partam.xml	2.072	51	48		41	85,42		
15	EP_2_1-partan.xml	2.432	81	76		53	69,74		
16	EP_2_1-partao.xml	2.160	60	55		45	81,82		
17	EP_2_1-partap.xml	1.976	48	47		34	72,34		
18	EP_2_1-partar.xml	2.418	83	78		60	76,92		

MUNDO_files.lst									
File Edit View Compute Settings Windows Help									
N	text file	file size	tokens (running words) in	tokens used for word list	sum of entries	types (distinct words)	type/token ratio	standard	disec TTR
1	Overall	1.368.102	91.067	87.700		11.422	13,02	45,80	
2	MUNDO_2_1-partaa.xml	8.182	557	546		299	54,76		
3	MUNDO_2_1-partab.xml	3.162	144	140		95	67,86		
4	MUNDO_2_1-partac.xml	7.382	468	462		271	58,66		
5	MUNDO_2_1-partad.xml	7.906	541	532		290	54,51		
6	MUNDO_2_1-partaf.xml	9.974	692	630		346	54,92		
7	MUNDO_2_1-partag.xml	10.502	719	684		342	50,00		
8	MUNDO_2_1-partah.xml	9.362	635	620		335	54,03		
9	MUNDO_2_1-partai.xml	13.098	865	783		406	51,85		
10	MUNDO_2_1-partaj.xml	3.318	169	161		97	60,25		
11	MUNDO_2_1-partak.xml	6.168	389	381		219	57,48		
12	MUNDO_2_1-partal.xml	6.594	412	410		240	58,54		
13	MUNDO_2_1-partap.xml	7.138	477	472		256	54,24		
14	MUNDO_2_1-partas.xml	3.884	187	187		132	70,59		
15	MUNDO_2_1-partaw.xml	6.162	395	395		231	58,48		
16	MUNDO_2_1-partay.xml	2.606	86	84		67	79,76		
17	MUNDO_2_1-partbb.xml	3.348	155	148		102	68,92		
18	MUNDO_2_1-partbe.xml	2.182	57	57		47	82,46		
19	MUNDO_2_1-partbf.xml	6.346	407	400		235	58,75		
20	MUNDO_2_1-partbg.xml	10.474	739	704		368	52,27		
21	MUNDO_2_1-partbm.xml	10.452	743	736		360	48,91		
22	MUNDO_2_1-partbn.xml	6.890	470	461		254	55,10		

Anexo 4

Verbos, auxiliares y modales en inglés	Verbos, auxiliares y modales diferentes
1. WAS	1. BE
2. SAID	2. SAY
3. IS	3. HAVE
4. HAD	4. GO
5. HAS	5. DO
6. WERE	6. SCORE
7. BEEN	7. PLAY
8. ARE	8. MAKE
9. GOING	9. GET
10. SCORED	10. WIN
11. PLAYED	11. PLAY
12. DID	12. LEAD
13. MADE	13. ADD
14. GOT	14. TAKE
15. WON	15. LOSE
16. PLAYING	16. COME
17. DON'T	17. TELL
18. WINNING	18. SEED
19. LED	19. KNOW
20. ADDED	20. INCLUDE
21. TOOK	21. MAKE
22. LOST	22. MISS
23. DIDN'T	23. FINISH
24. WENT	24. START
25. CAME	25. ASK
26. TOLD	26. TRY
27. SEEDED	27. USE
28. KNOW	28. COACH
29. PLAY	29. WANT
30. LOSING	30. RANK
31. GETTING	31. CALL
32. HAVING	32. LOOK
33. THERE'S	33. GIVE
34. LEADING	34. HELD
35. COMING	35. FEEL
36. INCLUDING	36. END
37. MAKING	37. SHOOT
38. MISSED	38. HELP
39. FINISHED	39. BECOME
40. STARTED	40. TURN
41. SCORING	41. VISIT
42. DOES	42. BEGIN
43. ASKED	43. EXPECT
44. TRYING	44. SEEM

45. CAN'T
46. WASN'T
47. WIN
48. DONE
49. USED
50. DOING
51. COACHING
52. WANTED
53. TAKE
54. RANKED
55. CALLED
56. BEING
57. DOESN'T
58. LOOKING
59. GAVE
60. STARTING
61. GO
62. MAKES
63. HELD
64. LOOKED
65. SEEN
66. ISN'T
67. FELT
68. ENDED
69. SHOOTING
70. HELPED
71. BECAME
72. COMES
73. TURNED
74. VISITING
75. BEGAN
76. EXPECTED
77. SEEMED
78. KNOWN
79. REMAINING
80. SEEMS
81. KNOWS
82. GIVEN
83. KNEW
84. SAW
85. AVERAGING
86. BROUGHT
87. RETURNED
88. TAKING
89. FELL
90. GOES
91. GETS
92. TALKING
93. WATCHING

45. REMAIN
46. SEE
47. AVERAGE
48. BRING
49. RETURN
50. FALL
51. TALK
52. WATCH
53. NEED
54. SUPPOSE
55. RECEIVE
56. SHOW
57. BEAT
58. COMBINE
59. SEND
60. REBOUND
61. LIKE
62. INJURE
63. APPEAR
64. ENTER
65. WORK
66. ANNOUNCE
67. BLOCK
68. DEFEND

94. NEEDED 95. COACHED 96. SUPPOSED 97. RECEIVED 98. SHOWED 99. WON'T 100. COULDN'T 101. SEED 102. BEAT 103. WATCHED 104. COMBINED 105. SENT 106. REBOUNDED 107. LIKE 108. INJURED 109. APPEARED 110. SAY 111. AVERAGED 112. LOOKS 113. SPENT 114. ENTERING 115. WORKED 116. ANNOUNCED 117. MISSING 118. BLOCKED 119. DEFENDING	
--	--

Anexo 5

Verbos, auxiliares y modales en español	Verbos, auxiliares y modales diferentes
1. HAN 2. HACE 3. HAY 4. SON 5. PUEDE 6. ESTABA 7. HECHO 8. PARECE 9. GANÓ 10. SEA 11. TENEMOS 12. HEMOS 13. ACABÓ 14. QUEDA 15. PUDO	1. HACER 2. HABER 3. SER 4. PODER 5. ESTAR 6. PARECER 7. GANAR 8. TENER 9. ACABAR 10. QUEDAR 11. JUGAR 12. DEBER 13. QUEDAR 14. ESPERAR 15. PASAR

16. ESTAMOS	16. DAR
17. HUBO	17. SABER
18. JUGANDO	18. IMPONER
19. DEBE	19. PERDER
20. QUEDÓ	
21. ESPERA	
22. HAYA	
23. PASÓ	
24. GANADO	
25. PUEDEN	
26. DADO	
27. SABE	
28. IMPUSO	
29. JUGAR	
30. JUGÓ	
31. JUEGA	
32. PERDIÓ	