

CARACTERIZACIÓN DE LOS ANUNCIOS DE EMPLEO A TRAVÉS DE UN ANÁLISIS SEMI-AUTOMÁTICO*

Victoria López Sanjuán

UNED, Madrid

vlopez@madrid.uned.es

This paper shows a description of the sublanguage of job ads, paying especial attention to its structure and grammatical and lexical features.

Sublanguages, defined as the language of a textual subgenre identified from a given domain and genre are characterized by several linguistic features. These features can be common or diverging from standard language. They can be lexical, semantic, morfosyntactic and discursive.

From a representative corpus of a linguistic entity, which is clearly enclosed and defined, a quantitative analysis is made to determine scientifically its identity and nature as a linguistic variant by processing the corpus with a set of computational tools.

Key words: *sublanguage, corpus, analysis tools, job ads, quantitative analysis.*

En este artículo se muestra una amplia descripción del sublenguaje de los anuncios de empleo, prestando especial atención a su estructura y a las características morfológicas y léxicas.

Hay que señalar que los sublenguajes, definidos como la lengua de un subgénero textual, identificado a su vez a partir de un dominio y género dados, se caracterizan por una serie de rasgos lingüísticos, tanto comunes como divergentes con respecto a la lengua estándar. Estas características pueden ser léxicas, semánticas, morfosintácticas, pragmáticas y discursivas.

A partir de un corpus representativo de una entidad lingüística claramente acotada y definida se ha

realizado un análisis cuantitativo para determinar científicamente su identidad y naturaleza como variante lingüística a través del procesamiento del corpus con un conjunto de herramientas computacionales estándar.

Palabras clave: *sublenguaje, corpus, herramientas de análisis, anuncios de empleo, análisis cuantitativo.*

1. INTRODUCCIÓN

En el tratamiento computacional de los corpus hoy en día destaca el de los sublenguajes. Es un tipo de variante lingüística que ha cobrado gran importancia en las últimas décadas debido, en primer lugar, a las múltiples dificultades en el tratamiento informatizado de las lenguas generales y, en segundo lugar, a su propio interés objetivo en el mundo científico y profesional de hoy. Todo ello ha conducido a la identificación de variantes lingüísticas temáticas y a la adopción de líneas de investigación *ex profeso*.

En el tratamiento computacional de los corpus es habitual el uso de una serie de herramientas computacionales. Estas herramientas se utilizan para obtener datos de un modo rápido, exhaustivo, flexible y fiable. La ventaja que tienen sobre las anotaciones manuales es que no dependen ni varían según la idiosincrasia de la persona que las hace. También se utilizan para llevar a cabo las anotaciones de los corpus.

Entre estas herramientas destacan las herramientas de análisis textual y las herramientas para la anotación de corpus. Las primeras facilitan sobre todo datos estadísticos, que sirven para realizar generalizaciones y establecer el comportamiento de estructuras o palabras en el texto. Estas herramientas se suelen aplicar a textos sin anotar (por ejemplo, TACT, Hyperbase o WordSmith Tools). Dentro de las herramientas de anotación de corpus destacan los etiquetadores (como CLAWS) y los analizadores sintácticos (como Link Grammar Parser). Los etiquetadores proporcionan el análisis de categorías léxicas y gramaticales y los analizadores el de las estructuras sintácticas.

En un estudio anterior (López Sanjuán 2003), se ha demostrado que en el estudio de sublenguajes no es suficiente la utilización de una herramienta aislada, por ejemplo de un analizador textual, para

identificar o tratar un sublenguaje determinado, ya que sólo proporciona resultados de tipo estadístico y léxico y no es suficiente. Para identificar un sublenguaje no llega con estudiar el vocabulario, hay que fijarse en otros planos lingüísticos y estos programas no proporcionan otro tipo de información. Lo mismo sucede con un etiquetador, que no suele ir más allá de la información morfológica o morfosintáctica. Sin embargo, la utilización de herramientas computacionales estándar en una aplicación conjunta puede aportar indicios de la presencia de una posible variante lingüística, que está acotada por tema y género.

Con la aplicación de un etiquetador y un analizador textual a un corpus de anuncios de empleo, se va a describir y caracterizar este tipo textual. Hay que señalar que la elección de un corpus de ofertas de empleo no ha sido fortuita: se han tenido en cuenta la disponibilidad de los textos de este tipo para su recopilación, la no excesiva especialización conceptual de este campo y que el género impone necesariamente rígidas normas por cuestiones como el espacio, el formato, etc. y el carácter abierto de su comunidad de hablantes.

Aunque puede argüirse que ha habido un gran número de estudios lingüísticos llevados a cabo con fines computacionales sobre este subgénero (Somers y otros 1997, Bárcena 1994), la aplicación conjunta de diferentes tipos de herramientas computacionales para obtener una caracterización global donde los resultados, positivos o negativos, muestren un conjunto suficientemente amplio y variado de evidencia para poder caracterizar este tipo textual e interpretarlo ofrece una perspectiva diferente y exhaustiva con el objetivo de describir este tipo textual, pero sin un fin computacional, sino meramente lingüístico. Un análisis de este tipo podría tomarse como base para la enseñanza de un curso de ESP en dónde se enseñase inglés de negocios, por ejemplo, o podría servir como ejemplo de lo que ofrecen las herramientas de análisis para investigar un sublenguaje y extraer, por ejemplo, información para la elaboración de materiales didácticos. Esto sin olvidar, claro está, su enfoque computacional.

Después de una breve introducción sobre los sublenguajes, y partiendo de la premisa de que en principio se ha recopilado un corpus representativo de una entidad lingüística claramente acotada y definida, se pretende determinar científicamente su identidad y

naturaleza como variante lingüística a través del procesamiento del corpus con un conjunto de herramientas computacionales estándar. Una vez procesado, estamos en posición de alcanzar una serie de conclusiones sobre su variante lingüística, basadas en los resultados obtenidos de ese procesamiento y que nos van a permitir describir los anuncios de empleo.

2. SUBLENGUAJES

Los sublenguajes, definidos como la lengua de un subgénero textual, identificado a su vez a partir de un dominio y género dados, se caracterizan por una serie de rasgos lingüísticos, tanto comunes como divergentes con respecto a la lengua estándar respectiva. Estas características pueden ser léxicas, semánticas, morfosintácticas, pragmáticas y discursivas.

En primer lugar, un sublenguaje tiene todas las propiedades "esenciales" de un sistema lingüístico, como la tendencia a la economía de expresión y el hecho de que sirve como vehículo de expresión de todos los mensajes que se pueden formular y emitir sobre el mundo de referencia correspondiente. Moskovich (1982:192) observa que incluso el más pequeño de los sublenguajes es capaz de producir un número prácticamente infinito de realizaciones, como sucede con las lenguas naturales. También Kittredge (2003:445) comenta que "*in theory a sublanguage is an infinite set of sentences.*" Naturalmente, mientras que una lengua natural puede describir por completo cualquier situación imaginable, cualquier sublenguaje sólo puede describir por completo una parte limitada de la realidad: aquélla a la que sirve como vehículo de expresión.

En segundo lugar, el sublenguaje se ha de referir a un dominio particular del discurso o a una familia de dominios relacionados. Pueden ser campos científicos que contienen otros subcampos más específicos. Por ejemplo, dentro de la biología se pueden identificar subcampos que se refieren a los animales y otros a las plantas, y así sucesivamente, pero el dominio ha de estar bien delimitado para que pueda ser homogéneo y sistemático desde un punto de vista lingüístico.

En tercer lugar, los subsistemas lingüísticos son máximos con respecto al dominio, en el sentido matemático de que sistemas mayores

tienen las mismas propiedades. Es decir, si se caracteriza un subsistema lingüístico, como los informes de avería de los servidores informáticos, esos mismos rasgos deben ser incluidos en los subsistemas más amplios que pertenezcan a ese ámbito como los informes de avería de equipamiento informático. Esta característica apunta hacia la relación que un sublenguaje mantiene con la lengua general. Los sublenguajes, según Kittredge (1987:61), comparten las propiedades básicas de las lenguas; no hay que olvidar que, de hecho, son subconjuntos de aquéllas. Es fácil identificar, por ejemplo, que una receta médica está expresada en tal o cual idioma. Lehrberger (1986:20) es contrario a la opinión de que un sublenguaje dado es meramente un subconjunto de la lengua estándar correspondiente ya que esto conllevaría que la gramática o el vocabulario de un sublenguaje deriva de la lengua estándar simplemente si se desecha un número de reglas o palabras que no son relevantes para el sublenguaje.

Mientras que el vocabulario de una lengua natural es casi ilimitado o al menos imposible de recopilar en su totalidad, el vocabulario de un sublenguaje es, salvo contadas excepciones, como el sublenguaje de los compuestos químicos, mucho más reducido. El carácter más o menos cerrado es una de las propiedades más importantes de los sublenguajes. Si el sublenguaje es completamente cerrado, se podrán inventariar mejor su vocabulario y sus construcciones gramaticales. No es suficiente si sólo se tienen en cuenta las restricciones cuantitativas que se observan en los vocabularios, ya que no definen por sí solas a los sublenguajes.

En el plano léxico, como se acaba de indicar, el vocabulario es un factor crucial a la hora de estudiar un sublenguaje. Cuanto más reducido y más ceñido a un determinado número de clases semánticas, evita las ambigüedades y permite crear patrones lingüísticos más fiables. Es importante saber si el sublenguaje presenta un carácter cerrado y si es sistemático, como por ejemplo, el sublenguaje de la química orgánica cuya combinabilidad de prefijos, raíces y sufijos es enorme, pero muy regular y predecible. También se puede analizar la relación que existe entre los tipos de palabras (*types*) y las apariciones (*tokens*). Cuanto menor sea la cifra de formas obtenida a mayor número de apariciones, mayor será la riqueza léxica, lo que en términos de sublenguajes probablemente indica que se trata de un sublenguaje parcialmente cerrado, si se usan sinónimos o paráfrasis.

En el plano morfológico, la simplificación de un sistema basado en un sublenguaje dado (que sin duda cabe esperar) facilita cuantitativamente el procesamiento general del sublenguaje y cualitativamente contribuye a un menor nivel de ambigüedad léxica. Además, sus correspondencias con otros niveles lingüísticos pueden servir como punto de referencia en el análisis.

En el plano sintáctico es importante analizar si la sintaxis es regular, reducida o sistemática, para ver si facilita la determinación de patrones lingüísticos asociados a otros niveles lingüísticos y permite la desambiguación semántica.

Los corpus son una fuente de referencia e información adecuada para llevar a cabo estudios lingüísticos específicos de los sublenguajes, fundamentalmente con el fin de elaborar programas computacionales adaptados que permitan maximizar las ventajas intrínsecas de dichas variantes, tales como las restricciones cuantitativas y cualitativas en los distintos niveles lingüísticos.

En la sección 4, se va a llevar a cabo un estudio para inventariar algunas de las características propias del sublenguaje de los anuncios de empleo, en el que se mostrarán alguna de estas propiedades establecidas a priori sobre los sublenguajes en general.

3. CARACTERIZACIÓN DEL CORPUS JADCOR

Para recopilar el corpus JadCor (*Job Ad Corpus*) de anuncios de empleo se realizó una búsqueda en Internet de textos de este tipo, siempre en portales especializados que ofreciesen esta información específica para que no quepa la posibilidad de segundas intencionalidades (por ejemplo, publicidad de la propia empresa contratadora). Internet es idóneo por la gran cantidad de textos disponibles y porque, actualmente, es una de las fuentes de información más utilizada por las personas que buscan empleo. Asimismo, los textos ya se encontraban en formato electrónico, lo que conlleva una postedición menor que la laboriosa utilización de un escáner y programa de ROC, cuyo resultado implica una gran revisión manual por parte del lingüista.

Dentro de los anuncios de empleo, se identificaron dos variedades: las ofertas de trabajo, publicadas por las empresas que

buscan personal y las demandas de trabajo, publicadas por personas que buscan empleo y ofrecen sus servicios. Para esta investigación se seleccionó el primer tipo, intentando recopilar anuncios procedentes de diferentes ámbitos laborales, tanto desde el punto de vista de la especialidad como de las cualificaciones requeridas. Podría argüirse que cuanto mayor es la cualificación solicitada a los candidatos, más cuidada es la elaboración de los anuncios, lo cual podría distorsionar los resultados de la investigación o al menos limitar su aplicabilidad.

Una mayor variedad y, por lo tanto, representatividad de textos procedentes del mismo subgénero permitirá valorar mejor sus rasgos lingüísticos. Es interesante mencionar que la fuente del JadCor es habitualmente privada y heterogénea y la autoría es ilimitada (los servicios de personal o departamentos de recursos humanos de empresas e instituciones). Además se han tomado de varias direcciones de Internet, no de una sola. Así, dentro de las ofertas publicadas en Internet por estos portales especializados, se buscó las que tenían un formato textual, por las mismas razones que las aludidas anteriormente. Se recopiló una selección de ofertas de empleo de tres portales distintos, que contienen anuncios escritos por compañías con presencia en un país de habla inglesa y que fueron publicados en mayo-junio del 2006. Son los siguientes:

- *Jobs.ac.uk* (www.jobs.ac.uk). Es un portal británico que publica ofertas de empleo académicas, principalmente en el Reino Unido.
- *The Guardian Jobs* (<http://jobs.guardian.co.uk/>). Es el portal dedicado a las ofertas de empleo de la edición digital del periódico británico *The Guardian*.
- *Monster* (www.monster.com). Es un portal de empleo de Estados Unidos con delegaciones en diversos países y que publica ofertas internacionales en diferentes lenguas.

Sin embargo, los textos que se encontraron no estaban disponibles en formato texto, al tratarse de textos en formato HTML y contener gráficos y tablas. Para evitar la opción de tener que posteditar los textos y que éstos mantuviesen su formato textual original, se realizó una exportación a un documento de texto evitando así cualquier tipo de referencia externa a los textos. La presencia de las etiquetas de los gráficos y las tablas numéricas debería ser eliminada ya que, aunque no aparecen en el texto con los códigos en HTML,

cuando se guardan en formato texto, sí quedan recogidas las etiquetas que precisan su ubicación y contenido, pudiéndose de esta manera alterar también el resultado del estudio.

Asumido que el formato de los textos que formaban el corpus no era el idóneo a las necesidades de la investigación y que una edición sería necesaria para su almacenamiento en el formato deseado, se procedió a examinar su contenido, para verificar era adecuado para el estudio a realizar.

Research Associate (2 years fixed-term)

Centre for Clinical Management Development
School for Health
£ 23, 643 per annum (RA1A Scale)

Duties will be to provide research and editorial support for a range of research projects, focusing on the capital investment and commissioning implications of clinical pathway based approaches to service delivery in primary and acute care settings. The successful candidate will have; a degree in the Social Sciences and/or Humanities, editorial experience and some appreciation of issues relating to quality improvement, the organization and management of clinical work and of the culture change issues that this entails.

Closing Date: 4 November 2004
Reference: 0528

Further details of the post and an application form are available on our website (<https://jobs.dur.ac.uk> telephone 0191 334 6499; fax 091 334 6495.

[Click here for Employer Profile](#)

If you apply for this position, please say you saw it on jobs.ac.uk

Figura 1. Extracto del JadCor con una oferta de empleo.

En la figura 1 se muestra un extracto del corpus JadCor de un anuncio de empleo procedente del portal *Jobs.ac.uk*. En él se puede observar la información que aparece, empezando por la descripción del

puesto vacante, requisitos de los candidatos, el horario, la localización, etc. Asimismo, al final de anuncio aparece información adicional de un modo breve y esquemático, por ejemplo, el salario y el tipo de contrato.

Una vez recopilado el corpus JadCor se puede definir como un corpus textual, con fines específicos, monolingüe (inglés), completo y sin anotar. Todos los textos son escritos, están en formato electrónico, han sido elaborados para ser publicados en Internet, corresponden a múltiples autores anónimos, y no están dirigidos a un público especializado.

El tamaño de un corpus es un factor importante a tener en cuenta a la hora de su diseño, tanto por su directa relación con el valor representativo del corpus como por cuestiones relativas a su tratamiento ya que, por ejemplo, no todas las herramientas funcionan con corpus de un número ilimitado de caracteres. Como Lehrberger y Bourbeau (1988:141) apuntan, un corpus para ser válido debe contener al menos 10.000 palabras y, dependiendo de la amplitud del subgénero, podría necesitar ser incluso mayor. Sin discrepar de esta afirmación, cabe añadir que el tamaño de un corpus varía también en función de las necesidades de la investigación. Cabe señalar, además, que el tamaño del corpus JadCor es reducido (20.679 palabras), si se compara con los grandes corpus que se están recopilando en la actualidad, como el *Bank of English* o el *British National Corpus*, que están formados por millones de palabras. Sin embargo, se ajusta sobradamente a los mínimos establecidos por Lehrberger y Bourbeau, por lo que en principio se considera lo suficientemente extenso para ajustarse a las necesidades de esta investigación.

Con el fin de diversificar los instrumentos de esta investigación, los anuncios de empleo de JadCor son de temática variada. Esto convierte a JadCor en un sublenguaje relativamente abierto. Kittredge (2003:436) señala que basar la descripción de un corpus en una sola fuente puede ser suficiente para “*prototyping, or even building, a simple processing application to serve only that source, but does not normally give a good perspective of the whole sublanguage, as used by a wider community*”. Por este motivo, los textos seleccionados en proceden de diferentes fuentes.

La elaboración del JadCor se realizó de un modo riguroso y objetivo. El criterio de selección estuvo claro desde un principio: se pretendía recopilar una muestra de ofertas de empleo lo

suficientemente amplia y representativa para realizar una descripción de los rasgos propios de estos sublenguajes en inglés.

Una vez determinado el objeto de la investigación y seleccionados los textos, no se ha realizado ninguna manipulación en los datos recogidos. Como se mencionaba a propósito del formato, se juzgó preferible suprimir las imágenes de JadCor para facilitar su tratamiento. Ha de hacerse hincapié en que los textos no se apoyan en modo alguno en dichas imágenes, ya que éstas consistían en anagramas o *slogans* publicitarios sobre la empresa, imágenes del equipo de profesionales, etc. Por lo tanto, queda garantizado que tras la eliminación de las imágenes el rigor y la fiabilidad del corpus han quedado intactos, ya que una manipulación de cualquier tipo sobre un corpus haría que éste dejara de ser una fuente de información lingüística fiable. En este sentido cabe afirmar los 35 textos del JadCor han sido tratados con la mayor imparcialidad, es decir, que no han sido acortados, extendidos, fusionados, manipulados, etc. en modo alguno. Además, estos textos han sido elaborados por autores diferentes, las propias empresas contratadoras, que quedan claramente fuera de la influencia de los distintos portales de empleo a los que los anuncios son enviados para su posterior publicación. De esta forma, se garantiza la representatividad del corpus⁴. También, sí hay un esquema aproximado de información *natural* común a todos los anuncios de empleo, como puede ser, por ejemplo, la descripción del puesto ofertado, los requisitos solicitados o la información de contacto (esto se explicará en más detalle cuando se hable de la estructura de los anuncios de empleo).

4. ANÁLISIS LINGÜÍSTICO DEL JADCOR

En esta sección se muestran los resultados del análisis realizado en el corpus JadCor de anuncios de empleo con el objetivo de describir este subgénero. Se trata básicamente de un análisis cuantitativo de las palabras del corpus. Este procesamiento se ha llevado a cabo de un modo semi-automático, pues se ha procesarlo con algunas herramientas computacionales, como el etiquetador *CLAWS* y la herramienta de análisis textual *WordSmith Tools*, pero ha sido necesario realizar una revisión manual complementaria para crear una base de datos a partir de los resultados globales.

En el proceso de asignar las etiquetas de *CLAWS*, se prestó atención, por ejemplo, a los homógrafos y se determinó de un modo manual a qué categoría pertenece cada una de las apariciones en el corpus, evitando así las ambigüedades o el recuento erróneo. La aplicación del etiquetador determinó qué categorías gramaticales están presentes en el corpus y en qué medida. Aunque fue necesaria una revisión de aquellas palabras que pudiesen resultar ambiguas, tal es el caso de *may* que puede ser el verbo modal o referirse al mes, y, por ejemplo, *IT*, que muestra dos apariciones diferentes en el corpus, por una parte el pronombre y, por otra, es el acrónimo de *Information Technology*. En este último caso, si el etiquetador pudiese distinguir el acrónimo por estar en mayúsculas dado que el pronombre se escribe todo en mayúscula en ocasiones aisladas, no haría falta esta desambiguación manual. Bien es cierto que *CLAWS* aporta la información de ambas categorías, pero se ha comprobado que en algunos casos, la desambiguación que hace en referencia al contexto, no es acertada y se debe hacer una postedición para asignar la etiqueta correcta (López Sanjuán, 2006). Asimismo, se solventaron también los errores de adjudicación errónea de la etiqueta de nombre común a los nombres propios que no empiezan por mayúscula en el corpus.

WordSmith Tools se utilizó para obtener un listado de los tipos de palabras que ocurren en el corpus ordenadas por su frecuencia. Sin embargo, en los resultados obtenidos de *WordSmith Tools* tampoco en ningún caso se indica que, por ejemplo, *may* se refiera a un sustantivo o a un verbo. Así, el programa solo contabiliza las apariciones totales de estas dos palabras conjuntamente, sin especificar cuando se trata del mes o del verbo modal, y se debe de hacer de un modo manual. Hay que señalar que, sin embargo, esta tarea de postedición manual no es difícil de realizar en un corpus de anuncios de empleo porque no es un corpus de gran tamaño, y apenas hay ambigüedades de este tipo. Además, se cuenta con la asistencia del programa de concordancias que muestra todas las concordancias y apariciones en una búsqueda refinada.

Una vez procesado el corpus, se procedió la descripción de los anuncios de empleo.

La descripción se inicia prestando atención al tipo textual de los anuncios de empleo para determinar su estructura y definir sus funciones retóricas, comprobando que son comunes a todos los textos

del corpus. A continuación, se muestran los resultados del análisis de las categorías gramaticales del corpus de anuncios de empleo. Aunque se ha denominado análisis categorial, y el punto de partida es un análisis morfológico, se ha añadido, además, información léxica y semántica, y en ocasiones, información sintáctica para hacer una descripción más exhausta de este sublenguaje.

4.1 Análisis estructural

Los anuncios de empleo constan de funciones retóricas fijas básicas como son:

- Encabezado en el que aparece el puesto a cubrir (puede aparecer la ubicación o no).
- Empresa contratante (que puede ser mencionada por su nombre o también por una breve descripción).
- Puesto ofertado (al que de nuevo se puede hacer referencia en mayor o menor detalle).
- Tareas del puesto (en qué consiste el empleo ofertado).
- Requisitos del candidato.
- Forma de contacto.

Sin embargo, también pueden aparecer otras funciones retóricas opcionales:

- Condiciones (sueldo, beneficios, perspectivas de promoción, etc.).
- Fecha de publicación.
- Plazo de recibo de cv.
- Información extra sobre la empresa ofertante (por ejemplo, la página web).

Puesto que toda la información relativa a una función retórica suele aparecer junta en el texto, es posible que un sistema anote automáticamente cada fragmento retórico en base simplemente a palabras claves para su procesamiento computacional.

Research Associate (2 years fixed-term) **[Encabezado]**

Centre for Clinical Management Development **[Empresa contratante]**
School for Health
£ 23, 643 per annum (RA1A Scale) **[Condiciones salariales]**

Duties will be to provide research and editorial support for a range of research projects, focusing on the capital investment and commissioning implications of clinical pathway based approaches to service delivery in primary and acute care settings. **[Tareas del puesto]** The successful candidate will have; a degree in the Social Sciences and/or Humanities, editorial experience and some appreciation of issues relating to quality improvement, the organization and management of clinical work and of the culture change issues that this entails. **[Requisitos del candidato]**

Closing Date: 4 November 2004 **[plazo para enviar cv]**
Reference: 0528

Further details of the post and an application form are available on our website (<https://jobs.dur.ac.uk> ; telephone 0191 334 6499; fax 091 334 6495) **[Forma de contacto]**

[Click here for Employer Profile](#)
If you apply for this position, please say you saw it on jobs.ac.uk

Figura 2. Estructura de un anuncio de empleo del JadCor.

En la figura 2 se puede observar que el anuncio de empleo de la figura 1 muestra todas las funciones retóricas indicadas para este tipo de anuncios e incluso alguna de las señaladas como opcionales como la fecha de entrega de cv, o un enlace a la página web de la empresa ofertante

4.2. Análisis léxico y categorial

En esta sección se muestran los resultados del estudio de las categorías gramaticales determinando su presencia o ausencia en el corpus de anuncios de empleo. Cabe señalar de nuevo que, a pesar de

que esta sección se ha denominado análisis categorial, contiene el análisis morfológico que se ha complementado con información léxica y semántica.

Antes de entrar a en este análisis morfológico en profundidad, conviene también ver algunos aspectos léxicos, como son la longitud de las palabras y la proporción entre tipos de palabras/apariciones (type/token ratio), que también nos puede proporcionar información interesante con respecto a los anuncios de empleo.

La longitud de las palabras se mide teniendo en cuenta los caracteres ortográficos de las palabras de un texto. A mayor longitud de las palabras, se interpreta que existe una mayor complejidad presente en el texto (Biber 1988:118). En este caso, se trata tanto de complejidad tanto léxica (palabras simples, compuestas, etc.) como sintáctica (oraciones simples o subordinadas, por ejemplo), y teniendo en cuenta, por ejemplo, la cohesión textual.

En el caso del JadCor, la longitud media de las apariciones del corpus es 5.29..

You will work closely with colleagues and partner organisations to develop a leading-edge provision across this broad range of professional programmes. Strong people management skills are essential, plus the ability to influence at senior level as you join the Executive Team in guiding and directing the development of the School as a whole.

En el ejemplo se puede observar que aunque existen palabras breves de dos y tres caracteres, como por ejemplo, *at*, *to*, *and* o *you* (todas ellas palabras gramaticales) abundan las más largas, lo que no es lo más habitual en inglés estándar. Esto puede deberse a que los términos de los negocios sean de origen romance, por ejemplo, *candidate*, *application* o *employer*; y estas palabras suelen ser más largas que las de origen germánico.

En lo que se refiere a la proporción de tipos de palabras/apariciones, este rasgo muestra la correspondencia que existe entre las diferentes palabras que conforman el texto y el número de sus apariciones. Esta proporción es muy importante para determinado tipo de textos, pues en un sublenguaje es un índice para interpretar si es

cerrado, por ejemplo. A mayor proporción de palabras diferentes, más abierto será el sublenguaje desde el punto de vista léxico. El valor medio de todos los textos del JadCor es de 0.62. En este caso, no se puede interpretar que estemos ante un léxico cerrado. Puede deberse, entre otros factores, a que la naturaleza de los anuncios de empleo utiliza un vocabulario común para determinados conceptos (requisitos, forma de contacto, etc.) pero otro más específico para describir las tareas del puesto (no son las mismas para un médico que para un fontanero, por ejemplo) y es que los campos profesionales del JadCor son abiertos.

La simplificación morfológica de un sistema basado en un sublenguaje dado (que sin duda cabe esperar) facilita cuantitativamente el procesamiento general del sublenguaje y cualitativamente contribuye a un menor nivel de ambigüedad léxica. Además, sus correspondencias con otros niveles lingüísticos pueden servir como punto de referencia en el análisis. Principalmente hay que analizar:

- Los morfemas de las categorías gramaticales presentes en el corpus.
- Los morfemas de las categorías gramaticales que no aparecen en el corpus.
- Los morfemas que aparecen más y los morfemas que aparecen menos.
- Las recurrencias que pueden aparecer si, por ejemplo, la presencia de determinado morfema, depende de algún factor contextual.
- Las divergencias que pueden aparecer con respecto a la lengua estándar.

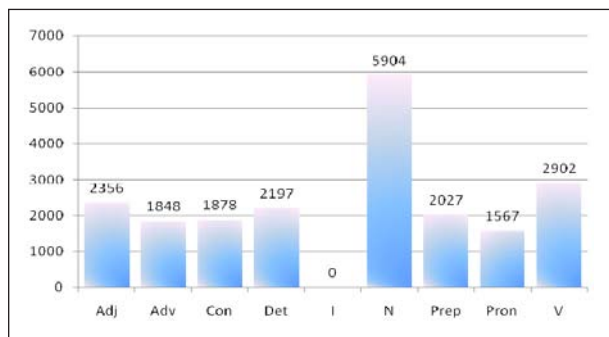


Figura 3. Análisis de las categorías del JadCor.

En la figura 3 se muestra el gráfico de la distribución de las categorías gramaticales del JadCor. En ella se puede observar que no todas las categorías están presentes, sino que no hay interjecciones en el corpus. El motivo puede ser que son textos escritos y las interjecciones son más propias de la lengua hablada.

Todas las demás categorías están presentes, siendo la más frecuente la categoría nominal seguida por la verbal. Los nombres duplican y triplican las apariciones de las otras categorías.

A partir de los datos obtenidos del procesamiento de cada categoría gramatical en el corpus, se ha realizado un estudio de cada categoría en cada uno de los textos que componen el JadCor. Una vez procesados todos los textos del corpus y obtenidos todos los resultados, éstos se *normalizan* a 1.000 palabras. Esta normalización se hace para que la diferencia de tamaño de los textos no desvirtúe la validez de los resultados del análisis.

Ilustremos este proceso: por ejemplo, el texto 1 contiene 1.785 palabras, y el rasgo que se quiere normalizar es el de verbos. Si en este texto aparecen, por ejemplo, 52 verbos, se aplica la siguiente fórmula para normalizar o equiparar los resultados: número de rasgos totales dividido entre las palabras del texto multiplicado por 1.000. El resultado sería:

$$52 : 1.785 \times 1.000 = 2.91$$

2.91 sería el número de apariciones de verbos en presente por cada 1.000 palabras. Este procedimiento permite homogeneizar los resultados de todos los textos, para que debido a su diferente extensión no se desvirtúen o invaliden los resultados para posteriores análisis contrastivos.

En la tabla 1 se pueden ver los resultados de las frecuencias normalizadas de las diferentes categorías gramaticales en el JadCor. En la primera columna se muestra el menor número de apariciones de cada categoría en el corpus, mientras que en la segunda está explícito el número mayor de apariciones. En la tercera columna, se señala la oscilación entre los textos con mayor y menor número de apariciones. Por último, en la cuarta columna se muestra el promedio de las apariciones de cada rasgo en el corpus.

	Valor mínimo	Valor máximo	Oscilación	Promedio
Adjetivos	21.89	185.81	163.92	93.84
Adverbios	9.7	59.59	49.89	30.19
Conjunciones	12.34	183.84	171,5	69.37
Determinantes	12.34	197.8	185,46	92.28
Nombres	173.17	555.55	382.38	393.83
Preposiciones	28.06	218.75	190.69	113.65
Pronombres	0	69.26	69.26	29.98
Verbos	31.63	186.81	155.18	93.80

Tabla 1. Frecuencias normalizadas de las categorías gramaticales del JadCor.

Estos resultados muestran, por ejemplo que el número de conjunciones o determinantes que aparecen en cada texto varía ostensiblemente, puesto que en algunos tan sólo hay 12.34 y en otros no sólo los duplican o triplican. Esto sucede con todas las categorías, pero se percibe con menor intensidad en el caso de los adverbios y de los pronombres, dado que el número de apariciones es también menor. Como se puede observar, en algunos textos ni siquiera hay alguna aparición de un pronombre. En cuanto al promedio de las diferentes categorías, se vuelve a determinar que el número de nombres es el que mayor promedio presenta en el JadCor, seguido por las preposiciones.

Una vez analizadas la presencia de las categorías en el JadCor se van a estudiar los principales rasgos de cada una de estas categorías gramaticales en los anuncios de empleo y la razón de su presencia en mayor o menor medida en este sublenguaje.

4.2.1. Nombres

Los nombres del JadCor representan casi el 50% de todas las palabras que ocurren en el corpus.

Los nombres propios que aparecen en este corpus se refieren a nombres de las empresas ofertantes y topónimos que indican la localización de la empresa o el lugar en donde se va a realizar el trabajo.

A partir de los resultados obtenidos automáticamente, se llevó a cabo el procesamiento semántico manual de los sustantivos que

aparecen en el JadCor. Se han determinado los siguientes:

- **Habilidades (H):** en este grupo se engloban los sustantivos relacionados con las habilidades que se pueden pedir para el trabajo, sin ser la cualificación específica. Por ejemplo, *confidence, credibility, ambition*, etc.
- **Empresariales (Em):** a este grupo pertenecen los sustantivos relacionados con la empresa y sus funciones. Por ejemplo, *accountability, finance, loan, partnerships*, etc.
- **Fechas/Tiempo (F/T):** por ejemplo, *Monday, September, January, week, year*, etc.
- **Dispositivos (D):** *computer, fax, internet, phone, consumables*, etc.
- **Educación (Ed):** *certificate, course, graduate, grants*, etc. así como las diferentes disciplinas específicas requeridas en cada oferta: *Chemicals, Cryptography, Sociology*, etc.
- **General (G):** en el corpus se encuentra un alto número de palabras que no se pueden catalogar dentro de una categoría específica dentro de los anuncios de empleo, por ejemplo, *attack, convergence, identification, number, problem* o *regeneration*, entre otras, que vienen determinadas por el contenido de cada anuncio.
- **Empleos (E):** los nombres de los diferentes puestos ofertados: *analyst, consultant, assistant, researcher*, etc.
- **Lugares/Direcciones (L/D):** en este grupo se incluyen los nombres propios de lugar, por ejemplo, *London, Birmingham, Wales*, etc. así como otros nombres comunes específicos de un lugar donde se va a realizar el trabajo, por ejemplo, *hospital, campus, centre*, etc.
- **Contacto (C):** por ejemplo, *e-mail, remit, mailto*, etc. así como las direcciones de correo electrónico.
- **Personas y entidades (P):** en este grupo se incluyen los sustantivos que tienen como referente a una persona o una entidad formada por personas, por ejemplo, *citizen, community, organisation, residents*, etc.
- **Puesto (Pu):** en este grupo se engloban las palabras relacionadas con información específica sobre la condiciones del trabajo, por ejemplo, *expectations*,

- holidays, leisure, minority, etc.*
- **Transporte (Tr):** por ejemplo, *transport, truck, vehicle, train, etc.*
 - **Trabajo (T):** en este grupo se recogen los sustantivos relacionados con el mercado laboral y con el trabajo en general, por ejemplo, *payments, term, vacancy, workload, labour, etc.*

En la figura 4 se muestra la distribución de los sustantivos del JadCor agrupados en campos semánticos. Como se puede observar, los que más apariciones tienen en el corpus son los relacionados con el trabajo y el mercado laboral y los generales, que no se engloban en ninguna categoría. Esto, a diferencia de otros sublenguajes, como el de los informes meteorológicos, indica que el vocabulario no es tancerrado como se podría desear en un sublenguaje.

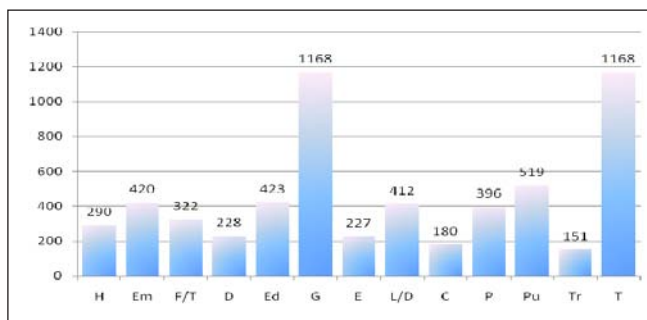


Figura 4. Distribución de los sustantivos del JadCor por campos semánticos.

En este ejemplo, *The **postholder** will be a **member** of the scientific **staff** working on **aspects** of antimicrobial **resistance**, in particular **infections** in **children** caused by **MRSA***, se puede observar que de las veinticuatro las palabras que componen la oración, ocho son nombres, siendo casi el 35 por ciento de la oración. En este caso, cinco de los nombres forman parte de sintagmas preposicionales y tres de ellos están modificados por un adjetivo atributivo. Esto indica una alta carga informativa y que el texto es muy descriptivo, pues tan sólo aparece un verbo principal y el resto son sintagmas nominales y preposicionales.

Cabe señalar también que al revisar todas las palabras que componen el campo semántico etiquetado como *Habilidades*, destaca el hecho de que la mayor parte de ellas son nominalizaciones, es decir, sustantivos formados por derivación a partir de verbos o adjetivos. El número de este rasgo es alto en el JadCor. Biber (1988:227) señala como nominalizaciones las palabras que terminan en sufijos como *-tion*, *-ment*, *-ness* o *-ity*, así como sus plurales. Las nominalizaciones denotan un alto grado de abstracción en la información de un texto y, por lo tanto, puede llevar a concluir que los anuncios de empleo son textos con una alta carga de abstracción. Además, las nominalizaciones se utilizan como una forma de economía de la lengua para integrar información en un menor número de palabras, por ejemplo, para convertir oraciones en sintagmas nominales más compactos y eficaces.

En el corpus, por ejemplo, se presenta el siguiente ejemplo de nominalización: **Awareness of security issues in a networked environment*. En este ejemplo se puede observar que se trata de un elemento de una lista de requisitos que se les pide a los candidatos a la plaza vacante. Utilizando la nominalización de *awareness* se evita tener que exponer una oración completa como, por ejemplo, *candidates should be aware of security issues...*

En el JadCor hay apariciones de nominalizaciones en todos los textos. Este hecho puede deberse a la intención de condensar la máxima información en el menor espacio, pues los anuncios de empleo deben de ser concisos y exponer de un modo claro la información para que el receptor la reciba correctamente. Con un mayor número de palabras en el texto, se puede perder el propósito comunicativo de estos anuncios. También responde a una falta de contenido dinámico, al ser eminentemente conceptual.

Hay que señalar que en este corpus se da una curiosa ambigüedad léxica con algunos nombres. Se trata de sustantivos que generalmente denominan partes del cuerpo, pero que en este corpus se utilizan en sentido figurativo. Aparecen, por ejemplo, *body*, *hands*, *head* y *heart*.

En el siguiente fragmento se muestran ejemplos de las apariciones de estos sustantivos en el JadCor y, como se puede observar, en ningún caso se refieren a la parte del cuerpo propiamente dicho:

- *We need a focussed, driven and ambitious individual who is comfortable getting involved in detail and be very **hands** on*
- *Conveniently located at the **heart** of the West Midlands, we have some superb rural scenery right on our doorstep,*
- *a public **body** dedicated to safeguarding children in the family courts.*
- *The Health Protection Agency (HPA) is an independent **body** that protects the health and well-being of everyone in England and Wales*
- *Reporting to the **Head** of Technical and Financial Advice and leading a small team*

En el fragmento se observa que en *hands* tiene un sentido figurado con *on*, *heart*, indica un lugar (en el “corazón”, centro de West Midlands), *body* se refiere a un organismo público. Por último *head* se utiliza en su acepción de director, la persona responsable de un departamento.

4.2.2 Adjetivos

Los adjetivos están por detrás de sustantivos y verbos en frecuencia en el JadCor, aunque se podría esperar un número más alto en textos tan descriptivos como son los anuncios de empleo. Cabe señalar que los adjetivos en grado comparativo y superlativo apenas aparecen en el corpus, siendo los más frecuentes los adjetivos en grado positivo.

Los únicos adjetivos en grado comparativo que hay en el corpus son: *higher* (4 apariciones), *larger* (1), *longer* (1), *lower* (1), *broader* (1) y *better* (1); y en grado superlativo: *largest* (4 apariciones), *broadest* (1), y *highest* (1).

En cuanto a su distribución, en su mayor parte los adjetivos acompañan a las descripciones de los puestos y a la cualificación requerida a los candidatos, aunque también hay un elevado número de adjetivos generales:

- **Descripción (D):** en este grupo se engloban los adjetivos relacionados con la descripción de la oferta, por ejemplo, *divisional*, *commercial*, *medical* y *corporate*, etc.
- **Nacionalidad (N):** a este grupo pertenecen los adjetivos relacionados con el lugar de origen del candidato que pueden determinar su exclusión del puesto de trabajo, por ejemplo, si la oferta es sólo para

ciudadanos de la Unión Europea, por motivos de permiso de residencia. Son, por ejemplo, *aboriginal*, *Canadian* o *European*, etc.

- **Cualificación (C):** calificaciones de los candidatos, por ejemplo, *capable*, *competitive*, *confident* y *hard-working*, etc.
- **Puesto (P):** referentes a las características del puesto de trabajo, por ejemplo, *full-time*, *negotiable*, *temporary*, *non-stipendiary* y *desirable*, etc.
- **Lugar (L):** en este grupo se incluyen los adjetivos que indican un lugar, por ejemplo, *nationwide*, *overseas*, *southern* y *local*, etc.
- **General (G):** al igual que sucedía con los nombres, en el corpus se encuentra un alto número de adjetivos que no se pueden catalogar dentro de una categoría específica dentro de los anuncios de empleo, por ejemplo, *radical*, *private*, *unrivalled*, etc.

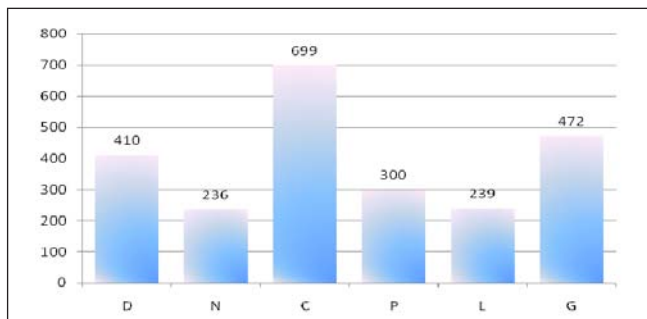


Figura 5. Distribución de los adjetivos del JadCor por campos semánticos.

En la figura 5 se muestra la distribución de los adjetivos del JadCor agrupados en campos semánticos. Como se puede observar, los adjetivos con mayor número de apariciones en el corpus son los relacionados con la cualificación de los candidatos y los generales, que no se engloban en una categoría determinada. Como ya se indicó cuando se trataron los campos semánticos de los nombres, este elevado número de adjetivos sin catalogar específicamente, puede indicar que un sublenguaje no es cerrado o no tanto como se desearía.

El siguiente ejemplo es una muestra de la adjetivación en el corpus: *Camden is a **vibrant** and **diverse** London borough. The Council is **proud** of its **public** service ethos and **strong** focus on performance.*

Hay que señalar también que los nombres de los anuncios de empleo, en muchas ocasiones, no permiten adjetivos calificativos, por ejemplo, *email* o *address*, *telephone* o *fax*. Esto no quiere decir que en otros contextos no puedan ir acompañando a estos sustantivos, pero no en el contexto de los anuncios de empleo, en donde sólo se proporciona información de un modo breve y conciso (por ejemplo, *telephone 0191 334 6499*; o *fax 091 334 6495*).

Otro rasgo destacable de los adjetivos en el JadCor, es que la mayor parte de ellos ocurren en posición atributiva. Esto puede deberse, como ya se ha comentado, al carácter descriptivo de estos textos.

4.2.3. Verbos

Los verbos son la segunda categoría en frecuencia en el JadCor. Sin embargo, es necesario ver que tiempo verbal es el predominante. En este caso, en todos los textos que componen el corpus hay verbos en presente, pero con una escasa incidencia. A pesar de esto, es el tiempo predominante en el corpus, ya que los verbos en pasado (en donde se incluyen los verbos de aspecto perfectivo) no aparecen en todos los textos y no resultan muy significativos dado que más de la mitad de los textos carecen de él.

En la figura 6 se muestra la distribución de los verbos del JadCor agrupados según su tiempo verbal. Como se puede observar, los que más apariciones tienen en el corpus son los verbos en presente, duplicando o incluso triplicando los verbos en pasado. En este caso no se ha llevado a cabo un análisis más profundo, para determinar cuando el presente simple o el presente continuo tienen valor de futuro, o cuando el modal *will* también conlleva este aspecto.

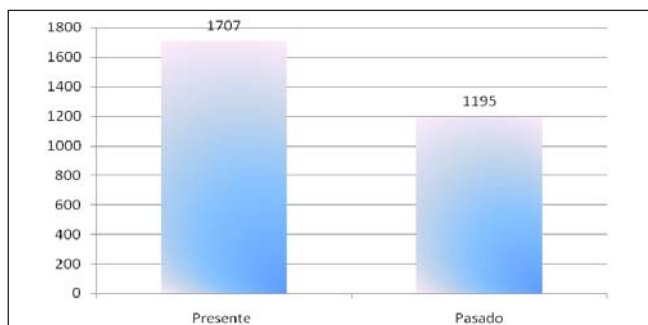


Figura 6. Distribución de los tiempos verbales.

Ha de puntualizarse también que en esta categoría se encuadran las seis personas del presente simple: la coincidente con la forma base del verbo y su correspondiente alternativa para la tercera persona, con *-s* o *-es* al final, así como los tiempos compuestos con *be*, como puede ser el presente continuo. Por ejemplo, *This service takes care of Development Control, Building Standards and Development Planning - and takes on the enormous responsibility of devising and implementing very large-scale regeneration projects*. El verbo en presente, *takes*, se utiliza para describir las funciones del departamento para el que está ofertado el anuncio.

Sin embargo, un mayor número de verbos en presente sería de esperar, pues por el contenido habitual de los anuncios de empleo se esperan acciones y situaciones tanto permanentes como puntuales, todas las cuales se expresan habitualmente con verbos en el tiempo presente. En general, en el corpus hay un número relativamente bajo de verbos. Esto puede ser un rasgo distintivo de este sublenguaje.

Por su parte, los verbos en pasado están asociados con los textos narrativos y descriptivos. Hay apariciones en 16 de los textos que componen el corpus. Por lo tanto, tampoco resulta un rasgo frecuente dado que aproximadamente la mitad de los textos del corpus carecen de él. Por ejemplo, *This Travelling Fellowship has been endowed with the object of assisting women graduates who have considerable experience of research*. En este caso, se utiliza el pretérito perfecto en pasiva para indicar que es una situación que empezó en el pasado, pero que tiene repercusiones que llegan prácticamente hasta el presente.

También es importante considerar los verbos modales en los anuncios de empleo. Hay que señalar que, aunque están presentes en la mitad de los textos del JadCor, no presentan tampoco ninguna excepcionalidad. Sin embargo, no alcanzan el 4% del total de los verbos en el corpus, lo viene a demostrar la escasa relevancia que se acaba de mencionar.

En el JadCor hay apariciones de modales de posibilidad, de predicción y de necesidad. En primer lugar, los modales de posibilidad se utilizan en algún caso para abrir la posibilidad de negociar algún aspecto de la oferta de empleo y, en otras ocasiones, lo deja abierto, sin concretar (por ejemplo, *these conditions may be modified according to the nature of the Project*). Además, la escasa presencia de los modales de predicción se justifica con el hecho de que en estos textos no se expone ningún tipo de predicción sino hechos reales, por ejemplo, *Only in a borough this diverse will you gain so much satisfaction and experience*. En este caso, *will* se utiliza para expresar lo que el candidato obtendrá de compensación derivada del puesto: *satisfaction* y *experience*. Por último, los modales de necesidad se utilizan en el corpus (aunque de un modo muy aislado, como se ha mencionado), por ejemplo, para indicar uno de los requisitos que necesita tener un candidato para poder ser tenido en cuenta y presentarse a la oferta vacante (*The candidate's areas of expertise should combine Health Sociology and any other substantive area in the discipline*).

Como se ha indicado anteriormente con respecto a los sublenguajes, es importante determinar qué rasgos o características estén presentes, pero es igualmente importante determinar cuáles no lo están y si hay alguna justificación para ello.

En el corpus JadCor de anuncios de empleo, por ejemplo, hay escasas apariciones de verbos de privacidadⁱⁱ y de persuasión (Quirk 1985:1180). Los primeros, los verbos de privacidad, indican la formación de una opinión, estado o sentimiento y por lo tanto expresa subjetividad, implicando, de algún modo, la participación del lector (por ejemplo, *If you believe that you have what our client is looking for then we would like to hear from you*). En el ejemplo se puede observar que se utiliza el verbo *believe* para dirigirse directamente al receptor y provocar una respuesta. Es un modo de involucrar al propio receptor, porque es él quien considera si está o no en posición de cubrir las expectativas del puesto. Es decir, lo anima a que si cumple las condiciones determinadas en el anuncio, se ponga en contacto con la empresa. Los

anuncios de empleo tienen un carácter altamente informativo (López Sanjuán 2006), y exponen una serie de datos que no están sujetos a ningún tipo de subjetividad, de ahí la explicación de su escasa presencia en el corpus.

Algo similar sucede con los verbos de persuasión. Todos estos verbos están formados al menos por dos argumentos e indican la implicación subjetiva del hablante. Tan sólo hay dos apariciones de estos verbos en el JadCor y se trata en ambas ocasiones del verbo *request* (por ejemplo, *Request an application pack here*). En este caso, *request* señala el lugar en donde obtener la información y documentación para solicitar el puesto.

Otros verbos a tener en cuenta son los verbos públicos, entre los que encontramos, por ejemplo, *acknowledge, admit, agree, assert, claim, complain, declare, deny, explain, insist, mention, promise, remark, reply, report, say, suggest o write*. Como se puede predecir de esta lista de verbos, la mayor parte de ellos están compuestos por tres argumentos. Este rasgo aparece en 16 de los 35 textos del corpus y, al igual que los anteriores, sus valores son poco frecuentes, dado que el promedio de todos los textos del corpus es de 1.86. En el ejemplo, *If you believe that you have what our client is looking for then we would like to hear from you*, se puede observar que se utiliza el verbo *believe* para dirigirse directamente al receptor y provocar una respuesta. Es un modo de involucrar al propio receptor, porque es él quien considera si está o no en posición de cubrir las expectativas del puesto.

4.2.4. Adverbios

En el corpus JadCor no hay una gran cantidad de adverbios, aunque destacan especialmente los adverbios de tiempo.

Los adverbios que aparecen en el corpus se pueden distribuir de la siguiente forma:

- **Adverbios de cantidad (C):** por ejemplo, *more, rather*, etc.
- **Adverbios de lugar (L):** por ejemplo, *abroad, here, geographically*, etc.
- **Adverbios de modo (M):** *operationally, informally commercially*, etc.
- **Adverbios de negativos (N):** por ejemplo, *not*.
- **Adverbios de tiempo (T):** por ejemplo, *never, monthly, now*, etc.

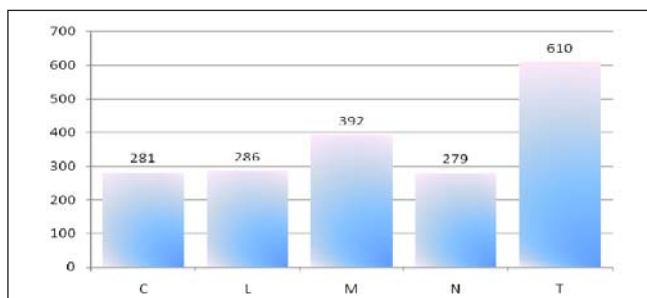


Figura 7. Distribución de los tipos de adverbios en el JadCor.

En la figura 7 se puede observar que los adverbios predominantes en el corpus son los adverbios de tiempo, seguido por los de modo y, en menor medida, por los de cantidad, lugar y negación.

Los adverbios de tiempo hacen referencia explícita al contexto situacional de un texto. Este rasgo sólo aparece en doce de los textos del JadCor y su relevancia, por lo tanto, parece limitada, a pesar de que son el tipo de adverbios más frecuentes. Este ejemplo extraído del corpus *The posts are tenable from 1 January 2005 (or as soon afterwards as possible)* muestra que *soon afterwards* parte de la fecha que aparece en el sintagma *1 January 2005*.

Al igual que los adverbios de tiempo, los adverbios de lugar expresan una referencia explícita al contexto. En el JadCor aparecen en catorce de los textos, pero con una frecuencia también baja. Por ejemplo, *As an alternative you can use the downloadable version of the application form below*. El adverbio *below* plantea una referencia interna en el texto, al modo de realizar la solicitud e indicando dónde se puede obtener la documentación pertinente para hacer la solicitud.

Las escasas apariciones de los adverbios negativos pueden llevar a considerar que las oraciones de este corpus no son en su mayoría afirmativas, quedando por determinar aún las oraciones interrogativas. La escasa aparición de cualquier tipo de negación en estos textos probablemente se deba a que el contenido de los anuncios de empleo no pretende ni prevenir ni prohibir, etc. sino que contienen listas de requisitos, condiciones, etc. expresadas de forma positiva y económica.

4.2.5. Pronombres

En el JadCor hay pronombres personales, relativos e indefinidos.

En la figura 8 se muestra la distribución de los tipos de pronombres en el JadCor. Los que más apariciones tienen son los pronombres personales, seguidos por los indefinidos y los relativos. Hay que señalar que los pronombres personales, cuadruplican al resto de los otros pronombres. Esto puede ser debido a que la escasa referencia pronominal que existe en el corpus es más bien de tipo específica y personal y que en este tipo de anuncios no cabe esperar muchas referencias indefinidas.

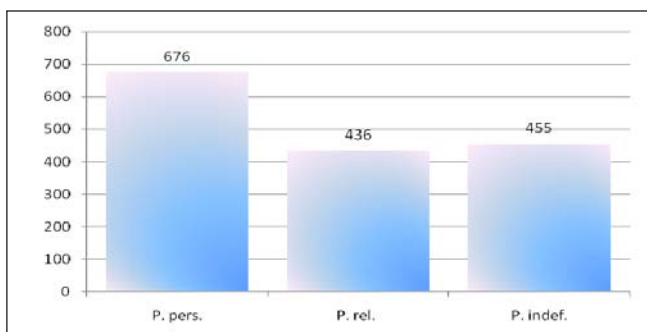


Figura 8. Distribución de los tipos de pronombres en el JadCor.

Sin embargo, en las ofertas de empleo sí que existen algunos pronombres de primera y segunda persona. En la figura 5 se muestran la distribución de este tipo de pronombres en el JadCor. Como se puede ver, los más abundantes son los pronombres de segunda persona y los de tercera persona, pues los de primera persona son poco significativos.

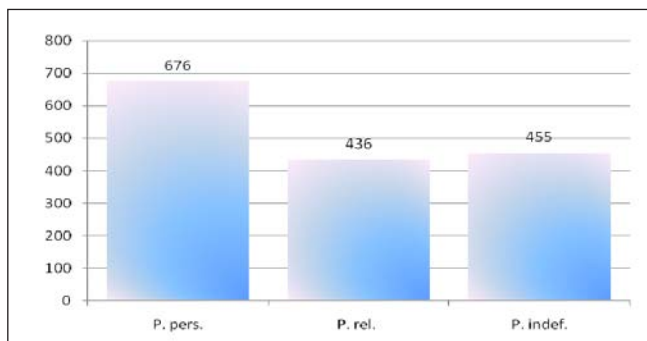


Figura 9. Distribución de los pronombres personales en el JadCor.

Los pronombres de primera persona en su forma singular no aparecen en ninguno de los textos que conforman el corpus JadCor, sólo la forma plural. En muchos casos, el pronombre *we* se refiere a la empresa que ofrece el puesto de trabajo para dar una imagen de equipo, cercana y accesible, que invita al posible candidato a integrarse, como en el ejemplo siguiente: *To deliver this vision we have created a key Divisional Manager post that will hold strategic and operational responsibility for specific services.*

Si se comparan en los textos individualmente los pronombres de primera persona y los pronombres de segunda persona, se puede comprobar que los primeros son inferiores en número a los segundos. En este caso se debe a que el agente principal del anuncio de empleo es el candidato más que la empresa ofertante. También se puede deducir que las ofertas de empleo al utilizar los pronombres de segunda persona, van dirigidas a alguien con unas características concretas (que son las que se redactan en cada oferta, donde se quiere despertar el interés del lector de una forma personalizada).

Los pronombres de segunda persona están presentes en la mayor parte de los textos que conforman el JadCor, pero su frecuencia en cada uno no es extremadamente alta. La existencia de un alto número de nombres en oraciones yuxtapuestas, pero independientes justifica la ausencia de pronombres personales o el bajo número de apariciones. En este ejemplo se muestra la utilización del pronombre *you*: *By joining CAFCASS you will become part of the largest children's and family service in the UK. But more than this, you will be working for an organisation unlike any other in the world - a public body dedicated to safeguarding children in the family courts.* Aquí, el pronombre de segunda persona se utiliza de un modo persuasivo, para instar al lector a reaccionar y optar al puesto que se oferta. En este sentido cabe señalar la utilización del pronombre de segunda persona para dirigirse de un modo sencillo, corto y directo al lector en vez de utilizar, por ejemplo, *the successful candidate* o incluso el más genérico *people who* (por ejemplo, *people who join CAFCASS will become...*). Cabe recordar que la economía del lenguaje es uno de los rasgos que define un sublenguaje.

Los pronombres de tercera persona suelen coocurrir con frecuencia con los verbos en pasado y con aspecto perfectivo, siendo característicos del estilo narrativo. Esta puede ser una de las causas por la que sólo aparecen en 4 de los textos del JadCor y no parecen ser muy

frecuentes. Sin embargo, el pronombre *it* muestra alguna divergencia con la lengua estándar, ya que en la lengua general la frecuencia de aparición de este pronombre suele ser más alta. Al ser textos con conceptos claramente definidos (a quien se dirige, quien se dirige, que quiere decir, etc.) no se utilizan demasiados pronombres de tercera persona como referencia indirecta a otras entidades o personas. Además, cada mensaje está contenido habitualmente en una sola oración y tampoco llega a haber un campo referencial suficientemente conocido entre el autor y el lector. Por ejemplo, *It also operates an independent service, providing specific solutions, for a large number of third party customers.* Este es uno de los pocos casos en el corpus en que la descripción de la empresa se extiende lo suficiente para permitir el uso del pronombre *it*.

4.2.6. Determinantes

Los determinantes tienen menos apariciones que nombres, verbos y adjetivos, pero más que el resto de categorías en el JadCor.

Todos los artículos determinados e indeterminados están presentes, así como los demostrativos. Además, también se encuentran en el corpus otros determinantes como pueden ser los indefinidos, por ejemplo, *several, some o each* o algunos posesivos, por ejemplo, *his o its*.

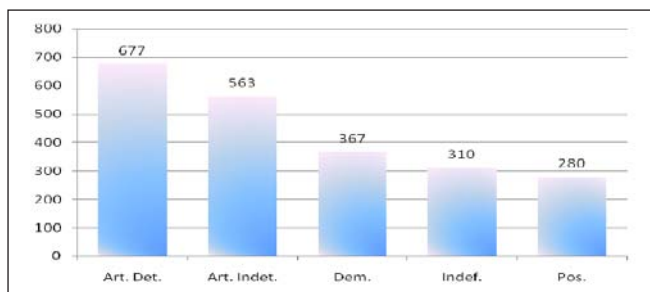


Figura 10. Distribución de los determinantes en el corpus.

La aparición de tantos determinantes, especialmente artículos, en el corpus los justifica la alta presencia de nombres que hay. Al igual que se señaló con respecto a los pronombres indefinidos, en este sublenguaje las referencias son concretas y bien definidas, debido al tipo de información que quieren hacer llegar al receptor.

4.2.7. Preposiciones

En todos los textos que conforman el JadCor están presentes las preposiciones, en la mayor parte de los casos, formando sintagmas preposicionales.

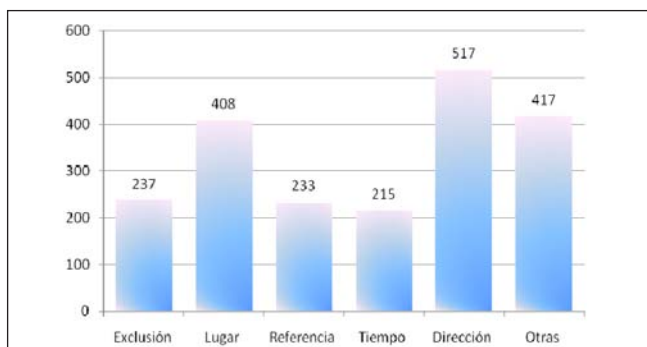


Figura 11. Distribución de las preposiciones en el JadCor.

Hay que apuntar que más de la mitad de las preposiciones del corpus indican una dirección (por ejemplo, *around*, *inside*, *towards*, *beyond*, etc.). Esto se puede justificar dada la gran cantidad de modificación que existe en este tipo de textos, para ser más descriptivos y ofrecer al receptor la mayor información posible sobre el puesto vacante y la empresa que lo ofrece.

En este ejemplo, *The postholder will be a member of the scientific staff working on aspects of antimicrobial resistance, in particular infections in children caused by MRSA*, se puede observar que el número de sintagmas preposicionales es alto, consecuencia en muchos casos de la gran cantidad de nombres relacionados en este tipo de textos. Además, los sintagmas preposicionales, ya de por sí modificadores nominales pesados (en el sentido de que constan al menos de dos palabras) son en este caso particularmente complejos y pesados, conllevando cadenas de sintagmas preposicionales insertadas en otras.

En escasas apariciones (sólo una en cada en nueve de los textos del corpus), las preposiciones ocupan en el corpus una posición final, pues se suele señalar que este tipo de estructura es más propio de la lengua oral y de estilos informales, lo que podría explicar su escasa presencia en este tipo de textos.

4.2.8. Conjunciones

En el JadCor hay apariciones de conjunciones coordinadas y también subordinantes, que indican esos tipos de estructuras respectivamente.

En la figura 12 se muestra un gráfico comparativo de las conjunciones del JadCor. Como se puede observar, es mayor el número de conjunciones coordinadas que de las subordinantes. Sin embargo, esto no quiere decir que haya mayor número de oraciones coordinadas que subordinantes en el corpus, ya que estas conjunciones no tienen porque coordinar sólo al nivel oracional, sino que también pueden hacerlo al nivel sintagmático.

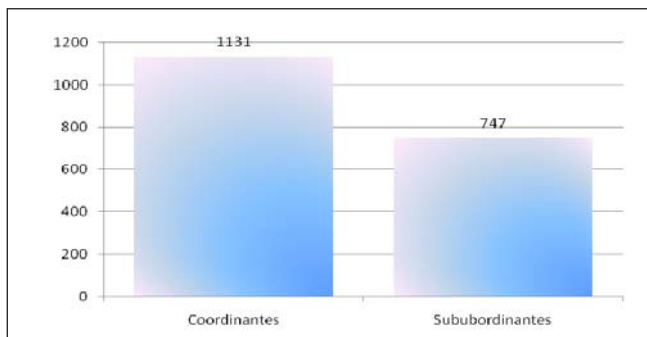


Figura 12. Distribución de las conjunciones en el JadCor.

En el JadCor se encuentran las conjunciones *or*, *but* y *and*, y coordinación sintagmática y oracional.

La coordinación oracional está presente en 16 de los 35 textos que componen el JadCor. Esto muestra que las oraciones del JadCor no son mayoritariamente oraciones compuestas y este dato ha de ir acompañado de un estudio de la presencia de oraciones subordinadas. En el ejemplo *The Travelling Fellowship will be available for all or part of the year 2005-2006 and the fund will pay research, travel and subsistence expenses normally up to £12,500 for the year* se puede ver un caso de coordinación oracional.

Por otro lado, la coordinación sintagmática engloba a todas las cadenas de verbo y verbo, sustantivo y sustantivo, adjetivo y adjetivo

o adverbio y adverbio que van unidas por *and* y *or*. *And* como conector sintagmático tiene una función integrativa, como se puede observar en el ejemplo *The successful candidate will have a degree in the Social Sciences and/or Humanities, editorial experience and some appreciation of issues relating to quality improvement, the organization and management of clinical work and of the culture change issues that this entails*. Aquí se puede observar que se trata de la coordinación de varios sintagmas nominales y en uno de los casos de los sintagmas preposicionales que postmodifican a uno de los sustantivos.

En cuanto a la subordinación, en el JadCor se encuentra subordinación temporal (por ejemplo, *until*, *when*), concesiva (*although*, *however* y *though*) y condicional (*if*), siendo la subordinación condicional la que está más presente, con mucha diferencia sobre los otros dos tipos que aparecen, aunque la incidencia no es alta. Este tipo de subordinación se utiliza en algunos casos para enmarcar el discurso y tiene funciones diferentes según sea su posición en la oración. Por ejemplo: *If you require further information before applying, contact Dr Laurence Kenney*. La oración condicional usada en el ejemplo expresa una condición real y ofrece al candidato, en caso de necesitar información extra, una dirección y una persona de contacto.

5. CONCLUSIÓN

Los resultados finales del análisis del corpus han servido para hacer una amplia descripción del sublenguaje de los anuncios de empleo.

En muchos casos, se trata de un análisis que se puede quedar en superficial, pues no se indican, por ejemplo, en el caso de los nombres, los tipos (por ejemplo, si son abstractos/concretos), ni el género ni el número de los nombres del corpus y tampoco se ha profundizado en el análisis sintáctico. Pero ese no era el objetivo de este artículo, en el que sólo se ha pretendido describir y poner de manifiesto algunas de las peculiaridades de este sublenguaje de los anuncios de empleo sin tener en cuenta sus posibles aplicaciones computacionales posteriores.

De este modo, en primer lugar se ha mostrado que la estructura retórica de los anuncios de empleo está bien definida, ya que muestra un esquema con elementos obligatorios y opcionales.

Además, tomando como punto de partida las categorías gramaticales presentes en el corpus se ha realizado la descripción de cada una de ellas, proporcionando no sólo información morfológica, sino también léxica y semántica y, en algunos casos, sintáctica y pragmática. En el plano morfológico se ha observado que no están todas las categorías presentes, pues no hay interjecciones, y en qué medida es cada una relevante en este tipo textual (abundando nombres y adjetivos, lo que demuestra el carácter descriptivo de este sublenguaje).

Por otra parte, en lo que se refiere al léxico, el análisis mostraba que los anuncios de empleo tampoco tiene un carácter muy cerrado. En el corpus JadCor, no se han acotado las profesiones de las ofertas, lo que conlleva que el vocabulario permanezca relativamente abierto. Además, los sustantivos no se pueden agrupar fácilmente en clases semánticas porque aparece una enorme y diversa cantidad de nombres de la lengua general. Esto se debe a que en este sublenguaje (y, por lo tanto, en el corpus) el dominio del subgénero se ha acotado de una forma muy amplia.

También se ha visto que los anuncios de empleo puede contener incursiones personales ocasionales en los anuncios más breves (generalmente para dar énfasis e integrar al receptor) y tiene una clara tendencia a la economía expresiva, aunque no es exactamente telegráfico. De ahí que no haya un gran número de pronombres personales ni de conjunciones subordinantes (que denotaría una mayor densidad oracional), ya que la mayor parte de las oraciones son simples y la subordinación, en su mayoría temporal, concesiva y condicional.

En conjunto, por los resultados obtenidos en el análisis, se puede concluir que los anuncios de empleo presentan características propias de los sublenguajes: sistematicidad, economía de expresión, léxico (relativamente) cerrado, etc., lo que se pone de relevancia a través del un análisis lingüístico a partir de las categorías gramaticales del corpus JadCor de anuncios de empleo.

NOTAS

¹ Según Leech (1991:27), un corpus es representativo cuando los resultados obtenidos de su contenido se pueden generalizar a una entidad lingüística mayor.

- ² Traducimos literalmente el término *privative verbs* de Biber (1998), tal y como hacen González Álvarez y Pérez Guerra (2003). Evidentemente estos verbos coinciden con las categorías tradicionales de “verbos sensoriales” y “verbos cognitivos”.

OBRAS CITADAS

- Bárcena, E. 1994. *Grammatical Representation of Examples on Interlingual Example-based MT*. Tesis doctoral. Manchester: UMIST
- Biber, D. 1988. *Variation across speech and writing*. Cambridge: Cambridge University Press
- González Álvarez, D. y J. Pérez Guerra. 2003. “La estadística de paseo con la Estilística o el estado de la cuestión en el análisis textual multidimensional”. *Revista Canaria de Estudios Ingleses* 46:131-159.
- Kittredge, R. 2003. “Sublanguages and Controlled Languages” en R. Mitkov, ed. *The Oxford Handbook of Computational Linguistics*. Oxford: Oxford University Press. 430-448.
- Kittredge, R. 1987. “The Significance of Sublanguage for Automatic Translation” en S. Niremburg, ed. *Machine Translation: Theoretical and Methodological Issues*. Cambridge: Cambridge University Press. 59-67.
- Leech, G. 1991. “The state of the Art in Corpus Linguistics” en K. Aijmer y B. Altenberg, eds. *English Corpus Linguistics. Studies in Honour of Jan Svartvik*. London: Longman. 8-29.
- Lehrberger, J. 1986. “Sublanguage Analysis”. En Grishman, R. y R. Kittredge, eds. *Analyzing Language in Restricted Domains. Sublanguage Description and Processing*. New Jersey: Lawrence Erlbaum. 19-38.
- Lehrberger, J. y L. Bourbeau 1988. *Machine Translation. Linguistic Characteristics of MT Systems and General Methodology of Evaluation. Linguisticae Investigationes: Supplementa 15*. Amsterdam: John Benjamins.
- López Sanjuán, V. 2003. *Tratamiento e identificación de sublenguajes a través de corpus electrónicos en lengua inglesa y de herramientas estándar de análisis con fines computacionales*. Sin publicar.
- López Sanjuán, V. 2006. *La identificación de sublenguajes con fines computacionales basada en corpus electrónicos en lengua inglesa y en la aplicación de herramientas de análisis y su comparación a partir del Modelo Multidimensional de Biber*. Tesis doctoral. Sin publicar.

- Moskovich, W. 1982. "What is a sublanguage? The notion of sublanguage in modern Soviet Linguistics" en R. Kittredge y J. Lehrberger, eds. *Sublanguages. Studies of Language in Restricted Semantic Domains*. Berlín: Walter de Gruyter. 191-205.
- Somers, H, B. Black, J. Nivre, T. Lager, A. Multari, L. Gilardoni, J. Ellman y Alex Rogers 1997. "Multilingual generation and summarization of job adverts: the TREE project". Fifth Conference on Applied Natural Language Processing. Washington. 269-276.
- Quirk, R., S. Greenbaum, G. Leech y J. Svartvik. 1985. *A Comprehensive Grammar of the English Language*. Harlow: Longman.