



Universidad
de La Laguna

Departamento de Didácticas Específicas
Área de Didáctica de las Ciencias Experimentales
Facultad de Educación

CONCEPTOS RELACIONADOS
CON ESTRELLA. LINGÜÍSTICA DE
CORPUS DE ASTRONOMÍA.

Cristina Sílvia Hansen Ruíz

2011

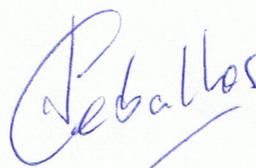
UNIVERSIDAD DE LA LAGUNA
DEPARTAMENTO DE DIDÁCTICAS ESPECÍFICAS
Área de Didáctica de las Ciencias Experimentales

Jesús Miguel Pérez Ceballos, con DNI 41890371B, Doctor en Ciencias (Sección Químicas) y Catedrático de Escuelas Universitarias en el Área de Didáctica de las Ciencias Experimentales, con destino en el Departamento de Didácticas Específicas de la Universidad de La Laguna, como Director del trabajo de Tesis Doctoral y Tutor de la Licenciada D^a Cristina Silvia Hansen Ruiz,

Autorizo

A que presente su trabajo de Tesis titulado “Conceptos relacionados con estrella. Lingüística de corpus de Astronomía”, en la Facultad de Educación de esta Universidad, considerando que reúne los requisitos académicos y formales previstos en la legislación vigente para optar al Grado de Doctor.

Lo que hago constar en La Laguna, a veintiséis de agosto de 2011.



Fdo.: Dr. Jesús Miguel Pérez Ceballos

CONCEPTOS RELACIONADOS CON *ESTRELLA*. LINGÜÍSTICA DE CORPUS DE ASTRONOMÍA.

Cristina Silvia Hansen Ruiz

Tesis presentada para la obtención del título de doctor en Ciencias de la Educación. Dirigida por el Doctor en Ciencias y Catedrático de Escuelas Universitarias en el Área de Didáctica de las Ciencias Experimentales:

Jesús Miguel Pérez Ceballos.

Departamento de Didácticas Específicas
Área de Didáctica de las Ciencias Experimentales
Facultad de Educación



Universidad
de La Laguna

Octubre de 2011

Para Erik

Para mi familia

Para mis maestros y profesores

AGRADECIMIENTOS

A mi director de tesis, el catedrático de Escuelas Universitarias Jesús Pérez Ceballos, por confiar en mis capacidades, enseñarme la rigurosidad con la que se debe llevar a cabo una investigación, por su paciencia y ese extraordinario dominio del lenguaje, básico para el entendimiento y explicación de cualquier tema.

A los compañeros del Departamento de Didácticas Específicas con los que coincidí a ratos a lo largo de todos estos años, especialmente a Angelo Galotti por sus largas discusiones sobre el funcionamiento de nuestra herramienta de trabajo y a María del Carmen Domínguez cuya tesis he usado de referencia. Al resto del profesorado de Didáctica y de la Facultad de Educación de la Universidad de La Laguna que en algún momento contribuyeron a la realización de la tesis.

A los investigadores que he conocido en distintos congresos y seminarios y en especial a los lingüistas que acudieron al AACL 2008 por la gran ayuda y formación en conceptos y terminología lingüística que me proporcionaron.

A todos los que han sido mis profesores, maestros o me han enseñado algo en esta vida y contribuido a mi formación tanto profesional como personal.

A Erik Stengler por estar siempre ahí, y apoyarme.

Y finalmente a mi padre, Niels Erik Hansen, por enseñarme que todo es posible y que cada uno tiene su propia felicidad.

ÍNDICE

AGRADECIMIENTOS	v
ÍNDICE	vii
LISTA DE TABLAS	xi
LISTA DE FIGURAS	xix
1.- INTRODUCCIÓN	1
1. 1.- Marco para la Realización de la Tesis	3
1. 2.- Elección del Tema	8
1. 3.- Objetivos de la Tesis	9
1. 4.- Estructura de la Tesis	10
2.- FUNDAMENTOS	13
2. 1.- Los Conceptos	15
2. 2.- Lenguaje y Adquisición de los Conceptos	18
2. 3.- La Lingüística Computacional y la Lexicometría.....	26
2. 3. 1.- El lenguaje de las ciencias desde el punto de vista de la lingüística	30
2. 3. 2.- Análisis de texto asistido por ordenador y análisis de Redes	33
3.- METODOLOGÍA	43
3. 1.- Antecedentes de la Metodología Utilizada	45
3. 2.- Metodología de Investigación	48
3. 2. 1.- Preparación del texto	49
3. 2. 2.- El procesamiento y análisis de los textos.....	58
3. 2. 2. 1.- Análisis de frecuencias con PAFE: lematización.....	59
3. 2. 2. 1. 1.- Clasificación de palabras	63
3. 2. 2. 1. 2.- La aplicación “Análisis de Frecuencias” del PAFE	67
3. 2. 2. 2.- Estudio de entorno con PAFE: relaciones	76
3. 2. 2. 2. 1.- La aplicación “Estudio de Entorno” del PAFE	80

ÍNDICE

3. 2. 2. 2. 2.- Colocaciones y uniones de ULL	83
3. 2. 2. 2. 3.- Vocabulario más frecuente y vocabulario específico	86
3. 2. 2. 2. 4.- Determinación del sistema óptimo en el que estudiar redes semánticas	90
3. 2. 2. 2. 5.- Redes semánticas: ULL relacionadas	98
3. 2. 3.- Esquema de la metodología	110
3. 2. 4.- Posibles fuentes de error	113
4.- UN ESTUDIO DE CASO: UN TEXTO ESCOLAR DE ASTRONOMÍA	115
4. 1.- Preparación del Texto.....	119
4. 2.- Análisis de Frecuencias del Texto Preparado	125
4. 3.- Análisis de Frecuencias del Texto Lematizado/Transformado	142
4. 4.- Estudios de Entorno con PAFE: Relaciones	157
4. 4. 1.- Búsqueda de colocaciones	157
4. 4. 2.- Vocabulario más frecuente y vocabulario específico	186
4. 4. 3.- Determinación del sistema a estudiar	193
4. 4. 4.- Redes semánticas.....	212
5.- DETERMINACIÓN DE ERRORES	247
5. 1.- Errores en la Preparación del Texto	249
5. 2.- Errores en el Análisis del Texto Preparado	251
5. 3.- Errores en el Análisis del Texto Lematizado/Transformado	252
5. 4.- Errores en el Estudio de Entorno.....	253
5. 4. 1- Errores en la búsqueda de colocaciones	253
5. 4. 2.- Errores en el vocabulario más frecuente y el vocabulario específico	254
5. 4. 3.- Errores la determinación del sistema óptimo y las redes semánticas	255
6.- CONCLUSIONES	257
6. 1.- Preparación del Texto.....	259
6. 2.- Análisis de Frecuencias del Texto Preparado y Lematizado.....	261
6. 3.- Estudio de Entorno	264
6. 3. 1.- Búsqueda de colocaciones	264
6. 3. 2.- Vocabulario más frecuente y vocabulario específico	265
6. 3. 3.- Determinación del sistema óptimo	266
6. 3. 4.- Redes semánticas.....	266
6. 3. 5.- Análisis de errores	268

6. 4.- Esquema de la Nueva Metodología.....	269
7.- CUESTIONES ABIERTAS.....	273
7. 1.- Preparación del Texto.....	275
7. 2.- Análisis de Frecuencias del Texto Preparado y Lematizado.....	276
7. 3.- Estudio de Entorno	278
7. 3. 1.- Búsqueda de colocaciones	278
7. 3. 2.- Vocabulario más frecuente y vocabulario específico	279
7. 3. 3.- Determinación del sistema óptimo	280
7. 3. 4.- Redes semánticas.....	280
7. 4.- Otros.....	281
GLOSARIO.....	283
BIBLIOGRAFÍA	297
ANEXOS	323
ANEXO A: Instalación y Administración del Programa PAFE	325
ANEXO B: “Medida de la Distancia y la Edad de un Cúmulo Globular de Estrellas”	330
ANEXO C: Texto Preparado FESTESE001V.....	336
ANEXO D: La Aplicación Informática PAFE.....	345
ANEXO E: Palabras y Locuciones Susceptibles de Ser Eliminadas	357
ANEXO F: Listado de Palabras Funcionales.....	359
ANEXO G: Texto Lematizado FESTESE001V	360
ANEXO H: Frecuencias del Texto Preparado FESTESE001V	371
ANEXO I: Frecuencias del Texto Lematizado FESTESE001V.....	386
ANEXO J: Hallando Colocaciones en FESTESE001V	399
ANEXO K: Nuevo Texto Transformado FESTESE001VD.....	406
ANEXO L: Texto con Unificación de Grafías: FESTESE001V-S.....	414
ANEXO M: Texto Transformado FESTESE001V-F	423
ANEXO N: Texto Transformado FESTESE001V-S.....	432
ANEXO O: Frecuencias del Texto Preparado FESTESE001V-S	441
ANEXO P: Frecuencias del Texto Transformado FESTESE001V-F	456
ANEXO Q: Frecuencias del Texto Transformado FESTESE001V-S	468
ANEXO R: Frecuencias del Texto Transformado FESTESE001VD.	480

LISTA DE TABLAS

Tabla 1.- <i>Cronología del aprendizaje del lenguaje</i>	23
Tabla 2.- <i>Cronología de la adquisición de conceptos</i>	24
Tabla 3.- <i>Algunas características de corpus del español</i>	28
Tabla 4.- <i>Diferenciación de las grafías con distintas funciones</i>	55
Tabla 5.- <i>Ejemplo de simbología matemática eliminada o no reconocida por el programa PAFE</i>	56
Tabla 6.- <i>Ejemplo de diferencias entre el texto original y el preparado</i>	57
Tabla 7.- <i>Clasificación de las palabras del primer párrafo del texto preparado mostrado en la Tabla 6</i>	60
Tabla 8.- <i>Ejemplo de plantilla usado por el programa PAFE</i>	67
Tabla 9.- <i>Clasificación en géneros, subgéneros y niveles de subgéneros de los textos a estudiar</i>	69
Tabla 10.- <i>Texto preparado a la izquierda y lematizado a la derecha</i>	71
Tabla 11.- <i>Fragmentos de las 20 frecuencias más altas en los textos preparado y lematizado, producidas por PAFE</i>	73
Tabla 12.- <i>Ejemplos de cambios en frecuencia absoluta entre el texto preprado y el texto lematizado, FESTEENSE001V</i>	75
Tabla 13.- <i>Ejemplos de valores de entorno</i>	79
Tabla 14.- <i>Características del primer Estudio de Entornos: hallando colocaciones</i>	83
Tabla 15.- <i>Ejemplo de propuestas de unión aceptadas para el texto preparado</i>	85
Tabla 16.- <i>Límites de la franja definida por el promedio</i>	87
Tabla 17.- <i>Características del Estudio de Entorno para determinar el sistema a estudiar</i>	91
Tabla 18.- <i>Sistemas a analizar</i>	94
Tabla 19.- <i>Sistemas a analizar en dos entornos</i>	94
Tabla 20.- <i>Características del Estudio de Entorno para el análisis de redes</i> ...	98
Tabla 21.- <i>Partes del ejercicio 4</i>	118

LISTA DE TABLAS

Tabla 22.- Características del texto escaneado (medido con Word 2007) FESTESE001V.....	118
Tabla 23.- Simbología matemática eliminada o no reconocida por el programa PAFE.....	120
Tabla 24.- Palabras que pueden ser confundidas con otra función gramatical o semántica.....	121
Tabla 25.- Unificaciones de grafías en FESTESE001V-S.....	123
Tabla 26.- Unificaciones de grafías por omisión o sustitución en FESTESE001V-S.....	124
Tabla 27.- Características del texto preparado (medido con Word 2007), FESTESE001V.....	125
Tabla 28.- Características de los textos preparados.....	126
Tabla 29.- Características de los textos lematizados (AO) y transformados (AF y AG).....	126
Tabla 30.- Franjas en frecuencia acumulada (N) del texto preparado.....	130
Tabla 31.- Otras características del texto preparado.....	131
Tabla 32.- Características por categoría gramatical del texto preparado FESTESE001V.....	132
Tabla 33.- Distribución del número de vocablos y frecuencias en todo el texto preparado en los tres análisis.....	134
Tabla 34.- Distribución del número de vocablos y frecuencias en cada franja en los tres análisis en el texto preparado.....	135
Tabla 35.- Distribución de cada categoría gramatical en todo el texto preparado en los tres análisis.....	135
Tabla 36.- Distribución de las subcategorías léxicas en los tres análisis en el texto preparado.....	137
Tabla 37.- Distribución de las subcategorías léxicas en todo el texto preparado en los tres análisis.....	139
Tabla 38.- Distribución de las subcategorías léxicas en cada franja en los tres análisis en todo el texto preparado.....	139
Tabla 39.- Distribución de cada subcategoría léxica en todo el texto preparado en los tres análisis.....	140
Tabla 40.- Importancia de las subcategorías léxicas.....	141
Tabla 41.- Importancia de las subcategorías léxicas en la franja alta.....	141
Tabla 42.- Franjas en frecuencia acumulada (N) del texto lematizado de FESTESE001V.....	144
Tabla 43.- Otras características del texto lematizado/transformado.....	146
Tabla 44.- Características por categoría gramatical del texto lematizado o transformado.....	148

Tabla 45.- Distribución del número de lemas y frecuencias en todo el texto lematizado/transformado en los tres análisis	150
Tabla 46.- Distribución de las subcategorías léxicas en los tres análisis en el texto lematizado/transformado	152
Tabla 47.- Distribución de las subcategorías léxicas en todo el texto lematizado/transformado en los tres análisis	154
Tabla 48.- Distribución de las subcategorías léxicas en cada franja en los tres análisis en todo el texto lematizado/transformado	154
Tabla 49.- Distribución de cada subcategoría léxica en todo el texto lematizado/transformado en los tres análisis	155
Tabla 50.- Hápax en las subcategorías léxicas en el texto lematizado/transformado	156
Tabla 51.- Importancia de las subcategorías léxicas del texto transformado .	156
Tabla 52.- Número de colocaciones en los pares de lemas relacionados	158
Tabla 53.- Listado de propuestas de uniones de lemas de FESTESE001V... 159	
Tabla 54.- Listado de propuestas de uniones de lemas de FESTESE001V-F161	
Tabla 55.- Listado de propuestas de uniones de lemas de FESTESE001V-S 162	
Tabla 56.- Porcentajes de uniones de FESTESE001V y FESTESE001V-S 164	
Tabla 57.- Porcentajes de uniones de FESTESE001V-F	165
Tabla 58.- Porcentaje del número de lemas en las colocaciones	166
Tabla 59.- Formación de colocaciones de más de 2 lemas en la Primera vez 167	
Tabla 60.- Porcentajes de las categorías gramaticales en las colocaciones de la Primera vez	168
Tabla 61.- Porcentaje de colocaciones entre las categorías gramaticales en la Primera vez	168
Tabla 62.- Colocaciones totales en los tres análisis	170
Tabla 63.- Porcentaje de colocaciones por número de lemas	170
Tabla 64.- Características de los textos transformados definitivos (medidos con la aplicación Análisis de Frecuencias del PAFE)	171
Tabla 65.- Franjas en frecuencia acumulada (N) de los textos transformados definitivos.....	173
Tabla 66.- Otras características de los textos transformados definitivos	174
Tabla 67.- Características por categoría gramatical de los textos transformados definitivos.....	176
Tabla 68.- Distribución del número de ULL y frecuencias en todo el texto transformado definitivo en los tres análisis	178
Tabla 69.- Distribución de las subcategorías léxicas en los tres análisis en el texto transformado definitivo.....	180

LISTA DE TABLAS

Tabla 70.- Distribución de las subcategorías léxicas en todo el texto transformado definitivo en los tres análisis	182
Tabla 71.- Distribución de las subcategorías léxicas en cada franja en los tres análisis en todo el texto transformado definitivo.....	183
Tabla 72.- Distribución de cada subcategoría léxica en todo el texto transformado definitivo en los tres análisis	183
Tabla 73.- Hápax en las subcategorías léxicas en el texto transformado definitivo.....	184
Tabla 74.- Importancia de las subcategorías léxicas del texto transformado definitivo.....	185
Tabla 75.- Franja de vocabulario más frecuente analizando las frecuencias de los textos transformados definitivos.....	187
Tabla 76.- Características de la franja determinada por $m=-1$ en las distribuciones de frecuencias absolutas.....	189
Tabla 77.- Vocabulario más frecuente y específico del texto transformado definitivo.....	190
Tabla 78.- Características del vocabulario más frecuente y específico.....	191
Tabla 79.- Categorías gramaticales del vocabulario más frecuente	192
Tabla 80.- Categorías gramaticales del vocabulario específico	192
Tabla 81.- Sistemas a analizar.....	193
Tabla 82.- Conservación de relaciones en cada ULL para distintos sistemas definidos en FESTESE001VD	198
Tabla 83.- Conservación de relaciones en cada ULL para distintos sistemas definidos en FESTESE001VD-F.....	199
Tabla 84.- Conservación de relaciones en cada ULL para distintos sistemas definidos en FESTESE001VD-S.....	204
Tabla 85.- Comparación de la conservación de relaciones	209
Tabla 86.- Comparación de categorías gramaticales en la conservación de relaciones	211
Tabla 87.- Rangos de frecuencias y relaciones en las gráficas.....	213
Tabla 88.- Características de las ULL representadas en las gráficas.....	226
Tabla 89.- Especificidad y relaciones de las ULL representadas en las gráficas	227
Tabla 90.- Comparación de categorías gramaticales representadas en las gráficas.....	227
Tabla 91.- Categorías gramaticales respecto a su especificidad y conservación de relaciones.....	228
Tabla 92.- ULL relacionadas con estrella.....	230
Tabla 93.- ULL relacionadas con estrella con el fragmento “punto y aparte”	245

LISTA DE TABLAS

Tabla 94.- Errores producidos por el investigador en FESTESE001V	251
Tabla 95.- Frecuencias del texto preparado FESTESE001V	371
Tabla 96.- Frecuencias del texto lematizado FESTESE001V	386
Tabla 97.- Listado de lemas con un mínimo de dos relaciones con el siguiente	399
Tabla 98.- Frecuencias del texto preparado FESTESE001V-S.....	441
Tabla 99.- Frecuencias del texto transformado FESTESE001V-F.....	456
Tabla 100.- Frecuencias del texto transformado FESTESE001V-S	468
Tabla 101.- Frecuencias del texto transformado FESTESE001VD.....	480

LISTA DE FIGURAS

<i>Figura 1.-</i> Definición de concepto. Los datos a partir de los cuales se forman los conceptos, ya sea directamente o mediante algún proceso, se muestran dentro de círculos en la parte superior de la figura. Los procesos que llevan de los datos al concepto se muestran en rectángulos con esquinas redondeadas. Las formas de expresar los conceptos aparecen debajo de los mismos en rectángulos.....	18
<i>Figura 2.-</i> Propiedades de las palabras.....	21
<i>Figura 3.-</i> Contextos de aprendizaje de las palabras.....	23
<i>Figura 4.-</i> Parte de un grafo obtenido mediante encuestas a 23 sujetos respecto al tema <i>Alimentos</i> . En este grafo los vocablos (nodos) están dentro de elipses que se hallan conectadas mediante aristas que tienen un peso. El peso de las aristas indica el número de veces que ha aparecido dicha relación. Se pueden apreciar 6 agrupaciones (categorías) así como la vecindad entre vocablos.....	37
<i>Figura 5.-</i> Red conceptual sobre el Sol. Se pueden unir las aristas procedentes de varios nodos para generar oraciones.....	37
<i>Figura 6.-</i> Mapa conceptual de mapa conceptual. Se parte del concepto general y se van añadiendo de forma jerárquica los conceptos cada vez más particulares. .	38
<i>Figura 7.-</i> Mapa semántico sobre el Sistema Solar.....	39
<i>Figura 8.-</i> Mapa mental del Sol. El mapa mental se parece a un árbol al que le salen ramas que a su vez se pueden ramificar.	39
<i>Figura 9.-</i> De los conceptos a las UL y de éstas a los conceptos.....	47
<i>Figura 10.-</i> Último párrafo del texto-ejemplo con salto de línea. Se han resaltado los espacios y saltos de línea con la ayuda de <i>Word 2007</i>	52
<i>Figura 11.-</i> Clasificación en texto preparado impreso de las palabras del primer párrafo del ejemplo de la Tabla 6.....	63
<i>Figura 12.-</i> Codificación alfanumérica de un texto.	68
<i>Figura 13.-</i> Distribución para un entorno de valor máximo 1. Las barras representan el número de relaciones y los puntos, el número de ULL, para dicha relación acumulada del eje de abscisas. Ra_i = relación absoluta acumulada	96
<i>Figura 14.-</i> Distribución para un entorno de valor máximo 3. Las barras representan el número de relaciones y los puntos, el número de ULL, para dicha relación acumulada del eje de abscisas. Ra_i = relación absoluta acumulada	97
<i>Figura 15.-</i> Gráfica del estudio de caso.....	102

LISTA DE FIGURAS

<i>Figura 16.-</i> Gráfica del caso de divulgación (PAFE).	104
<i>Figura 17.-</i> Gráfica del caso de divulgación (Excel2003).	104
<i>Figura 18.-</i> Gráfica del sistema $f_{\text{mínima}}=0.81$	106
<i>Figura 19.-</i> Gráfica del sistema $f_{\text{mínima}}=0.68$	106
<i>Figura 20.-</i> Gráfica del sistema de todas las ULL (completo).	107
<i>Figura 21.-</i> Primera parte del esquema de la metodología.	111
<i>Figura 22.-</i> Segunda parte del esquema de la metodología.	112
<i>Figura 23.-</i> Tercera y última parte del esquema de la metodología.	113
<i>Figura 24.-</i> Frecuencia absoluta de FESTEELSE001V.	129
<i>Figura 25.-</i> Frecuencia acumulada de FESTEELSE001V.	130
<i>Figura 26.-</i> Número de vocablos por categoría gramatical FESTEELSE001V.	131
<i>Figura 27.-</i> Frecuencias por categoría gramatical FESTEELSE001V.	131
<i>Figura 28.-</i> Número de vocablos por categoría gramatical en todo el texto de FESTEELSE001V.	133
<i>Figura 29.-</i> Frecuencias por categoría gramatical en todo el texto de FESTEELSE001V.	133
<i>Figura 30.-</i> Número de vocablos por categoría gramatical en cada franja de FESTEELSE001V.	133
<i>Figura 31.-</i> Frecuencias por categoría gramatical en cada franja de FESTEELSE001V.	133
<i>Figura 32.-</i> Número de vocablos de cada categoría gramatical en todo el texto de FESTEELSE001V.	134
<i>Figura 33.-</i> Frecuencias de cada categoría gramatical en todo el texto de FESTEELSE001V.	134
<i>Figura 34.-</i> Número de vocablos por subcategoría léxica de FESTEELSE001V.	136
<i>Figura 35.-</i> Frecuencias por subcategoría léxica de FESTEELSE001V.	136
<i>Figura 36.-</i> Número de vocablos por subcategoría léxica en todo el texto de FESTEELSE001V.	138
<i>Figura 37.-</i> Frecuencias por subcategoría léxica en todo el texto de FESTEELSE001V.	138
<i>Figura 38.-</i> Número de vocablos por subcategoría léxica en cada franja de FESTEELSE001V.	138
<i>Figura 39.-</i> Frecuencias por subcategoría léxica en cada franja de FESTEELSE001V.	138
<i>Figura 40.-</i> Número de vocablos de cada subcategoría léxica en todo el texto de FESTEELSE001V.	138
<i>Figura 41.-</i> Frecuencias de cada subcategoría léxica en todo el texto de FESTEELSE001V.	138

<i>Figura 42.-</i> Frecuencia absoluta del texto lematizado de FESTEELSE001V... 143	143
<i>Figura 43.-</i> Frecuencia acumulada del texto lematizado de FESTEELSE001V.	143
<i>Figura 44.-</i> Número de lemas por categoría gramatical en todo el texto de FESTEELSE001V.....	149
<i>Figura 45.-</i> Frecuencias por categoría gramatical en todo el texto de FESTEELSE001V.....	149
<i>Figura 46.-</i> Número de lemas por categoría gramatical en cada franja de FESTEELSE001V.....	149
<i>Figura 47.-</i> Frecuencias por categoría gramatical en cada franja de FESTEELSE001V.....	149
<i>Figura 48.-</i> Número de lemas de cada categoría gramatical en todo el texto de FESTEELSE001V.....	149
<i>Figura 49.-</i> Frecuencias de cada categoría gramatical en todo el texto de FESTEELSE001V.....	149
<i>Figura 50.-</i> Número de lemas por subcategoría léxica de FESTEELSE001V..	151
<i>Figura 51.-</i> Frecuencias por subcategoría léxica de FESTEELSE001V.....	151
<i>Figura 52.-</i> Número de lemas por subcategoría léxica en todo el texto de FESTEELSE001V.....	152
<i>Figura 53.-</i> Frecuencias por subcategoría léxica en todo el texto de FESTEELSE001V.....	152
<i>Figura 54.-</i> Número de lemas por subcategoría léxica en cada franja de FESTEELSE001V.....	153
<i>Figura 55.-</i> Frecuencias por subcategoría léxica en cada franja de FESTEELSE001V.....	153
<i>Figura 56.-</i> Número de lemas de cada subcategoría léxica en todo el texto de FESTEELSE001V.....	153
<i>Figura 57.-</i> Frecuencias de cada subcategoría léxica en todo el texto de FESTEELSE001V.....	153
<i>Figura 58.-</i> Frecuencia absoluta de FESTEELSE001VD.	172
<i>Figura 59.-</i> Frecuencia acumulada de FESTEELSE001VD.....	173
<i>Figura 60.-</i> Número de ULL por categoría gramatical en todo el texto de FESTEELSE001VD.....	177
<i>Figura 61.-</i> Frecuencias por categoría gramatical en todo el texto de FESTEELSE001VD.....	177
<i>Figura 62.-</i> Número de ULL por categoría gramatical en cada franja de FESTEELSE001VD.....	177
<i>Figura 63.-</i> Frecuencias por categoría gramatical en cada franja de FESTEELSE001VD.....	177
<i>Figura 64.-</i> Número de ULL de cada categoría gramatical en todo el texto de FESTEELSE001VD.....	177

LISTA DE FIGURAS

<i>Figura 65.</i> - Frecuencias de cada categoría gramatical en todo el texto de FESTESE001VD.....	177
<i>Figura 66.</i> - Número de ULL por subcategoría léxica de FESTESE001VD.	179
<i>Figura 67.</i> - Frecuencias por subcategoría léxica de FESTESE001VD.....	179
<i>Figura 68.</i> - Número de ULL por subcategoría léxica en todo el texto de FESTESE001VD.....	180
<i>Figura 69.</i> - Frecuencias por subcategoría léxica en todo el texto de FESTESE001VD.....	180
<i>Figura 70.</i> - Número de ULL por subcategoría léxica en cada franja de FESTESE001VD.....	180
<i>Figura 71.</i> - Frecuencias por subcategoría léxica en cada franja de FESTESE001VD.....	180
<i>Figura 72.</i> - Número de ULL de cada subcategoría léxica en todo el texto de FESTESE001VD.....	181
<i>Figura 73.</i> - Frecuencias de cada subcategoría léxica en todo el texto de FESTESE001VD.....	181
<i>Figura 74.</i> - Ajuste de la distribución de frecuencias absolutas del texto lematizado FESTESE001VD.....	186
<i>Figura 75.</i> - Ajuste de la distribución de frecuencias acumuladas del texto lematizado FESTESE001VD.....	187
<i>Figura 76.</i> - Determinación del vocabulario más frecuente mediante las frecuencias absoluta y acumulada.....	190
<i>Figura 77.</i> - Distribución del número de ULL (símbolos) y relaciones (barras) respecto al valor de relación en FESTESE001VD. r_i =valor de relaciones absoluto, N^o = número.....	196
<i>Figura 78.</i> - Distribución del número de ULL (símbolos) y relaciones (barras) respecto al valor de relación acumulado en FESTESE001VD. Ra_i =valor de relaciones acumulado, N^o = número.....	197
<i>Figura 79.</i> - Parte izquierda de la gráfica con relaciones relativas al listado completo de FESTESE001VD.....	214
<i>Figura 80.</i> - Parte derecha de la gráfica con relaciones relativas al listado completo de FESTESE001VD.....	215
<i>Figura 81.</i> - Parte izquierda de la gráfica con relaciones relativas a la ULL origen de FESTESE001VD.....	216
<i>Figura 82.</i> - Parte derecha de la gráfica con relaciones relativas a la ULL origen de FESTESE001VD.....	217
<i>Figura 83.</i> - Parte izquierda de la gráfica con relaciones relativas al listado completo de FESTESE001VD -F.....	218
<i>Figura 84.</i> - Parte derecha de la gráfica con relaciones relativas al listado completo de FESTESE001VD -F.....	219

<i>Figura 85.-</i> Parte izquierda de la gráfica con relaciones relativas a la ULL origen de FESTESE001VD-F.	220
<i>Figura 86.-</i> Parte derecha de la gráfica con relaciones relativas a la ULL origen de FESTESE001VD-F.	221
<i>Figura 87.-</i> Parte izquierda de la gráfica con relaciones relativas al listado completo de FESTESE001VD -S.	222
<i>Figura 88.-</i> Parte derecha de la gráfica con relaciones relativas al listado completo de FESTESE001VD -S.	223
<i>Figura 89.-</i> Parte izquierda de la gráfica con relaciones relativas a la ULL origen de FESTESE001VD-S.	224
<i>Figura 90.-</i> Parte derecha de la gráfica con relaciones relativas a la ULL origen de FESTESE001VD-S.	225
<i>Figura 91.-</i> Parte izquierda de la gráfica con relaciones relativas al listado completo, fragmento="punto y aparte", de FESTESE001VD.	232
<i>Figura 92.-</i> Parte derecha de la gráfica con relaciones relativas al listado completo, fragmento="punto y aparte", de FESTESE001VD.	233
<i>Figura 93.-</i> Parte izquierda de la gráfica con relaciones relativas a la ULL origen, fragmento="punto y aparte", de FESTESE001VD.	234
<i>Figura 94.-</i> Parte derecha de la gráfica con relaciones relativas a la ULL origen, fragmento="punto y aparte", de FESTESE001VD.	235
<i>Figura 95.-</i> Parte izquierda de la gráfica con relaciones relativas al listado completo, fragmento="punto y aparte", de FESTESE001VD-F.	236
<i>Figura 96.-</i> Parte derecha de la gráfica con relaciones relativas al listado completo, fragmento="punto y aparte", de FESTESE001VD-F.	237
<i>Figura 97.-</i> Parte izquierda de la gráfica con relaciones relativas a la ULL origen, fragmento="punto y aparte", de FESTESE001VD-F.	238
<i>Figura 98.-</i> Parte derecha de la gráfica con relaciones relativas a la ULL origen, fragmento="punto y aparte", de FESTESE001VD-F.	239
<i>Figura 99.-</i> Parte izquierda de la gráfica con relaciones relativas al listado completo, fragmento="punto y aparte", de FESTESE001VD-S.	240
<i>Figura 100.-</i> Parte derecha de la gráfica con relaciones relativas al listado completo, fragmento="punto y aparte", de FESTESE001VD-S.	241
<i>Figura 101.-</i> Parte izquierda de la gráfica con relaciones relativas a la ULL origen, fragmento="punto y aparte", de FESTESE001VD-S.	242
<i>Figura 102.-</i> Parte derecha de la gráfica con relaciones relativas a la ULL origen, fragmento="punto y aparte", de FESTESE001VD-S.	243
<i>Figura 103.-</i> Primera parte del esquema de la metodología.	269
<i>Figura 104.-</i> Segunda parte del esquema de la metodología.	270
<i>Figura 105.-</i> Estructura de las carpetas de instalación.	325
<i>Figura 106.-</i> Iconos de las aplicaciones PAFE-AF y PAFE-EE.	326
<i>Figura 107.-</i> Estructura final de la carpeta PAFE-BD usuarios.	327

LISTA DE FIGURAS

<i>Figura 108.-</i> Usuarios nuevos del PAFE.....	327
<i>Figura 109.-</i> Submenú de Bases de Datos.....	328
<i>Figura 110.-</i> Ventana de relleno de datos del investigador.....	328
<i>Figura 111.-</i> Listado de usuarios.	329
<i>Figura 112.-</i> Primera página del texto escolar “Medida de la distancia y la edad de un cúmulo globular de estrellas”.....	330
<i>Figura 113.-</i> Segunda página del texto escolar “Medida de la distancia y la edad de un cúmulo globular de estrellas”.....	331
<i>Figura 114.-</i> Tercera página del texto escolar “Medida de la distancia y la edad de un cúmulo globular de estrellas”.....	332
<i>Figura 115.-</i> Cuarta página del texto escolar “Medida de la distancia y la edad de un cúmulo globular de estrellas”.....	333
<i>Figura 116.-</i> Quinta página del texto escolar “Medida de la distancia y la edad de un cúmulo globular de estrellas”.....	334
<i>Figura 117.-</i> Sexta página del texto escolar “Medida de la distancia y la edad de un cúmulo globular de estrellas”.....	335
<i>Figura 118.-</i> Entrada en el Análisis de Frecuencias del PAFE: introducción de la clave de usuario.....	345
<i>Figura 119.-</i> Análisis de Frecuencias: selección del texto a procesar.	345
<i>Figura 120.-</i> Análisis de Frecuencias: plantilla de identificación del texto a procesar.	346
<i>Figura 121.-</i> Análisis de Frecuencias: menú para procesar el texto.....	346
<i>Figura 122.-</i> Análisis de Frecuencias: elección de considerar los verbos dentro del análisis.	347
<i>Figura 123.-</i> Análisis de Frecuencias. selección de la base <i>Eliminar</i> para la palabra que se indica. Esta es también la ventana que se abre para la clasificación previa en cualquiera de las bases de datos. Sólo en el caso de elegir <i>Cambiar a o Verbos</i> cambia la ventana abierta.	347
<i>Figura 124.-</i> Análisis de Frecuencias. ventana de lematización de formas no verbales (en <i>Cambiar a</i>) y de formas verbales (en <i>Verbos</i>).	348
<i>Figura 125.-</i> Análisis de Frecuencias: menú de las bases de datos.	348
<i>Figura 126.-</i> Análisis de Frecuencias: modificación de las bases de datos. ...	348
<i>Figura 127.-</i> Análisis de Frecuencias: texto lematizado o transformado, menú de frecuencias y salto al Estudio de Entorno.....	349
<i>Figura 128.-</i> Análisis de Frecuencias: listado de la frecuencia absoluta y la frecuencia relativa.	349
<i>Figura 129.-</i> Estudio de Entorno: acceso a la aplicación desde su icono.	350
<i>Figura 130.-</i> Estudio de Entorno: selección del texto transformado a estudiar.	350

Figura 131.- Estudio de Entorno: determinación de las características del estudio..... 351

Figura 132.- Estudio de Entorno: selección del valor de relación mínimo a ver 351

Figura 133.- Estudio de Entorno: listado de ULL del sistema definido. 352

Figura 134.- Estudio de Entorno: *Lista de relaciones*, archivo *.grp con el listado de ULL del sistema definido. 352

Figura 135.- Estudio de entorno: crear lista de palabras..... 353

Figura 136.- Estudio de entorno: crear lista de palabras usando la frecuencia mínima..... 353

Figura 137.- Estudio de entorno: lista de palabras en el cuadro situado a la derecha..... 354

Figura 138.- Estudio de entorno: características de la gráfica. 354

Figura 139.- Estudio de entorno: selección de intervalos de frecuencias y relaciones. 355

Figura 140.- Estudio de entorno: datos de las ULL del sistema estudiado. 355

Figura 141.- Estudio de entorno: relaciones en el sistema estudiado..... 356

Figura 142.- Estudio de entorno: cambio de representación gráfica. 356

1.- INTRODUCCIÓN



1.- INTRODUCCIÓN

1. 1.- Marco para la Realización de la Tesis

La enseñanza de las ciencias en los colegios se puede resumir en aprendizaje de conceptos, resolución de problemas y aprendizaje de habilidades.

Son diversos los estudios realizados hasta el momento con el fin de resolver la enseñanza-aprendizaje de algún tema concreto de una ciencia en particular. Sin embargo son estudios dirigidos a resolver cuestiones que afectan a la práctica docente de una asignatura en un determinado curso o nivel educativo (Devetak, Vogrinc, & Glazar, 2010; Furió & Guisasola, 1998; Glenn, 1990). En algunos casos se abordan consideraciones más amplias, como las de un curso completo (Gil & Solbes, 1993) o incluso todo un ciclo educativo (García Barros, Martínez Losada, Mondelo Alonso, & Vega Marcote, 1997). Todos los estudios anteriores están condicionados por el enfoque dominante del Sistema Educativo del momento; pretenden dar respuestas inmediatas y globales, sin haber tenido la oportunidad ni contado con la metodología adecuada para manejar las numerosas y complejas variables que intervienen en el sistema investigado, difíciles o imposibles de controlar. Si bien sus resultados son de suma importancia de cara a la innovación educativa, no lo son tanto desde la perspectiva del conocimiento fundamental de los procesos de enseñanza-aprendizaje de los contenidos de las ciencias.

INTRODUCCIÓN

Este tipo de trabajos es abundante, puesto que responden a las necesidades de los responsables de Sistemas Educativos en constante cambio, lo cual mantiene una permanente demanda de estudios relacionados con la puesta en práctica de las frecuentes innovaciones didácticas. Además, los propios investigadores del Área de Didáctica de las Ciencias, en su mayoría, compaginan su labor investigadora con la práctica docente, lo cual tiende a supeditar sus análisis al enfoque de la docencia que el ejercicio de la enseñanza les ha imbuido en su concreto entorno geográfico y cultural. El carácter mayoritario de trabajos de ésta índole puede comprobarse tanto en los índices de reconocidas revistas especializadas como el *Journal of Chemical Education* o *The Physics Teacher*, muchos de cuyos artículos exponen estudios como los mencionados, interesantes para docentes en ejercicio, y que abordan contenidos curriculares puntuales en la disciplina a la que se dedican o cuestiones sobre trabajos de campo y de laboratorio (Barberá & Valdés, 1996; Caamaño, 1992; Kirk & Layman, 1996; Küçüközer, Korkusuz, Küçüközer, & Yürümezoğlu, 2009; Lillo Beviá, 1994).

Todos estos planteamientos son necesarios y útiles, pero no cubren aspectos básicos para el desarrollo de la Didáctica de las Ciencias, que requiere la consecución de otros objetivos fundamentales e imprescindibles para el aprendizaje de las ciencias. La tesis se desarrolla dentro de la línea de investigación creada por el grupo de trabajo GICEC (Grupo de Investigación sobre Conceptos en la Enseñanza de las Ciencias) del Departamento de Didácticas Específicas, Área de Didáctica de las Ciencias Experimentales de la Universidad de La Laguna (Ceballos, Galotti, & Varela, 1998; Ceballos, Varela, & Galotti, 1999).

El GICEC centra su trabajo en la investigación de la enseñanza de conceptos relacionados con las ciencias, el diseño y desarrollo de recursos didácticos

adecuados a dicho fin y el lenguaje de las ciencias. La línea de investigación creada por el GICEC tiene por objeto estudiar los conceptos propios de cada ciencia, para contribuir a facilitar su aprendizaje. Así mismo aporta datos sobre la elaboración de conceptos científicos y las dificultades de aprendizaje, proporciona criterios de selección y secuenciación de contenidos curriculares, y asume el diseño, producción y evaluación de recursos que optimicen los correspondientes sistemas de enseñanza-aprendizaje. Por otra parte, trata de aportar datos acerca de la estructuración del conocimiento humano, mediante el análisis del conocimiento científico en expertos.

Se utilizan las conclusiones y técnicas de trabajo acreditadas en otras ramas del saber (Lingüística de Corpus, Psicología, Tecnología de la Educación, etc.) a los condicionantes específicos de la tarea, valorando especialmente su sencillez y el bajo coste del instrumental necesario. El trabajo investigador puede describirse mediante las siguientes fases: la detección de los conceptos más frecuentes así como sus redes semánticas, el análisis de dichos conceptos y la enseñanza de los mismos, que incluye el diseño y desarrollo de los recursos didácticos más adecuados para el aprendizaje de ellos. Desde el punto de vista metodológico cada una de ellas recibe un tratamiento diferenciado que se reseña a continuación.

Para la **Detección de conceptos** de un tema de Ciencias, se establece una metodología de investigación, en la que se utiliza un Programa de Análisis de Frecuencias y Entornos (PAFE), diseñado y desarrollado en el GICEC, que facilita la toma de datos, su tratamiento automatizado y la presentación de resultados. Con relación a un contenido temático predeterminado, se prepara una muestra de textos seleccionados de los géneros divulgativo, académico e investigador, y se realiza su análisis mediante el PAFE. Como resultado se obtienen listados de frecuencia,

INTRODUCCIÓN

vocabularios específicos y relaciones cuantificadas entre vocablos, que permiten determinar redes semánticas. Mediante otras medidas, es posible determinar más parámetros, como densitometrías de información, criterios para la selección y secuenciación de contenidos curriculares, etc.

El estudio sobre el **Análisis de conceptos** se inicia con la elección de algunos de las redes semánticas identificadas en la primera fase, para lo que se tienen en cuenta determinados criterios (frecuencia correspondiente a los conceptos que lo conforman, el valor de las relaciones de éstos, etc.). Con los datos de la primera fase es posible estudiar la conformación del concepto en función de sus relaciones, cómo se genera y elabora a lo largo del tiempo en los ámbitos en los que se utiliza, las ideas previas que le afectan, los requisitos curriculares y las dificultades conocidas para su aprendizaje. En relación con ello, es esencial aprovechar la abundante información bibliográfica generada en los últimos años sobre estas cuestiones, así como la referida al desarrollo psicoevolutivo que precisa el aprendizaje de cada concepto.

Con los resultados de la fase anterior se puede matizar y mejorar la propuesta inicial de selección y secuenciación de contenidos, así como fundamentar suficientemente los aspectos docentes para abordar la **Enseñanza de conceptos**. Ésta se inicia con un estudio de las posibles estrategias de enseñanza-aprendizaje de cada concepto, seguida de la determinación de los recursos más apropiados para el aprendizaje de los mismos, y de su diseño y realización en caso necesario. Los **recursos** se someten a análisis y evaluación, con las técnicas pertinentes a cada caso, y se establecen sus características y límites de uso. Se finaliza con recomendaciones didácticas sobre los conceptos estudiados, sin que esto presuponga ningún tipo de propuesta metodológica de enseñanza de las ciencias.

En el GICEC se han llevado a cabo estudios con esta línea de investigación de distintos temas: ondas y sonido (Ceballos, Galotti, González, & Varela, 1997; Ceballos, Galotti, Santana, & Varela, 2000a; Ceballos, Galotti, Santana, & Varela, 2000b; Ceballos, Galotti, Santana, & Varela, 2000c; Ceballos, Galotti, Santana, & Varela, 2001; Ceballos, Galotti, Varela, & Leal, 1998; Ceballos, Galotti, Varela, & Talavera, 1998; Ceballos, Galotti, Varela Calvo, & Talavera Sosa, 1999) estrellas (Ceballos & Hansen-Ruiz, 2002; Ceballos & Hansen-Ruiz, 2005; Ceballos, Hansen, & Stengler, 1999; Ceballos, Hansen-Ruiz, Galotti, & Stengler, 2005; Hansen-Ruiz, Ceballos, & Stengler, 2005a; Hansen-Ruiz, Ceballos, & Stengler, 2005b), Sistema Solar (Domínguez & Galotti, 2003; Domínguez & Varela, 2001; Domínguez & Varela, 2005a; Domínguez & Varela, 2005b; Domínguez Herrera, 2006; Domínguez Herrera, 2007; Domínguez Herrera, 2009; Domínguez Herrera, Galotti, & Varela Calvo, 2002), física cuántica (Galotti & Ceballos, 2001; Galotti, Ceballos, & Matus, 2003), evolución biológica, medio ambiente (Fajardo Rodríguez & Varela Calvo, 1999; Varela Calvo & Fajardo Rodríguez, 2001) y ecología (Ceballos, Santos, Varela, & Galotti, 2000; Ceballos, Santos, Varela, & Galotti, 2001).

En concreto esta tesis, se centra en la primera fase de la línea de investigación del GICEC, concretamente en mejorar la metodología y aplicarla a un estudio de caso: conceptos relacionados con *estrella* en un texto de secundaria.

1. 2.- Elección del Tema

La investigación en didáctica de la ciencia está directamente relacionada con cómo mejorar la educación y la divulgación de la ciencia. La mejora en ambos aspectos permitirá a futuros ciudadanos, incluidos los dirigentes, tomar decisiones más acertadas sobre distintos aspectos económicos, ecológicos, sanitarios, energéticos, etc. La educación y la divulgación pueden considerarse como la misma cuestión tratada desde dos puntos de vista distintos: la enseñanza formal, a la que todo ciudadano ha tenido acceso, y la enseñanza informal, a la que todo ciudadano puede tener acceso.

La Astronomía se considera como una asignatura interdisciplinar en educación, pues contiene conceptos de distintas ramas de la ciencia (Física, Química, Geología e incluso Biología) y las Matemáticas, y hasta permite abordar conceptos históricos y filosóficos. La Astronomía es también usada recurrentemente en la divulgación de la ciencia, ocupando un 12% de las páginas de las revistas de divulgación (Ceballos & Hansen-Ruiz, 2005).

El tema elegido para esta tesis, de “estrellas” reúne principalmente conceptos tanto de la Física como de la Química. Las estrellas y el Sol se encuentran en el currículo legislado desde la educación primaria hasta el bachillerato (Hansen-Ruiz, Ceballos, & Stengler, 2005a). Los contenidos sobre estrellas constituyen más del 15% de los contenidos sobre astronomía en las revistas de divulgación (Ceballos & Hansen-Ruiz, 2005).

1. 3.- Objetivos de la Tesis

El principal objetivo de esta tesis es mejorar la metodología utilizada en el GICEC, que hace uso del programa PAFE, para el análisis de vocabulario y redes semánticas. Se analiza la metodología aplicándola a un caso particular: un texto de secundaria sobre *estrella*. Se pretende determinar las limitaciones de los resultados cuantificando los posibles errores de cada una de las fases del proceso. Para ello se consideran los siguientes subobjetivos:

- Aplicación y mejora de técnicas provenientes de la lingüística de corpus a la investigación en Didáctica:
 - Uso de las frecuencias para determinar vocabularios frecuentes y específicos.
 - Determinación de colocaciones.
 - Enriquecimiento de la información proporcionada por las redes semánticas.
- Determinación de la dependencia de la metodología de la intervención de un experto en el tema abordado.
- Acotación de los errores posibles en la metodología de investigación.
- Mejora de la interpretación de los resultados para su uso en la Enseñanza de las Ciencias.

1. 4.- Estructura de la Tesis

La tesis está dividida en distintos capítulos:

- **Introducción.** Se presenta el marco de investigación en el que se realiza la tesis, las razones por las que se eligió un tema de astronomía, los objetivos y la estructura de la tesis.
- **Fundamentos.** Se comienza estableciendo la relación entre conceptos y lenguaje. Se define la lingüística de corpus y las distintas utilidades para investigar textos escritos. Además se proporciona información de los resultados de investigaciones lingüísticas sobre el lenguaje de la ciencia. Por último se comentan los distintos tipos de análisis de texto asistidos por ordenador.
- **Metodología.** Ésta recoge aspectos de la lingüística computacional que se aplican al lenguaje de las ciencias, con el objetivo de investigar los conceptos etiquetados por la palabra escrita. Se exponen los pasos a seguir para analizar un texto o corpus, así como las aplicaciones informáticas a usar. Además se detalla cómo funciona la aplicación informática PAFE, constituyendo, junto con varios de los anexos, un manual de instalación y uso de todas las funciones del mismo que hasta ahora no existía.
- **Un estudio de caso: un texto escolar de astronomía.** Aquí se aplica la metodología del capítulo anterior, con tres variaciones, a un texto de más de 1000 palabras. Las variaciones metodológicas son: un análisis hecho como lo harían los lingüistas, sin despreciar ni una sola palabra (análisis ortodoxo); otro más ajustado a la investigación sobre conceptos que se pretende realizar, despreciando la parte del lenguaje que carece de significado (análisis sin palabras funcionales); y el último, donde además de la eliminación de palabras

funcionales, se agrupan las distintas grafías de un mismo concepto en una sola (unificación de grafías). Se presentan los resultados y se comparan en cada uno de los pasos, discutiéndose las consecuencias para la metodología y para los resultados que interesan para la Enseñanza de las Ciencias.

- **Análisis de errores.** Se discuten las fuentes de errores posibles, cuantificando el error generado. Además se discute si la metodología depende de que el investigador sea también experto en el tema estudiado.
- **Conclusiones.** Se propone una metodología mejorada para poder analizar corpus para la obtención de datos para la Enseñanza de las Ciencias.
- **Cuestiones abiertas.** Se comentan los nuevos interrogantes generados a lo largo de la investigación pero que no han formado parte de la misma.
- **Glosario.** Con terminología lingüística y relacionada con la metodología.
- **Bibliografía.**
- **Anexos.** Con imágenes de la aplicación PAFE así como resultados que ilustran distintos pasos de la metodología.

2.- FUNDAMENTOS



2.- FUNDAMENTOS

2. 1.- Los Conceptos

Según el diccionario de la Real Academia de la Lengua Española (2001) un concepto es “una idea que concibe o forma el entendimiento”, también se entiende como un “pensamiento expresado con palabras”. Además el diccionario ofrece una definición para formar concepto como “determinar algo en la mente después de examinadas las circunstancias”.

“Los conceptos y las categorías son los bloques de construcción del conocimiento. Los conceptos nos permiten interpretar nuestras experiencias, conectarlas con conocimiento previo, razonar y hacer predicciones.” (Vosniadou, Pagondiotis, & Deliyi, 2005, p. 115).

A pesar de las numerosas investigaciones en torno a las concepciones alternativas y cómo se cambia de concepto en la enseñanza se encuentran pocas definiciones de “concepto” en la bibliografía:

- Concepto como relación entre los datos dentro de una red de significados que explique por qué se producen y qué consecuencias tienen (Pozo Muncio & Gómez Crespo, 1998). Por ejemplo conocer el hecho de que los días y las noches tienen distinta duración a lo largo del año, no implica conocer por qué se

FUNDAMENTOS

produce dicho fenómeno. A lo sumo permite predecir que a la noche le sigue el día, pero eso no significa que se haya interpretado o comprendido.

- Concepto como generalización a partir de datos relacionados, que posibilita responder a, o pensar en, estímulos específicos o percepciones de una manera determinada (Lovell, 1977). Es decir, concepto y juicio son equivalentes y concepto se utiliza como un criterio. Cuando oímos la palabra “ave” o la vemos impresa, no pensamos en todas las especies posibles de aves existentes, sino que su significado se asocia a una clase de animales que tienen plumas y dos patas, de los que la gran mayoría puede volar.
- Concepto como base cognoscitiva para asignar un rótulo o un término a una categoría, la cual, a su vez, se refiere a algunos casos específicos (Ellis, 1972, p. 183). Un concepto se refiere a una clase de estímulos o eventos que comparten una o más características en común. Por ejemplo Júpiter, Marte, Mercurio y Tierra comparten la característica común de ser planeta.
- Concepto como significado de las palabras, las palabras nombran los conceptos (Klausmeier, 1990; Vygotsky, 1986).
- Concepto como consenso cultural relativo a los atributos esenciales (criterio e identificación), abstracción y representación simplificada y generalizada de la realidad (Ausubel, 1968).
- El concepto como la estructura fundamental para el pensamiento durante la vida del ser humano, y el lenguaje como parte importante en el desarrollo del concepto (Prater, 1993). El concepto considerado como estructura mental consiste en “la información organizada de una persona de un dato o una clase de datos que permite a la persona discriminar el dato o la clase de datos de otros

datos y también relacionarlo con otros datos o clases de datos”. (Klausmeier, 1990, p. 94).

- El concepto identificado de tres formas distintas de acuerdo a sus características funcionales (Ozdemir, 2004):
 - Elementos: entidades abstractas como fuerza, masa, velocidad,... que muchas veces forman parte del lenguaje cotidiano.
 - Principios: abstracción de sucesos naturales o artificiales. Estas abstracciones se adquieren al comprender las relaciones entre elementos diferentes mediante el razonamiento inductivo. Por ejemplo: “la fuerza es proporcional a la masa”.
 - Modelos: explicaciones que el individuo ofrece a los problemas encontrados. Son el resultado de construir una representación a partir de la transferencia apropiada de principios en una situación-problema.

- Los conceptos científicos como entes no exclusivamente matemáticos.

Las ideas en Física (o en otras disciplinas) no son puramente matemáticas. Las ideas se pueden representar matemáticamente así como en palabras, imágenes o gráficos. De hecho, el número de representaciones a las que un individuo tiene acceso es una medida de su dominio. Idealmente, los estudiantes deberían poder representar cualquier idea a través de representaciones múltiples, incluido las matemáticas. (Lasry, Finkelstein, & Mazur, 2009, p. 420).

- Los conceptos científicos entendidos en múltiples representaciones: textual, visual, matemática, figurativa y gestual (Hubber, Tytler, & Haslam, 2010; Klein

FUNDAMENTOS

& Kirkpatrick, 2010; Kohl & Finkelstein, 2005; Roth, 1995; Yore, Bisanz, & Hand, 2003).

Estas definiciones se pueden agrupar como muestra la *Figura 1*.

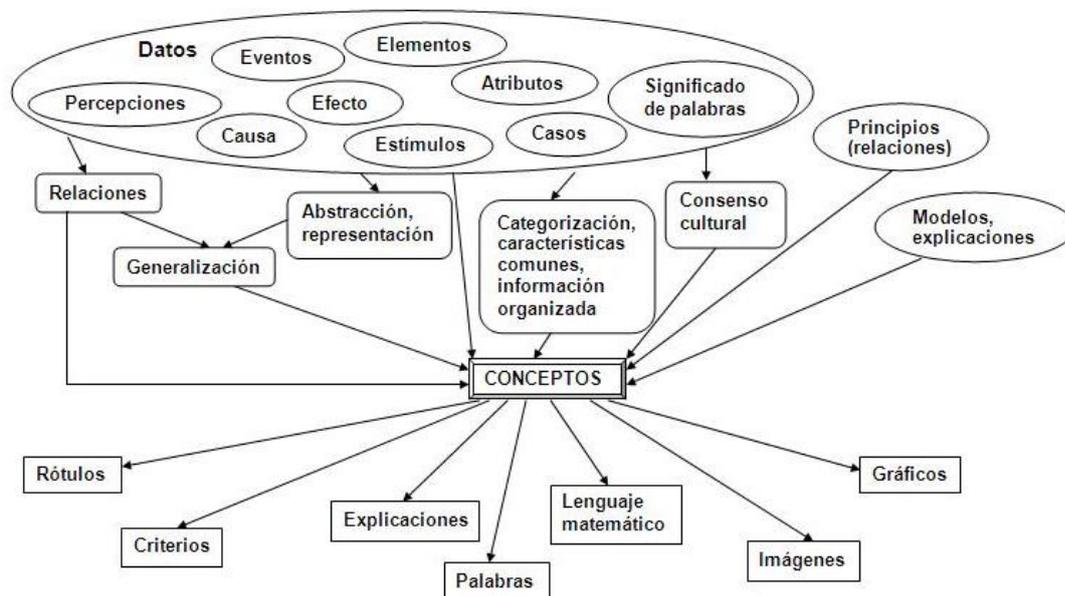


Figura 1.- Definición de concepto. Los datos a partir de los cuales se forman los conceptos, ya sea directamente o mediante algún proceso, se muestran dentro de círculos en la parte superior de la figura. Los procesos que llevan de los datos al concepto se muestran en rectángulos con esquinas redondeadas. Las formas de expresar los conceptos aparecen debajo de los mismos en rectángulos.

2. 2.- Lenguaje y Adquisición de los Conceptos

El lenguaje es una tecnología esencial y por tanto una parte integral de la ciencia y de la alfabetización científica, particularmente el lenguaje escrito. El lenguaje es un medio de hacer ciencia y de construir el entendimiento científico; el lenguaje es también un fin, un objetivo fundamental de la alfabetización científica, pues es usada para comunicar sobre las indagaciones, procedimientos y

entendimiento científico a otra gente para que puedan tomar decisiones y acciones informadas. (Yore et al., 2004).

Aunque existen pruebas de que los conceptos se pueden alcanzar sin llegar a tener palabras y de que se puede tener palabras sin llegar a comprender el concepto, es innegable la importancia del lenguaje en la formación del concepto.

Es sabido que los bebés ya tienen un repertorio de conceptos previo a la producción del lenguaje, siendo adeptos a la percepción de las entidades descritas por nombres (Arunachalam & Waxman, 2010; Mandler, 2000; Spelke, 2000).

Además tanto niños como adultos pueden haber alcanzado un concepto lo suficientemente válido y, sin embargo, ser incapaces de definir este concepto en términos verbales. Es el caso de las emociones, importantes para la salud mental y las relaciones interpersonales (Schachner, Shaver, & Mikulincer, 2005).

Morrison & Conway (2010) concluyen que la adquisición de las palabras está íntimamente relacionada con la formación del concepto. Mediante dos experimentos se requirieron las memorias autobiográficas de la infancia que indicaran palabras de objetos, lugares, actividades y emociones. La palabra asociada a una memoria se adquiría sistemáticamente antes, con una diferencia de hasta varios meses. Este patrón se observó para todo tipo de palabras y tiempo de adquisición (temprano o tardío) así como para personas de distintas edades (jóvenes o adultos más viejos). El trabajo sugiere que el intervalo de tiempo entre la adquisición de la palabra y la memoria asociada refleja la formación del concepto, abstraído de detalles presentes en memorias tempranas episódicas. Una vez que el conocimiento se ha formado la palabra que lo indica y el conocimiento conceptual al que corresponde se pueden usar para acceder a memorias episódicas específicas.

FUNDAMENTOS

El aprendizaje de palabras es un problema especial debido a que las mismas constituyen una colección compleja de propiedades (Koenig & Woodward, 2007):

- Una palabra es un símbolo. Los bebés de hasta 20 meses parecen aprender de igual manera símbolos no verbales que términos verbales para los nombres de los objetos (Campbell & Namy, 2003). A partir de los 20 meses empieza a emerger una preferencia por aprender los términos verbales (Woodward & Hoyne, 1999).
- Una palabra es una unidad lingüística con propiedades sintácticas, morfológicas y fonológicas. Desde finales de los años 50 distintos investigadores han mostrado como la interpretación de una nueva palabra por parte de los niños varía en función de la clase sintáctica en la que se dice. Los niños de 3 años son capaces de diferenciar la categoría gramatical de la palabra, así como usar la categoría para interpretar el significado de la misma (Hall & Bélanger, 2005).
- Una palabra es una convención social. Los niños de 2 años ya son capaces de entender que el conocimiento de una palabra es compartido (Henderson & Graham, 2005). Son capaces de compartir el conocimiento de una nueva palabra con hablantes que no se la enseñaron, es decir, entienden que la palabra es una convención social y que el nuevo hablante conoce las convenciones sociales.
- Una palabra es un tipo de acción intencional que se refiere a un concepto con significado. Con 18 meses se puede aprender el significado de una palabra cuando la persona que la pronuncia señala el objeto (acción intencional) (Baldwin, Markman, Bill, Desjardins, Irwin, & Tidball, 1996). Muchos experimentos muestran la conexión entre palabra y concepto desde muy temprano: los bebés de 13 meses presuponen una propiedad común invisible a

distintos objetos que comparten el mismo nombre (Graham, Kilbreath, & Welder, 2004). Lo que se discute es el origen y desarrollo de esta conexión.

La complejidad de la palabra se resume en la *Figura 2*:

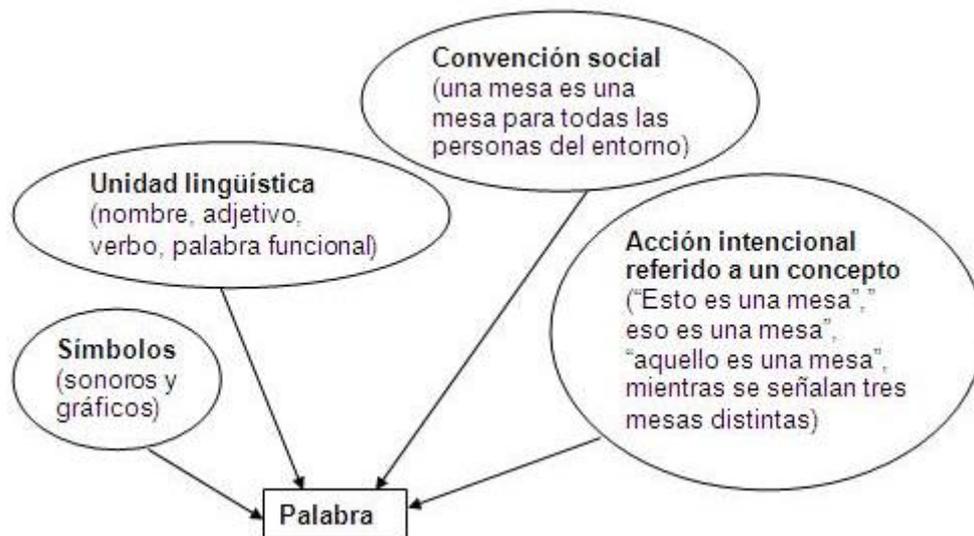


Figura 2.- Propiedades de las palabras.

Los niños aprenden palabras de muy distintos contextos (Akhtar, 2004). Muchos de estos contextos incluyen contextos de aprendizaje indirectos donde las palabras se aprenden sin instrucción explícita de un adulto. Según Akhtar, los niños intentan entender aquello a lo que el hablante está prestando atención más que establecer el significado abstracto de la palabra en sí. Los estudios que revisa muestran como los niños usan los contextos discursivos en los que se hablan las palabras para determinar a qué se refiere el hablante. Todos los estudios se refieren a palabras como términos de referencias: nombres de objetos, verbos de acción o adjetivos que describen propiedades visibles. Contextos mencionados por distintos investigadores (Akhtar, 2004; Baldwin, 1995; Linebarger & Vaala, 2010) son:

- **El contexto ostensivo/didáctico.** Mostrar un objeto diciendo su nombre es la forma más simple de enseñanza del lenguaje a un niño (Baldwin, 1995). El marco de rotulación es “Esto es un X”. La atención del niño está generalmente

FUNDAMENTOS

en el objeto cuando se nombra, coincidiendo la atención conjunta, la contigüidad temporal del referente y el nombre. Este contexto apenas se usa en el aprendizaje de verbos.

- **El contexto directivo** (Akhtar, 2004). Se requiere del niño una acción de un objeto que se ha mostrado previamente, es decir, se dirige el comportamiento del niño en vez de centrar la atención sobre el objeto y su nombre. Es la forma más común de aprender verbos, aunque también se pueden aprender del conocimiento de acontecimientos previos.
- **Contexto de rutina (repetitivo)**. Los niños pueden aprender palabras de contextos o acontecimientos repetitivos en los que la palabra y el referente no coincidan (Akhtar, 2004). Por ejemplo: mencionar una acción, pero no hacerla.
- **Relevancia del contexto discursivo**. Los niños parecen prestar atención a lo que es más relevante en el discurso de los adultos, aprendiendo palabras que son nuevas en dicho discurso sin que el adulto haga referencia al objeto real (Akhtar, 2004). También son capaces de asociar palabras dentro del tema del discurso. Por ejemplo en la frase “Trae el bermejo, no el rojo” interpretarían que bermejo es un color. Se aprende el vocabulario a través del lenguaje.
- **El contexto de oír una conversación ajena**. Se ha demostrado que los niños son capaces de aprender palabras de conversaciones que oyen y en las que no están participando (Akhtar, 2004). Este contexto puede ser importante a la hora de adquirir pronombres personales.
- **Aprendizaje a través de contextos**. Los niños aprenden palabras que se repiten en distintos contextos (Akhtar, 2004). Este mecanismo se puede ejemplificar con adjetivos, los cuales se asocian en distintos contextos (por ejemplo el color rojo, asociado a todas las cosas que tienen dicho color).

- **El contexto de los medios audiovisuales** (Linebarger & Vaala, 2010). Sobre todo aquellos que reflejan las experiencias de la vida real del infante, como por ejemplo las historias simples y los objetos cotidianos. Este aprendizaje se ve reforzado con la presencia de un espectador competente.

Los contextos de aprendizaje de las palabras se resumen en la *Figura 3*.

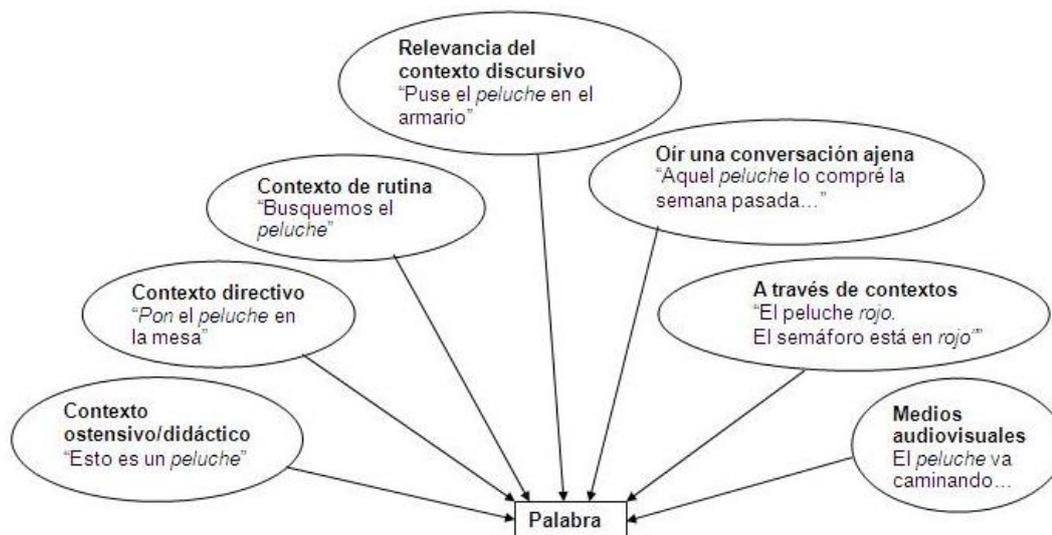


Figura 3.- Contextos de aprendizaje de las palabras.

Tabla 1

Cronología del aprendizaje del lenguaje

Edad	Adquisición del lenguaje
5-6 meses	Reconocimiento de su nombre.
6-7 meses	Reconocimiento de objetos y categorización.
9-10 meses	Relacionan una palabra nueva al objeto al cual el hablante está prestando atención.
12 meses	Reconocimiento de palabras y comienzo de producción de las mismas.
24 meses	Se producen cientos de palabras y se comienza a combinarlas sistemáticamente para formar frases.
2-3 años	Descubrimiento de las distintas categorías gramaticales relacionadas con los distintos tipos de significados.

FUNDAMENTOS

La cronología del aprendizaje del lenguaje, y por tanto de los primeros conceptos, se puede resumir en las Tablas 1 y 2 (Behl-Chadha, 1996; Clark, 2004; Vosniadou, 2002; Waxman & Lidz, 2006). La adquisición de los adjetivos parece descansar en la adquisición previa de los sustantivos. El aprendizaje de verbos requiere haber aprendido a reconocer argumentos sintácticos, por lo que es más tardío que el de los sustantivos.

Tabla 2

Cronología de la adquisición de conceptos

Edad	Adquisición de los conceptos
3-4 meses	Representaciones básicas de categorías. Distinción de objetos tridimensionales con un fondo uniforme. Las superficies se mueven juntas si están en contacto e independientemente si están separadas.
6-8 meses	Empieza la distinción espacial. Los objetos sin soporte caen hacia abajo (principio de la gravedad) o continúan en movimiento en ausencia de obstáculos (principio de inercia). Aparecen indicios de percibir la causalidad.
9-12 meses	Las palabras nuevas estimulan la categorización.
11 meses	Las palabras nuevas dirigen la atención hacia categorías o propiedades comunes.
12-14 meses	Relación entre palabras nuevas y organización conceptual. Se entiende la causalidad.
13 meses	Las palabras nuevas estimulan la inducción.
13-14 meses	Son sensibles a las categorías gramaticales aunque no se acaben de relacionar con los conceptos (caso de los adjetivos y verbos).
14 meses	Se distinguen los nombres de los adjetivos.
18 meses	Se relacionan los verbos con las acciones.
21 meses	Producción de adjetivos.
2-4 años	Uso de la sintaxis de la frase para aprender los verbos.

El uso del lenguaje está ligado a manejar significados (Janicki, 2006). El lenguaje se usa en distintos momentos de nuestra vida, como medio de expresión de significados, valores y tradiciones, así como para adquirir conceptos. El lenguaje de la ciencia tiene menor imprecisión que otros, manejando conceptos más precisos.

Deese (1965) concluye que las asociaciones libres de palabras de los adultos reflejan las propiedades de contigüidad, semántica y frecuencia de las palabras del lenguaje (citado en Hills, Maouene, Riordan, & Smith, 2010). Las palabras que forman concordancia con muchas otras palabras en el medio de aprendizaje son las que después tendrán mayor probabilidad de activarse en cualquier contexto.

Distintas investigaciones (Waxman & Gelman, 2009) revelan que a medida que los bebés y niños construyen un repertorio de conceptos y adquieren palabras para describirlos, se aprovechan tanto de la información perceptible como de la información conceptual, apoyando en ambas sus teorías rudimentarias. Los niños parecen ser sensibles a las repeticiones de la información perceptible.

Durante los años de escolarización se adquiere una cantidad de vocabulario a una media de miles de palabras por año, lo cual se traduce en varias palabras al día. Muchas investigaciones han intentado determinar la cantidad de vocabulario que se adquiere con la edad. Una de estas investigaciones (Anglin, 1993) se hizo sobre niños ingleses de 1º, 3º y 5º, que corresponden a las edades medias de 6, 8 y 10 años respectivamente. Para asegurar que los niños conocían el significado de la palabra (y por tanto el concepto) se usaron tres criterios: definición, frase y preguntas cerradas con respuesta de alternativa múltiple. Los resultados dieron una estimación de unas 10.000, 19.000 y 40.000 palabras para el vocabulario de las edades medias de 6, 8 y 10 años. Se estima además que entre los 6 y los 10 años los niños aprenden una media de 20 palabras al día. La velocidad de adquisición del

FUNDAMENTOS

vocabulario crece con la edad. Calderero (2003) considera que el dominio del vocabulario es casi sinónimo del dominio de la disciplina escolar.

2. 3.- La Lingüística Computacional y la Lexicometría

La Lingüística Computacional surge con la utilización de los ordenadores para elaborar modelos computacionales que reproduzcan distintos aspectos del lenguaje humano. El interés de la Lingüística Computacional se centró desde el principio en programas de traducción. Luego surgieron otros campos como el análisis morfológico y sintáctico o el análisis de grandes colecciones de textos o corpus (Domínguez Burgos, 2002). El análisis de corpus permite analizar grandes cantidades de texto teniendo en cuenta el contexto, cuantificar una palabra o un conjunto de palabras, así como el uso de una palabra en el contexto y las asociaciones de palabras, etc. (Biber, Conrad, & Reppen, 1998).

Es necesario trabajar estadísticamente los corpus ya que el ser humano tiende a fijarse más en las palabras y asociaciones infrecuentes que en las frecuentes, por lo que no son fiables las conclusiones basadas en la intuición (Biber, Conrad, & Reppen, 1998). Los datos estadísticos proporcionados por el estudio de los corpus también se deben interpretar cualitativamente.

La lexicometría es el análisis cuantitativo del uso de palabras, así como de las relaciones de una palabra con otra dada (Schöttler, 1989; Williams, 1999). La medida más obvia que se puede obtener de un corpus es el número de apariciones

de cada palabra, es decir, su frecuencia absoluta. Sin embargo la frecuencia absoluta no nos dice nada sobre la probabilidad de encontrar una palabra en un corpus o en un texto. Para ello es mejor utilizar la frecuencia relativa como una estimación puntual de la probabilidad de que una palabra aparezca en el corpus. La frecuencia relativa nos permite además comparar resultados en distintos corpus o textos.

En un listado de frecuencias, la frecuencia de una palabra es representativa de los textos con los que se ha formado el corpus. Además las palabras con frecuencias más altas son mayoritariamente las palabras funcionales (Renouf, 1992; Sinclair, 1991; Zipf, 1972 citado en Velasco, Díaz, Lloréns, de Amescua, & Martínez, 1999), es decir, aquellas que carecen de contenido descriptivo y actúan como instrumentos (relación, determinantes, cuantificadores,...) (Escandell Vidal & Leonetti, 2000; Morera, 2007). Las palabras funcionales forman la estructura en la que aparecen otro tipo de palabras, las palabras léxicas, es decir, aquellas que tienen significado, que representan un concepto.

La lingüística de corpus se ha ocupado tradicionalmente de la frecuencia de cada palabra y de las colocaciones. Las colocaciones son aquellas palabras que aparecen siempre juntas (Firth, 1957 citado en Baker, Hardie, & McEnery, 2006). Es decir, las colocaciones se refieren al hecho de que ciertas combinaciones de palabras aparezcan con cierta frecuencia en ciertos contextos.

De acuerdo con los lingüistas, en un léxico especializado como el del lenguaje científico se encuentra vocabulario general de uso frecuente, formado por unidades léxicas de la lengua general que sin perder su significado propio está dentro de la especialidad (Alcaraz Varó, 2000). El vocabulario específico es el que hace referencia a aquellas unidades léxicas usadas en un dominio del conocimiento,

FUNDAMENTOS

peculiares en virtud de su especialidad, con un significado altamente específico y un único referente conceptual (Gómez González-Jover, 2007).

La comparación de corpus generales con corpus específicos se usa para determinar el vocabulario específico, que se realiza comparando la frecuencia de una unidad léxica (formada por una o varias palabras) entre un corpus general de referencia y un corpus específico usando distintas técnicas estadísticas (Chung, 2003; Drouin, 2004; Lemay, L'Homme, & Drouin, 2005; Peñas, Verdejo, & Gonzalo, 2001; Vogel, 2003). En el vocabulario específico es posible encontrar palabras que sólo se encuentran en el corpus específico y otras que son usadas también en el corpus general. Por ejemplo, es posible encontrar la palabra *estrella* tanto en el corpus específico como en el corpus general CREA (Real Academia Española: Banco de datos (CREA) [en línea], 2008), sin embargo la palabra *Hertzprung-Russell* sólo existe en nuestro corpus específico. Existen varios corpus generales del español con los que se podría hacer la comparación (Tabla 3).

Tabla 3

Algunas características de corpus del español

Corpus	Nº	Procedencia cronológica ^a
Juilland & Chang-Rodríguez, 1964	500.000	1920-1940
Alameda & Cuetos, 1995	2.000.000	1978-1993
Justicia, 1995	528.544	1979-80, 1983-84
LEXESP, 2000	5.020.930	1978-1995
Davies, 2002-	22.822.256	Siglo XX
CUMBRE, 2005	20.662.306	1950-1995
CREA, 2007	154.279.050 ^b	1975-2004

Nota. Nº= número de palabras.

^aFechas de publicación de los textos que forman los corpus. ^bLas 154 millones de formas se refieren al corpus escrito. Además hay un corpus oral de casi 9 millones de formas.

El corpus de Juilland & Chang-Rodríguez (1964) se presenta dividido en cinco universos de 100,000 palabras cada uno: obras de teatro (lo más parecido al lenguaje hablado), ficción, prosa, artículos (prensa) y literatura técnica.

El corpus de Alameda & Cuetos (1995) contiene un 50% de novela, un 25% de prensa, un 15% de ensayo y un 10% de divulgación científico-técnica.

El corpus de Justicia Justicia (1995) es un corpus ya lematizado procedente de textos escritos por 3402 niños con edades comprendidas entre 6 y 13 años.

El corpus LEXESP (Sebastián Gallés, Martí Antonín, Carreiras Valiña, & Cuetos Vega, 2000) está formado en un 40% por narrativa, en un 10% por divulgación científica, en un 10% por ensayo, en un 25% por prensa, en un 10% por semanarios y en un 5% por prensa deportiva.

El corpus de Davies (2002-) contiene textos desde el siglo XII hasta el siglo XX. Aquí sólo se menciona la parte del siglo XX, en la cual se han recopilado tanto textos de España (43%) como de Latinoamérica (57%).

El corpus CUMBRE (Almela, Cantos, Sánchez, Sarmiento, & Almela, 2005) contiene un 30% de lenguaje oral y un 70% de lenguaje escrito procedente en un 35% de Hispanoamérica.

El corpus CREA (Real Academia Española: Banco de datos (CREA) [en línea], 2008) incluye un 10% de lenguaje oral y un 90% de lenguaje escrito. En esta tesis se usa este último corpus como referencia debido a que se puede consultar separadamente la parte de lenguaje escrito, está formado por textos más recientes y contiene el mayor número de palabras.

2. 3. 1.- El lenguaje de las ciencias desde el punto de vista de la lingüística

La ciencia no se comunica a través del lenguaje verbal solamente, sino simultáneamente a través de lenguaje verbal, matemático, visual, gráfico y operacional (Lemke, 1998). En una muestra de 20 artículos del *Physical Review Letters*, revista que sólo acepta informes breves de los principales descubrimientos, Lemke encontró que había una media de 1.2 gráficos y 2.7 ecuaciones por página. En esta muestra se consideraron como ecuaciones aquellas que venían en una línea aparte del texto y no aquellas que venían insertadas como parte de las frases del mismo. Se observó además que el texto no tenía sentido sin las ecuaciones matemáticas integradas, y que las mismas debían ser leídas como parte del texto. En esta tesis se estudia el lenguaje escrito que contiene simultáneamente lenguaje verbal y matemático.

El lenguaje es parte de la ciencia y de la alfabetización científica (Yore, Bisanz, & Hand, 2003). A pesar del énfasis que se hace durante la escolarización en las fórmulas, las matemáticas no son el sistema de lenguaje exclusivo de la ciencia. Es más bien el lenguaje oral y escrito el sistema de símbolos que más usan los científicos para construir, describir, argumentar, comunicar procedimientos y divulgar ciencia (Yore, Hand, & Prain, 2002). La ciencia no puede prescindir del lenguaje ordinario (Galán Rodríguez & Montero Melchor, 2002), sino que es un cuerpo de ideas expresado en diversos lenguajes que se necesitan y superponen. Como construcción lingüística, los textos de ciencias son susceptibles de ser analizados de manera similar a como se analiza un discurso narrativo o de ficción.

El estudio del lenguaje de las ciencias es de interés en la comunidad científica (Ash, 2004; Freeman & Taylor, 2006; Young, 2005). Las características del lenguaje científico se pueden resumir como siguen:

- **La nominalización**, es decir, la transformación de un verbo en un nombre (Shanahan & Shanahan, 2008; Yore, Bisanz, & Hand, 2003). En los textos de ciencias la nominalización se usa para crear vocabulario técnico. Por ejemplo, en vez de escribir “la sal se disolvió...”, un científico escribiría “la disolución de la sal...”. La nominalización se usa para trasladar un fenómeno de lo particular o específico a lo general o abstracto. Es por ello que muchas veces se alude como característica de los textos científicos la abstracción, a través de la nominalización (Fang, 2005). La nominalización es menor en la divulgación (Galán Rodríguez & Montero Melchor, 2002).
- Los textos de ciencias tienen **una densidad léxica mayor** que los textos de matemáticas o historia (Fang, 2005; Shanahan & Shanahan, 2008). La densidad léxica es el número de palabras con significado respecto al número total de palabras. Las palabras con significado son términos que deben ser entendidos para poder entender la ciencia a la que se refieren. En una oración hablada hay unas 2 o 3 palabras con contenido (significado), en una oración escrita de 4 a 6 palabras, mientras que en el lenguaje científico escrito hay entre 10 y 13 palabras con significado por oración (Halliday, 1993).
- **El uso de analogías, conectores lógicos e intertextualidad** para justificar los procedimientos y construir conocimiento (Yore, Bisanz, & Hand, 2003). Los conectores lógicos son palabras o expresiones que marcan la relación entre ideas. La intertextualidad es la citación de otros trabajos científicos relevantes y

FUNDAMENTOS

se usa para demostrar cómo los procedimientos y el conocimiento se conectan con otros ya establecidos.

- **Vocabulario técnico o específico** (Fang, 2005). En los textos usados en la enseñanza este vocabulario suele aparecer en negrita. En los textos de divulgación se reduce el vocabulario técnico sustituyéndolo por un vocabulario más común (Galán Rodríguez & Montero Melchor, 2002).
- **Autoridad** (Fang, 2005): el autor se presenta como un experto que proporciona información objetiva. La neutralidad y objetividad se expresan mediante el uso de oraciones enunciativas en modo indicativo, oraciones en voz pasiva e impersonales, construcciones nominales, densidad léxica y empleo del artículo con valor generalizador (Galán Rodríguez & Montero Melchor, 2002).
- **Lenguaje conciso**, evitando la redundancia (Snow, 2010).
- **Univocidad**, se tiende a la monosemia de los términos científicos, es decir, un significado por término (Galán Rodríguez & Montero Melchor, 2002).
- **Precisión** en la expresión (Snow, 2010). El significado de los términos está condicionado por la relación entre esos términos con un ámbito específico del conocimiento y no por la situación comunicativa (Galán Rodríguez & Montero Melchor, 2002). Por ejemplo “*raíz*” tiene una definición específica en cada uno de los ámbitos de las Matemáticas, Botánica y Lengua. Estas definiciones no son intercambiables.
- **Oraciones complejas**, mediante el uso de subordinadas (Fang, 2006).

Se pueden distinguir tres niveles de comunicación en ciencias (Sevilla Muñoz & Sevilla Muñoz, 2003):

- A través de publicaciones científicas (lenguaje escrito) o de conferencias, coloquios, sesiones clínicas, seminarios (lenguaje oral).

- Un contexto informal, como el que se aprecia en diarios de laboratorio, cartas, correo electrónico (lenguaje escrito) y discusiones informales (lenguaje oral).
- Al público en general, en publicaciones divulgativas (lenguaje escrito) o en programas divulgativos de radio y televisión (lenguaje oral).

El lenguaje usado no es el mismo cuando se dirige a otros científicos que cuando se dirige al público en general (Galán Rodríguez & Montero Melchor, 2002). En este último caso no se pueden tratar los temas en profundidad. El discurso de la divulgación depende tanto del medio que se emplee como de la palabra y prestigio del científico que divulga. También cambia el lenguaje a la hora de destinarlo para una determinada función, como lo es la enseñanza. Por tanto, se pueden distinguir tres tipos de discursos (Marimón Llorca & Santamaría Pérez, 2007): el altamente especializado (dirigido a otros científicos), el medianamente especializado o didáctico (dirigido a los que están en formación) y el de bajo nivel de especialización (dirigido al público en general). Estos tres tipos de discursos son los que se tomarán en esta tesis para clasificar los textos en los géneros científico, educativo y divulgativo.

2. 3. 2.- Análisis de texto asistido por ordenador y análisis de redes

El análisis de texto asistido por ordenador es el nombre que reciben distintas técnicas que van desde las ciencias sociales cuantitativas al análisis de contenido, incluyendo encuestas, entrevistas transcritas y discursos (Brier & Hopp, 2011). A la

FUNDAMENTOS

hora de analizar textos asistidos con ordenador pueden ser interesantes las frecuencias de las palabras, la búsqueda de palabras claves, las apariciones de palabras y frases contiguas (co-ocurrencias) en altas frecuencias, o la exclusión de palabras. La preparación de los textos que se analizan, así como su almacenamiento, es una tarea que consume mucho tiempo. Esto se debe a imperfecciones en la tipografía, en la transcripción y en la identificación de las palabras. La manera más simple de tratar los textos es tenerlos en formato de sólo texto, habiendo hecho la exclusión de todo material que no sea relevante para el propósito de la investigación. A la hora de trabajar con una aplicación informática se hace necesario probar primero con una muestra del texto o corpus.

Las frecuencias de las palabras son importantes pues indican la medida de uso de dicha palabra en el lenguaje. Para la enseñanza de una lengua extranjera, la frecuencia de uso de las palabras es un indicador de cuáles se deben enseñar (Leech, 2001) tanto desde el punto de vista gramatical como léxico. De la misma manera, en esta tesis, la frecuencia de uso de los lemas y sus relaciones es un indicador de los conceptos relevantes en la enseñanza de las ciencias.

La recurrencia, es decir, una frecuencia de 2 o más, es el mínimo para considerar una palabra como unidad independiente del lenguaje que se estudia (Sinclair, 2005). Las palabras que sólo ocurren una vez sólo sirven para estudios de autoría. Por lo tanto el investigador debe fijar una frecuencia mínima a partir de la cual las palabras son objeto de análisis.

Dado un corpus los ordenadores permiten el análisis del uso de las palabras en su contexto natural mediante sus frecuencias (Biber, Conrad, & Reppen, 1998). De esta manera es posible identificar el significado de las palabras y sus distintos usos. El estudio de la frecuencia de una palabra en relación con otras muestra

palabras comunes (entre textos) y colocaciones (grupos de palabras que siempre aparecen juntas). Al comparar las apariciones de una palabra en distintos tipos de textos se usa la frecuencia relativa.

También es útil estudiar las co-ocurrencias entre palabras o frases (Brier & Hopp, 2011). Para ello se debe definir la unidad de contexto en la que se estudia, es decir, si dentro de una frase o conjuntos de frases. La aplicación de distintas unidades de contexto llevará a diferentes conclusiones sobre las relaciones entre las palabras y los conceptos que representan. El estudio de las co-ocurrencias se hace mediante representaciones en términos de análisis de redes.

Desde la perspectiva de la lingüística de corpus, Stuart & Botella (2009) proponen una nueva metodología donde se analizan corpus usando redes semánticas. Su intención no es otra que la representación del discurso mediante redes de significados.

Para poder analizar textos se han desarrollado numerosos programas con distintos fines (Llisterri, 2007) entre los que se encuentran:

- **Wordsmith Tools.** Conjunto de programas con los que se obtienen listas de palabras o grupos de palabras de un texto en orden alfabético o en orden de frecuencias (Wordlist), listas de palabras en contexto (Concord), búsquedas de palabras clave en un texto (KeyWords) y con las que se pueden hallar concordancias, es decir, al lado de qué palabras aparece una dada (Scott, 2010).
- **Concordance.** Permite hacer índices y listas de palabras, halla frecuencias de aparición de palabras y concordancias.
- **MonoConc Pro.** Busca palabras, frases y halla concordancias y colocaciones.
- **Collocate.** Permite buscar y analizar estadísticamente colocaciones.

FUNDAMENTOS

- **Para Conc.** Busca concordancias en corpus paralelos bilingües o multilingües, posibilitando traducciones.

La teoría de grafos, o análisis de redes, se puede aplicar a cualquier estructura que contenga nodos conectados unos a otros. Estos nodos pueden ser palabras y las relaciones entre los nodos pueden indicar el grado de asociación entre las palabras de un corpus (Hills, Maouene, Riordan, & Smith, 2010).

Existen distintos tipo de redes (Calderero, 2003):

- **Redes semánticas.**

Una red semántica representa el conocimiento como un grafo en forma de red. Una idea, un evento, una situación u objeto tiene casi siempre una estructura compuesta; esto se representa en una red semántica mediante una correspondiente estructura de nodos que representan unidades conceptuales, y aristas direccionadas que representan las relaciones entre las unidades. (Lehmann, 1992, p. 2 citado en Echeverría, Vargas, Urzúa, & Ferreira, 2008).

Ejemplos de estas redes (*Figura 4*) se pueden ver en el trabajo de Echeverría, Vargas, Urzúa, & Ferreira (2008), donde las vecindades y agrupaciones definidas en los grafos expresan valores semánticos tanto de unidades (vocablos) como de conjuntos (categorías). En este trabajo los datos procedían de encuestas a individuos centradas en determinados temas.

- **Red conceptual.** Red de conceptos esenciales del tema, donde éstos se colocan en los nodos y se conectan mediante aristas terminadas en punta de flecha que indican el sentido en el que hay que leer (Galagovsky, 1993). Las aristas llevan un verbo preciso que hace que se generen oraciones nucleares entre nodos. Se

puede empezar a leer en cualquier nodo. Se conciben como guías de enseñanza para el docente y de aprendizaje para el alumno. No se cuantifican ni los nodos ni las aristas. Un ejemplo se puede ver en la *Figura 5*.

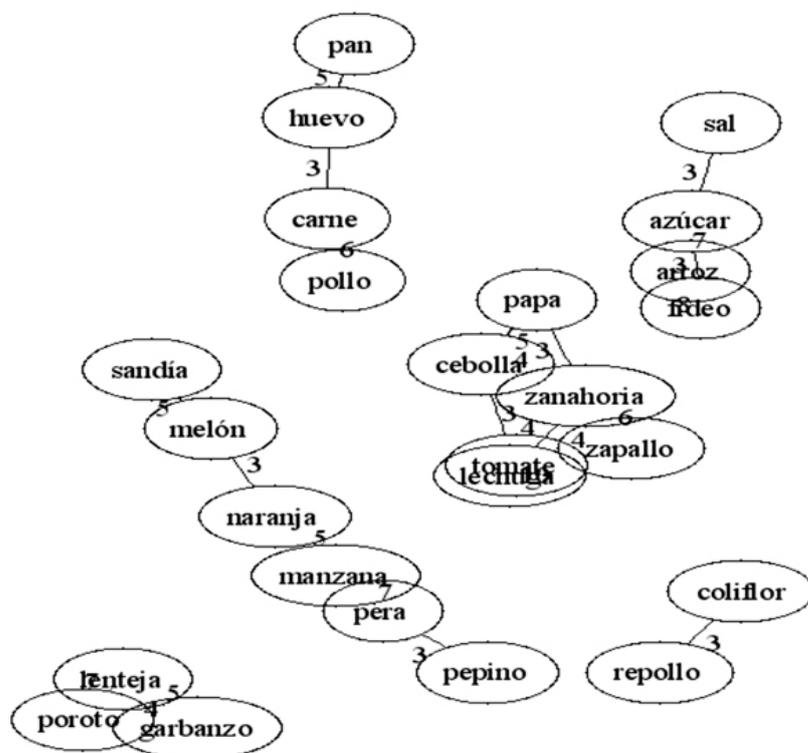


Figura 4.- Parte de un grafo obtenido mediante encuestas a 23 sujetos respecto al tema *Alimentos*. En este grafo los vocablos (nodos) están dentro de elipses que se hallan conectadas mediante aristas que tienen un peso. El peso de las aristas indica el número de veces que ha aparecido dicha relación. Se pueden apreciar 6 agrupaciones (categorías) así como la vecindad entre vocablos.

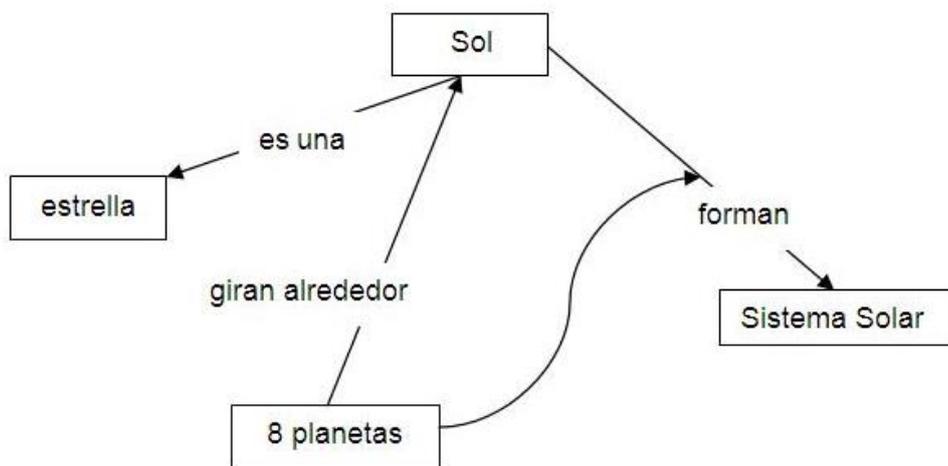


Figura 5.- Red conceptual sobre el Sol. Se pueden unir las aristas procedentes de varios nodos para generar oraciones.

FUNDAMENTOS

- **Mapa conceptual.** Diagrama jerárquico que procura reflejar la organización conceptual de una disciplina, o parte de ella (Novak y Gowin, 1988 citado en Calderero, 2003). En un mapa conceptual aparecen los conceptos en los nodos y palabras que los relacionan en las aristas. En los mapas conceptuales no se pretende cuantificar ni los nodos ni las aristas. Un ejemplo de mapa conceptual se presenta en la figura *Figura 6*.

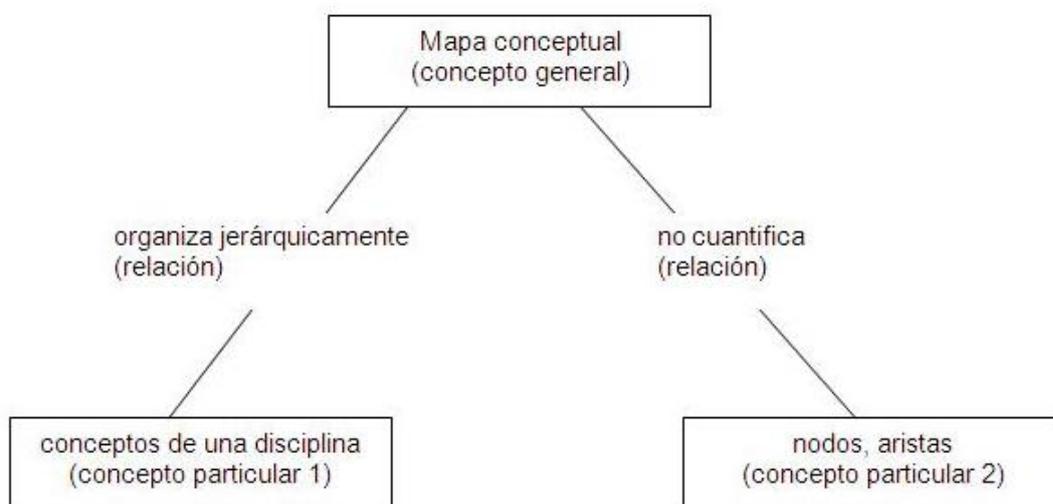


Figura 6.- Mapa conceptual de mapa conceptual. Se parte del concepto general y se van añadiendo de forma jerárquica los conceptos cada vez más particulares.

- **Mapa semántico.** Organizador gráfico de categorías de información respecto a un concepto central que se usa para el análisis conceptual de textos (Heimlich & Pittelman, 1990). Ayuda a los alumnos a desarrollar una estrategia simple para desenmarañar un texto y para crear estructuras de conocimiento. No pretende ser un grafo ni están cuantificados sus posibles nodos o aristas. Un ejemplo de mapa semántico se presenta en la figura *Figura 7*.
- **Mapa mental.** Gráfico en el que se refleja un punto central y los pensamientos asociativos que se conectan a éste. Los nodos se pueden expresar mediante imágenes o palabras. Los mapas mentales estimulan la expresión en todas sus facetas, despiertan la imaginación, desarrollan la capacidad de síntesis y de

análisis y contribuyen a un mejor manejo del tiempo. Se utilizan hojas blancas, ideas-centro, colores, flechas, símbolos, dibujos, palabras claves, códigos, que permiten recordar con facilidad. (Buzan & Buzan, 1996; Calderero, 2003). No se cuantifican nodos ni aristas. Un ejemplo de mapa mental se puede ver en la *Figura 8*.

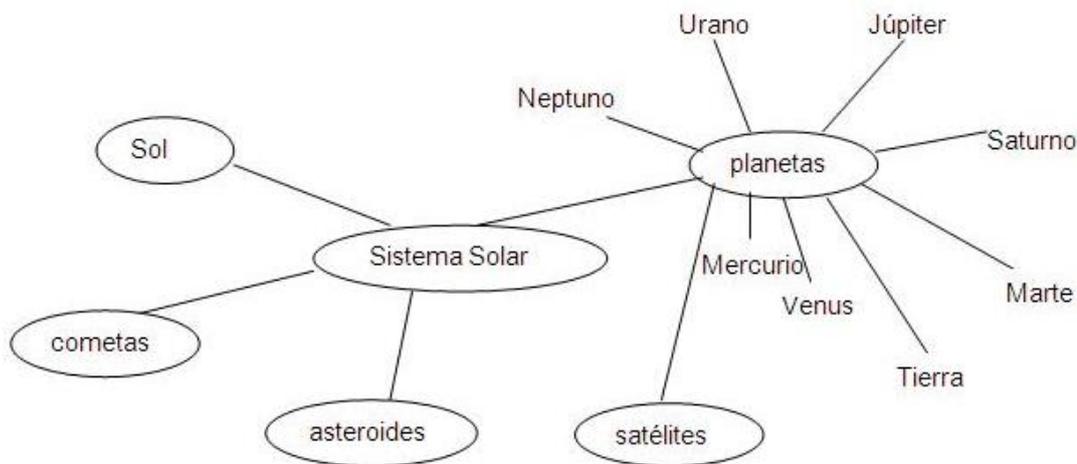


Figura 7.- Mapa semántico sobre el Sistema Solar.

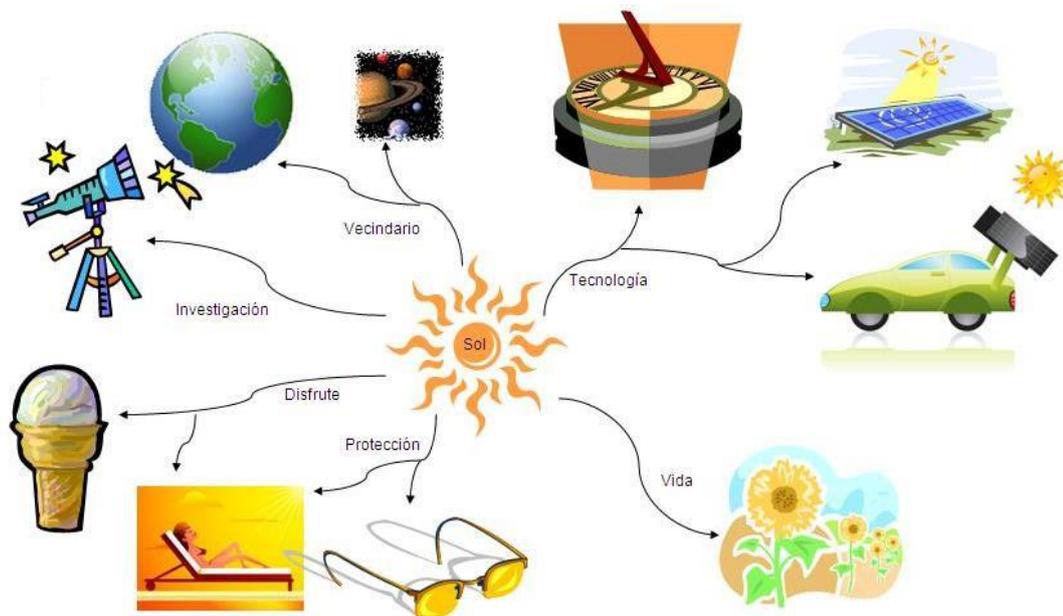


Figura 8.- Mapa mental del Sol. El mapa mental se parece a un árbol al que le salen ramas que a su vez se pueden ramificar.

- **Grafo textual.** Grafo donde los nodos y las aristas pueden representar palabras, vocablos o conceptos. Los nodos y las aristas se cuantifican y se pueden

FUNDAMENTOS

analizar mediante las técnicas matemáticas de análisis de grafos. El grafo textual permite (Calderero, 2003): distinguir ideas centrales y secundarias, organizar y estructurar ideas, jerarquizar contenidos, identificar palabras-clave y distinguir entre conceptos y conexiones entre los mismos.

Ya existen aplicaciones informáticas que permiten dibujar redes semánticas como el DispoGrafo (Echeverría, Vargas, Urzúa, & Ferreira, 2008). Esta aplicación informática se diseñó para el análisis de términos mediante encuestas de léxico disponible. Esta aplicación permite eliminar las conexiones débiles (de menor peso) para así dejar sólo las relaciones más robustas y visualizar de este modo las relaciones más relevantes. DispoGrafo permitirá diversas observaciones cualitativas tales como la categorización, a través de las agrupaciones nodales claramente visibles en el grafo; la configuración de núcleos y vecinos inmediatos; la fuerza de la relación establecida por el peso de las aristas, etc.

El análisis con redes semánticas se ha usado ya en libros escolares (Doyle, 2005; Phillips, 1985; Phillips, 1989) y artículos de investigación. Phillips (1985), usando libros de ciencias universitarios, muestra que hay grupos de colocaciones (estructuras) dentro del texto. La organización léxica de estas estructuras contribuye a la estructuración del texto y son a la vez un elemento vital en el entendimiento del texto por parte del lector. La organización concebida por el autor se confirma con un análisis de redes semánticas. Las redes semánticas articulan un área crucial del contenido conceptual del texto. Phillips (1989) argumenta que el significado de los términos es una función de las relaciones léxicas obtenidas en el texto. Las redes semánticas no sólo organizan el texto en la escala del capítulo pero también en la escala global al aparecer en distintos capítulos. La macroestructura, redes semánticas que aparecen en todo el texto, es pues una categoría convencional en los

textos de ciencias. Los libros de texto que no eran de ciencias presentan colocaciones más débiles, y por tanto poca tendencia a organizaciones en macroestructuras. Las estructuras (redes semánticas) de un texto de ciencias se pueden identificar mediante criterios formales aunque no sean susceptibles de ser observadas directamente.

El análisis con grafos textuales de 20 libros de secundaria de 4 editoriales distintas en torno a las propiedades físicas de la materia ha permitido establecer la homogeneidad de los textos (Calderero, 2003). En este caso los nodos con sintagmas nominales y los vértices unen dos sintagmas nominales. Para ello se codifica el texto en oraciones simples (Calderero, 2003; Sánchez Gómez, Cervelló, & Martín, 2001). De esta manera se han analizado cuestionarios de alumnos universitarios de la carrera de químicas en torno al tema de “*cómo se disponen los electrones en un átomo*” (Sánchez Gómez, Morcillo, Martín, & Silván, 2002).

3.- METODOLOGÍA

Las estrellas
Una estrella es una bola gigantesca de gas que tiene luz propia con propiedades físicas tales como la masa, la temperatura o el radio. También es de interés para los astrónomos la distancia de la estrella a la Tierra. La estrella más cercana— y por ende, más estudiada — es, por supuesto, nuestro propio Sol.

Analisis de frecuencias.(C:\Documents and Settings\user\Mis documentos\doctorado)

Una estrella es una bola gigantesca de gas que tiene luz propia con propiedades físicas tales como la masa, la temperatura o el radio. También es de interés para los astrónomos la distancia de la estrella a la Tierra. La estrella más cercana— y por ende, más estudiada— es, por supuesto, nuestro propio Sol.

Procesamiento

Palabra no encontrada

Palabra: FISICAS

Cambiar por: FISICO

Eliminar Cambiar... Dejar Verbo

Analisis de frecuencias.(C:\Documents and Settings\Datos\Mi

INCIDENTAL AEREA ESTRELLA CAPA ?
FORMAR PARTIS RESTO CAPA EXTERNO
ESAL ESOL * NUCLEO COLAPSADO EST
FICCIÓN * ESTRELLA DENSOHARIA RUY
VIDA ESTRELLA NORMAL TIPO SOLAR
LUMINOSIDAD TEMPERATURA SUPERFICIAL TODO ESTRELLA SECUENCIA PRINCIPAL
VEZ DIAGRAMA HERTZSPONGERHELL FIGURA 6 * LINEA MAS MENOSO RECTO SECUENCIA
PRINCIPAL EXTENDER APROXIMADAMENTE 1 ORDEN MAGNITUD TEMPERATURA 3 MULTIPLICAR
1046 KELVIN HASTAN 3 MULTIPLICAR 1046 KELVIN * RANGO LUMINOSIDAD EXTENDER
APROXIMADAMENTE 6 ORDEN MAGNITUD 1046 LUMINOSIDAD SOLAR HASTAN 1046 LUMINOSID
SOLAR * TANTO ESTIMAR FORMA APROXIMADA LUMINOSIDAD PROPORCIONAL TEMPERATURA
SUPERFICIAL 46 ESTRELLA SECUENCIA PRINCIPAL ** DAR ALGUNO EJEMPLO ESTRELLA
GRANDE MASA SECUENCIA PRINCIPAL TEMPERATURA SUPERFICIAL APROXIMADAMENTE
TEMPERATURA SUPERFICIAL ESTRELLA IGUAL 1046 MULTIPLICAR 1046 KELVIN
TENEN LUMINOSIDAD APROXIMADAMENTE LUMINOSIDAD ESTRELLA IGUAL 1046 DIVIDIR
100046 46 MULTIPLICAR LUMINOSIDAD SOLAR APROXIMADAMENTE 2 6 VEZ LUMINOSIDAD
SOL * LUMINOSIDAD SOL TENER VALOR ESTANDAR 1 ESCALA LUMINOSIDAD ** ESTRELLA

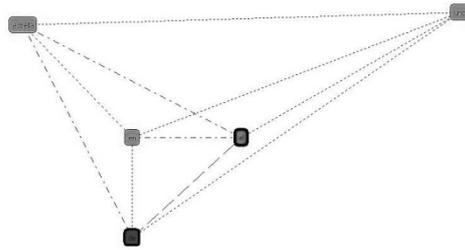
Analisis de frecuencias.(C:\Documents and

Palabra	Frecuencia		
632	TAN	1	0.419
633	TELESCOPIO	1	0.419
634	TEORIAS	1	0.419
635	TERMONUCLEAR	1	0.419
636	TIERRA	1	0.419
637	TIPO	1	0.419
638	TOOL	1	0.419
639	TODOS	1	0.419
640	TRANSFORMADO	1	0.419
641	TRAYESOO	1	0.419
642	TIPO	1	0.419
643	TERMICA	1	0.419
644	TERMINO	1	0.419
645	TIPICOO	1	0.419
646	UNAS		
647	UNIVERSO		
648	USA		
649	USARCE		
650	USAR		
651	VALIOSA		
652	VALIOSO		
653	VANI		

Gráfica de relaciones

Estudio de entorno: "FESTESE001VD".

LINEAS:
[2,000,120,000]... G=1, [120,000,238,001]... G=1, [238,001,356,001]... G=1;



Cargar Estudio del Entorno

Palabras a procesar: Frecuencia Mínima: 0.416

Fragmento: Izquierdo: 0 palabra; Derecho: punto y seguido.

Entorno: Retorno: más+1; Decremento: 1; Procesar: aritmética.

Aceptar Cancelar

Navegador de estudio del entorno

Nº Palabras	Visualizadas	Procesadas
Estrella	12,000	
de	371,000	
o	23,000	
luminosidad	16,000	
nieves	16,000	
temperaturasuperficial	16,000	
sol	14,000	
ray	13,000	
secuenciaprincipal	13,000	

Lista de palabras: estrella, de, o, ocnologibular, luminosidad, nieves, temperaturasuperficial, sol, ray, secuenciaprincipal

Print Preview

Estudio de entorno: "FESTESE001VD".

Frecuencia mínima: 64.000 Nº Palabras: 2

Palabras a procesar: Fragmento: Izquierdo: Punto y seguido; Derecho: Punto y seguido; Entorno: Valor máximo: 3; Decremento: 1; Tipo de sucesión: Aritmética

Palabras a visualizar: Frecuencia mínima: 64; Relación mínima: 6; Nº Palabras: 5

Palabras filtradas: Frecuencia mínima: 64; Relación: Entre Palabras Origen con todas las palabras.

Palabra Origen	Palabra	Relación Absoluta	Relación Relativa a Palabra Origen
de	de	356	5651
de	estrella	163	238,001
de	de	133	203,003
de	de	130	196,473
de	de	115	161,115
de	de	87,439	14,201

3.- METODOLOGÍA

3. 1.- Antecedentes de la Metodología Utilizada

La metodología empleada se utiliza en la línea de investigación que el grupo GICEC (Grupo de Investigación de Conceptos en la Enseñanza de las Ciencias) desarrolla desde 1997. En esta línea de investigación se determinan las unidades léxicas (UL) que aparecen en diferentes temas de ciencias y las relaciones de cada una de ellas con otras UL (Ceballos, Galotti, Santana, & Varela, 2000a; Ceballos, Galotti, & Varela, 1998; Ceballos, Galotti, Varela Calvo, & Talavera Sosa, 1999; Ceballos, Hansen, & Stengler, 1999; Domínguez Herrera 2009; Domínguez Herrera, Galotti, & Varela Calvo, 2002; Domínguez Herrera & Varela Calvo 2009; Galotti, Ceballos, & Matus, 2003).

Como ya se ha indicado en una tesis anterior (Domínguez, 2006), la metodología de investigación aplicada se enmarca dentro de las técnicas de Lingüística Computacional para el Análisis Textual, mediante las cuales se pueden obtener datos estadísticos de léxicos específicos temáticos, así como los valores de sus relaciones (Bécue, 1991; Etxeberría, García, Gil, & Rodríguez, 1995). Este enfoque lexicométrico y de estadística textual está inicialmente fundamentado en las técnicas estadísticas desarrolladas por la Escuela Francesa de Análisis de Datos (Benzécri, 1973; Benzécri, 1976).

METODOLOGÍA

En líneas generales, las técnicas empleadas para este tipo de análisis textual se identifican con procedimientos lexicométricos orientados a lematizar el contenido, contar las frecuencias de uso de las UL, y después operar con algún tipo de tratamiento estadístico, a partir de los resultados de tales recuentos, para establecer órdenes de frecuencias de aparición y relaciones entre UL (Lebart, Salem, & Bécue, 2000).

En concreto, con la lematización del contenido se trata de seleccionar convencionalmente una forma léxica para remitir a ella todas las de su misma familia o grupo de palabras vinculadas por la modificación de morfemas dependientes (género, número y tiempo verbal). Esta forma léxica recibe el nombre de lema, la cual en ocasiones se puede unir a otra para formar una UL independiente. Como resultado de la lematización, se obtiene el conjunto de UL que configuran el repertorio léxico y que, al igual que las palabras originales, etiquetan o representan a los conceptos que describen (*Figura 9*).

Las actuales posibilidades informáticas de cálculo y de gestión están ayudando a realizar este tipo de operaciones con programas diseñados específicamente para el Análisis Textual (Llisterri, 2007) tales como Wordcruncher 4.5, Concordance, Concordar Pro, MonoConc Pro, Ultra Find, Word List Maker, Collocate. Cabe destacar el programa SPAD.T, Système Portable pour L'Analyse des Données Textuelle, que cuenta con gran difusión entre los aplicadores del Análisis Estadístico Textual (De Rosa, 2002; Dugac, Kern, & Majdancic, 2005; Lebart, Morineau, & Bécue, 1989; Lebart, Morineau, Bécue, & Haeusler, 1994).

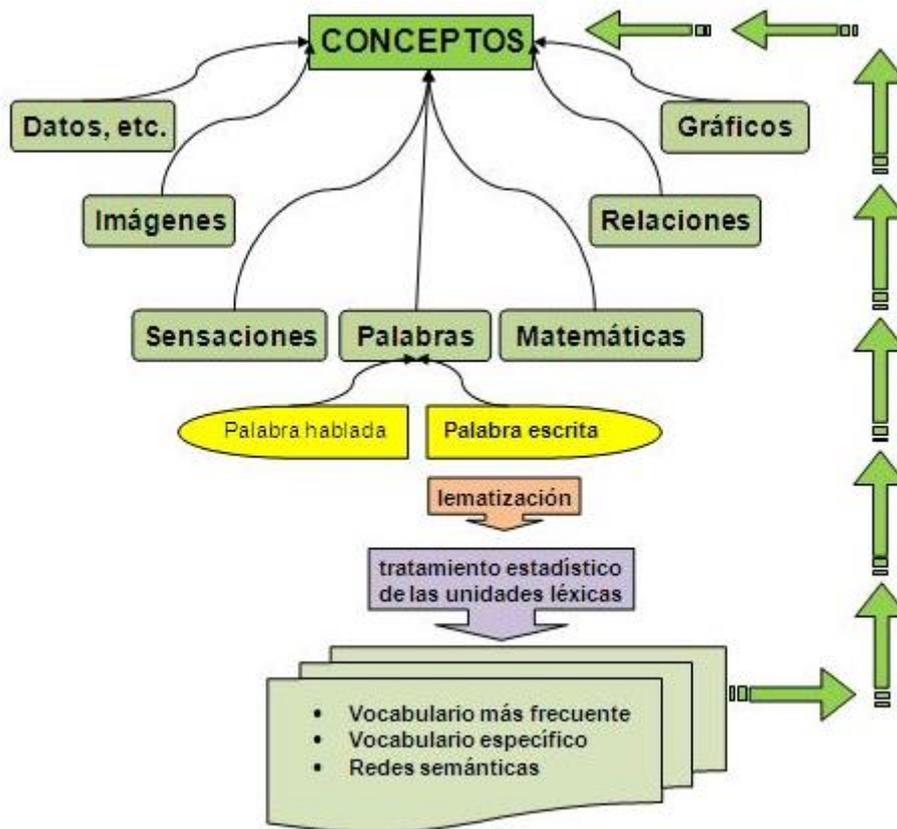


Figura 9.- De los conceptos a las UL y de éstas a los conceptos.

El desarrollo de las técnicas de la estadística textual ha hecho que el análisis estadístico de textos se haya constituido en una herramienta interdisciplinar, integrada por la Estadística, el Análisis del Discurso, la Lingüística, la Informática y la Investigación Documental (Lebart & Salem, 1994). Este tipo de análisis es muy utilizado en diversos campos de las Ciencias Sociales, como por ejemplo, Historia (Dugac, Kern, & Majdancic, 2005); Política (Gattino & Miglietta, 2004); Economía; Antropología (Blazekovic-Mitakovic, Kern, & Vukovic, 2000); Psicología (Gattino & Miglietta, 2004); Sociología, específicamente en el análisis de los discursos sociales del sujeto mediático (Delgado & Gutiérrez, 1999; Galindo, 1998; Gattino & Miglietta, 2004; Lebart & Salem, 1994); Educación y en concreto en Didáctica (Bécue-Bertaut, Lebart, & Rajadell, 1992).

METODOLOGÍA

En el presente trabajo se aplica este análisis en el ámbito de la Didáctica de las Ciencias, sustentándose en el estudio lexicométrico del discurso empleado en textos de ciencias y en procesos de estadística textual.

3. 2.- Metodología de Investigación

La metodología de investigación consiste en un análisis de las frecuencias de las palabras y lemas y de las frecuencias de las relaciones entre ellas, realizado con el programa PAFE (Ceballos, Galotti, Santana, & Varela, 2000b) desarrollado en el grupo de investigación GICEC (Ceballos, Varela, & Galotti, 1999). La instalación del programa PAFE se puede ver en el Anexo A. Para trabajar con la aplicación informática PAFE es necesario previamente preparar los textos.

Para ello se elige previamente el tema a estudiar o registro de acuerdo con los intereses del investigador, en nuestro caso Astronomía: estrella. Posteriormente se realiza una selección representativa de documentos escritos de un mismo registro en distintos géneros: académico, divulgativo y científico (Ceballos, Galotti, Santana, & Varela, 2000b).

Para ejemplificar e investigar sobre la metodología usaremos un texto escolar de astronomía: “Medida de la distancia y la edad de un cúmulo globular de estrellas” cuyos datos bibliográficos pueden verse en la Tabla 8 y el texto original en el Anexo B.

3. 2. 1.- Preparación del texto

La preparación del texto es la primera y una de las fases más importantes del proceso, pues de su coherencia y rigor depende el grado de fiabilidad de los resultados; por tanto, debe realizarse siguiendo una metodología sistematizada y escrupulosa (Galotti, Ceballos, & Matus, 2003).

Para poder analizar el texto se debe preparar de modo conveniente a los intereses de la investigación y para que sea reconocido por el programa informático a utilizar (PAFE). El texto seleccionado se digitaliza con un escáner y se pasa a formato Microsoft Word (*.doc) mediante un programa de reconocimiento óptico de caracteres (OCR). Para la investigación se han usado distintos escáneres (HP ScanJet 5p, Canon LIDE) y programas OCR, tales como Omni Page y TextBridge.

Los textos de los distintos géneros a considerar son diferentes en cuanto a la presentación de su contenido. Hay que recordar que lo que interesa al investigador es todo el texto que tiene que ver con la exposición y explicación del contenido. Un texto escolar, como el que se ha escogido de ejemplo, es el que presenta la información de forma más variada: títulos, subtítulos, contenido en forma de texto, fórmulas, imágenes, gráficos, tablas, notas al pie, notas al margen, pies de figuras, pies de tablas, pies de imágenes, pies de gráficos, resumen, esquemas, mapas conceptuales, experimentos, problemas, actividades y para saber más (biografías, reproducción de artículos, etc.).

Las imágenes, gráficos, tablas, esquemas, mapas conceptuales, experimentos, problemas y actividades no forman parte de la exposición del contenido, sino que se usan para ilustrar, describir, relacionar con la realidad,

METODOLOGÍA

reflexionar y mejorar la conceptualización. Por tanto, todas estas características quedan fuera del marco de los textos a estudiar.

Las partes expositivas del contenido son: títulos, subtítulos, contenido en forma de texto, fórmulas, notas al pie, notas al margen, pies de figuras, pies de tablas, pies de imágenes, pies de gráficos, resumen y para saber más. Los títulos y subtítulos no forman en realidad parte del discurso, sino que sirven para estructurar el texto. Además todos los pies de figuras, tablas, imágenes y gráficos no son expositivos, sino que algunos se limitan a sólo describir la figura, tabla, imagen o gráfico a la que se refieren por lo que tampoco se consideran en el análisis.

Por tanto, se seleccionan para el OCR el contenido en forma de texto, fórmulas, notas al pie y al margen, resumen, para saber más y pies de figuras, tablas, imágenes o gráficos de carácter expositivo. En el ejemplo del Anexo B se han eliminado las imágenes, los gráficos, los pies de las figuras 2, 5, 6 y 7, los títulos y subtítulos.

Tras obtener un documento reconocible para un procesador de textos se debe revisar el texto y comprobar que efectivamente es una reproducción fiel de la edición original, ya que los programas OCR no son totalmente fiables. Además se revisará que los textos adicionales, tales como notas al margen o al pie y pies expositivos de figuras, tablas, gráficos o imágenes, están colocados inmediatamente después del párrafo desde el cual se hace referencia a éstos. El documento resultante se guarda en formato *.doc como una reproducción fiel de la edición impresa original, y se seguirá trabajando con una copia.

Seguidamente se procede a preparar el texto obtenido (documento con formato *.doc) para su posterior procesamiento en la aplicación informática PAFE.

Es decir, se acondiciona el texto de modo sistemático para que la aplicación del programa se realice correctamente. Durante este proceso no se va a perder información ni se va a manipular el contenido del texto, sino que se modificará desde un punto de vista formal.

En esta fase de preparación inicial del texto se usará un procesador de textos, en nuestro caso Microsoft Word, para realizar las siguientes acciones:

1. Comprobación de que detrás de cada punto y seguido aparece un espacio, por ejemplo: “palabra. (espacio) Palabra”. Este proceso puede hacerse rápidamente utilizando las herramientas de Microsoft Word: Buscar y Reemplazar. Si antes del punto y seguido hay un número, separar éste del punto y seguido con un espacio, por ejemplo: “número (espacio) . (espacio) Palabra”, ya que de otro modo el programa PAFE confunde dicho punto y seguido con una coma matemática y lo elimina. Además se deben eliminar todos los puntos que aparecen detrás de una abreviatura, puesto que el programa PAFE lo considera como un punto y seguido.
2. Comprobación de que entre el punto y aparte y el marcador de salto de línea no hay ningún espacio, así como que antes de estos marcadores siempre aparece un punto. Se facilita esta acción pulsando el botón “Mostrar todo” o “Mostrar y Ocultar” del menú “Párrafo”.
3. Comprobación de que al punto final del texto le sigue un salto de línea, como se puede ver en la *Figura 10*.
4. Adaptación del texto para el procesado:
 4. 1. Comprobar que ninguna palabra exceda los 25 caracteres. Si no acortarla para cumplir con el límite mencionado.

METODOLOGÍA

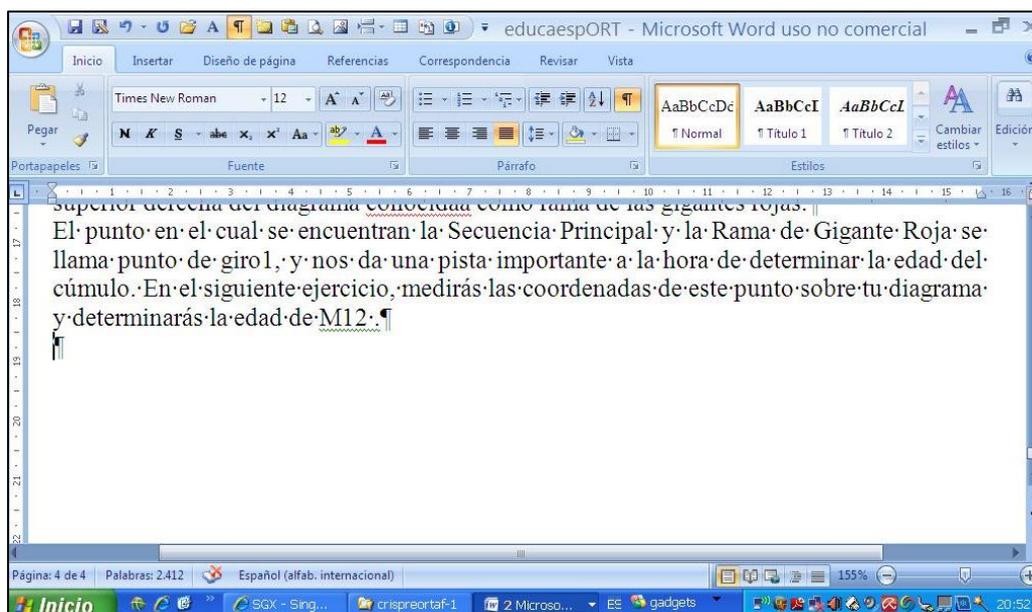


Figura 10.- Último párrafo del texto-ejemplo con salto de línea. Se han resaltado los espacios y saltos de línea con la ayuda de *Word 2007*.

4. 2. En el texto transformado posterior, producido por el programa informático PAFE, todas las letras están en mayúsculas. Si aparece la misma grafía con distinto significado en mayúsculas y en minúsculas se debe marcar una de ellas. Así por ejemplo, en el estudio de caso aparece el acrónimo de la institución científica *ESO* que debe ser distinguida del demostrativo *eso*; por lo tanto marcamos a la institución añadiéndole un número (*1*) y quedaría como *ESOI*, mientras que el demostrativo quedaría como *ESO*. Ver texto original en el Anexo B.
4. 3. Los puntos suspensivos y etc. deben sustituirse por otra palabra, por ejemplo etcétera, para no crear conflictos en la interpretación de los marcadores gramaticales.
4. 4. Comprobar los signos de puntuación antes y después de los signos de interrogación y admiración. Añadir un punto y seguido en el caso en que no lo tuviesen y fuese procedente. Por ejemplo es distinto tener una

pregunta en medio de una oración, que una pregunta que haga las veces de oración:

- *Pregunta ¿qué es? cuando lo veas.* En este caso el signo de interrogación no implica un punto y seguido.
 - *¿Se hace así? Pregúnta siempre.* En este caso el signo de interrogación implica un punto y seguido. Habría que añadirlo manualmente en la preparación del texto.
4. 5. Quitar el punto de las abreviaturas, ya que es confundido como punto y seguido por el programa.
 4. 6. Sustitución del punto o de la coma matemática. Por ejemplo 27.2 se sustituye por $27coma2$.
 4. 7. Si una frase acaba en número, añadir un espacio entre el número en el que acaba la frase y el punto y seguido siguiente.
 4. 8. Separación de los factores de una ecuación matemática para que sean analizados por separado. Por ejemplo $F=ma$ por $F = m a$.
 4. 9. Modificación de superíndices y subíndices matemáticos. El texto se guarda en formato **.txt* para poder ser usado por el programa y este formato no distingue superíndices ni subíndices. Se deben señalar cuando sea procedente, como en las operaciones exponenciales, por ejemplo $E=mc^2$ por $E = m c a2$. Hay subíndices que no hace falta señalar, es el caso de m_B (nombre de una variable) cuya transformación al formato **.txt* produce la grafía mB que es fácilmente identificable.
 4. 10. Sustitución de algunos signos matemáticos y otros símbolos no reconocidos. Por ejemplo $1/3$ por $1dividido3$, m_B-m_V por mb menos mv .

METODOLOGÍA

4. 11. En las ecuaciones matemáticas muchas veces se obvia escribir el símbolo de multiplicar. Aquí lo añadiremos como una “x” siempre que no aparezca explícitamente. Por ejemplo $E = m c a^2$ por $E = m x c a^2$.
 4. 12. Sustitución de los caracteres en alfabeto griego por sus nombres latinos. Por ejemplo σ por *sigma*.
 4. 13. Eliminación de los signos de las viñetas, como es el caso de los puntos rellenos o vacíos que aparecen en los listados.
 4. 14. Sustitución de los guiones largos obtenidos en el escaneado por guiones cortos reconocibles por el programa PAFE.
5. Identificación de las grafías a fin de diferenciarlas según sus contenidos gramaticales y semánticos. Se señalan las grafías que tengan dos o más funciones gramaticales o semánticas en la oración. El criterio arbitrario y fijo que se ha empleado se muestra en la Tabla 4. En la primera columna de la izquierda aparece la función gramatical o semántica que hay que distinguir. En la segunda columna aparece cómo se va a modificar la grafía. En la tercera un ejemplo de grafías a modificar y en la cuarta la modificación realizada. Para la distinción de una posible ambigüedad de una forma gráfica se ha usado el Diccionario de la Real Academia de la Lengua Española (2001). Existen participios que pueden actuar de adjetivo o sustantivo. En estos casos señalaremos el participio para clasificarlo como “no verbo” en la aplicación de Análisis de Frecuencias. Las formas verbales de los verbos auxiliares son señaladas para posteriormente optar a ser clasificados en la base *Eliminar*.
 6. Guardar el documento en formato solo texto con salto de línea (*.txt). Revisar el documento en formato *.txt, puesto que si en el formato *.doc hubiera alguna impureza (marcas tales como puntos o rayas procedentes de la digitalización y

que resultan imposibles de eliminar en el procesador de textos Word) serán transformados en interrogaciones (?) en el formato *.txt. Por tanto, hay que eliminar los símbolos de interrogación de más que aparezcan en el formato *.txt.

Tabla 4

Diferenciación de las grafías con distintas funciones

Función	Añadimos	Ejemplo	Modificación
Adjetivo	+”a”	medido, pesado	medidoa, pesadoa
Verbo	+”v”	forma	formav
Verbo auxiliar	+ “x”	puede	puedex
Sustantivo	+”1”, +”2”, ...	paso, e	paso1, el
“un, uno, una, unos, unas” con significado numérico	+”aa” (en “un”), “+a”	uno	unoa

7. Puede darse el caso de que la digitalización produzca el mismo carácter con distinto formato, siendo imposible la unificación de formato en el procesador de textos Word. Es el caso del símbolo de la multiplicación en el estudio de caso, que se digitaliza con distinto formato hacia el final del documento ($T_{estrella} = Icoma0 \times 10^{a4} K$). Se puede optar por borrar el carácter con distinto formato en el procesador de textos en el archivo *.doc o *.txt y ponerlo con el formato general del resto del texto. Curiosamente el archivo *.txt conserva en este caso concreto la diferencia de formatos de carácter, lo cual genera posteriormente la duplicidad del símbolo en el análisis con el PAFE.
8. Hacer un listado de los cambios para hacerlos siempre de la misma manera en todos los textos.

Las palabras unidas por guiones, como es el caso de *Hertzprung-Russell*, son entendidas por la aplicación PAFE como si se tratara de una sólo palabra, sin guiones. En el ejemplo que se acaba de poner PAFE suprime el guión y entiende

METODOLOGÍA

que se trata de la palabra *HertzsprungRussell*. Para esta investigación se considerará que las palabras unidas por un guión corto constituyen una única UL.

Un ejemplo de la simbología matemática eliminada o no reconocida por el programa se da en la Tabla 5, donde se indica con qué grafía se sustituye. La primera columna de la izquierda muestra el símbolo matemático que aparecía integrado en el texto. La segunda columna presenta el cambio hecho y la tercera el significado del símbolo matemático. Se trata de la simbología matemática del ejemplo del Anexo B (ver capítulo 4). Si bien el símbolo % es reconocido por la primera aplicación del programa (*Análisis de frecuencias*), no es reconocido por la segunda (*Estudio de entorno*) por lo que se ha incluido en la Tabla 5.

Tabla 5

Ejemplo de simbología matemática eliminada o no reconocida por el programa PAFE

Símbolos matemáticos	Se cambia a	Significado
...	hastam	hasta más o menos
σ	sigma	constante de Stefan-Boltzmann
π	pi	3,14...
\propto	proporcional	proporcional
/	dividido	dividido
-	menos	menos
%	porciento	tanto por ciento

A este texto modificado se le llama texto preparado (ver Anexo C). En la Tabla 6 se muestra un ejemplo de la diferencia entre el texto original y el preparado.

Los textos se han coloreado para mostrar las diferencias entre ambos:

 En rojo los títulos que se suprimieron.

 En verde los participios con función de adjetivo (aplicación del punto 5).

 En azul las adaptaciones del texto para su procesado (aplicación del punto 4).

- En rosa sustantivos de igual grafía que una forma verbal (aplicación del punto 5).
- En marrón se señala “uno” cuando es adjetivo numeral (aplicación del punto 5).
- En naranja los verbos auxiliares (aplicación del punto 5).

Tabla 6

Ejemplo de diferencias entre el texto original y el preparado

Texto original, *.doc	Texto preparado, *.doc
La combustión del hidrógeno	
<p>La luz emitida por la mayoría de las estrellas es un subproducto del proceso de fusión termonuclear que se produce en el núcleo interno de las estrellas. Una estrella normal como el Sol se compone de aproximadamente un 74% de hidrógeno y un 25% de helio, siendo el restante 1% una mezcla de elementos más pesados.</p> <p>...</p> <p>La masa de un átomo de helio es sólo el 99.3% de la masa de los cuatro núcleos originales de hidrógeno. El proceso de fusión convierte el 0.7% residual de la masa en energía — mayoritariamente en forma de luz. La cantidad de energía se puede calcular usando la famosa ecuación de Einstein $E = Mc^2$. Ya que c^2 es un número muy grande, esta ecuación implica que incluso una pequeña cantidad de materia puede convertirse en una formidable cantidad de energía. El 0.7% residual de la masa de los cuatro núcleos de hidrógeno involucrados en una sola reacción puede parecer pequeño, pero cuando se considera el número total de reacciones de todo el proceso de fusión, hay implicada una masa total (y por tanto energía) considerable.</p>	<p>La luz emitidaa por la mayoría de las estrellas es un subproducto del proceso de fusión termonuclear que se produce en el núcleo interno de las estrellas. Una estrella normal como el Sol se compone de aproximadamente un 74 por ciento de hidrógeno y un 25 por ciento de helio, siendo el restante 1 por ciento una mezcla1 de elementos más pesadosa.</p> <p>...</p> <p>La masa de unaa átomo de helio es sólo el 99coma3 por ciento de la masa de los cuatro núcleos originales de hidrógeno. El proceso de fusión convierte el 0coma7 por ciento residual de la masa en energía- mayoritariamente en forma1 de luz. La cantidad de energía se puedex calcular usando la famosa ecuación de Einstein $E = M \times c \ a2$. Ya que $c \ a2$ es un número muy grande, esta ecuación implica que incluso una pequeña cantidad de materia puedex convertirse en una formidable cantidad de energía. El 0coma7 por ciento residual de la masa de los cuatro núcleos de hidrógeno involucradosa en unaa sola reacción puedex parecer pequeño, pero cuando se considera el número total de reacciones de todo el proceso de fusión, hayx implicadaa una masa total (y por tanto energía) considerable.</p>

METODOLOGÍA

Los textos se analizan mediante aplicaciones informáticas, una de las cuales, PAFE, ha sido desarrollada en el GICEC para dicho fin específico. A continuación se describe cómo utilizar la aplicación informática PAFE dentro de la metodología a usar.

3. 2. 2.- La aplicación informática: PAFE

La herramienta que se usa para el tratamiento textual es el programa informático de Análisis de Frecuencias y Estudio de Entorno, PAFE¹ (Ceballos et al., 2000; Pérez García & Pérez Ceballos, 1997), desarrollado por el grupo GICEC para el entorno Windows. Los requisitos mínimos de ejecución son un procesador Pentium a 100 MHz, 16 Mb de RAM, 15 Mb de espacio libre y una resolución de 800x600 píxeles. El programa se divide en dos partes consecutivas y dependientes: Análisis de Frecuencias y Estudio de Entorno. Estos programas sirven para analizar los términos que aparecen en los textos seleccionados, ya preparados como se ha indicado en el apartado 3. 2. 1.

El procesamiento del texto preparado se hace en dos fases en estas dos partes del PAFE. En la primera fase se hace un análisis de frecuencias y en la segunda un estudio de entorno.

¹ **Dirección:** Dr. Jesús Pérez Ceballos, **verificación:** D. Ángelo Galotti, **versión 2.0:** D. David Pérez García.

3. 2. 2. 1.- Análisis de frecuencias con PAFE: lematización

En esta fase se lleva a cabo la lematización del texto preparado y se obtiene un texto lematizado. Para ello usaremos el programa Análisis de Frecuencias del PAFE, donde deberemos clasificar cada una de las palabras y UL en cuatro bases de datos diferentes llamadas *Eliminar*, *Cambiar a*, *Verbos* y *Dejar*. Los criterios a seguir en la clasificación de las palabras son los siguientes:

- Los sustantivos y adjetivos en su forma masculina singular irán a la base de datos *Dejar* (ver *Figuras 123 y 124* en el Anexo D). En esta base de datos entran todas las formas que no se cambian. Por lo tanto también formarán parte de esta base sustantivos en género femenino, cuyo género masculino no exista y UL tales como las resultantes de la transformación del lenguaje simbólico en discursivo, como *porciento* en la Tabla 7. En la columna de la derecha de la Tabla 7 se muestran las palabras cuya forma se deja intacta correspondientes al primer párrafo del ejemplo expuesto en la Tabla 6. Nótese que se incluye aquí el lenguaje matemático que no ha sido necesario transformar, como 74, y que forma parte del discurso a analizar. También se incluirán en esta base los adverbios. Las cuatro columnas de la Tabla 7 muestran los nombres de las cuatro bases de datos.
- Los sustantivos y adjetivos que se encuentren en género femenino, y cuyo género masculino exista sin que cambie su significado, o se encuentren en número plural se irán a la base de datos *Cambiar a*. En esta base de datos entran todas las formas no verbales que se cambian a su forma en número singular y género masculino si lo hubiera. En la segunda columna desde la izquierda de la

METODOLOGÍA

Tabla 7 se muestran las palabras del primer párrafo del ejemplo expuesto en la Tabla 6 que se deben cambiar al masculino singular. En este caso también hay una palabra que se debe cambiar a su forma en femenino singular (*estrella*), puesto que no existe su forma masculina. En esta base de datos se introduce la forma lematizada que sustituirá a la forma original del texto preparado. Por ejemplo, en el caso de *emitidaa*, además de clasificarla en la base *Cambiar a* se introducirá su forma lematizada *emitidoa* que sustituirá a la primera en el texto lematizado (ver *Figuras 124* y *126* en el Anexo D). La base de datos *Cambiar a* copia la forma lematizada en la base de datos *Dejar*.

Tabla 7

Clasificación de las palabras del primer párrafo del texto preparado mostrado en la Tabla 6

Eliminar	Cambiar a	Verbos	Dejar
la	emitidaa	es	luz
por	estrellas	produce	mayoría
de	elementos	compon	subproducto
las	pesadosa	e	proceso
una		siendo	fusión
un			termonuclear
del			núcleo
que			interno
se			estrella
en			normal
el			Sol
como			aproximadamente
y			74
			porciento
			hidrógeno
			25
			helio
			restante
			1
			mezcla 1
			más

- Todas las formas verbales, incluido el infinitivo, participio y gerundio, se clasificarán dentro de la base de datos *Verbos*. En esta base de datos se

introduce la forma lematizada que sustituirá a la forma original del texto preparado. Por ejemplo, en el caso de *es*, además de clasificarla en la base *Verbos* se introducirá su forma lematizada *ser* que sustituirá a la primera en el texto lematizado (ver *Figuras 123 y 124* en el Anexo D). No entran los participios que estén actuando como adjetivos, pues previamente se habrán señalado en la preparación del texto. Todas las formas verbales incluidas se cambian a su forma en infinitivo en el texto lematizado. Las formas verbales de los verbos auxiliares (ver siguiente apartado), previamente señaladas con una x, se transforman a su forma infinitiva añadiéndole una x para distinguir la función de verbo auxiliar. Se hace necesario incluir las formas infinitivas ya que el programa ofrece posteriormente la posibilidad de hacer un estudio incluyendo los verbos o excluyéndolos. En la segunda columna desde la derecha de la Tabla 7 se muestran las palabras clasificadas como *Verbos* del primer párrafo del ejemplo expuesto en la Tabla 6. Nótese que los participios usados como adjetivos ya fueron clasificados en *Cambiar a* o en *Dejar*, y no entran en esta base *Verbos*.

- Las palabras funcionales (ver Análisis sin palabras funcionales en capítulo 4) y las formas de los verbos auxiliares se pueden clasificar en la base de datos *Eliminar* (*Figura 123*, Anexo D). Las palabras funcionales son todas aquellas palabras necesarias gramaticalmente para la correcta construcción de la frase, pero que no aportan ningún significado o que sólo matizan el significado de otra palabra. En la primera columna de la izquierda de la Tabla 7 se muestran las palabras clasificadas en *Eliminar* del párrafo del ejemplo expuesto en la Tabla 6. Como se menciona en las cuestiones abiertas, esta base de datos podría ser usada para eliminar cualquier tipo de palabras que se considere no van a aportar

METODOLOGÍA

nada a la investigación. En esta tesis se hace una prueba al respecto en el capítulo 4. La supresión de palabras para resaltar las que son representativas de un texto se hace para la construcción automática de tesauros (Velasco, Díaz, Lloréns, de Amescua, & Martínez, 1999).

Cada vez que se clasifica una palabra en una de las bases de datos, el programa es capaz de reconocer la misma palabra cuando aparece más adelante en el texto. Es decir, cada palabra se clasifica una sólo vez y el programa no vuelve a preguntar por ella.

En textos con un alto número de palabras se asegura su correcta clasificación señalándolas sobre el texto preparado impreso, esto es, se marcan las que se modificarán y se tachan las que se van a eliminar. De esta manera se comprueba que el texto transformado que sale tras aplicar el programa ha sido correctamente lematizado. Hay que decir que una vez introducido el texto preparado y clasificadas las palabras el único error posible de la metodología estaría en este último paso. Podría existir un posible error del investigador al clasificar las palabras, el cual se minimiza comprobando que la clasificación manual es la que luego se metió en el programa. Esta clasificación en el texto preparado impreso permite a la vez encontrar posibles errores en la preparación del texto. El resultado de la clasificación manual para el primer párrafo del texto del ejemplo de la Tabla 6 se muestra en la *Figura 11*. En ésta se han tachado todas las palabras que se clasificarán como *Eliminar*. Se han dejado intactas las palabras que irán a la base de datos *Dejar*. Y se han subrayado con una línea recta los verbos que irán a la base *Verbos*, y con una línea ondulada aquellas que serán modificadas mediante la base *Cambiar a*.

La luz emitida por la mayoría de las estrellas es un subproducto del proceso de fusión termonuclear que se produce en el núcleo interno de las estrellas. Una estrella normal como el Sol se compone de aproximadamente un 74 por ciento de hidrógeno y un 25 por ciento de helio, siendo el restante 1 por ciento una mezcla de elementos más pesados

Figura 11.- Clasificación en texto preparado impreso de las palabras del primer párrafo del ejemplo de la Tabla 6.

3. 2. 2. 1. 1.- Clasificación de palabras

Los lingüistas clasifican las palabras en dos grandes categorías: **palabras léxicas**, que poseen significación denotativa o conceptual, y **palabras funcionales**, que carecen de contenido descriptivo y actúan como instrumentos (relación, determinantes, cuantificadores,...) (Escandell Vidal & Leonetti, 2000; Morera, 2007). Esta categorización es clave a la hora de investigar los conceptos presentados a través de los textos, ya que los mismos estarán representados por palabras léxicas.

De acuerdo con Alarcos Llorach (2009) existen cuatro clases de palabras autónomas o independientes, es decir, con contenido semántico: sustantivo, adjetivo, adverbio y verbo. Estas cuatro clases de palabras serán consideradas en esta investigación como palabras léxicas.

Las clases de palabras definidas por Alarcos Llorach (2009) y Gómez Tórrego (2007) son las siguientes:

METODOLOGÍA

- Sustantivo: toda palabra capaz de cumplir la función de sujeto explícito o de objeto directo sin necesidad de ningún otro elemento.
- Adjetivo: funcionan como adyacentes del sustantivo o aparecen solos cumpliendo la función de atributo. Hay dos tipos de adjetivos: los calificativos que admiten cualquier posición respecto del sustantivo y los determinativos que exigen estar antepuestos (*este, mi, un, algún...*). Los determinativos a su vez se agrupan en demostrativos, posesivos, numerales e indefinidos:
 - Demostrativos: muestran los objetos señalando su situación respecto de determinada persona, y no requieren la aparición del artículo para su sustantivación. Son: *este, ese, aquel, esta, esa, aquella, estos, esos, aquellos, estas, esas, aquellas, tal, tales*. Pueden actuar de sustantivos (pronombres demostrativos): *éste, ése, aquél, ésta, ésa, aquélla, éstos, éstos, aquéllos, éstas, éstas, aquéllas, esto, eso, aquello*.
 - Posesivos: cumplen la función de atributo de un núcleo verbal y adyacente del sustantivo. Son: *mío, mía, míos, mías, tuyo, tuya, tuyos, tuyas, suyo, suya, suyos, suyas, nuestro, nuestra, nuestros, nuestras, vuestro, vuestra, vuestros, vuestras*. También están los átonos: *mi, tu, su, mis, tus, sus*.
 - Numerales: algunos pueden actuar como sustantivos (cardinales). Los ordinales, partitivos (como doceavo) y multiplicativos (como triple) siempre funcionan de adjetivos.
 - Indefinidos: también pueden actuar de sustantivos. Son: *alguien, algo, nadie, nada, quienquiera, quienesquiera, uno, una, unos, unas, algún, alguna, algunos, algunas, cualquier, cualquiera, cualesquiera, más, menos, mucho, mucha, muchos, muchas, poco, poca, pocos, pocas, todo,*

toda, todos, todas, mismo, ningún, ninguno, ninguna, ningunos, ningunas, misma, demás, otro, otra, otros, otras, varios, varias, cierto, cierta, ciertos, ciertas, bastante, bastantes, cada.

- Adverbio: son adyacentes de un verbo, adjetivo u otro adverbio.
- Verbo: funcionan como núcleo de la oración.
- Verbos auxiliares: aquellos que se unen al participio, gerundio e infinitivo formando una perífrasis verbal. El verbo auxiliar ha perdido su significado total o parcialmente. Ejemplos de verbos que pueden actuar como auxiliares son:
haber, ser, estar, acabar de, dejar de, tener, llegar a, ir a, empezar a, ponerse a, echarse a, explotar a, soler, volver a, andar, seguir, llevar, tener que, haber de, haber que, deber, poder, deber de, querer, venir a.
- Pronombres relativos, interrogativos y exclamativos: *que, cual, cuales, quien, quienes, cuyo, cuya, cuyos, cuyas, como, donde, cuando, cuanto, cuanta, cuantos, cuantas, cuál, cuáles, qué, cuán, cuándo, cómo, dónde, quién, quiénes, cuánto, cuánta, cuántos, cuántas.*
- Pronombres personales: con la función de los sustantivos. Son: *yo, nosotros, nosotras, tú, vosotros, vosotras, vos, él, ella, ello, ellos, ellas, me, nos, te, os, lo, la, le, los, las, les, se, mí, ti, sí, usted, ustedes, conmigo, contigo, consigo.*
- Artículo, llamado definido o determinado: *el, la los, las, lo.*
- Preposiciones: *a, ante, bajo, con, contra, de, desde, durante, en, entre, hacia, hasta, mediante, para, por, según, sin, sobre, tras.* Incluiremos aquí también las antiguas *cabe, so.*
- Conjunciones: *y, ni, o, u, e, pero, sino, empero, conque, luego, que, porque, si.* Existen locuciones que actúan de conjunciones pero no serán tenidas en cuenta como tales locuciones en la investigación sino como palabras separadas.

METODOLOGÍA

Ejemplos: *de forma que, de manera que, así que, de modo que*. Estas locuciones conjuntivas han sido detectadas en otros trabajos del GICEC como candidatas a meter en la base *Eliminar* (Anexo E). En las cuestiones abiertas se considera la posibilidad de hacer un estudio de caso donde se eliminen estas locuciones.

- Interjecciones: se limitan a imitar ruidos de la realidad o a expresar sentimientos del hablante. Dadas las características de los textos a analizar es poco probable que aparezcan.

Si bien los lingüistas discuten sobre la clasificación de una clase de palabras en funcionales o léxicas, aquí se adopta la siguiente:

- Palabras léxicas: sustantivos; adjetivos calificativos, numerales e indefinidos (excepto *un, uno, una, unos y unas*); adverbios y verbos.
- Palabras funcionales: adjetivos demostrativos, posesivos y los indefinidos *un, uno, una, unos y unas*; artículos; pronombres personales, relativos, interrogativos y exclamativos; preposiciones; conjunciones; interjecciones; contracciones (formadas por preposiciones y artículos: *del, al*) y verbos auxiliares (ver listado en el anexo F).

Los adjetivos indefinidos se consideran, con la excepción de *un* y sus variantes, como palabras léxicas pues muchos dan información sobre cantidades.

3. 2. 2. 1. 2.- La aplicación “Análisis de Frecuencias” del PAFE

En esta primera fase se entra mediante una clave de acceso (anexo D, *Figura 118*) en el programa Análisis de Frecuencias del PAFE y se selecciona *Fichero: Abrir* en el Menú Principal (anexo D, *Figura 119*). Una vez ahí, se abre el fichero de extensión *.txt que se quiere procesar.

La primera vez que se trabaja con el fichero se rellena una plantilla de identificación (anexo D, *Figura 120*) con los datos del texto estudiado (Tabla 8). En el caso de los libros de texto se tomará como año de publicación el del depósito legal dado que en muchos de ellos no aparece el mismo.

Tabla 8

Ejemplo de plantilla usado por el programa PAFE

Descriptor	Datos
Código	FESTESE001V
Autor(es)	Lars Lindberg Christensen y Richard West
Título	Medida de la distancia y la edad de un cúmulo globular de estrellas
Editorial / Revista	ESA y ESO, El Centro de Información de la Agencia Espacial Europea del Hubble
Ciudad (País)	Garching bei München (Alemania)
Año	2002
Nº Edición / volumen, nº (o mes)	2ª
Idioma	Español
ISBN /ISSN	
Nº páginas del libro	32
Nº página inicial y final sometida al análisis	5-10
Localización	www.astroex.org
Descriptor del tema tratado	estrellas, física de estrellas, evolución estelar, cúmulos estelares, cúmulo globular

METODOLOGÍA

El código es el nombre que el programa dará a todos los archivos generados. De este modo, con la plantilla se puede identificar fácilmente los textos estudiados. El código se construye como se muestra en la *Figura 12*.

F	E	S	T	E	E	3	S	0	0	1	V
---	---	---	---	---	---	---	---	---	---	---	---

Figura 12.- Codificación alfanumérica de un texto.

- **Posición 1:** Una letra inicial para el área. Por ejemplo: F de física.
- **Posiciones 2, 3, 4:** Tres letras que identifican el tema. En este caso: EST referido a estrellas.
- **Posición 5:** Una letra que indica el idioma en el que está escrito el texto. Por ejemplo: E de español.
- **Posición 6:** Una letra que concreta el género: divulgativo (D), científico (C) o educativo (E). En este caso: E.
- **Posiciones 7 y 8:** Dos letras para el subgénero o nivel de subgénero (en el caso el género educativo). Por ejemplo: 3S de 3º de Secundaria.
- **Posiciones 9, 10 y 11:** Tres números para identificar el texto. En este caso: 001
- **Posición 12:** La letra C si no se consideran verbos en el análisis y la letra V si se consideran verbos en el mismo. Por ejemplo: V

Los géneros y subgéneros que se van a considerar y las abreviaturas a usar en la codificación, entre paréntesis, se muestran en la Tabla 9. El género divulgativo se divide en tres subgéneros:

- **Divulgación alta**, es decir, divulgación entre pares. Es el caso de las revistas *Mundo científico e Investigación y ciencia*. Se requiere un nivel cultural de universitario de ciencias para poder entender los textos.

- **Divulgación media**, es decir, divulgación a un público interesado por los temas con un nivel cultural medio. Es el caso de las revistas *Newton*, *Discovery*, *Ciencia y vida*.
- **Divulgación baja**, es decir, divulgación a público general. Es el caso de los periódicos y de la revista *Muy interesante*.

Tabla 9

Clasificación en géneros, subgéneros y niveles de subgéneros de los textos a estudiar

Género	Subgénero	Nivel de subgénero
Divulgativo (D)	Alta (DA)	
	Media (DM)	
	Baja (DB)	
Científico(C)	General (CG)	
	Especializada (CE)	
Educativo (E)	Primaria (EP)	1° (1P)
		2° (2P)
		3° (3P)
		4° (4P)
		5° (5P)
		6° (6P)
	Secundaria (ES)	1° (1S)
		2° (2S)
		3° (3S)
		4° (4S)
	Bachillerato (EB)	1° (1B)
		2° (2B)
		Otros (EO)

En el género científico se consideran dos subgéneros:

- **Científico general:** las revisiones temáticas y revistas de revisiones, como *Annual Review of Astronomy and Astrophysics*.
- **Científico especializado:** las investigaciones publicadas en revistas indexadas.

En el género educativo se consideran los subgéneros correspondientes a las distintas etapas educativas y sus niveles.

METODOLOGÍA

Tras rellenar la plantilla de la Tabla 8 y pulsar *Aceptar*, ésta se guarda automáticamente como un archivo con la extensión **.PLA*, mientras que el texto aparece en la pantalla y podemos comenzar a analizarlo. Para ello, primero hay que clasificar las palabras en las distintas bases de datos del PAFE.

Hay que recordar que en el subprograma del PAFE para el análisis de frecuencias se clasifican las palabras en cuatro grupos o bases de datos diferentes, con el objetivo de obtener un texto lematizado. El proceso se muestra en la *Figura 123* del Anexo D:

1. En el menú principal, en la opción *Procesar*, seleccionar *Configuración* (anexo D, *Figuras 121* y *122*). La aplicación permite analizar el texto usando o no la base *Verbos*. Por defecto, la aplicación analiza el texto incluyendo la base *Verbos*, que además es lo que se hace en esta metodología.
2. En el menú principal, en la opción *Procesar*, seleccionar *Procesar Documento Actual* (anexo D, *Figura 121*)
3. El programa pide que se clasifique cada una de las palabras en una de las cuatro bases de datos, teniendo en cuenta si deben ser eliminadas, cambiadas, incluidas como verbos o mantenidas en la forma en que se encuentran (ver apartado 3. 2. 2. 1. 1. de este capítulo). Las palabras quedarán incluidas en las bases de datos *Eliminar*, *Cambiar a*, *Verbos* y *Dejar*, respectivamente (anexo D, *Figura 123* y *124*). Estas bases de datos pueden ser consultadas, revisadas y modificadas en caso de error (anexo D, *Figuras 125* y *126*).

Una vez las palabras están clasificadas, es decir, una vez el texto se ha lematizado o transformado, éste aparecerá en la pantalla del programa (anexo D, *Figura 127*).

En el texto lematizado (Anexo G) todas las palabras con el mismo significado tienen la misma forma gráfica. Tal y como se ha explicado el método más rápido para evitar tener que clasificar las palabras una y otra vez consiste en leer el texto preparado decidiendo cómo debería quedar el texto lematizado. Una vez que éste se ha obtenido se compara con la clasificación manual; en caso de error se corrigen las bases de datos y se vuelve a procesar el texto preparado. En el caso del texto ejemplo las diferencias entre el texto preparado y el lematizado pueden verse en la Tabla 10.

Tabla 10

Texto preparado a la izquierda y lematizado a la derecha

Texto preparado, extensión *.txt	Texto lematizado, extensión *.at1
La luz emitidaa por la mayoría de las estrellas es un subproducto del proceso de fusión termonuclear que se produce en el núcleo interno de las estrellas. Una estrella normal como el Sol se compone de aproximadamente un 74 por ciento de hidrógeno y un 25 por ciento de helio, siendo el restante 1 por ciento una mezcla1 de elementos más pesadosa.	EL LUZ EMITIDOA POR EL MAYORÍA DE EL ESTRELLA SER UNO SUBPRODUCTO DEL PROCESO DE FUSIÓN TERMONUCLEAR QUE SE PRODUCIR EN EL NÚCLEO INTERNO DE EL ESTRELLA * UNO ESTRELLA NORMAL COMO EL SOL SE COMPONER DE APROXIMADAMENTE UNO 74 PORCIENTO DE HIDRÓGENO Y UNO 25 PORCIENTO DE HELIO SER EL RESTANTE 1 PORCIENTO UNO MEZCLA1 DE ELEMENTO MÁS PESADOA *

Los puntos seguidos se sustituyen por una estrella, y los puntos y aparte por dos estrellas seguidas. El párrafo que aparece en la Tabla 10 es un párrafo incompleto y por esta razón no finaliza con dos estrellas en el texto lematizado.

Tras obtener el texto lematizado se procede a obtener las frecuencias de los textos preparado y lematizado (ver Figuras 127 y 128) mediante la opción *Estadística* de la barra del menú principal. Las palabras aparecen listadas por orden

METODOLOGÍA

lexicométrico, estando los números antes de la “a” y las palabras que comienzan por vocal con tilde detrás de la “z”.

El análisis de frecuencias produce los siguientes archivos:

- Código.txt con el texto preparado.
- Código.at1 con el texto lematizado.
- Código.PLA con la plantilla.
- Código.f1 con las frecuencias absolutas y relativas del texto preparado, número de palabras en el texto preparado y número de palabras diferentes del mismo.
- Código.f2 con las frecuencias absolutas y relativas del texto lematizado, número de palabras en el texto lematizado y número de palabras diferentes del mismo.

Las frecuencias obtenidas para los dos tipos de textos difieren debido a la lematización, pudiéndose ver un ejemplo en la Tabla 11 (el listado completo se puede ver en la tercera y cuarta columna de los anexos H e I). La frecuencia relativa nos proporciona una medida de la importancia de uso de la palabra o del lema en el caso del texto lematizado. Las palabras/lemas de la Tabla 11 se encuentran por orden lexicométrico.

También se calcula la frecuencia acumulada a partir de la frecuencia absoluta usando el programa *Excel* (ver quinta columna de los anexos H e I). La frecuencia acumulada sirve para distinguir distintas franjas de palabras, es decir, palabras agrupadas según dicha frecuencia. En el gráfico de frecuencias acumuladas del texto preparado se pueden distinguir tres franjas (obtenidas en estudios anteriores no publicados del GICEC):

- Una franja alta con palabras de alta frecuencia, que cubre aproximadamente un 40% de la frecuencia acumulada.

- Una franja media con palabras que tienen frecuencias medias, cuyo límite superior está en un 75% de la frecuencia acumulada.
- Una franja baja con palabras que tienen las frecuencias más bajas, que cubre el restante 25 % de la frecuencia acumulada y que contiene más del 50% de las palabras diferentes.

Tabla 11

Fragmentos de las 20 frecuencias más altas en los textos preparado y lematizado, producidas por PAFE

Fragmento del listado de frecuencias, extensión *.f1				Fragmento del listado de frecuencias, extensión *.f2			
OL	Palabra	n _i	f _i	OL	Lema	n _i	f _i
1	DE	222	92.346	1	EL	273	113.561
2	LA	128	53.245	2	DE	222	92.346
3	EL	67	27.87	3	ESTRELLA	76	31.614
4	EN	64	26.622	4	UNO	74	30.782
5	SE	54	22.463	5	EN	64	26.622
6	ESTRELLAS	48	19.967	6	SE	54	22.463
7	Y	48	19.967	7	Y	48	19.967
8	LAS	41	17.055	8	SER	41	17.055
9	QUE	39	16.223	9	CÚMULO	39	16.223
10	UNA	38	15.807	10	QUE	39	16.223
11	DEL	30	12.479	11	DEL	30	12.479
12	LOS	29	12.063	12	A	28	11.647
13	A	28	11.647	13	ESTO	26	10.815
14	ES	28	11.647	14	NÚCLEO	19	7.903
15	ESTRELLA	28	11,647	15	SU	19	7,903
16	UN	27	11,231	16	MÁS	18	7,488
17	CÚMULOS	23	9,567	17	GLOBULAR	17	7,072
18	MÁS	18	7,488	18	PARA	17	7,072
19	PARA	17	7,072	19	POR	16	6,656
20	CÚMULO	16	6,656	20	TEMPERATU RA	15	6,240

Nota. OL= orden lexicométrico; n_i = frecuencia absoluta; f_i = frecuencia relativa a 1000.

Para el texto preparado del ejemplo podemos ver las franjas señaladas en la *Figura 25*, la Tabla 31 (apartado 4.2 del estudio de caso) y la Tabla 94 (Anexo H).

Asimismo en el gráfico de la frecuencia acumulada del texto lematizado o transformado se distinguen tres franjas:

METODOLOGÍA

- Una franja alta con palabras de alta frecuencia, que cubre aproximadamente un 15-20% de la frecuencia acumulada.
- Una franja media con palabras que tienen frecuencias medias, cuyo límite superior está en un 60% de la frecuencia acumulada.
- Una franja baja con palabras que tienen las frecuencias más bajas, que cubre el restante 25-20 % de la frecuencia acumulada y que contiene más del 70% de las palabras diferentes.

Esto quiere decir que con un 30 % de palabras diferentes se cubre un 75-80 % de la frecuencia acumulada. Ello sugiere que dominando el significado del 30% de las palabras diferentes se podría entender la mayor parte del texto.

Para el texto lematizado del ejemplo podemos ver las franjas señaladas en la *Figura 43*, la *Tabla 43* (apartado 4.3 del estudio de caso) y la *Tabla 95* (Anexo I).

Para determinar los límites de cada franja se aplica el porcentaje arriba mencionado a la suma de todas las frecuencias absolutas (frecuencia acumulada total). La palabra, lema UL a la que corresponde el valor calculado nos indica la frecuencia absoluta o relativa límite de la franja. Es decir, en el caso de la franja media del texto preparado del estudio de caso (*Tabla 31*) se observa que la misma llega hasta la frecuencia acumulada 1803. Esto se corresponde con la palabra *cientos* que tiene de frecuencia absoluta 2. El orden en el listado de frecuencias (ver Anexo H) es lexicométrico, por lo que la frecuencia acumulada 1803 podría haber estado en cualquier otra palabra de frecuencia absoluta 2. Por ello, el grupo de palabras de frecuencia absoluta 2 es el que delimita la franja media y no una sola palabra del mismo.

En la Tabla 12 se muestran algunos ejemplos de los cambios que sufren las palabras en el proceso de lematización. El artículo (*el*) ha quedado el primero en la lista de frecuencias del texto lematizado (Tabla 11). El verbo *formar* se encuentra en la franja de frecuencias medias debido a la lematización de conjugaciones que de otra manera se habrían quedado en la franja baja (Anexo H, Tabla 94). Los lemas pueden sufrir cambios de posición respecto a la frecuencia dentro de la misma franja o incluso cambiar de franja. La lematización resalta la importancia del significado en términos de frecuencias.

Tabla 12

Ejemplos de cambios en frecuencia absoluta entre el texto preparado y el texto lematizado, FESTESE001V

Texto lematizado		Texto preparado	
Lema	n_i	Palabras	n_i
el	273	la	128
		el	67
		las	41
		los	29
		lo	8
núcleo	19	núcleo	14
formar	7	núcleos	5
		forman	4
		formar	2
		formaron	1

Nota. n_i = frecuencia absoluta.

Los datos obtenidos con la aplicación Análisis de Frecuencias permiten además hallar el cociente tipo/caso, así como la densidad léxica. El cociente tipo/caso se calcula como el cociente entre el número de vocablos (tipos) y el número de palabras (casos). Cuando el cociente tipo/caso es 1 indica que no se repite ninguna palabra, mientras que cuanto más cerca de 0 está indica que se repiten más palabras. Es también una medida de la variabilidad léxica. Es decir, da idea del dominio del lenguaje que se debe tener a la hora de entender un texto y por

METODOLOGÍA

tanto la dificultad de lectura que pueden tener. Cuanto más cerca de 1 está el cociente tipo/caso mayor será la riqueza léxica del texto, es decir, más variabilidad léxica hay. El cociente tipo/caso depende del tamaño de la muestra, por ejemplo, cuanto más grande es la muestra más repeticiones habrá de palabras funcionales (Baker, Hardie, & McEnery, 2006). A pesar de que las muestras de texto en esta investigación son de tamaño variable, se usará dicho cociente como medida de la variabilidad léxica.

La densidad léxica (LD) se calcula como el cociente entre el número de palabras léxicas y el número total de palabras (Ure, 1971). La densidad léxica indica la densidad de los textos y por tanto la dificultad de lectura que pueden tener. Es decir, da idea de la cantidad de información que contiene un texto, puesto que mide cómo se empaquetan las palabras léxicas dentro de la estructura gramática (Johansson, 2008). En esta tesis se usará el siguiente cociente para calcular la densidad léxica: $LD = N_{\text{léxicas}}/N$, siendo $N_{\text{léxicas}}$ el número de palabras léxicas y N el número de palabras totales. La densidad léxica se representa en porcentaje, por lo que el resultado del cociente se multiplica por 100.

3. 2. 2. 2.- Estudio de entorno con PAFE: relaciones

Con el texto lematizado accedemos al estudio de entorno del PAFE, donde obtendremos la frecuencia de aparición de unidades léxicas lematizadas (ULL) en un entorno determinado de una dada. Esta frecuencia nos determina y mide las

relaciones entre las ULL. El entorno de una ULL dada está integrado por n ULL a cada lado de la misma, pudiéndosele atribuir un valor diferente de acuerdo a su proximidad. El conjunto de relaciones cuantificadas permite atribuir a cada ULL un valor de capacidad de relación.

Al estudiar el entorno se pueden hacer las siguientes consideraciones:

- ¿Interesa estudiar todas las ULL del texto, o sólo un subconjunto?
- Dada una ULL situada al comienzo o final de una frase ¿se debe considerar dentro de su entorno ULL que están antes o después de un punto y seguido? Es decir, se debe determinar la unidad de contexto.
- ¿Qué valor se le debe dar a la proximidad de distintas ULL a una dada?

La aplicación *Estudio de Entorno* del PAFE se ha desarrollado para tener en cuenta todas estas consideraciones. Para ello la aplicación permite que el investigador pueda determinar las siguientes variables:

- La frecuencia mínima de las ULL a estudiar.
- Las características que definen la extensión del entorno.
- Los valores de entorno, es decir, los valores asignados a cada ULL por proximidad a una ULL origen.

La frecuencia mínima de las ULL a estudiar va a determinar un sistema o conjunto de ULL a analizar. Esta frecuencia mínima se puede dar en términos de la frecuencia absoluta o frecuencia relativa obtenida en el Análisis de Frecuencias del PAFE para el texto lematizado (ver en la Tabla 11 las cuatro columnas de la derecha, correspondientes a la extensión *.f2). En el estudio de caso, mostrado en los anexos G e I, elegir una frecuencia mínima absoluta de 1 implica elegir todas las

METODOLOGÍA

ULL del texto lematizado. Sin embargo elegir una frecuencia mínima absoluta de 2 implica eliminar los hápax en el estudio de las relaciones.

La extensión del entorno en términos de frases (separadas por un punto y seguido) o párrafos (separados por un punto y aparte) permite estudiar las relaciones entre ULL que suelen aparecer juntas en la unidad de contexto (frase o párrafo) y elegir si la extensión es simétrica, para hallar relaciones entre ULL, o asimétrica, para hallar colocaciones en frases (ver apartado 3.2.2.2.2.). En el PAFE la unidad de contexto y la simetría de la extensión se definen dentro de la característica *Fragmento*.

Los valores de entorno permiten determinar cuántas ULL se consideran alrededor de la ULL origen a estudiar, así como el valor que se le da por cercanía a dicha ULL origen. Por ejemplo, en un entorno de 1 se consideran las relaciones con la ULL inmediatamente anterior y/o posterior y en un entorno de 2 se estudiarán las relaciones con las dos ULL inmediatamente anteriores y/o posteriores. Por otro lado es posible tratar la cercanía de las ULL con distintos valores. Ejemplos de entornos se pueden ver en la Tabla 13 donde se presentan entornos alrededor de la palabra *ser* en la frase transformada: EL LUZ EMITIDO A POR EL MAYORÍA DE EL ESTRELLA SER UNO SUBPRODUCTO DEL PROCESO DE FUSIÓN TERMONUCLEAR QUE SE PRODUCIR EN EL NÚCLEO INTERNO DE EL ESTRELLA. En la primera columna de la izquierda se muestran cuatro entornos de la ULL *ser* (señalada en negrita). La segunda columna muestra el valor máximo que tendrá la ULL más cercana a *ser* (variable *Valor máximo* en el Estudio de Entorno). Y la tercera columna indica cómo decrece dicho valor máximo al irse alejando de *ser* (variable *Decremento* en el Estudio de Entorno).

Tabla 13

Ejemplos de valores de entorno

Fragmento	Valor máximo	Decremento
ESTRELLA SER UNO	1	1
EL ESTRELLA SER UNO SUBPRODUCTO	4	2
DE EL ESTRELLA SER UNO SUBPRODUCTO DEL	3	1
DE EL ESTRELLA SER UNO SUBPRODUCTO DEL	6	2

En la aplicación Estudio de Entorno las variables *Valor máximo* y *Decremento* definen conjuntamente el entorno como el número de ULL que se van a considerar alrededor de una ULL origen. En el primer caso que se presenta en la Tabla 13 (ESTRELLA **SER** UNO) se estudia alrededor de la ULL origen *ser* un entorno de 1 a ambos lados; en este entorno el valor que tiene la ULL inmediatamente anterior y posterior es 1. Es decir, la frecuencia de aparición será multiplicada por 1. El último caso de la Tabla 13 presenta un entorno de 3 a ambos lados de *ser*, determinado por un *Valor máximo* de 6 y un *Decremento* de 2. Es decir, el valor de la frecuencia de aparición de la ULL inmediatamente precedente o siguiente se multiplica por 6, el de la segunda se multiplica por 4 y el de la tercera por 2. De este modo se da mayor valor a las ULL más cercanas que a las más lejanas. La aplicación del Estudio de Entorno permite analizar un entorno de máximo 7.

3. 2. 2. 2. 1.- La aplicación “Estudio de Entorno” del PAFE

A la aplicación Estudio de Entorno del PAFE se puede acceder a través del icono del escritorio destinado a tal fin o desde la aplicación *Análisis de Frecuencias*, en la opción *Estadística* del menú principal (anexo D, *Figuras 129 y 127* respectivamente).

Al entrar en el Estudio de Entorno podemos elegir entre:

- **Estudio de entorno**, para comenzar un nuevo estudio de entorno.
- **Navegador**, para acceder a un estudio de entorno ya hecho y guardado en una base de la aplicación.
- **Borrar bases**, para eliminar todas las bases generadas por estudios de entorno anteriores. Los estudios de entorno se pueden volver a repetir a partir del archivo con el texto lematizado o transformado.
- **Salir**, para salir de la aplicación.

Tras elegir la opción de Estudio de Entorno se debe seleccionar el texto lematizado, de extensión *.at1, que se desea estudiar y abrirlo (anexo D, *Figura 130*). Para que la aplicación funcione correctamente, el archivo *.at1 ha de tener ya formados sus correspondientes *.f1 y *.f2, es decir, habría que haber obtenido las frecuencias del texto preparado y transformado mediante la opción *Estadística* de la barra del menú principal en el Análisis de Frecuencias. Una vez seleccionado el texto a estudiar debemos especificar las variables con las que deseamos hacer nuestro estudio. Estas variables son presentadas por el programa de la siguiente manera (anexo D, *Figura 131*):

- **Palabras a procesar:** donde se determina en términos de la frecuencia absoluta o relativa el número de ULL que se van a analizar o sistema.
- **Fragmento:** donde se decide si el entorno está a ambos lados de la ULL o sólo a la derecha (ULL siguientes). En el caso de que el entorno esté a ambos lados hay que especificar la unidad de contexto: si lo está dentro de una frase (punto y seguido) o dentro de un párrafo (punto y aparte). Cuando se elige el fragmento a la derecha la aplicación PAFE automáticamente selecciona como unidad de contexto la frase separada por punto y seguido.
- **Entorno:** donde se determina el valor máximo del entorno y el decremento a usar por proximidad.

Una vez especificadas las variables de entorno, la aplicación analiza y guarda los datos en una nueva base. Al investigador puede interesarle ver todo o parte de los datos del análisis. Por ello, seguidamente la aplicación pregunta por la frecuencia mínima de las ULL a mostrar, y además por el valor mínimo de relación entre dichas ULL que se quiere considerar (anexo D, *Figura 132*). Cada vez que el investigador propone una frecuencia mínima distinta varía el número de ULL que se muestran, es decir, el sistema a visualizar. Posteriormente, y dentro de ese mismo sistema, el valor mínimo de relación se puede ir cambiando en el menú *Opciones* del menú principal. Es decir, dado un sistema de ULL a estudiar, determinado por la frecuencia mínima, se pueden ver cómo son las relaciones variando el valor mínimo de relación.

Tras especificar el valor mínimo de relación se obtiene el listado de ULL del sistema junto con las relaciones que presentan (anexo D, *Figura 133*). En la parte superior se puede ver el número de ULL que forman el sistema definido (*Nº Palabras Visualizadas*). Las ULL se listan por orden lexicométrico. Al pulsar sobre

METODOLOGÍA

el signo más situado a su izquierda se obtiene el listado de ULL con las que se relaciona así como los valores de relación. Además se puede obtener un archivo en formato *.qrp pulsando en el botón inferior *Lista de relaciones* (anexo D, *Figura 134*) y guardando el listado correspondiente. Este archivo contiene la información relativa a las características del sistema en la parte superior, así como los valores de relación entre las ULL. En la parte superior izquierda aparecen las características que definieron el sistema (*Palabras a procesar*), mientras que a la derecha aparecen las características del sistema a visualizar o que interesa ver (*Palabras a visualizar*).

A continuación aparece el listado de ULL del sistema a visualizar junto con las ULL con las que se relacionan. En la primera columna de la izquierda está la ULL alrededor de la cual se estudió el entorno (*Palabra origen*). En la segunda columna aparecen las ULL con las que la ULL origen se relaciona. En la tercera y cuarta columnas aparecen la frecuencia absoluta de relación y la frecuencia relativa de relación respectivamente. En la cuarta columna titulada *Relación relativa a esta lista*, la relación relativa rr_i . Si llamamos ra_i a la relación absoluta de la tercera columna, entonces $rr_i = ra_i / \sum ra_i$. Los datos del archivo *.qrp se pueden exportar a Excel para poder trabajar con ellos matemáticamente. En esta investigación se han exportado los archivos *.qrp a Excel usando el programa OneView (Gnostice Information Technologies Private Limited, 2002).

3. 2. 2. 2.- Colocaciones y uniones de ULL

El primer estudio de entorno se hace con una extensión de entorno igual a 1 y un entorno sólo a la derecha dentro de la frase. Es decir, sólo se va a considerar el lema siguiente al lema a analizar. Con este primer estudio se pretende hallar de una forma lo más automatizada posible aquellos conceptos formados por más de un lema o aquellos lemas que van siempre juntos, es decir, lo que en lingüística de corpus se denominan colocaciones. El sistema a estudiar viene definido por las siguientes características en el programa PAFE (Estudio de Entorno):

Tabla 14

Características del primer Estudio de Entorno: hallando colocaciones

Variable	Valor
Unidades léxicas a procesar	frecuencia mínima (absoluta) =1
Fragmento	a la derecha
Entorno	valor máximo=1, decremento =1
Valor de relación mínima	2

Se trata de hallar qué lemas aparecen con mayor frecuencia juntos. Para ello hay que fijarse en los lemas que contribuyen a valores de relación mayor o igual a 2. No se consideran relaciones menores de 2 para evitar las relaciones con hápax o entre éstos. La lista de relaciones se obtiene en el botón inferior *Lista de relaciones*, de donde se puede grabar un archivo en formato *.grp. Dados dos lemas, de los cuales uno es el origen (*i*), se calcula el cociente entre el valor de relación entre dichos lemas (ra_{ij}) y la frecuencia absoluta del lema origen (n_i , dado por el listado de frecuencias del texto lematizado en la aplicación de Análisis de Frecuencias del PAFE). Los cálculos se hacen usando el programa *Excel 2007* a partir de los datos extraídos del archivo *.grp. Si el cociente es superior a 0.5, significa que más del

METODOLOGÍA

50% de las veces en que aparece el lema origen lo hace relacionado con el otro lema. Cuando el cociente vale 1 aparecen siempre juntos. En esta investigación se considerarán las uniones de aquellas ULL cuyo cociente ra_{ij}/n_i sea mayor o igual a 0.5. De este modo es posible detectar lemas que juntos forman un nuevo significado (como por ejemplo “*secuencia principal*” en la Tabla 54).

El criterio del cociente no siempre es válido para detectar lemas que juntos forman un nuevo significado, puesto que puede darse el caso de que dos lemas que cumplen esta condición previamente conocida por el experto presente un bajo cociente (como “*diagrama hr*” en la Tabla 54). A la hora de proponer uniones entre lemas, interesa que la condición de formar un nuevo significado se refleje en las uniones. Es decir, se hace necesario que el experto revise el listado de relaciones para determinar las propuestas de lemas a unir. La revisión de un listado como el que se presenta en el Anexo J (Tabla 97), es mucho más rápida y eficaz que la revisión del texto completo, donde los lemas a unir están en medio de otros que no se unen.

Una vez obtenido el listado de propuestas a unir, se unen dichos lemas en el texto lematizado y se sustituyen las estrellas por punto y seguido y las dos estrellas por punto y aparte. Para ello se usa el buscador del procesador de textos *Word 2007*. A la vez hay que revisar que ninguna de las uniones exceda los 25 caracteres. Si hubiera alguna unión que excediera los 25 caracteres se borrarían letras hasta conseguir que esté dentro del límite de caracteres establecido. Tras unir los lemas en el texto se vuelve a pasar éste por el PAFE, comenzando por la clasificación de las nuevas uniones en la base *Dejar* y se vuelve a repetir el estudio de entorno mencionado anteriormente hasta que no salgan más propuestas de unión.

De esta manera se obtiene un texto lematizado y transformado por la unión de lemas como el del Anexo K. Un ejemplo de la transformación se muestra en la Tabla 15. En el texto lematizado los lemas aparecen separados, mientras que en el texto transformado ya están unidos. En ambos textos los lemas afectados aparecen en negrita. Dado que el proceso de hallar propuestas de unión se repite hasta que no se encuentren más, se puede dar el caso de llegar a unir varios lemas. Además de las uniones de dos lemas, en el párrafo de la Tabla 15 se puede ver un caso de unión de tres lemas (“*cantidad de energía*”) y otro de unión de seis lemas (“*0coma7 por ciento residual de la masa*”).

Tabla 15

Ejemplo de propuestas de unión aceptadas para el texto preparado

Texto lematizado, extensión *.at1	Texto transformado con palabras unidas, extensión *.txt
EL MASA DE UNOA ÁTOMO DE HELIO SER SÓLO EL 99COMA3 PORCIENTO DE EL MASA DE EL CUATRO NÚCLEO ORIGINAL DE HIDRÓGENO * EL PROCESO DE FUSIÓN CONVERTIR EL 0COMA7 PORCIENTO RESIDUAL DE EL MASA EN ENERGÍA MAYORITARIAMENTE EN FORMA1 DE LUZ * EL CANTIDAD DE ENERGÍA SE PODERX CALCULAR USAR EL FAMOSO ECUACIÓN DE EINSTEIN $E1 = M X C A2$ * YA QUE C A2 SER UNO NÚMERO MUY GRANDE ESTO ECUACIÓN IMPLICAR QUE INCLUSO UNO PEQUEÑO CANTIDAD DE MATERIA PODERX CONVERTIR EN UNO FORMIDABLE CANTIDAD DE ENERGÍA * EL 0COMA7 PORCIENTO RESIDUAL DE EL MASA DE EL CUATRO NÚCLEO DE HIDRÓGENO INVOLUCRADOA EN UNOA SOLO REACCIÓN	EL MASA DE UNOA ÁTOMO DE HELIOSER SÓLO EL 99COMA3 PORCIENTO DE EL MASA DE EL CUATRONÚCLEO ORIGINAL DE HIDRÓGENO. EL PROCESO DE FUSIÓN CONVERTIR EL 0COMA7PCIENTORESIDL MASA EN ENERGÍA MAYORITARIAMENTE EN FORMA1 DE LUZ. EL CANTIDADDEENERGÍA SE PODERX CALCULAR USAR EL FAMOSO ECUACIÓN DE EINSTEIN $E1 = M X$ CA2 . YA QUE CA2 SER UNO NÚMERO MUY GRANDE ESTO ECUACIÓN IMPLICAR QUE INCLUSO UNO PEQUEÑO CANTIDADDE MATERIA PODERX CONVERTIR EN UNO FORMIDABLE CANTIDADDEENERGÍA . EL 0COMA7PCIENTORESIDL MASA DE EL CUATRONÚCLEO DE HIDRÓGENO INVOLUCRADOA EN UNOA SOLO REACCIÓN PODERX PARECER PEQUEÑO PERO CUANDO SE CONSIDERAR EL NÚMERO TOTAL DE REACCIÓN DE TODO EL

METODOLOGÍA

Texto lematizado, extensión *.at1	Texto transformado con palabras unidas, extensión *.txt
PODERX PARECER PEQUEÑO PERO CUANDO SE CONSIDERAR EL NÚMERO TOTAL DE REACCIÓN DE TODO EL PROCESO DE FUSIÓN HABERX IMPLICADOA UNO MASA TOTAL Y POR TANTO ENERGÍA CONSIDERABLE **	PROCESO DE FUSIÓN HABERX IMPLICADOA UNO MASA TOTAL Y POR TANTO ENERGÍA CONSIDERABLE.

3. 2. 2. 2. 3.- Vocabulario más frecuente y vocabulario específico

El texto lematizado modificado con las colocaciones halladas anteriormente se renombra añadiendo una D a su código y se vuelve a pasar por la aplicación Análisis de Frecuencias del PAFE. Es a partir de este estudio cuando se establece el vocabulario más frecuente y se analizan las relaciones entre ULL pertenecientes a dicho vocabulario.

Los análisis de corpus no muestran una uniformidad en distinguir qué se entiende por vocabulario más frecuente (Almela, Cantos, Sánchez, Sarmiento, & Almela, 2005; Davies, 2002-; García Hoz, 1976; Juilland & Chang-Rodríguez, 1964; Justicia Justicia, 1995). Lo más general es encontrar listados de las 500, 1000, 5000 o 10000 palabras más frecuentes, según la finalidad de dicho vocabulario (escolar, diccionarios, etc.). En algunos casos se tienen en cuenta datos relativos a la aparición de las palabras en distintos universos además de su frecuencia total a la hora de decidir la inclusión de una palabra dentro del vocabulario más frecuente.

Establecemos el vocabulario más frecuente estudiando la distribución de frecuencias absolutas y la de frecuencias acumuladas del texto transformado. En un primer momento se podría pensar en hallar el promedio de las distribuciones para hallar el vocabulario más frecuente. Un análisis de las distribuciones de frecuencias absolutas y acumuladas del texto transformado en términos del promedio se presenta en la Tabla 16.

Tabla 16

Límites de la franja definida por el promedio

Frecuencia		Franja (en f)	Palabras límites	% N°	% $\sum n_i$
n	3,82	1,369-120,949	el-vez	23,91	74,40
N	1805,09	0,913-120,949	el-él	45,72	85,81

Nota. n=frecuencia absoluta; N=frecuencia acumulada; \bar{f} = promedio; f=frecuencia relativa; N°= número de unidades léxicas.

En ambos casos, las franjas son grandes incluyendo ULL con frecuencia absoluta muy baja. Esto se debe a que hay pocas ULL con frecuencias altas y muchas con frecuencias bajas. En el caso de las frecuencias absolutas la franja contiene un 24% de las ULL, que producen un 74% de las frecuencias. Mientras que en el caso de las frecuencias acumuladas, la franja contiene un 46% de las ULL (casi el doble que el anterior), que producen un 86% de las frecuencias (un 16% más que en el anterior). Debido al alto número de frecuencias que se incluyen en la franja definida por el promedio, éste no permite hallar el conjunto de ULL más frecuentes. Se hace necesario analizar las distribuciones de otra manera.

A priori consideraremos que las dos distribuciones presentan asíntotas verticales y horizontales a la hora de ajustarlas. Es decir, que se produce un cambio claro de tendencia en los datos de la distribución. Este cambio producido cuando el valor de la tangente a la curva de ajuste vale -1, en el caso de la distribución de

METODOLOGÍA

frecuencias absolutas, $y + 1$, en el caso de la distribución de frecuencias acumuladas, nos indica qué vocabulario es el más frecuente (ver *Figuras 74 y 75*). Las ULL que se queden a la izquierda de dicho punto se considerarán vocabulario más frecuente. Como dicho punto puede caer dentro de una franja de ULL con la misma frecuencia, se cogerá toda la franja como perteneciente a las ULL más frecuentes. Si el punto es un valor intermedio entre dos franjas de distinto valor en frecuencia, se cogerá la franja de valor superior como límite del vocabulario frecuente. En el caso de la *Figura 74* dicho punto cae entre las franjas de frecuencias de valores 17 y 18. Por lo tanto se cogerá como franja límite de vocabulario más frecuente a la franja de frecuencia de valor 18. Esta franja contiene 13 ULL. Si cogiéramos el valor inferior este número se vería aumentado, es decir, el vocabulario más frecuente sería todavía mayor.

En el caso de un corpus de unas 33.333 palabras, se puede estimar que la franja de vocabulario más frecuente tendría unas 159 ULL. Hay que tener en cuenta que los resultados deben poder ser aplicables en clases lectivas. Un vocabulario frecuente grande podría implicar más horas lectivas de las que realmente se tienen. De ahí que se haya decidido elegir como límite el valor mayor en frecuencia al aplicar el método de la derivada.

El vocabulario más frecuente no tiene por qué ser el vocabulario específico. La determinación del vocabulario específico se puede hacer usando la técnica descrita en Chung (2003). Se compara la frecuencia normalizada (frecuencia relativa) de una ULL del corpus específico con un corpus general mediante el cociente f_{ce}/f_{cg} , siendo f_{ce} la frecuencia relativa en el corpus específico y f_{cg} la frecuencia relativa en el corpus general. Se espera que la frecuencia del vocabulario no específico sea más o menos la misma en ambos corpus, mientras que la

frecuencia del vocabulario específico será mayor en el corpus específico. Al usar el concepto del cociente tendremos en cuenta las siguientes consideraciones:

- Si $f_{ce}/f_{cg} < 1$, la ULL aparece más veces en el corpus general y no es específica.
- Si $f_{ce}/f_{cg} = 1$, las frecuencias son iguales, la ULL no es específica.
- Si $f_{ce}/f_{cg} > 1$, la ULL aparece más veces en el corpus específico, por lo que podría ser específica.
- Si $f_{ce}/f_{cg} \rightarrow \infty$, la ULL aparece muchísimo más en el corpus específico y por tanto es específica.

El cociente f_{ce}/f_{cg} debe ser bastante grande para una ULL específica. De acuerdo con Chung (2003) un cociente f_{ce}/f_{cg} mayor o igual a 50 hace que se identifique el mayor número posible de ULL específicas. Este método es capaz de identificar un 86% de las ULL específicas. Un ejemplo de cómo identificar vocabulario específico usando este criterio se da en la Tabla 77, donde se compara el vocabulario más frecuente del estudio de caso con los datos del corpus general CREA (Real Academia Española: Banco de datos (CREA) [en línea], 2008). En este caso se resalta en **negrita** la única ULL que es específica siguiendo el criterio del cociente.

De los corpus generales del español existentes (Tabla 3 en el apartado 2. 3.) se ha escogido el CREA (Real Academia Española: Banco de datos (CREA) [en línea], 2008) para hacer la comparación y así identificar el vocabulario específico. El corpus general CREA es el más grande existente en línea, y además permite hacer la consulta sólo en la parte procedente de textos escritos (libros, revistas, periódicos y miscelánea), que son aquellos objeto de nuestro estudio. La consulta del CREA para cada palabra genera la frecuencia absoluta en el mismo, que luego

METODOLOGÍA

normalizamos a un texto de 1000 palabras mediante el programa *Excel 2007*. Al hacer la consulta en el CREA hay que tener en cuenta que la base de datos del CREA distingue entre minúsculas y mayúsculas. Por lo tanto a la hora de hacer una búsqueda hay que buscar la misma palabra comenzando en minúscula y mayúscula. Además habrá que buscar la palabra en las formas no lematizadas con las que apareció en el texto preparado (singular, plural, masculino, femenino), puesto que llegados a este punto se está trabajando con un texto lematizado. La base de datos del CREA permite también buscar conjunto de palabras que aparezcan juntas, como es el caso de las ULL formadas por más de una palabra. Por ejemplo, para buscar *estrella* se buscará *estrella, estrellas, Estrella y Estrellas*, sumándose las frecuencias en *Excel 2007*. Con este mismo programa calculamos el cociente f_{ce}/f_{cg} e identificamos el vocabulario específico de acuerdo a los criterios ya explicados.

3. 2. 2. 2. 4.- Determinación del sistema óptimo en el que estudiar redes semánticas

A la hora de determinar un sistema de ULL en el cual estudiar las relaciones se considera el que contenga el máximo número de relaciones de alta frecuencia y el mínimo de ULL. Para ello se analiza el número de ULL que contribuyen a los distintos números de relaciones en distintos sistemas determinados por una frecuencia límite inferior. Se utiliza la aplicación Estudio de Entorno del PAFE con las características de la Tabla 17.

Por número de relaciones se entiende el número de veces que un determinado valor de relación aparece en el texto. Supongamos que dos ULL i y j aparecen juntas 6 veces en un entorno 1, por lo que tienen un valor de relación 6. Otras dos ULL distintas pueden tener ese mismo valor de relación, es decir, aparecer juntas 6 veces. Si no hay más parejas distintas de ULL que tengan el valor de relación 6, el número de relaciones es 2 para dicho valor de relación en el sistema que se estudia. En esta tesis, debido a que los listados provenientes del PAFE duplican la información relativa a las parejas de ULL, puesto que cada una de ellas aparece como ULL origen, el número de relaciones aparece multiplicado siempre por 2.

Tabla 17

Características del Estudio de Entorno para determinar el sistema a estudiar

Variable	Valor
Unidades léxicas a procesar	n_{\min}, f_{\min}
Fragmento	a ambos lados, punto y seguido
Entorno	valor máximo=1, decremento =1
Número de relaciones mínima	1

Nota. n = frecuencia absoluta; f = frecuencia relativa.

Se eligen distintos sistemas a estudiar definidos por distintas frecuencias relativas o absolutas mínimas. Como valores de frecuencias mínimas se eligen aquellas que limitan las tres franjas definidas por la distribución de frecuencias acumulada, así como la definida por el vocabulario más frecuente. También se pueden elegir otras frecuencias relativas mínimas para comparar. Por ejemplo en la Tabla 81 aparecen estas cuatro franjas con otros valores que intentan definir franjas equiespaciadas en términos de la frecuencia acumulada. A partir de aquí se estudia la distribución del número de ULL y del número de relaciones respecto a los

METODOLOGÍA

valores de relaciones absolutas o acumuladas, para así poder determinar el tamaño óptimo del sistema en el que analizar las relaciones.

En la aplicación Estudio de Entorno, en la ventana *Navegador de estudio del entorno*, se pueden variar la frecuencia mínima y el valor de relación absoluta mínima a visualizar en el cuadro de diálogo que aparece al pulsar *Opciones* en el menú (anexo D, *Figuras 133 y 132*). En la misma ventana se puede obtener la *Lista de relaciones* mediante el botón situado en la parte inferior izquierda. Dicha lista (*Figura 134*) contiene la información relativa a las características del sistema, así como los valores de relación absolutos entre una ULL respecto a una ULL origen (ver tercera columna) y los relativos a las relaciones totales de la ULL origen en el listado (ver cuarta columna). La *Lista de relaciones* se puede guardar en un archivo de formato **.grp*, que mediante el programa Oneview (Limited, 2002) transformamos en un archivo **.xls*.

De la *Lista de relaciones* se pueden obtener los siguientes datos:

- El número de relaciones que contribuyen a cada valor de relación (ra). Para ello se suman las relaciones, dadas por cada dos ULL relacionadas, que contribuyen a un valor de relación dado.
- El número de relaciones que contribuyen a cada valor de relación acumulado (Ra). Para ello se suman las relaciones, dadas por cada dos ULL relacionadas, que contribuyen a un valor de relación dado y superiores.
- El número de ULL (N°) que contribuyen a cada valor de relación absoluto (ra) o acumulado (Ra). Para ello se suman las distintas ULL origen que contribuyen a un valor de relación dado en el primer caso (ra) o a uno dado y superiores en el segundo (Ra).

Los datos se obtienen con ayuda de la hoja de cálculo Excel 2007 para los distintos sistemas mencionados. Luego se comparan los resultados para los distintos sistemas mediante las distribuciones de relaciones y ULL respecto al valor de relación acumulado. La comparación permite la selección de un sistema que contenga el mínimo número de ULL con el máximo número de relaciones (Ceballos, Hansen-Ruiz, Galotti, & Stengler, 2005). La selección de dicho sistema hace que luego sea más rápido el estudio de las relaciones más importantes, definidas como aquellas más frecuentes, entre las ULL más relevantes, también definidas como aquellas más frecuentes.

Aunque posteriormente el estudio de redes semánticas se realiza en un entorno de valor máximo 3, se escoge el valor máximo 1 para la determinación del sistema óptimo. Este criterio se puede razonar a partir del estudio del caso presentado en Ceballos, Hansen-Ruiz, Galotti, & Stengler (2005). Este estudio de caso se basa en el análisis del texto extraído del artículo “Los monstruos del universo” publicado en la revista de divulgación científica *Ciencia y vida* en abril de 1998. Se trata de un texto del género divulgativo de 1598 palabras con 564 palabras diferentes. De acuerdo con Biber (1990) muestras de 1000 palabras son suficientes para el estudio de las características gramaticales más comunes. Por lo tanto, consideraremos que el texto escogido es representativo respecto al comportamiento de dichas características. En el análisis de este texto no se tuvieron en cuenta las palabras funcionales. En este caso se analizaron los sistemas definidos por la frecuencia relativa mínima de las ULL en la Tabla 18.

El sistema definido por $f_{\min}=0.14$ ($n_{\min}=1$) es el que incluye todas las ULL, es decir, todas con una $f \geq 0.14$. La proporción de ULL no varía de igual modo que el de las frecuencias absolutas. Unas pocas ULL producen gran cantidad de

METODOLOGÍA

frecuencias. Por ejemplo, 14 (3.72%) de las 376 (100%) ULL generan un 27% de las frecuencias absolutas.

Tabla 18

Sistemas a analizar

Sistema (f_{\min})	% N°	% (Σn_i)
0.14	100.00	100.00
0.27	35.37	69.00
0.41	18.35	53.70
0.54	11.17	44.10
0.81	5.05	31.70

Nota. n= frecuencia absoluta; f= frecuencia relativa; N°=número de ULL.

Para determinar qué sistema es el más adecuado para estudiar relaciones se analizaron los cinco sistemas en dos entornos de valores máximos distintos: 1 y 3. Se trata de observar si existen diferencias al intentar determinar el sistema óptimo cuando varía el valor máximo. Las características de los cinco sistemas en los dos entornos respecto a las relaciones del sistema (Σra_i) se muestran en la Tabla 19. Las relaciones del sistema (Σra_i) se han calculado de la *Lista de relaciones*, por lo que están multiplicadas por 2. El botón inferior *Valor-N° Relaciones* muestra el número de relaciones con un determinado valor, pero los datos no son exportables a ningún tipo de fichero.

Tabla 19

Sistemas a analizar en dos entornos

Sistema (f_{\min})	Entorno 1		Entorno 3	
	Σra_i	% Σra_i	Σra_i	% Σra_i
0.14	657	100.00	7248	100.00
0.27	415	63.17	4690	64.71
0.41	309	47.03	3490	48.15
0.54	239	36.38	2717	37.49
0.81	159	24.20	1846	25.47

Nota. f= frecuencia relativa; ra= relación absoluta.

Las relaciones del sistema se multiplican aproximadamente por 11 al pasar de un entorno de valor máximo 1 (columnas 2 y 3) a un entorno de valor máximo 3 (columnas 4 y 5). Las variaciones se deben a la diferencia de entorno.

Consideremos dos ULL i y j , siendo la ULL origen i . Supongamos además que el número máximo de relaciones está en función del máximo de posiciones que puede ocupar la ULL j en torno a la ULL origen. En el entorno de valor máximo 1 la ULL j puede estar a la izquierda o a la derecha de i , variando su número de relación de 1 a 2. En el entorno de valor máximo 3 la ULL j puede estar a la izquierda o a la derecha de i en tres posiciones distintas con tres valores distintos, luego su valor de relación va de 1 a 12. Es decir, entre el mínimo valor de relación del entorno 1 y el máximo valor de relación en el entorno 3 hay una diferencia en los valores de 11, y un factor de multiplicación de 12. Estas consideraciones hipotéticas no tienen en cuenta que el entorno se pare en un punto y seguido o en un punto y aparte, ni que la ULL j no tiene por qué aparecer en todas las posiciones posibles alrededor de la ULL i . Si además se tiene esto último en cuenta, al pasar de un entorno 1 a un entorno 3 el valor de relaciones en el entorno 3 será inferior al del entorno 1 multiplicado por 12.

Independientemente de cómo varía el número de relaciones de los cinco sistemas en los dos entornos, el porcentaje de dichas relaciones es semejante (ver Tabla 19). La diferencia en porcentaje de las relaciones de un sistema en los dos entornos es de un 1% para todos los sistemas donde la $f_{\min} > 0.14$ ($n_{\min} > 1$), siendo el porcentaje mayor en el entorno 3. Por tanto, en cuanto a las relaciones del sistema no parece haber diferencia al intentar determinar el sistema óptimo entre el entorno 1 y el entorno 3.

METODOLOGÍA

Las distribuciones de los números de relaciones y el número de palabras (en el eje de las ordenadas) respecto a las relaciones acumuladas (eje de las abscisas) se muestran en las *Figuras 13 y 14* en escala logarítmica. En la *Figura 13* (entorno de valor máximo 1), en el caso del sistema con $f_{\min}=0.14$ ($n_{\min}=1$), todas las relaciones entre hápax o de una ULL con los mismos aparecen en las relaciones acumuladas de valor 1. En el resto de los sistemas los hápax han desaparecido. El eje de las abscisas está formado por sólo 8 valores de relaciones acumuladas. El valor de relación acumulada a partir del cual todos los sistemas contienen el mismo número de relaciones y palabras, 5, es cercano al punto medio de valores de relación acumuladas. Esto quiere decir que el sistema de 14 palabras ($f_{\min}=0.81$) contiene todos los valores de relación mayores o iguales a 5.

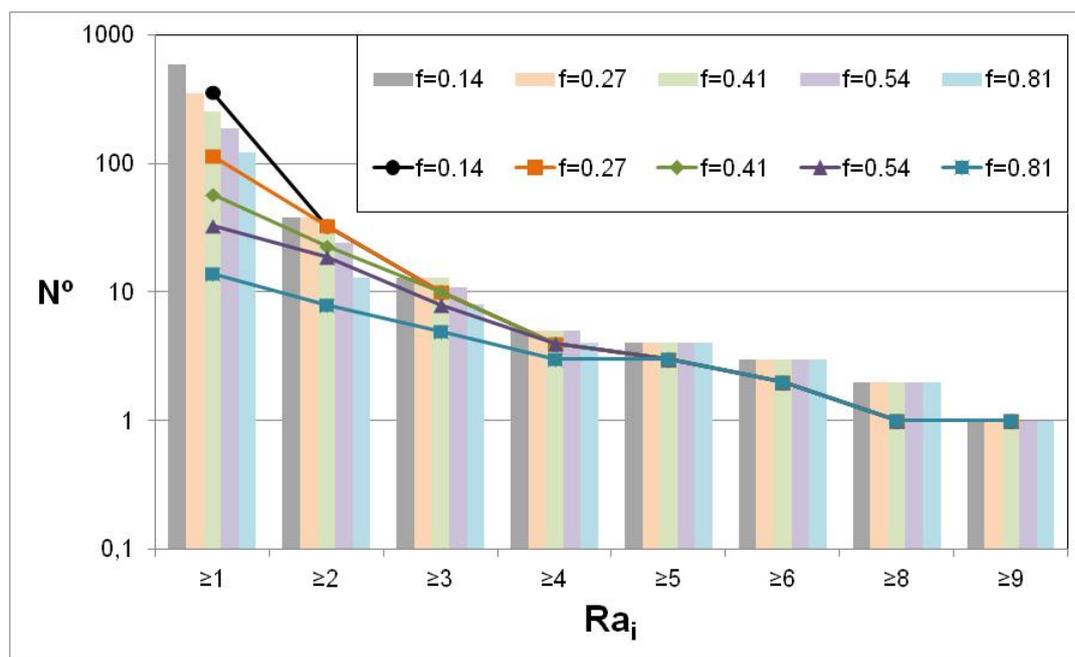


Figura 13.- Distribución para un entorno de valor máximo 1. Las barras representan el número de relaciones y los puntos, el número de ULL, para dicha relación acumulada del eje de abscisas. Ra_i = relación absoluta acumulada

En la *Figura 14* (entorno de valor máximo 3) las relaciones entre hápax o de una ULL con los mismos se encuentran para valores de relación menores que 4.

Debido al decremento de 1, una ULL que aparece un mínimo de dos veces cerca de otra genera un valor mínimo de relación de 2. Sin embargo un hápax puede generar un valor de relación de hasta 3.

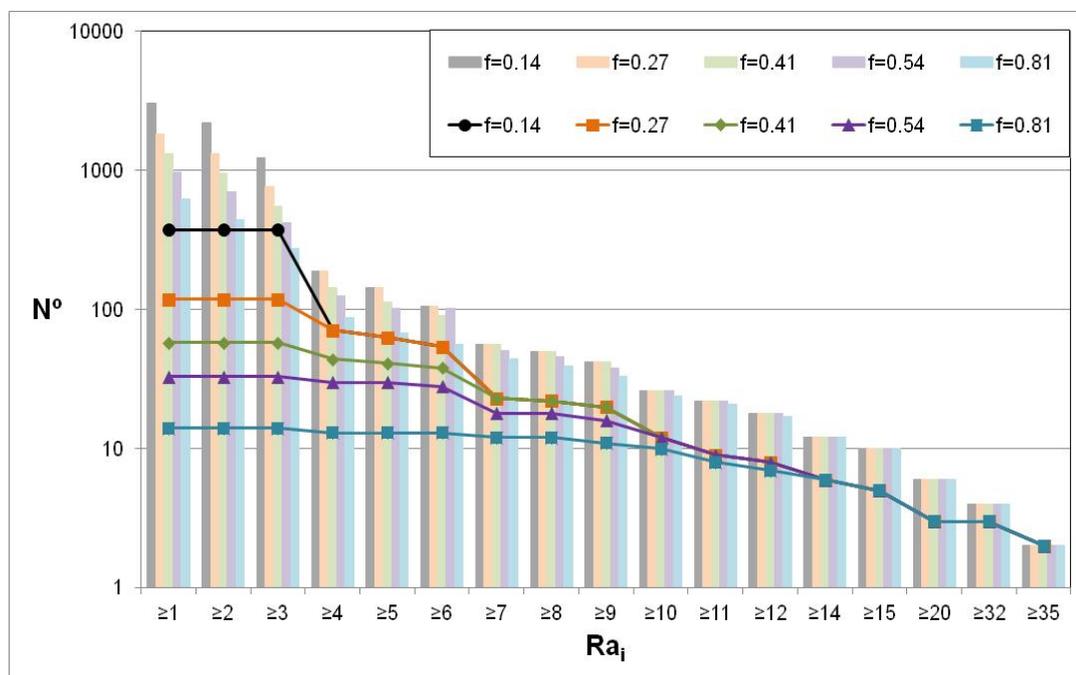


Figura 14.- Distribución para un entorno de valor máximo 3. Las barras representan el número de relaciones y los puntos, el número de ULL, para dicha relación acumulada del eje de abscisas. Ra_i = relación absoluta acumulada

El eje de las abscisas contiene 17 valores de relaciones acumuladas, un 200% más que en el caso del entorno de valor máximo 1. Al igual que en dicho entorno el valor de relación acumulada a partir del cual todos los sistemas contienen el mismo número de relaciones y palabras, 14, es cercano al punto medio de valores de relación acumuladas. Es decir, el sistema de 14 palabras ($f_{\min}= 0.81$) contiene todas las relaciones mayores o iguales a 14.

Las Figuras 13 y 14 son similares respecto al comportamiento de los distintos sistemas. En cuanto al número de relaciones acumuladas el análisis para determinar el sistema óptimo es parecido en un entorno de valor máximo 1 y en un entorno de valor máximo 3. Teniendo en cuenta que el trabajo requerido para

METODOLOGÍA

analizar los sistemas es menor en el entorno 1, parece razonable elegir este entorno para determinar el sistema óptimo en el que se van a estudiar posteriormente las redes semánticas.

3. 2. 2. 2. 5.- Redes semánticas: ULL relacionadas

A la hora de presentar resultados en nuestra metodología usaremos redes semánticas representadas por grafos donde los nodos estarán ocupados por ULL. La caracterización de los nodos dará una idea de la frecuencia de las ULL y la de las aristas indicará el número de conexiones o relaciones entre las ULL.

Tras elegir el sistema óptimo y por tanto los límites del mismo, se analiza con la aplicación Estudio de Entorno del PAFE con las siguientes características:

Tabla 20

Características del Estudio de Entorno para el análisis de redes

Variable	Valor
Lemas a procesar	f_{\min} del sistema óptimo
Fragmento	a ambos lados, punto y seguido
Entorno	valor máximo=3, decremento =1
Valor de relación mínima	6

Se pretende analizar gráficamente las relaciones entre ULL, es decir, obtener las redes semánticas. Debido al decremento de 1, una ULL que aparece un mínimo de dos veces cerca de otra produce como mínimo un valor de relación de 2. Sin embargo un hápax puede generar un valor de relación de hasta 3. Hasta ahora se

han definido franjas y sistemas con el fin de estudiar las ULL más frecuentes y por tanto no se consideran los hápax. De igual modo interesa analizar las relaciones más frecuentes. Dadas dos ULL i y j , siendo la ULL origen i , en un entorno de valor máximo 3 la ULL j puede producir valores de relación con i que varían de 1 a 3. Si la ULL j aparece dos veces en el entorno de la ULL i , puede generar valores de relación que varían entre 2 y 6 dependiendo de la posición que tiene dentro del entorno. El valor de relación mínimo de 6 significa que como mínimo las ULL i y j aparecen inmediatamente juntas 2 veces o que la ULL j aparece en el entorno de i más de 2 veces. Es decir, el valor de relación mínimo de 6 hace que en el análisis gráfico sólo estén las relaciones que aparecen más de 2 veces y aquellas que apareciendo sólo 2 veces lo hacen en la primera posición adyacente a la ULL i . En las características del Estudio de Entorno se escoge pues observar las relaciones entre ULL de valor igual o mayor a 6.

Para realizar la gráfica de un determinado número de ULL del listado, se selecciona la opción *Crear lista* del submenú de *Palabras* (anexo D, *Figura 135*) y se escoge una frecuencia de corte, que puede coincidir o ser superior a la frecuencia mínima que caracterizó al sistema (anexo D, *Figura 136*). También se puede hacer la *Lista de palabras* eligiendo el número de palabras, que la aplicación selecciona empezando por las más frecuentes y siguiendo el orden lexicométrico mostrado. En la *Lista de palabras* creada (anexo D, *Figura 137*) es posible añadir o eliminar ULL del listado usando los botones de la izquierda de la ventana. Esto es particularmente útil a la hora de separar el vocabulario específico del vocabulario más frecuente. También se puede querer aislar un grupo de palabras para analizarlas.

Finalizada la lista definitiva, se selecciona *Gráfica* en el submenú de *Palabras*. En este caso se decide no sólo el número de intervalos de frecuencias y

METODOLOGÍA

valores de relación, sino los colores y tipos de líneas, correspondientes a cada uno de ellos (anexo D, *Figura 138*). Además se puede elegir el intervalo de valores para cada frecuencia y relación (anexo D, *Figura 139*). En las investigaciones de esta tesis se dividen las frecuencias del texto y las relaciones del sistema en 4 intervalos iguales, que en otros estudios se han mostrado suficientes (Ceballos, Hansen-Ruiz, Galotti, & Stengler, 2005). En el mismo submenú de *Palabras*, se pueden escoger las opciones de *Tabla* y de *Relaciones*, que muestran las relaciones totales de cada ULL en el sistema y de las ULL respecto a una ULL origen, respectivamente (anexo D, *Figuras 140 y 141*). Las relaciones mostradas por las opciones de *Tabla* y de *Relaciones* se pueden guardar en archivos *.grp, presentando en la parte superior las características del sistema estudiado. La *Figura 140* muestra los datos de las ULL que forman el sistema estudiado. En la primera columna de la izquierda (*Palabra*) aparece la ULL. La segunda y tercera columnas muestran su frecuencia absoluta y relativa (a 1.000), respectivamente, en el texto transformado. La cuarta y quinta columna muestran el número de relaciones que presenta cada ULL dentro del sistema estudiado como relación absoluta y relativa (a 1.000) en dicho sistema. La *Figura 141* muestra las relaciones que hay entre las ULL que forman el sistema. La primera columna de la izquierda (*Palabra origen*) muestra la ULL origen, mientras que la segunda muestra la ULL con la que se relaciona. En la tercera columna aparece la relación absoluta entre las dos ULL anteriores. Y en la cuarta columna se ha calculado la relación relativa (a 1.000) respecto a la ULL origen, es decir, si la relación absoluta de la ULL j respecto a la ULL origen i es ra_{ij} , la relación relativa rr_{ij} se halla dividiendo por la suma de todas las relaciones de la ULL origen con otras: $rr_{ij}=ra_{ij}/\sum_j ra_{ij}$. En el listado de la *Figura 141* pueden aparecer relaciones de una ULL consigo misma (ver cuarta fila). Estas relaciones no tienen sentido para

los propósitos de la investigación y por tanto no se computan como relaciones para calcular ni la relación relativa a la ULL origen, ni en la relación absoluta de una ULL proporcionada en la cuarta columna de la *Figura 140*. En el listado de la *Figura 141* aparecen todas las relaciones de las palabras del sistema escogido con la ULL origen, independientemente de la relación mínima que se use. Esto quiere decir, que dada una relación mínima, en este listado pueden aparecer relaciones de valor inferior, que no se tienen en cuenta en la gráfica de relaciones. Este listado permite tener una visión más global de todas las relaciones que se producen dentro del sistema escogido.

La representación gráfica de los resultados obtenidos proporciona una red, en la que cada ULL viene determinada por su frecuencia y su valor de relación (*Figura 15*). En esta figura las ULL vienen caracterizadas por colores de relleno y grosores de la línea de la caja que están en función de su frecuencia, la cual aparece en el listado obtenido en *Tabla* en el submenú de *Palabras*. Las líneas que unen las ULL están caracterizadas por el valor de relación entre las mismas, el cual viene en el listado obtenido en *Relaciones* en el submenú de *Palabras*. En este ejemplo se seleccionan 5 intervalos de caracterización de las ULL y 3 intervalos para las relaciones entre las mismas; sin embargo las ULL del sistema sólo están en tres de los intervalos de frecuencia (1º, 2º y 5º).

La representación gráfica del PAFE es limitada, permite colocar las palabras sólo en dos tipos de rejillas: de 3 por 5 (15 ULL como máximo) o de 5 por 8 (40 ULL como máximo). Esto limita las posiciones de las ULL, así como la posición de las relaciones, pudiéndose crear imágenes con líneas que atraviesan las cajas de las ULL no quedando claro en ocasiones si la línea continúa hacia otra ULL o finaliza en dicha caja. La *Figura 16* presenta la gráfica producida con el programa PAFE,

METODOLOGÍA

mientras que la *Figura 17* presenta la gráfica generada con Excel2003 a partir de los datos del PAFE para el estudio del caso presentado en Ceballos, Hansen-Ruiz, Galotti, & Stengler (2005).

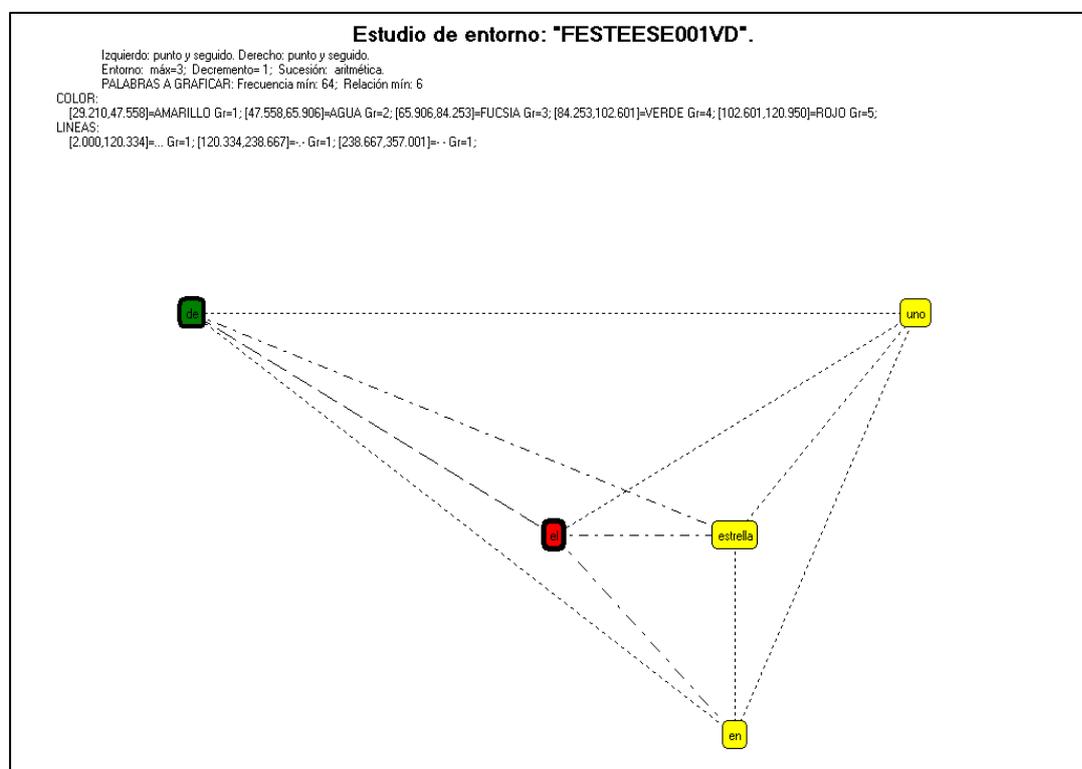


Figura 15.- Gráfica del estudio de caso.

En la *Figura 16* las ULL *convertir* y *masasolar* parecen estar relacionados. Mientras que en la *Figura 17* estas dos ULL no aparecen relacionadas. En caso de duda, para comprobar la apariencia de las líneas en la gráfica se puede repetir la misma pulsando en *Nueva distribución* en el menú de *Opciones* de la ventana *Gráfica de relaciones* (anexo D, *Figura 142*), o comprobar los datos de las relaciones en las tablas previamente mencionadas. En esta tesis se toma como punto de partida la gráfica generada por el PAFE y se reproduce la misma en Excel 2007. En las gráficas hechas con *Excel* se omiten aquellas ULL que han quedado aisladas, puesto que la gráfica se construye con el propósito de visualizar las relaciones. Además en Excel es posible modificar la gráfica para poner la palabra de mayor

frecuencia y número de relaciones en el centro e ir construyendo los distintos grupos de relaciones en torno a ésta. En el *PAFE* no hay modo de centrar la posición de una ULL dada. Otra ventaja a la hora de usar *Excel* es la posibilidad de poder modificar el grosor de las líneas de relación de modo que se vea el número de relaciones entre dos ULL respecto al número total de relaciones de una ULL (*Figuras 17, 140 y 141*). En la *Figura 17* los números representan la relación relativa aproximada entre dos ULL, redondeada a números naturales. La ULL *estrella* se relaciona relativamente 7 veces con *convertir*, pero el número de relaciones relativas de *convertir* respecto a *estrella* es de 39, es decir, *convertir* se relaciona más con *estrella* de lo que ésta se relaciona con *convertir*. Se trata de relaciones asimétricas, que se representan por líneas que se estrechan o se ensanchan dependiendo de la relación relativa correspondiente. Los colores de las relaciones corresponden a los intervalos de las mismas. También se puede establecer el mismo código de colores en los intervalos de frecuencias y relaciones, de modo que sea fácil identificar rápidamente tanto las frecuencias como las relaciones más o menos altas.

Para mostrar la diferencia en las gráficas respecto a los distintos sistemas se muestra el caso presentado en Ceballos, Hansen-Ruiz, Galotti, & Stengler (2005). Las *Figuras 18, 19 y 20* presentan cómo varía la gráfica en tres sistemas distintos estudiados con un entorno 1.

METODOLOGÍA

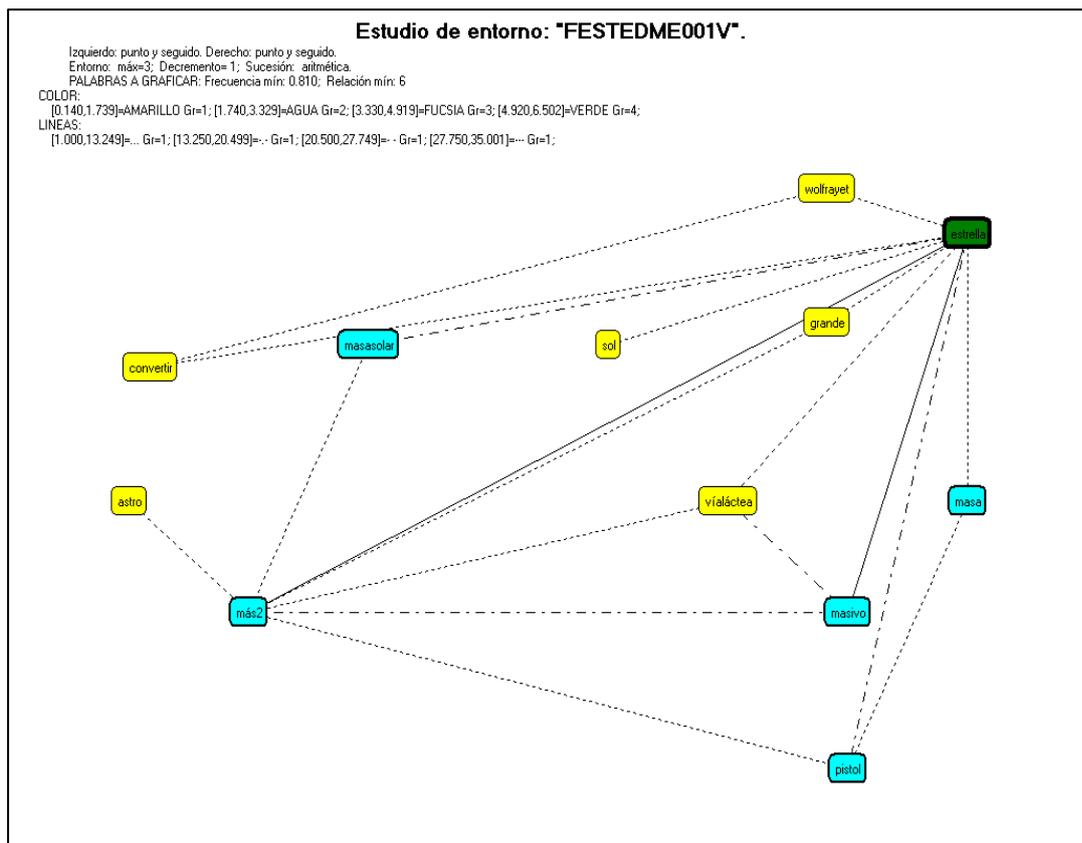


Figura 16.- Gráfica del caso de divulgación (PAFE).

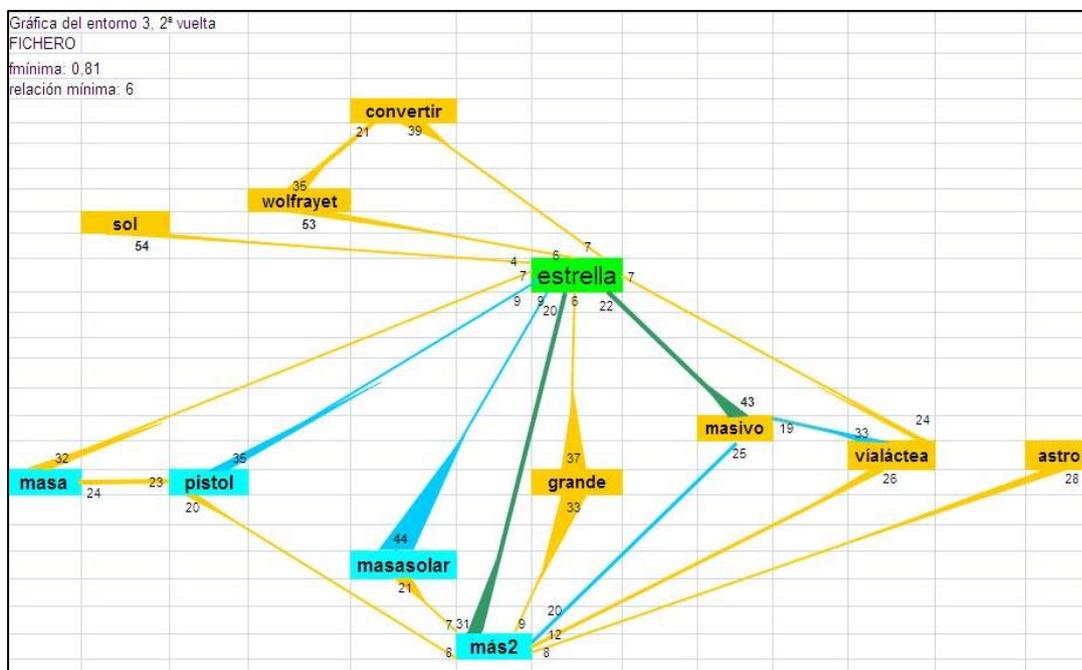


Figura 17.- Gráfica del caso de divulgación (Excel 2003).

El sistema de la *Figura 18* donde $f_{\text{mínima}}=0.81$, tiene un 27% de las frecuencias acumuladas. El de la *Figura 19* tiene un 34% de las mismas. El primer sistema ha perdido respecto al segundo sólo dos ULL de baja frecuencia (*pérdida y encontrar*), según la clasificación de colores presentada en la *Figura 20*. En ésta se representan sólo 54 ULL del total de 376, repartidas no sólo en torno a *estrella*, sino también en grupos aislados (que se ven a la derecha). El sistema de la *Figura 18* ha perdido 44 ULL respecto al completo, de las cuales 31 se hallan en grupos aislados del principal. Todas las ULL perdidas se encuentran en el intervalo de frecuencias bajas, y sólo se ha perdido una relación de mayor frecuencia (*pérdida-masa*). Siete de las ULL perdidas se relacionaban con ULL de mayor frecuencia y sólo tres con la ULL de mayor frecuencia. A pesar de la pérdida de ULL, el sistema de la *Figura 18* nos muestra las ULL claves de la información contenida en el artículo de divulgación, que se puede resumir como: una estrella de gran masa, llamada Pistol, que se encuentra en la Vía Láctea; las estrellas masivas se pueden convertir en estrellas de tipo Wolf-Rayet. En la medida en que las ULL sirven de etiquetas a los conceptos, son un indicativo de los conceptos principales que aparecen en el texto y de las relaciones entre los mismos.

El número de ULL representadas frente al número de ULL que componen el sistema depende del corte en el número de relaciones. En los casos de las *Figuras 18, 19 y 20* la proporción de ULL varía de un 71 a un 12% del total de ULL que componen cada sistema. El número de ULL representadas siempre es menor al número de ULL que componen el sistema y depende de las relaciones entre ellas y del corte en el número de relaciones a representar.

De igual modo que se han visto las diferencias entre los distintos sistemas en función de la pérdida de ULL, podemos ver cómo desde el sistema de menor ULL

METODOLOGÍA

(Figura 18) hacia el completo (Figura 20) se añaden ULL. Es decir, las ULL se incorporan al esquema más básico, haciendo crecer la estructura sin modificarla.

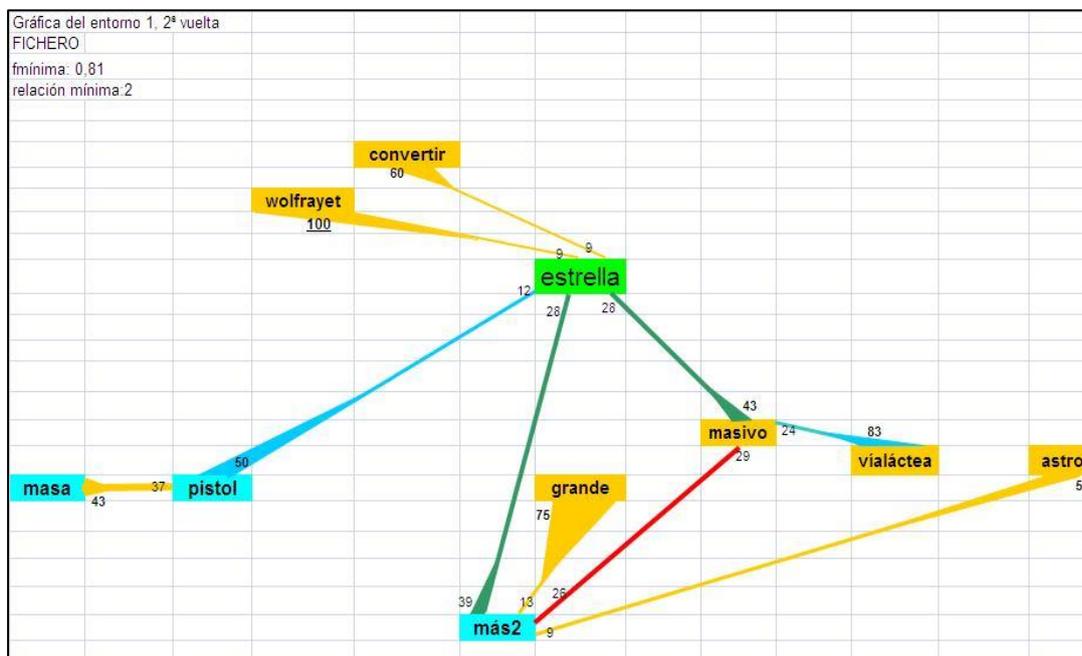


Figura 18.- Gráfica del sistema $f_{\text{mínima}}=0.81$.

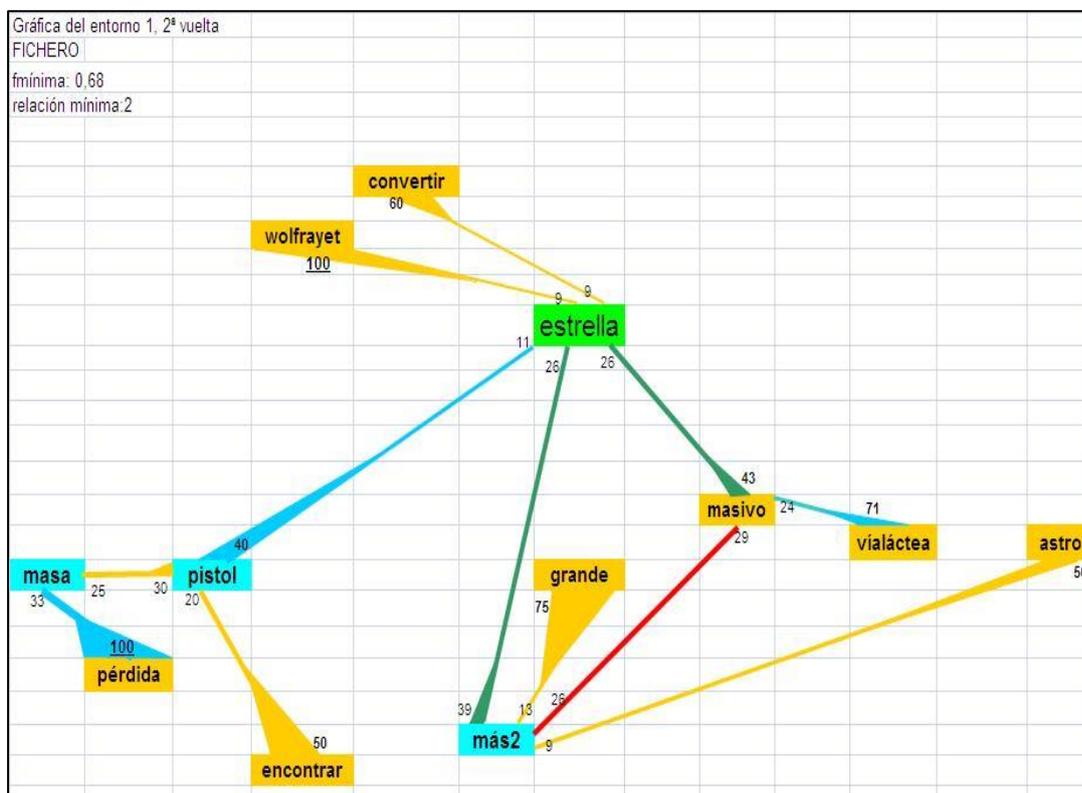


Figura 19.- Gráfica del sistema $f_{\text{mínima}}=0.68$.

METODOLOGÍA

Nótese cómo varían las relaciones relativas de la *Figura 18* a la *Figura 20*. La relación relativa de una ULL respecto a otra ULL origen se calcula dividiendo el valor de relaciones absolutas entre ambas ULL por el valor de relación total de la ULL origen en el sistema. En la *Figura 20* la relación total de *estrella* es 449, mientras que en la *Figura 18* su relación total es de 162. Teniendo en cuenta que las relaciones absolutas de *estrella* con otras ULL son siempre las mismas, la relación relativa de una ULL con *estrella* como ULL origen es mayor en la *Figura 18* que en la *Figura 20*. Lo mismo ocurre con el resto de relaciones relativas que aumentan de la *Figura 20* hacia la *Figura 18*. Las variaciones de las relaciones relativas de las distintas ULL no son comparables de un sistema a otro puesto que la suma de las relaciones absolutas de cada ULL origen varía con la variación del sistema.

Supongamos el caso del par de ULL *Pistol-estrella*, que en la *Figura 18* presentan las relaciones relativas 50-12 respectivamente ($ra_{Pistol, estrella} / \sum ra_{Pistol} f_{min=0.81} - ra_{Pistol, estrella} / \sum ra_{estrella} f_{min=0.81}$), mientras que en la *Figura 20* muestran los valores 15-5 ($ra_{Pistol, estrella} / \sum ra_{Pistol} f_{min=0.14} - ra_{Pistol, estrella} / \sum ra_{estrella} f_{min=0.14}$). Al comparar los valores mediante cocientes en estos dos sistemas se observa que $15/50 \neq 5/12$. En general, esto ocurre para cualesquiera ULL comparadas de un sistema a otro, debido a que la variación de la suma de relaciones absolutas de las ULL origen es diferente. Por otro lado tampoco se pueden comparar mediante cociente pares de ULL relacionadas en el mismo sistema por la misma razón. En el caso anterior se observa que $12/50 \neq 5/15$.

Para economizar tiempo de computación se elige visualizar el sistema óptimo en el entorno 3. Esto facilita la comparación de las gráficas de distintos textos usando las frecuencias y relaciones relativas. Se comparan las ULL dentro de los cuatro intervalos de frecuencia relativa definidos para cada texto o corpus. De

igual modo se procede con las relaciones relativas en sus cuatro intervalos. En las *Figuras 18, 19 y 20* estos intervalos se codifican mediante cuatro colores distintos: amarillo, azul, rojo y verde, según aumentan los valores de los intervalos.

Los parámetros iniciales de estudio de distintos textos son iguales, excepto en la selección del sistema óptimo determinado por su f_{\min} (Tabla 20). Es decir, aunque se procede a seleccionar el sistema de la misma manera, el sistema óptimo no tiene por qué ser el mismo para distintos textos. Por ejemplo, se podrían obtener sistemas óptimos cuyas frecuencias acumuladas varíen entre un 40 y un 20% del total de sus frecuencias.

¿Hasta qué punto el análisis con dos entornos diferentes produce resultados distintos? Para ello se comparan las gráficas de la *Figura 18*, entorno 1, con la de la *Figura 17*, entorno 3. Se puede observar que la gráfica de la *Figura 17* contiene a la de la *Figura 18*. La gráfica del entorno 3 contiene dos ULL más (*sol, masasolar*) de las que ya estaban seleccionadas en el sistema óptimo ($f_{\min}=0.81$) y 8 nuevas relaciones. En el entorno 1 dichas ULL no presentaban suficiente número de relaciones, pues la relación mínima a visualizar es 6, y por eso no aparecen en la gráfica (ver *Figura 18*). Por lo tanto se opta por analizar las redes semánticas en el entorno 3, recordando que se utilizan las ULL correspondientes al sistema óptimo (ver apartado anterior).

La comparación de gráficas del entorno 3 permite observar diferencias tanto de vocabulario usado, como de relaciones entre el mismo para distintos niveles educativos o géneros.

El estudio del entorno 3 en el fragmento “punto y seguido”, tal y como aparece en la Tabla 20, permite el análisis de relaciones dentro de las frases,

METODOLOGÍA

perdiéndose relaciones entre frases contiguas del párrafo. Desde el punto de vista de la lingüística un párrafo indica la expresión de una idea distinta a la de otro párrafo. La elección del fragmento “punto y aparte” hace que el PAFE considere las siguientes ULL después del punto y seguido como parte del entorno de la ULL final de la frase anterior. Después del análisis con los parámetros de la Tabla 20 se hará otro análisis con los mismos parámetros excepto el fragmento donde se elegirá “punto y aparte”. Este análisis servirá para determinar la influencia del fragmento en las gráficas de relaciones.

3. 2. 3.- Esquema de la metodología

En la columna de la izquierda de las *Figuras 21, 22 y 23* se observan los distintos pasos de la metodología que se van a seguir a lo largo de este capítulo. En cada paso se usan distintas aplicaciones informáticas, presentadas en la columna central, y a su derecha los resultados obtenidos a través de dichas aplicaciones.

En la primera parte del esquema de la metodología (*Figura 21*) se parte del texto original y se llega a un texto transformado que contiene todas las colocaciones halladas. En la segunda parte (*Figura 22*) se determinan el vocabulario más frecuente, el vocabulario específico, el sistema óptimo de ULL y se analizan las relaciones dentro del fragmento del “punto y seguido”. En la tercera parte (*Figura 23*) se analizan las relaciones dentro del fragmento del “punto y aparte”.

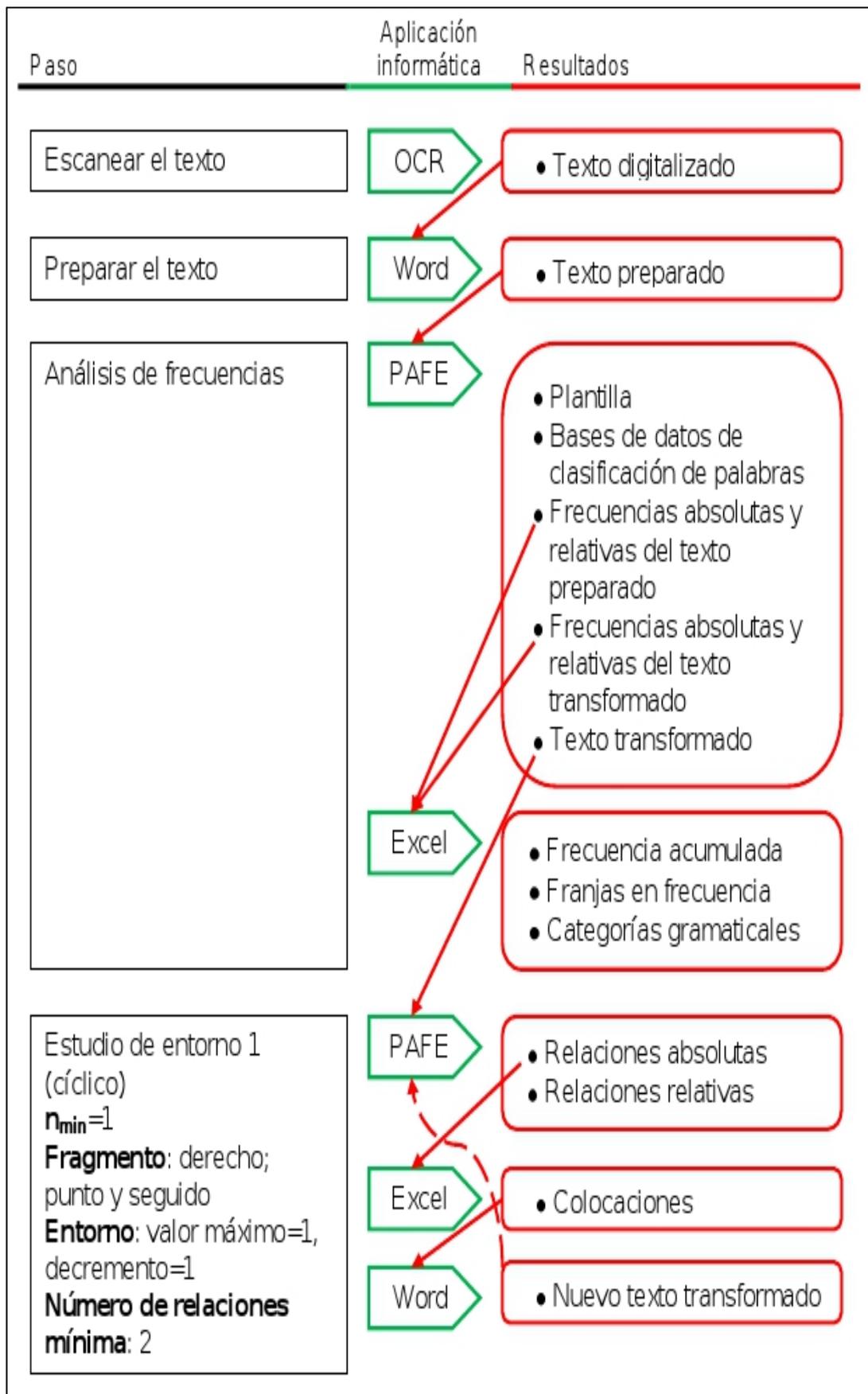


Figura 21.- Primera parte del esquema de la metodología.

METODOLOGÍA

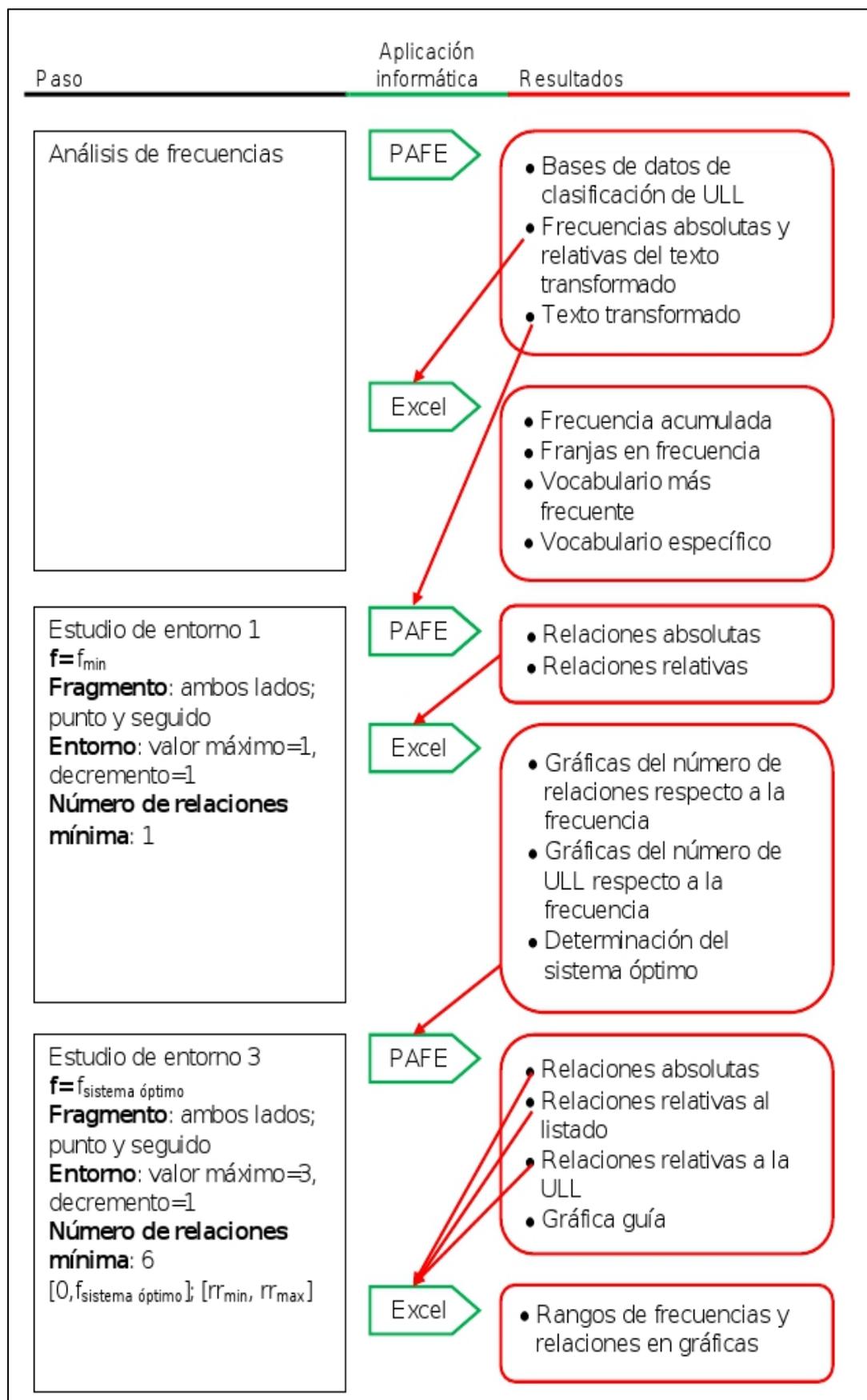


Figura 22.- Segunda parte del esquema de la metodología.

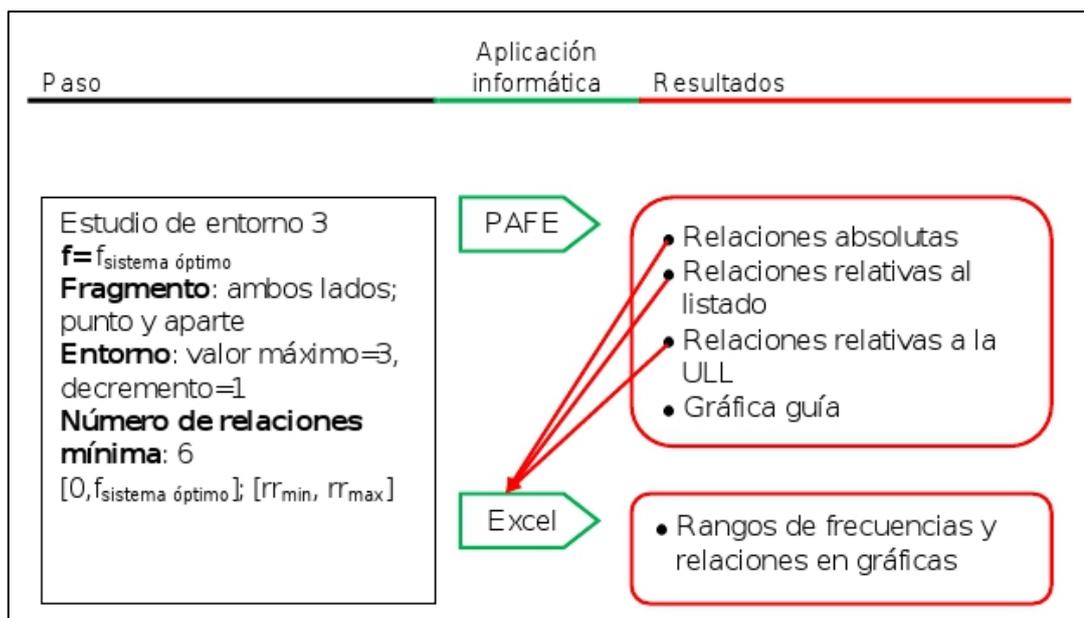


Figura 23.- Tercera y última parte del esquema de la metodología.

3. 2. 4.- Posibles fuentes de error

El error cometido por la aplicación informática Análisis de Frecuencias del PAFE al contar las palabras, las ULL o las relaciones, es decir, al hallar las frecuencias, es 0.

Las posibles fuentes de error de esta metodología provienen de la intervención del investigador:

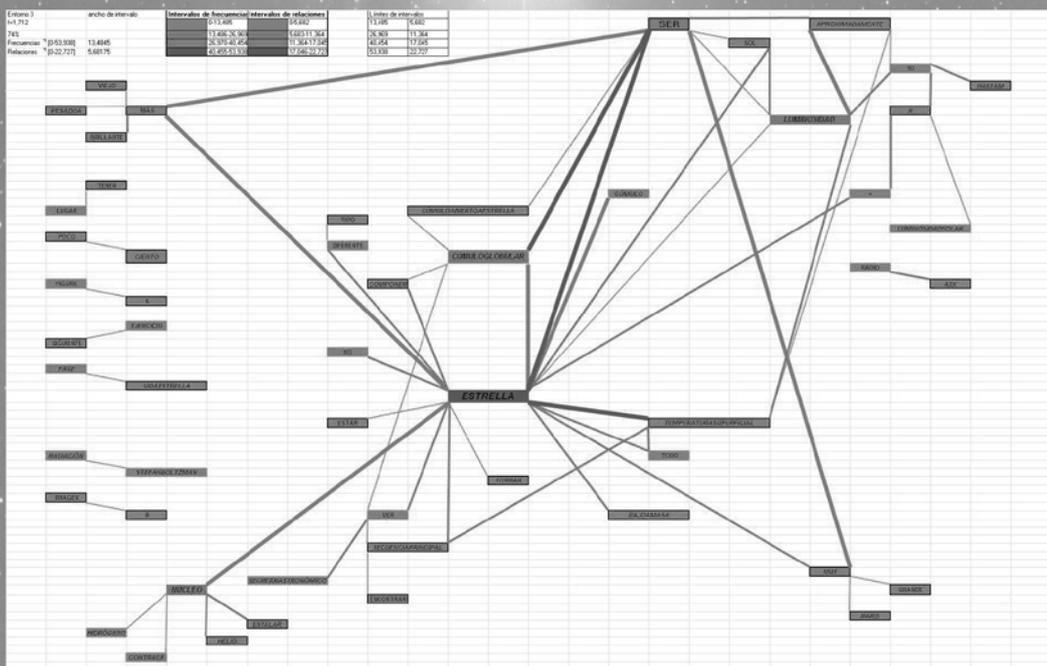
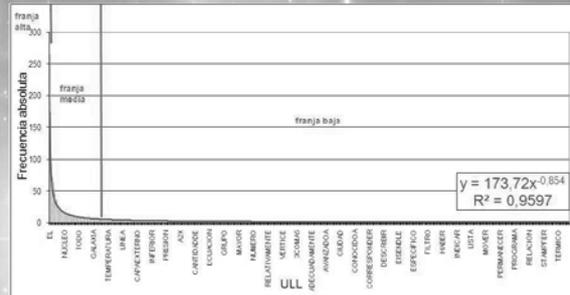
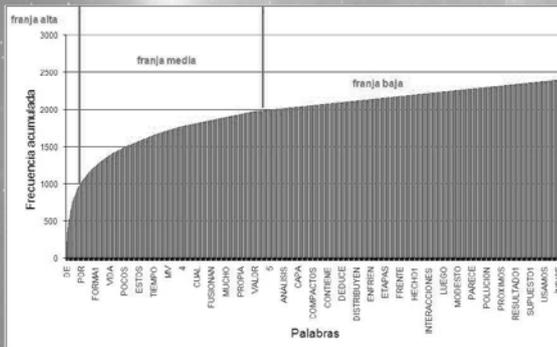
- Se pueden producir errores en la preparación del texto, al no diferenciar significados o no unir en una misma grafía sinónimos.
- Se pueden cometer errores al clasificar las palabras en las bases de datos de la aplicación PAFE.

METODOLOGÍA

- Se puede errar en la determinación de las colocaciones que no cumplen con el criterio del cociente y que son estimadas por el investigador.

Cuando estos errores se dan en las palabras o ULL de baja frecuencia, su influencia en los resultados a analizar (ULL de frecuencia alta) es despreciable.

4.- UN ESTUDIO DE CASO: UN TEXTO ESCOLAR DE ASTRONOMÍA



4.- UN ESTUDIO DE CASO: UN TEXTO ESCOLAR DE ASTRONOMÍA

Se ha escogido un estudio de caso para discutir y mejorar la metodología empleada y señalar la cantidad de información que puede obtenerse mediante la estadística y el estudio de las frecuencias de UL en textos. Se trata de un texto escolar de Astronomía: “*Medida de la distancia y la edad de un cúmulo globular de estrellas*” (ver el texto original en el Anexo B y la bibliografía en la Tabla 8). Dicho texto pertenece al género educativo, subgénero secundaria, y aparece en seis páginas con secciones separadas mediante títulos, acompañado de figuras con pies descriptivos o explicativos que contienen fotografías, gráficos e ilustraciones. El texto se presenta como una parte del ejercicio 4 del Programa de Ejercicios de Astronomía producido por El Centro de Información de la *Agencia Espacial Europea del Hubble* y el *Observatorio Europeo Austral*. El ejercicio 4 es un archivo en formato pdf con 32 páginas. Las partes del archivo se explican en la Tabla 21.

En primer lugar se han eliminado las imágenes y los gráficos del texto escolar. Este texto tiene 2724 palabras, medidas con el programa Word 2007. De acuerdo con Biber (1990), muestras de 1000 palabras son suficientes para el estudio de las características gramaticales más comunes. Por tanto, se considera que el texto escogido es representativo respecto al comportamiento de dichas características.

UN ESTUDIO DE CASO: UN TEXTO ESCOLAR DE ASTRONOMÍA

Tabla 21

Partes del ejercicio 4

Parte	Páginas	Notas
Portada	1	
Índice	3	
Prefacio	4	Texto con imagen.
Texto escolar	5-10	Texto con títulos; fotografías, gráficos e ilustraciones con pie.
Tareas (actividades)	11-20	Actividades; imágenes, tablas y gráficos con pie.
Lecturas	21	Listado de lecturas adicionales.
Edición	22	Texto sobre la edición.
Guía del profesor	23-31	Texto con tablas y gráficos con pie.
Contraportada	32	

Nota. Las páginas son las del archivo pdf.

Otras características, medidas con el programa Word 2007, se presentan en la siguiente tabla:

Tabla 22

Características del texto escaneado (medido con Word 2007) FESTESE001V

Característica	Nº	
Palabras	2723	
Líneas	234	12
Párrafos	49	56
Páginas	5	545

Nota. Nº = número; = media de palabras.

Dado que el interés del estudio se centra en el contenido léxico del discurso del texto, en la preparación del mismo se desechan los títulos y pies de figuras descriptivos (tal y como se explica en el apartado 3. 2. 1.). Es decir, consideramos pies de figuras descriptivos aquellos que meramente describen la figura, como por ejemplo el pie de la figura 2. Fueron desechados los pies de figuras de las figuras 2 a la 7 y 9 títulos que aparecían en el texto. En este caso los títulos contenían el 2.13% de las palabras y los pies de figuras descriptivos el 11.57% de las mismas.

En total se desechan el 13.70% de las palabras (373). Si bien los pies de figuras desechados representan algo más de un 10% del texto original, los títulos apenas suponen un 2% del mismo. Estos datos son del mismo orden de magnitud que los obtenidos en otros estudios del GICEC.

Para ver cómo las variaciones en la metodología influyen sobre los resultados se analiza el texto de tres maneras distintas:

- análisis ortodoxo (AO): en el que no se clasifican palabras en la base de datos *Eliminar*. El texto de partida se denomina FESTESE001V.
- análisis sin palabras funcionales (AF): en el que se clasifican todas las palabras funcionales y verbos auxiliares en la base de datos *Eliminar*. El texto de partida se denomina FESTESE001V-F.
- análisis con unificación de grafías (AG): en el que la preparación del texto incluye la unificación de sinónimos y se explicitan las sustituciones de palabras. Además se clasifican todas las palabras funcionales y verbos auxiliares en la base de datos *Eliminar*. El texto de partida se denomina FESTESE001V-S.

4. 1.- Preparación del Texto

Siguiendo la metodología explicada anteriormente, el texto original sufrió distintas modificaciones para poder ser reconocido por el programa PAFE o por el investigador. Las modificaciones se presentan en las Tablas 23, 24, 25 y 26.

Tabla 23

Simbología matemática eliminada o no reconocida por el programa PAFE

Símbolos matemáticos	Se cambia a	Significado
...	hastam	hasta más o menos
σ	sigma	constante de Stefan-Boltzmann
π	pi	3,14...
\propto	proporcional	proporcional
/	dividido	dividido
-	menos	menos
%	porciento	tanto por ciento

Los siete primeros símbolos matemáticos de la Tabla 23 no fueron reconocidos por la aplicación Análisis de Frecuencias, mientras que el último símbolo (%) no fue reconocido por la aplicación Estudio de Entorno. Representan un 15% del lenguaje matemático encontrado en el texto preparado. En otros textos con otra simbología se pueden encontrar problemas análogos de reconocimiento.

Se encontraron 17 participios ejerciendo la función de adjetivo, que constituyen un 9% de los adjetivos del texto preparado. Todos los participios fueron modificados añadiéndose una “a” para señalar su condición de adjetivo. De esta manera se evita clasificarlos erróneamente en la base *Verbos*. Los 17 participios modificados son: *abierto, ejercido, emitido, ligado, pesado, avanzado, estudiado, repetido, llamado, colapsado, implicado, medida, complicado, indicado, aproximado, conocido e involucrado*.

Además se encontraron 25 palabras (Tabla 24) que podían ser confundidas con otras funciones gramaticales o semánticas a la hora de clasificarlas correctamente en las bases de datos del programa PAFE. De estas 25 palabras, 19 no deben ser confundidas con una forma verbal. Cuando se añaden los participios usados como adjetivos se observa que 36 modificaciones (un 71%) se hicieron para

UN ESTUDIO DE CASO: UN TEXTO ESCOLAR DE ASTRONOMÍA

no confundir palabras con una forma verbal. Un 82% de las modificaciones se hicieron de acuerdo al punto 5 de la preparación del texto (ver apartado 3. 2. 1.), mientras que el 18% restante eran caracteres no admitidos por la aplicación PAFE.

Tabla 24

Palabras que pueden ser confundidas con otra función gramatical o semántica

Palabra	Función	Confusión posible con	Se cambia a
aumento	Sustantivo	verbo aumentar	aumento1
constantes	Adjetivo	sustantivo constante	constantesa
poder	Sustantivo	verbo poder	poder1
hecho	Sustantivo	verbo hacer	hecho1
resultado	Sustantivo	verbo resultar	resultado1
muestra	Sustantivo	verbo mostrar	muestra1
fue	Verbo ser	verbo ir	fuevser
listado	Sustantivo	verbo listar	listado1
medida	Sustantivo	adjetivo o verbo	medida1
estándar	Adjetivo	sustantivo	estándara
baja	Adjetivo	verbo bajar	bajaa
estudio	Sustantivo	verbo estudiar	estudio1
giro	Sustantivo	verbo girar	giro1
prueba	Sustantivo	verbo probar	prueba1
supuesto	Sustantivo	verbo suponer	supuesto1
mezcla	Sustantivo	verbo mezclar	mezcla1
paso	Sustantivo	verbo pasar	paso1
forma	Sustantivo	verbo formar	forma1
constante	Sustantivo	adjetivo constante	constante1
escala	Sustantivo	verbo escalar	escala1
reserva	Sustantivo	verbo reservar	reserva1
precisa	Adjetivo	verbo precisar	precisaa
E	Energía	conjunción e	E1
final	Adjetivo	sustantivo final	finala
suministro	Sustantivo	verbo suministrar	suministro1

También hubo que unificar bajo el mismo signo “x” el signo de la multiplicación que aparecía con dos tipografías distintas pero muy parecidas. Este error se generó en el OCR, manteniéndose en los archivos en formato *.doc y *.txt, que es el archivo de entrada en el PAFE.

En total se realizaron 51 modificaciones, es decir, un 7% de las palabras diferentes del texto preparado fueron modificadas previamente al análisis con el programa PAFE. Este texto preparado fue el que se usó en el AO y AF.

UN ESTUDIO DE CASO: UN TEXTO ESCOLAR DE ASTRONOMÍA

En el AG se añadió a la preparación del texto la modificación de distintos tipos de sinonimias y sustituciones de palabras léxicas:

- Sinónimos propiamente dichos, es decir, palabras o expresiones distintas que significan lo mismo. Es el caso de *Vía Láctea* y *nuestra galaxia*. En los dos casos se hace referencia al mismo objeto astronómico: la galaxia en la cual vivimos. Se unifica hacia una de las expresiones.
- Abreviaturas. Es el caso de *Fig* y *figura*, que aparecen en el texto indistintamente. Se sustituye la abreviatura por la palabra completa.
- Matematización de conceptos, símbolos matemáticos de conceptos. Es el caso de E , que aparece en la ecuación de Einstein en el texto (tercer párrafo) y que aparece como *energía* en otras partes del texto. Para la unificación se sustituirá el símbolo por la palabra si se menciona dicha palabra en el texto.
- Números escritos en letras o numéricamente. Es el caso de *dos* y 2 . Para la unificación se usará sólo el símbolo matemático.
- Omisión de sujeto. Es el caso de la segunda frase del quinto párrafo, donde se omite el sujeto *cúmulos abiertos de estrellas*, del cual se venía hablando en la frase anterior. Para la unificación se reescribirá dicho sujeto.
- Sustitución de palabras mediante pronombres. Es el caso de la tercera frase del quinto párrafo donde se omite *abiertos de estrellas* con el pronombre *Estos*. Para unificar se sustituye el pronombre por las palabras correspondientes.
- Cuando la unificación de grafías produce la repetición de grafías una de las dos es suprimida. Es el caso del párrafo 13 en el que aparecía *El cúmulo globular Messier 12 (o M12)...*, donde se suprime todo el paréntesis.

En definitiva, se trata de que al hallar las frecuencias, el programa usado tenga en cuenta el “concepto” al que se refieren los sinónimos, las elipsis y los

UN ESTUDIO DE CASO: UN TEXTO ESCOLAR DE ASTRONOMÍA

pronombres. El programa sólo diferencia grafías, por lo que el mismo concepto debe estar expresado por la misma grafía. No se han tenido en cuenta las oraciones coordinadas ni subordinadas para este análisis.

La Tabla 25 muestra el listado de grafías que se unifican o modifican. La Tabla 26 muestra el listado de las grafías que se añadieron por estar omitidas o haber sido sustituidas por un pronombre en el texto.

Tabla 25

Unificaciones de grafías en FESTESE001V-S

Variaciones originales	Grafía elegida	Nº	Nº repeticiones
fig. figura	figura	4	
dos. 2	2	2	
cuatro. 4	4	3	
energía. E	energía	1	
masa. M	masa	1	
unoa. unaa. 1	1	2	
nuestra galaxia	Vía Láctea	2	1
Messier 12. M12	Messier 12	3	1
NGC. Nuevo Catálogo General	NGC	1	1
diagrama de Hertzsprung-Russell. diagrama H-R. diagrama	diagrama H-R	6	1
luminosidad. L	luminosidad	11	3
magnitud absoluta. M	magnitud absoluta	1	1
temperatura superficial. T. temperatura de su superficie. temperatura	temperatura superficial	11	4
radio. R	radio	4	1
mB	magnitud aparente del índice de color B	3	1
mV	magnitud aparente del índice de color V	3	

Nota. Nº= número de variaciones modificadas; Nº repeticiones=número de grafías repetidas que luego fueron suprimidas.

Debido a la unificación de grafías puede ocurrir que una grafía aparezca al lado de sí misma. Estas repeticiones se han suprimido puesto que no tienen sentido

UN ESTUDIO DE CASO: UN TEXTO ESCOLAR DE ASTRONOMÍA

a la hora de leer el texto. Nótese que el aumento de frecuencia de las grafías no es grande si las mismas previamente no se encontraban en la parte superior de la franja media. Destacan las modificaciones de *luminosidad* y *temperatura superficial*, que podrían hacer que pasaran a formar parte del vocabulario más frecuente.

Tabla 26

Unificaciones de grafías por omisión o sustitución en FESTESE001V-S

Omisión o sustitución	Grafía elegida	Nº
omisión de sujeto	cúmulos abiertosa de estrellas	1
	las Pléyades	1
	estrellas	2
sustitución por pronombre	cúmulos globulares	5
	cúmulos abiertosa de estrellas	1

Nota. Nº= número de veces que aparecían las omisiones o sustituciones.

A las 51 modificaciones que se hicieron en la preparación del texto para el AO y AF, se han añadido otras 68 en el AG. En el AG un 18% de las palabras diferentes del texto preparado fueron modificadas previamente al análisis con el programa PAFE.

Debido a las modificaciones que sufre el texto en su preparación (ver capítulo 3) sus características difieren del texto escaneado (ver Tablas 22 y 27). La media de palabras por línea, párrafo y página del texto preparado ha aumentado respecto al texto escaneado. Esto es consecuencia de la eliminación de títulos y pies de figuras descriptivos. El texto preparado conserva un 89% de las palabras totales en el caso del AO y AF y un 90% de las mismas en el caso del AG. La unificación de sinónimos y la escritura de sujetos omitidos hace que el número de palabras del texto preparado en el AG aumente ligeramente, provocando una conservación superior del número de palabras respecto al texto escaneado que es ficticia. Los

textos preparados se pueden ver en el Anexo C para el AO y el AF (cuyos textos son iguales), y en el Anexo L para el AG.

Tabla 27

Características del texto preparado (medido con Word 2007), FESTESE001V

Característica	AO y AF		AG	
	Nº		Nº	
Palabras	2412		2460	
Líneas	176	14	176	14
Párrafos	30	80	30	82
Páginas	4	603	4	615

Nota. Nº = número; = media de palabras.

4. 2.- Análisis de Frecuencias del Texto Preparado

En el AO no se usa la base *Eliminar* en la clasificación de palabras. En el AF y AG no se tienen en cuenta las palabras funcionales ni verbos auxiliares a partir del texto lematizado/transformado ni en los estudios de entorno. Para facilitar la eliminación de las palabras funcionales se confecciona un texto que contiene todas las palabras funcionales mencionadas en el capítulo 3 y se procesa en el PAFE clasificándolas en la base *Eliminar*. A continuación se procesa en el PAFE el texto a analizar, y se clasifican también los verbos auxiliares, previamente marcados con una x al final, en la base *Eliminar*.

Los textos preparados y lematizados (caso de AO) o transformados (para AF y AG) presentan las siguientes características medidas con el programa PAFE:

UN ESTUDIO DE CASO: UN TEXTO ESCOLAR DE ASTRONOMÍA

Tabla 28

Características de los textos preparados

Texto	V	TTR	N _{léxicas}	LD (%)	Anexo
AO y AF	683	0.28	1378	57.32	C
AG	668	0.27	1411	57.57	L

Nota. V= número de vocablos; TTR= cociente tipo/caso; N_{léxicas}= número de palabras/lemas léxicos; LD= densidad léxica.

Tabla 29

Características de los textos lematizados (AO) y transformados (AF y AG)

Texto	V	TTR	N _{léxicas}	LD (%)	Anexo
FESTEELSE001V	556	0.23	1378	57.32	G
FESTEELSE001V-F	511	0.37	1378	100	M
FESTEELSE001V-S	496	0.35	1411	100	N

Nota. V= número de vocablos; TTR= cociente tipo/caso; N_{léxicas}= número de palabras/lemas léxicos; LD= densidad léxica.

El número de palabras (N) y el de vocablos (V) han sido obtenidos con la aplicación Análisis de Frecuencias del PAFE. Con estos datos se ha calculado el cociente tipo/caso (TTR= V/N) de las Tablas 28 y 29, donde los vocablos son los tipos y las palabras son los casos. El cociente tipo/caso de los textos preparados indica que los mismos tienen una baja riqueza léxica. La alta repetición de palabras sugiere además que existirá un alto número de relaciones entre cada dos palabras.

En el AO el texto lematizado presenta sólo un 81% de los vocablos del texto preparado, integrando el porcentaje restante variantes gráficas de un mismo lema, ya sea en número, género o conjugación verbal. Ello hace que su cociente tipo/caso sea aún más bajo. Algo parecido ocurre con los AF (75%) y AG (74%), donde los transformados difieren del caso del AO en que se han perdido además las palabras funcionales y verbos auxiliares. Por esta misma razón el cociente tipo/caso del texto transformado en los AF y AG es mayor que el del texto preparado, y su riqueza

léxica es mayor. En los AF y AG el cociente mide la riqueza léxica respecto a las palabras léxicas. En estos análisis el texto transformado presenta menos repeticiones que cuando las palabras funcionales y verbos auxiliares están incluidos, y por tanto menor número de relaciones.

El texto transformado en los AF y AG presenta sólo un 57% y 58% de los casos del texto preparado respectivamente. Los porcentajes restantes están constituidos por palabras funcionales.

La unificación de sinónimos no parece influir en el cociente tipo/caso de los textos preparados ni en los textos transformados (ver Tablas 27 y 28). La eliminación de palabras funcionales y verbos auxiliares aumenta el cociente tipo/caso (Tabla 28).

El trabajo de Butler et al. (2004) sobre un corpus de 12 textos de ciencias del curso de 5º de primaria donde se seleccionan cuatro temas (materia, plantas, tormentas y ciclo del agua) encuentra un cociente tipo/caso medio de 0.41. En esta investigación hecha en lengua inglesa no se tuvieron en cuenta ni los números ni los símbolos. En textos universitarios sobre ecología (Conrad, 1996) el cociente tipo/caso es todavía mayor (0.54). Estudios hechos en 8 muestras de texto científico (Tagliacozzo, 1976) indican que los mismos tiene un cociente tipo/caso (≈ 0.30 - 0.38) más bajo que los textos que provienen de revistas y periódicos dirigidos al público general. Asimismo una selección de los artículos de investigación recogidos en el corpus Dascitex (Teich & Fankhauser, 2010) presentan un cociente tipo/caso en torno a 0.34. El texto que se ha escogido para el estudio de caso, aunque escrito para alumnos de secundaria, ha sido escrito desde una institución que se dedica a la

UN ESTUDIO DE CASO: UN TEXTO ESCOLAR DE ASTRONOMÍA

investigación científica. Esto podría explicar que su cociente tipo/caso sea más cercano al de los textos científicos que al de textos dedicados a la educación.

La densidad léxica ($LD = N_{\text{léxicas}}/N$) en los textos preparados y el texto lematizado del AO es prácticamente la misma. La unificación de grafías no supone una gran variación de la cantidad de información. El número de palabras léxicas como el número de palabras totales no varían con la lematización. La eliminación de las palabras funcionales y verbos auxiliares en los AF y AG produce un texto transformado que contiene exclusivamente toda la información del texto preparado.

La media de densidad léxica de un texto escrito es 42% (Wang, 2006). Tanto el texto preparado como el texto lematizado tienen mayor densidad léxica que la media de un texto escrito. En una investigación (Donovan & Smolkin, 2001) hecha sobre libros no escolares seleccionados por profesores para sus clases de ciencias, el libro *The Earth and Sky (Discovery Book)* con contenidos de Astronomía, dio una densidad léxica de un 51%. Este libro fue escogido por profesores de primero a quinto de primaria. El corpus de artículos científicos Dascitex da una media de 65% como densidad léxica (Tools and Deliverables Dascitex). El texto escogido para el caso de estudio presenta más información que los libros de divulgación usados en clase, pero bastante menos que los artículos científicos. La densidad léxica está más cerca de los datos sobre libros de divulgación que los de los artículos científicos, al contrario que lo que ocurre con el cociente tipo/caso.

Los gráficos de frecuencias del texto preparado para todos los análisis son parecidos. En las *Figuras 24* y *25* se pueden ver los correspondientes al AO. La *Figura 24* muestra la frecuencia absoluta tal y como aparece en el listado del Anexo H. Tanto en la *Figura 24* como en la *25* las palabras se encuentran en orden

lexicométrico. En la *Figura 24* la frecuencia absoluta cae bruscamente en un número reducido de palabras, mientras que la distribución hacia la derecha parece tender hacia una asíntota a lo largo de un gran número de palabras. En la *Figura 25* la frecuencia acumulada tiene en el eje de las ordenadas un comportamiento análogo pero invertido. La unificación de grafías produce la variación de la frecuencia de algunas palabras (comparar Anexos H y O). De acuerdo con la metodología (apartado 3. 2. 2. 1. 2.), se pueden distinguir tres franjas de palabras atendiendo a la frecuencia acumulada (ver *Figura 25* y Tabla 30), que se representan también en el gráfico de la frecuencia absoluta (*Figura 24*). Los tres análisis presentan franjas similares en cuanto al número de vocablos y frecuencias, siendo las diferencias mayores las de los vocablos y frecuencia relativa límite en la franja alta y media.

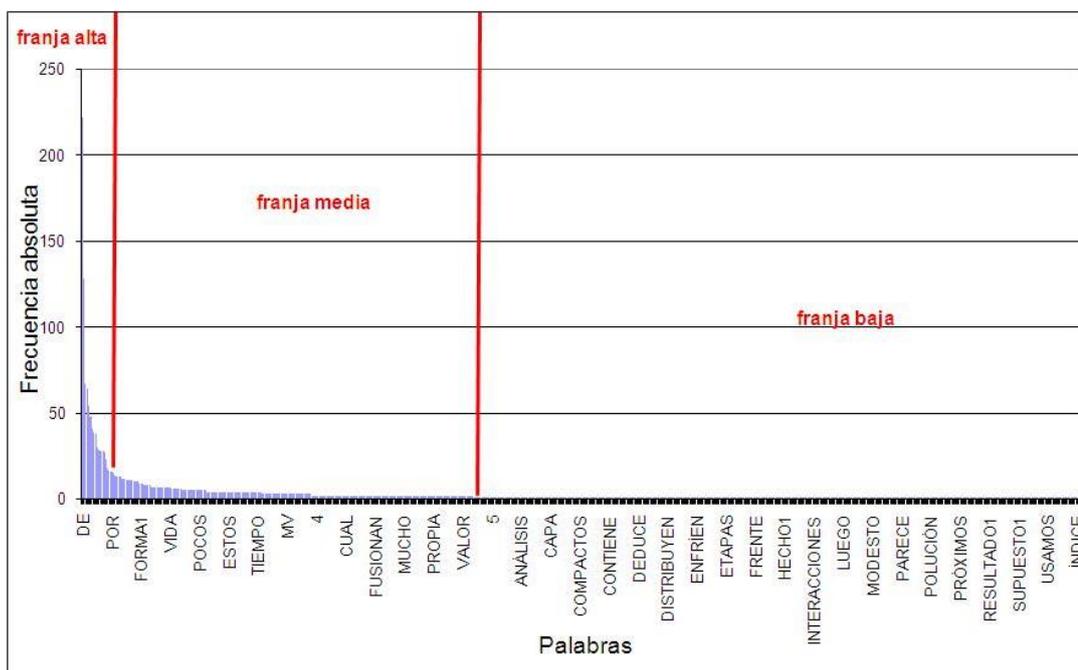


Figura 24.- Frecuencia absoluta de FESTEESE001V.

Otras características de los textos preparados se presentan en la Tabla 31. La franja alta, de frecuencias mayores, está formada por un grupo muy pequeño de vocablos que se repiten mucho y es de esperar que presenten gran número de

UN ESTUDIO DE CASO: UN TEXTO ESCOLAR DE ASTRONOMÍA

relaciones (ver Tablas 30 y 31). No parece haber variaciones entre los AO y AF, y el AG.

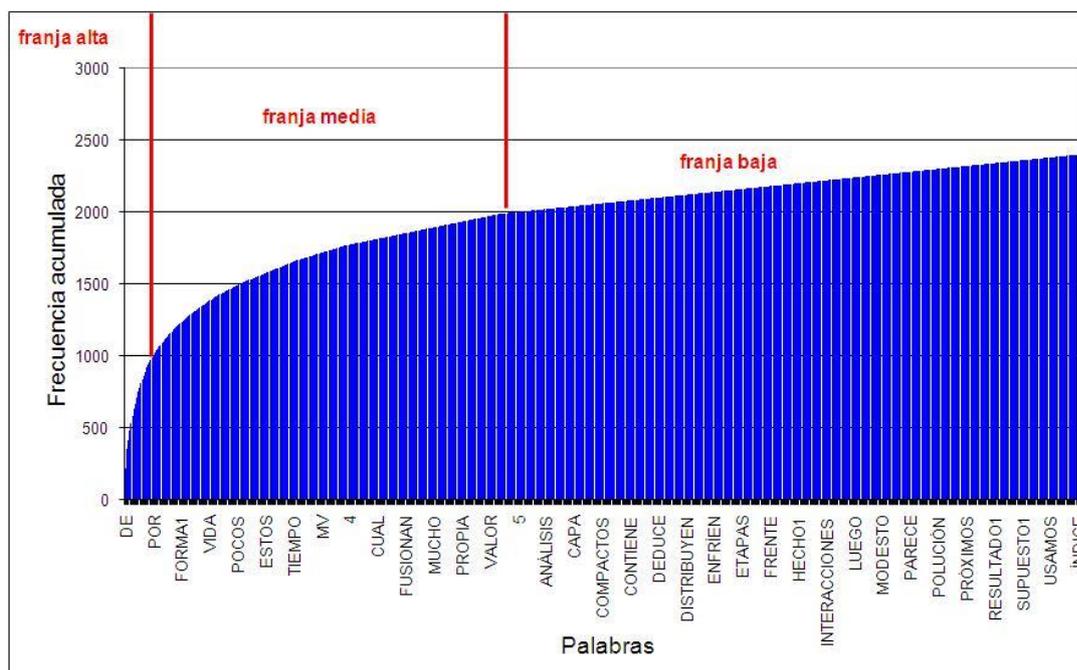


Figura 25.- Frecuencia acumulada de FESTEES001V.

Tabla 30

Franjas en frecuencia acumulada (N) del texto preparado

Análisis	Límites (f)	Vocablos límite	% N°	% $\sum n_i$
Franja alta ($\leq 40\%$ de N)				
AO y AF	7.488-92.346	de-más	2.64	39.93
AG	10.200-93.023	de-cúmulos	2.54	39.53
Franja media ($\leq 75\%$ de N)				
AO y AF	0.832-7.072	para-él	36.60	42.80
AG	0.816-8.568	temperatura-él	35.93	43.70
Franja baja (=100% de N)				
AO y AF	0.416	100-útil	60.76	17.26
AG	0.408	100-útil	61.53	16.77

Nota. N= frecuencia acumulada; f= frecuencia relativa; N°= número de vocablos; n_i = frecuencia absoluta del vocablo i.

Al mirar las categorías gramaticales se observa que los 67 vocablos funcionales producen proporcionalmente muchas más frecuencias que los 616 vocablos léxicos (ver Figuras 26 y 27 del AO y AF). De los 67 vocablos

funcionales 14 son verbos auxiliares (un 2% del total de los vocablos). Los resultados para los distintos análisis se presentan en la Tabla 31.

Tabla 31

Otras características del texto preparado

Análisis	V_{alta}	$V_{hápx}$	$V_{n \leq 2}$	$N^{\circ}_{léxicos}$	$\% \sum n_{i, léxicos}$	$\% N^{\circ}_{func}$	$\% \sum n_{i, func}$
AO y AF	18	415	77.01	90.19	57.32	9.81	42.68
AG	17	411	76.95	89.97	57.57	10.03	42.43

Nota. V_{alta} = número de vocablos en la franja alta; $V_{hápx}$ = número de vocablos que son hápx; $V_{n \leq 2}$ = número de vocablos con $n \leq 2$; $N^{\circ}_{léxicos}$ = número de vocablos léxicos; $n_{i, léxicos}$ = frecuencia absoluta del vocablo léxico i ; N°_{func} = número de vocablos funcionales; $n_{i, func}$ = frecuencia absoluta del vocablo funcional i .

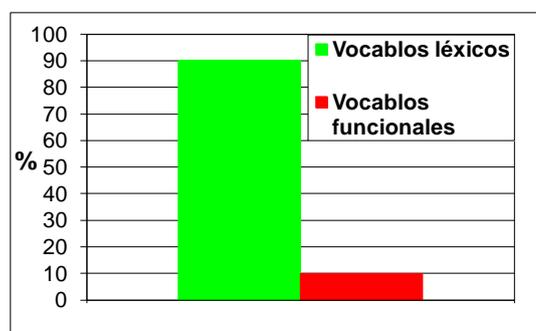


Figura 26.- Número de vocablos por categoría gramatical FESTESE001V.

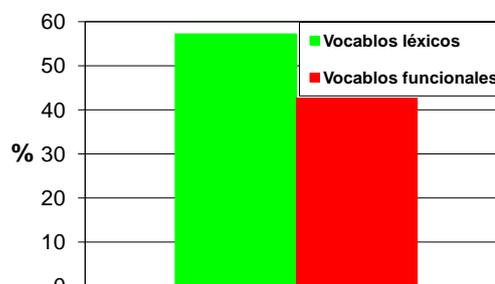


Figura 27.- Frecuencias por categoría gramatical FESTESE001V.

Como las palabras funcionales carecen en su gran mayoría de significado (Aarcos Llorach, 2009; Gómez Tórrego, 2007), los resultados indican la enorme influencia de conocer cómo se construye la lengua para la comprensión del texto.

Cuando se calcula el cociente tipo/caso y la densidad léxica para cada categoría gramatical se obtienen los distintos valores de la Tabla 32. Las cinco últimas filas presentan los resultados de cada subcategoría léxica. Los resultados del TTR indican la alta repetición de vocablos funcionales frente a los léxicos. Por tanto los vocablos funcionales serán parte del vocabulario más frecuente y presentarán mayor número de relaciones. En las subcategorías léxicas, los verbos y los adjetivos son los que menos se repiten, mientras que las repeticiones de sustantivos, adverbios y lenguaje matemático son equivalentes. Estas últimas serán

UN ESTUDIO DE CASO: UN TEXTO ESCOLAR DE ASTRONOMÍA

las más frecuentes y con mayor número de relaciones. La densidad léxica, como indicador de la cantidad de información del texto, se reparte desigualmente por las subcategorías léxicas. Los sustantivos llevan la mayor parte de la información muy por encima de otras subcategorías, por lo que se espera que estén más representados tanto en los vocabularios frecuentes como en las redes semánticas.

Tabla 32

Características por categoría gramatical del texto preparado FESTESE001V

Categorías gramaticales	AO y AF			AG		
	V	TTR	LD (%)	V	TTR	LD (%)
palabras funcionales	67	0.07		67	0.06	
palabras léxicas	616	0.45	57.32	601	0.43	57.57
verbos	126	0.61	8.65	126	0.61	8.49
adverbios	29	0.30	3.99	29	0.30	3.88
adjetivos	182	0.57	13.35	187	0.54	13.46
lenguaje matemático	48	0.35	5.70	42	0.35	4.94
sustantivos	231	0.38	25.62	226	0.34	26.81

Nota. V= número de vocablos; TTR=cociente tipo/caso; LD=densidad léxica.

La unificación de grafías ha provocado una disminución del número de vocablos léxicos, habiendo afectado a sustantivos, adjetivos y lenguaje matemático. Esto produce por un lado un ligero aumento en la frecuencia de adjetivos y sustantivos, que a la vez se repiten más que en el AO, y por otro lado ligeras variaciones en la cantidad de información que llevan las subcategorías, donde aumenta la cantidad de información de adjetivos y sustantivos. Las diferencias más importantes respecto al TTR se dan en sustantivos y adjetivos, lo que indica una mayor cantidad de unificación de grafías en estas dos subcategorías. Como ya se vio en la Tabla 25 varias palabras del lenguaje matemático han sido sustituidas por sustantivos, influyendo en las diferencias respecto a la LD en sustantivos y lenguaje matemático. Los valores de TTR y LD apenas varían entre textos preparados.

UN ESTUDIO DE CASO: UN TEXTO ESCOLAR DE ASTRONOMÍA

A continuación se presentan los resultados de las categorías gramaticales y subcategorías léxicas en las tres franjas de frecuencia. En las siguientes figuras se puede observar la importancia de las categorías gramaticales tanto en número de palabras como en frecuencias producidas en el AO. Las distribuciones en número de vocablos se presentan en las *Figuras 28, 30 y 32* y las de frecuencias en las *Figuras 29, 31 y 33*. En los datos que corresponden a la *Figura 28* los verbos auxiliares se sitúan a partes iguales en la franja media y en la franja baja.

Las gráficas para los AF y AG son parecidas, presentándose los resultados en las Tablas 33, 34 y 35, para facilitar la comparación. Las variaciones de valores en número de vocablos y frecuencias entre los distintos análisis no son importantes.

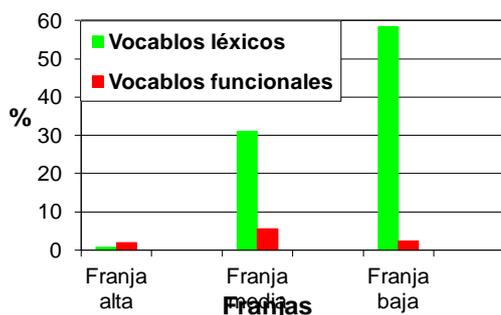


Figura 28.- Número de vocablos por categoría gramatical en todo el texto de FESTESE001V.

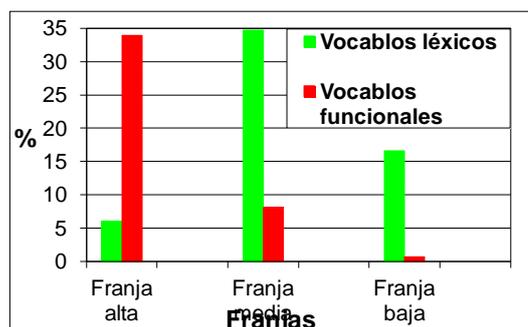


Figura 29.- Frecuencias por categoría gramatical en todo el texto de FESTESE001V.

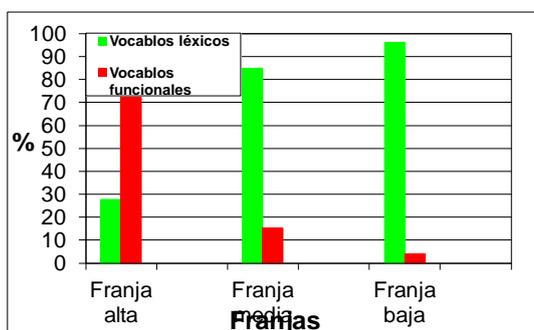


Figura 30.- Número de vocablos por categoría gramatical en cada franja de FESTESE001V.

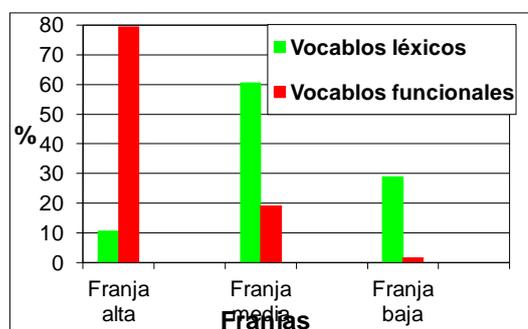


Figura 31.- Frecuencias por categoría gramatical en cada franja de FESTESE001V.

UN ESTUDIO DE CASO: UN TEXTO ESCOLAR DE ASTRONOMÍA

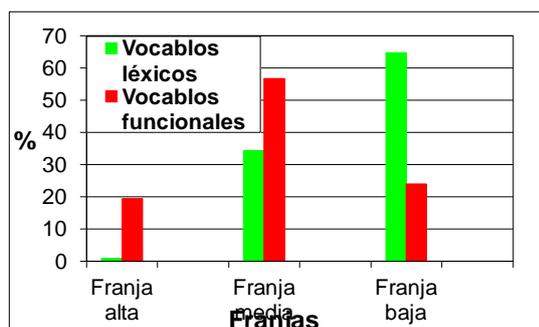


Figura 32.- Número de vocablos de cada categoría gramatical en todo el texto de FESTESE001V.

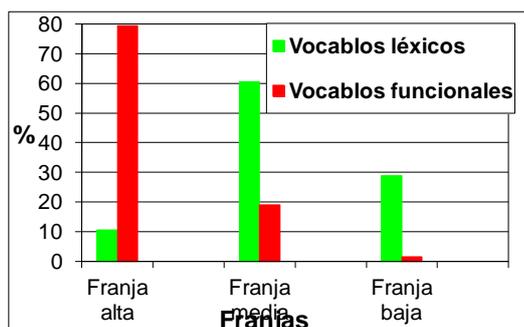


Figura 33.- Frecuencias de cada categoría gramatical en todo el texto de FESTESE001V.

Tabla 33

Distribución del número de vocablos y frecuencias en todo el texto preparado en los tres análisis

Análisis	% N ^o léxicos	% N ^o funcionales	% $\sum n_i$ léxicas	% $\sum n_i$ funcionales
Franja alta ($\leq 40\%$ de N)				
AO y AF	0.73	1.90	6.03	33.90
AG	0.60	1.95	5.43	34.11
Franja media ($\leq 75\%$ de N)				
AO y AF	31.04	5.56	34.69	8.11
AG	30.39	5.54	36.07	7.63
Franja baja (=100% de N)				
AO y AF	58.42	2.34	16.60	0.67
AG	58.98	2.54	16.08	0.69

Nota. N^o léxicos = número de vocablos léxicos; N^o funcionales = número de vocablos funcionales; n_i léxicas = frecuencia absoluta del vocablo léxico i ; n_i funcionales = frecuencia absoluta del vocablo funcional i .

Los vocablos funcionales predominan en la franja alta debido al alto número de frecuencias que produce un pequeño número de los mismos. Los vocablos léxicos muestran una tendencia a concentrarse en las franjas media y baja tanto en número como en frecuencias. En una metodología donde se va a precisar un sistema óptimo en función de la frecuencia para estudiar relaciones, el mismo estará dominado por vocablos funcionales.

Tabla 34

Distribución del número de vocablos y frecuencias en cada franja en los tres análisis en el texto preparado

Análisis	% N ^o léxicos	% N ^o funcionales	% $\sum n_i$, léxicas	% $\sum n_i$, funcionales
Franja alta ($\leq 40\%$ de N)				
AO y AF	27.78	72.22	15.10	84.90
AG	23.53	76.47	13.73	86.27
Franja media ($\leq 75\%$ de N)				
AO y AF	84.80	15.20	81.05	18.95
AG	84.58	15.42	82.54	17.46
Franja baja (=100% de N)				
AO y AF	96.14	3.86	96.14	3.86
AG	95.86	4.14	95.86	4.14

Nota. N^o léxicos = número de vocablos léxicos; N^o funcionales = número de vocablos funcionales; n_i, léxicas = frecuencia absoluta del vocablo léxico i; n_i, funcionales = frecuencia absoluta del vocablo funcional i.

Tabla 35

Distribución de cada categoría gramatical en todo el texto preparado en los tres análisis

Análisis	% N ^o léxicos	% N ^o funcionales	% $\sum n_i$, léxicas	% $\sum n_i$, funcionales
Franja alta ($\leq 40\%$ de N)				
AO y AF	0.81	19.40	10.52	79.43
AG	0.67	19.40	9.43	80.38
Franja media ($\leq 75\%$ de N)				
AO y AF	34.42	56.72	60.52	19.01
AG	33.78	55.22	62.65	17.98
Franja baja (=100% de N)				
AO y AF	64.77	23.88	28.96	1.56
AG	65.56	25.37	27.92	1.63

Nota. N^o léxicos = número de vocablos léxicos; N^o funcionales = número de vocablos funcionales; n_i, léxicas = frecuencia absoluta del vocablo léxico i; n_i, funcionales = frecuencia absoluta del vocablo funcional i.

Dentro de los vocablos léxicos hay cinco subcategorías bien diferenciadas: verbos, adverbios, adjetivos, sustantivos y lo que aquí se llama lenguaje matemático. Este último se entiende como los signos matemáticos, incluidos los

UN ESTUDIO DE CASO: UN TEXTO ESCOLAR DE ASTRONOMÍA

números, y símbolos de magnitudes y de sistemas de unidades de medidas. El símbolo, al igual que la palabra correspondiente, no son más que diferentes tipos de representación de un mismo concepto. Para la clasificación de los vocablos léxicos en las subcategorías mencionadas se ha consultado el Diccionario de la RAE online (2001). Se ha cogido la primera clasificación morfológica que aparecía en dicho diccionario; en caso de duda, se ha consultado el texto para decidir la clasificación más conveniente. Los nombres propios han sido clasificados automáticamente como sustantivos. La clasificación de los vocablos léxicos se puede ver en la quinta columna de las Tablas 95 y 98 (Anexos H y O). Se encontraron cuatro vocablos que no están en el diccionario: *adecuadamente*, *gravitatoriamente*, *mayoritariamente* y *protogaláctico*. Dos de éstos tienen su origen en el lenguaje científico. En el caso de los verbos se ha distinguido entre auxiliares y no auxiliares, es decir, los verbos auxiliares se han considerado vocablos funcionales. La importancia relativa de cada subcategoría en el AO se observa en las *Figuras 34 y 35*.

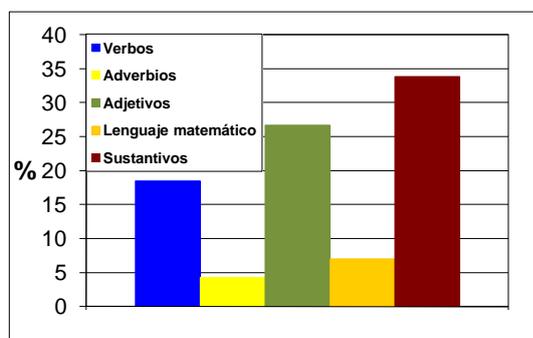


Figura 34.- Número de vocablos por subcategoría léxica de FESTESE001V.

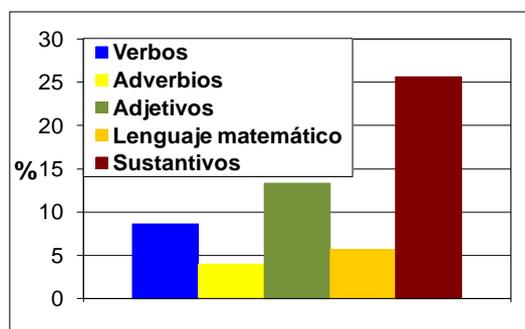


Figura 35.- Frecuencias por subcategoría léxica de FESTESE001V.

Las gráficas para los AF y AG son parecidas, por lo que se presentan los resultados en la Tabla 36 para facilitar la comparación. No hay grandes variaciones de valores para el número de vocablos y frecuencias entre los distintos análisis.

Tabla 36

Distribución de las subcategorías léxicas en los tres análisis en el texto preparado

Subcategorías léxicas	AO y AF		AG	
	% N°	% $\sum n_i$	% N°	% $\sum n_i$
verbos	18.45	8.65	18.86	8.49
adverbios	4.25	3.99	4.34	3.88
adjetivos	26.65	13.35	26.65	13.46
lenguaje matemático	7.03	5.70	6.29	4.94
sustantivos	33.82	25.62	33.83	26.81

Nota. N°= número de vocablos; n_i = frecuencia absoluta del vocablo i.

Las subcategorías léxicas presentan una menor diferencia entre el número de vocablos y las frecuencias producidas, comparado con los vocablos funcionales. Esto se debe a la escasa repetición de los vocablos léxicos en comparación con los funcionales (ver *Figuras 26 y 27*, Tabla 31). A pesar de que verbos, adjetivos y sustantivos tienen mayor número de casos que los vocablos funcionales, éstos producen más frecuencias (un 43%, ver *Figura 27* y Tabla 32).

Las distribuciones en número de vocablos en el AO se presentan en las *Figuras 36, 38 y 40*, y las de frecuencias en las *Figuras 37, 39 y 41*. El comportamiento de cada subcategoría léxica no está correlacionado con la tendencia general de su categoría, excepto sustantivos y adjetivos, que en general lo están tanto en número de palabras como en frecuencias. Los sustantivos no difieren mucho de la tendencia general (ver *Figura 40*, donde los valores en la franja media y baja son 34.40% y 34.22% respectivamente). El predominio de adjetivos y sustantivos se puede deducir del hecho que estas dos subcategorías constituyen un 67% del número de vocablos y un 68% de las frecuencias de la categoría léxica (Tabla 40). Los verbos se desvían de la tendencia general en cuanto a las frecuencias producidas, mientras que el lenguaje matemático y los adverbios lo hacen tanto en número de palabras como en frecuencias. Las distribuciones en

UN ESTUDIO DE CASO: UN TEXTO ESCOLAR DE ASTRONOMÍA

número de vocablos y frecuencias del lenguaje matemático parecen estar correlacionadas en las franjas, cosa que no ocurre con el resto de subcategorías.

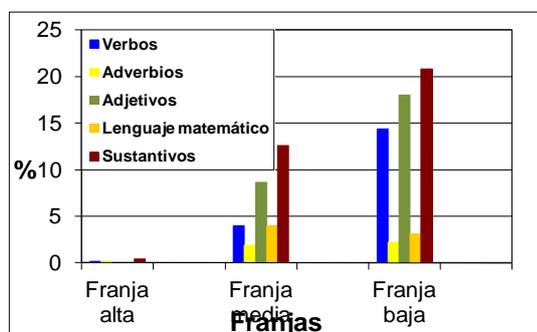


Figura 36.- Número de vocablos por subcategoría léxica en todo el texto de FESTESE001V.

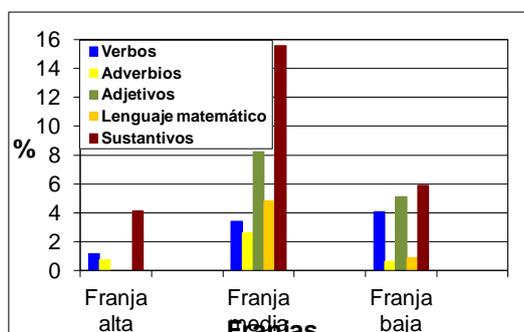


Figura 37.- Frecuencias por subcategoría léxica en todo el texto de FESTESE001V.

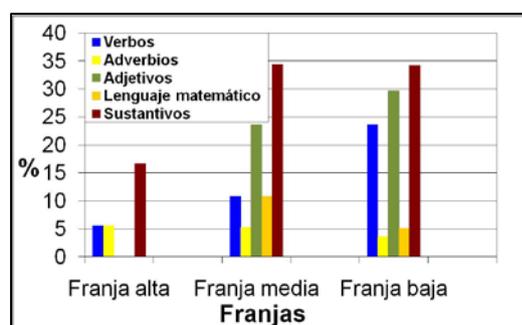


Figura 38.- Número de vocablos por subcategoría léxica en cada franja de FESTESE001V.

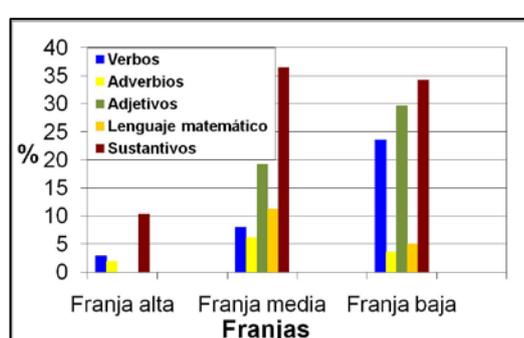


Figura 39.- Frecuencias por subcategoría léxica en cada franja de FESTESE001V.

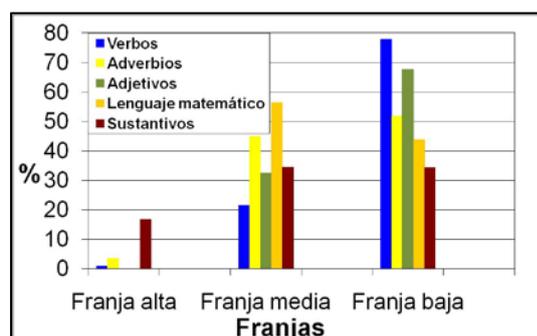


Figura 40.- Número de vocablos de cada subcategoría léxica en todo el texto de FESTESE001V.

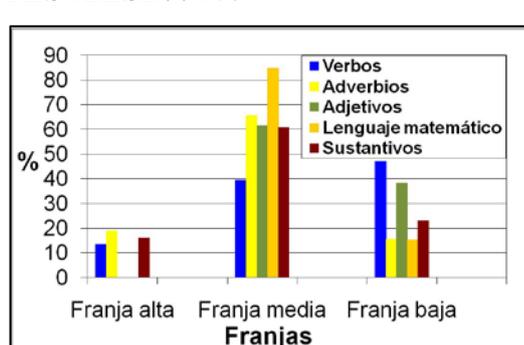


Figura 41.- Frecuencias de cada subcategoría léxica en todo el texto de FESTESE001V.

Las gráficas para los AF y AG son parecidas, por lo que se presentan los resultados en las Tablas 37, 38 y 39, para facilitar la comparación.

UN ESTUDIO DE CASO: UN TEXTO ESCOLAR DE ASTRONOMÍA

Tabla 37

Distribución de las subcategorías léxicas en todo el texto preparado en los tres análisis

Análisis	AO y AF		AG	
	% N°	% $\sum n_i$	% N°	% $\sum n_i$
Franja alta ($\leq 40\%$ de N)				
verbos	0.15	1.16	0.15	1.14
adverbios	0.15	0.75	0.00	0.00
adjetivos	0.00	0.00	0.00	0.00
lm	0.00	0.00	0.00	0.00
sustantivos	0.44	4.12	0.45	4.28
Franja media ($\leq 75\%$ de N)				
verbos	3.95	3.41	4.04	3.35
adverbios	1.90	2.62	2.10	3.26
adjetivos	8.64	8.24	8.38	8.49
lm	3.95	4.83	3.14	4.08
sustantivos	12.59	15.60	12.72	16.89
Franja baja (=100% de N)				
verbos	14.35	4.08	14.67	4.00
adverbios	2.20	0.62	2.25	0.61
adjetivos	18.01	5.12	18.26	4.98
lm	3.07	0.87	3.14	0.86
sustantivos	20.79	5.91	20.66	5.63

Nota. N°= número de vocablos; n_i = frecuencia absoluta del vocablo i ; lm= lenguaje matemático.

Tabla 38

Distribución de las subcategorías léxicas en cada franja en los tres análisis en el texto preparado

Análisis	AO y AF		AG	
	% N°	% $\sum n_i$	% N°	% $\sum n_i$
Franja alta ($\leq 40\%$ de N)				
verbos	5.56	2.92	5.88	2.89
adverbios	5.56	1.88	0.00	0.00
adjetivos	0.00	0.00	0.00	0.00
lm	0.00	0.00	0.00	0.00
sustantivos	16.67	10.31	17.65	10.84
Franja media ($\leq 75\%$ de N)				
verbos	10.80	7.97	11.25	7.66
adverbios	5.20	6.12	5.83	7.47
adjetivos	23.60	19.24	23.33	19.42
lm	10.80	11.27	8.75	9.34
sustantivos	34.40	36.44	35.42	38.66

UN ESTUDIO DE CASO: UN TEXTO ESCOLAR DE ASTRONOMÍA

Análisis	AO y AF		AG	
	% N°	% $\sum n_i$	% N°	% $\sum n_i$
Franja baja (=100% de N)				
verbos	23.61	23.61	23.84	23.84
adverbios	3.61	3.61	3.65	3.65
adjetivos	29.64	29.64	29.68	29.68
lm	5.06	5.06	5.11	5.11
sustantivos	34.22	34.22	33.58	33.58

Nota. N°= número de vocablos; n_i = frecuencia absoluta del vocablo i ; lm= lenguaje matemático.

Tabla 39

Distribución de cada subcategoría léxica en todo el texto preparado en los tres análisis

Análisis	AO y AF		AG	
	% N°	% $\sum n_i$	% N°	% $\sum n_i$
Franja alta ($\leq 40\%$ de N)				
verbos	0.79	13.46	0.79	13.46
adverbios	3.45	18.75	0.00	0.00
adjetivos	0.00	0.00	0.00	0.00
lm	0.00	0.00	0.00	0.00
sustantivos	1.30	16.07	1.33	15.98
Franja media ($\leq 75\%$ de N)				
verbos	21.43	39.42	21.43	39.42
adverbios	44.83	65.63	48.28	84.21
adjetivos	32.42	61.68	31.46	63.03
lm	56.25	84.67	50.00	82.64
sustantivos	37.23	60.88	37.61	63.01
Franja baja (=100% de N)				
verbos	77.78	47.12	77.78	47.12
adverbios	51.72	15.63	51.72	15.79
adjetivos	67.58	38.32	68.54	36.97
lm	43.75	15.33	50.00	17.36
sustantivos	61.47	23.05	61.06	21.00

Nota. N°= número de vocablos; n_i = frecuencia absoluta del vocablo i ; lm= lenguaje matemático.

El comportamiento de las subcategorías léxicas es análogo en los 3 análisis para verbos, adjetivos y sustantivos. En el caso del AG sólo los adjetivos están correlacionados con la tendencia general de su categoría. Sustantivos, adverbios y lenguaje matemático difieren de la tendencia de la categoría léxica en número de

vocablos y frecuencias, mientras que los verbos sólo lo hacen en frecuencias. El número de vocablos del lenguaje matemático se encuentra repartido entre las franjas media y baja a partes iguales. La mayor diferencia entre los AO y AF respecto al AG es la desaparición de los adverbios de la franja alta, debido a la redistribución de frecuencias.

Otros datos relativos a la importancia de cada subcategoría léxica respecto a su categoría se presentan en la Tabla 40, y respecto a la franja alta en la Tabla 41. En las 126 formas verbales, no se han tenido en cuenta las correspondientes a los verbos auxiliares. Si las añadimos se tendrían 140 formas verbales, de los cuales un 10% son auxiliares.

Tabla 40

Importancia de las subcategorías léxicas

Subcategorías léxicas	AO y AF			AG		
	Nº	%	%	Nº	%	%
	Nº	% N ^o léxicas	$\sum n_{i,léxicas}$	Nº	N ^o léxicas	$\sum n_{i,léxicas}$
verbos	126	20.45	15.09	126	20.97	14.74
adverbios	29	4.71	6.97	29	4.83	6.73
adjetivos	182	29.55	23.29	178	29.62	23.39
lenguaje matemático	48	7.79	9.94	42	6.99	8.58
sustantivos	231	37.5	44.70	226	37.60	46.56

Nota. N^o= número de vocablos de la subcategoría; N^o léxicas= número de vocablos de la subcategoría respecto a la categoría léxica; $\sum n_{i,léxicas}$ = frecuencias de la subcategoría

Tabla 41

Importancia de las subcategorías léxicas en la franja alta

Subcategorías léxicas	AO y AF		AG	
	V _{alta}	% $\sum n_{i,subcategoría}$	V _{alta}	% $\sum n_{i,subcategoría}$
verbos	<i>es</i>	13.46	<i>es</i>	13.46
adverbios	<i>más</i>	18.75		0.00
sustantivos	<i>estrellas,</i> <i>estrella,</i> <i>cúmulos</i>	16.07	<i>estrellas,</i> <i>estrella,</i> <i>cúmulos</i>	15.98

Nota. V_{alta} = vocablos en la franja alta; $\sum n_{i,subcategoría}$ = frecuencias de la subcategoría en la franja alta.

4. 3.- Análisis de Frecuencias del Texto Lematizado/Transformado

Las diferencias en los gráficos de frecuencias del texto lematizado (AO) o transformado (AF y AG) en los distintos análisis se deben a la supresión de las palabras funcionales y verbos auxiliares en los AF y AG (ver listados lexicométricos en Anexos I, P y Q). Las diferencias entre estos dos últimos se deben a la unificación de grafías. Los gráficos de frecuencias del texto lematizado en el AO se pueden ver en las *Figuras 42 y 43*. La *Figura 42* muestra la frecuencia absoluta tal y como aparece en el listado del Anexo I. Tanto en la *Figura 42* como en la *Figura 43* los lemas se encuentran en orden lexicométrico. Ambas figuras son parecidas a las del texto preparado (ver *Figuras 24 y 25*). En la *Figura 42* la frecuencia absoluta cae bruscamente en un número reducido de palabras, mientras que la distribución hacia la derecha parece tender hacia una asíntota a lo largo de un gran número de lemas. En la *Figura 43* la frecuencia acumulada tiene en el eje de las ordenadas un comportamiento análogo pero invertido. De acuerdo con la metodología (apartado 3. 2. 2. 1. 2.) se pueden distinguir tres franjas de lemas atendiendo a la frecuencia acumulada (ver *Figura 43* y Tabla 42), que se representan también en el gráfico de la frecuencia absoluta (*Figura 42*). La eliminación de palabras hace que todas las franjas en los AF y AG sólo tengan lemas léxicos. Nótese cómo cambian de lugar los lemas en la lista del texto lematizado/tranformado en comparación con la lista de frecuencias de las palabras del texto preparado (ver Anexos H y I para el AO, Anexos H y P para el AF y Anexos O y Q para el AG).

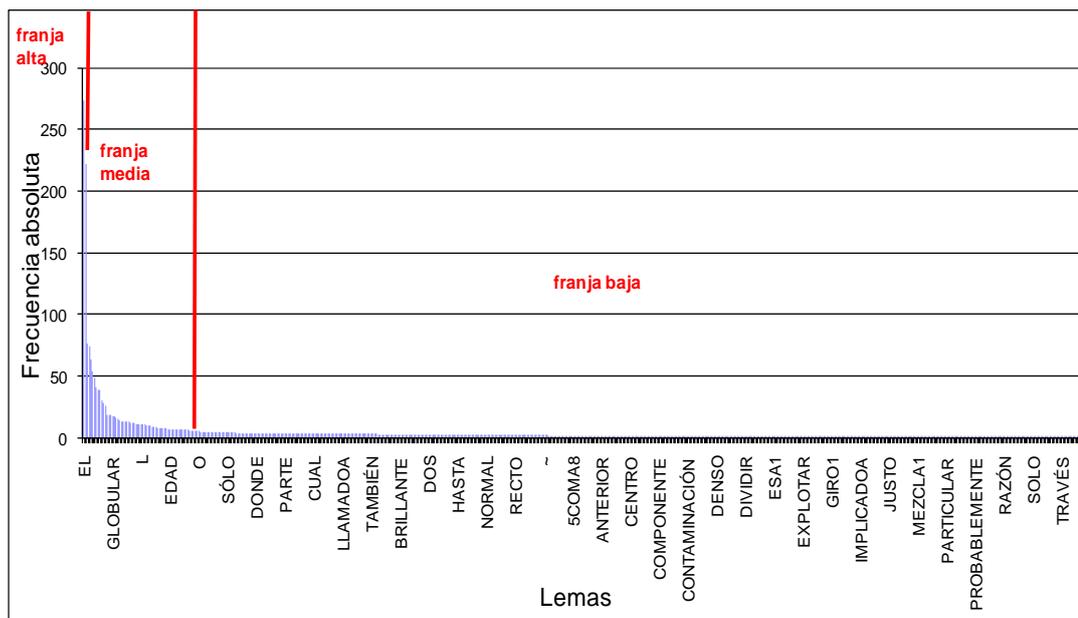


Figura 42.- Frecuencia absoluta del texto lematizado de FESTESE001V.

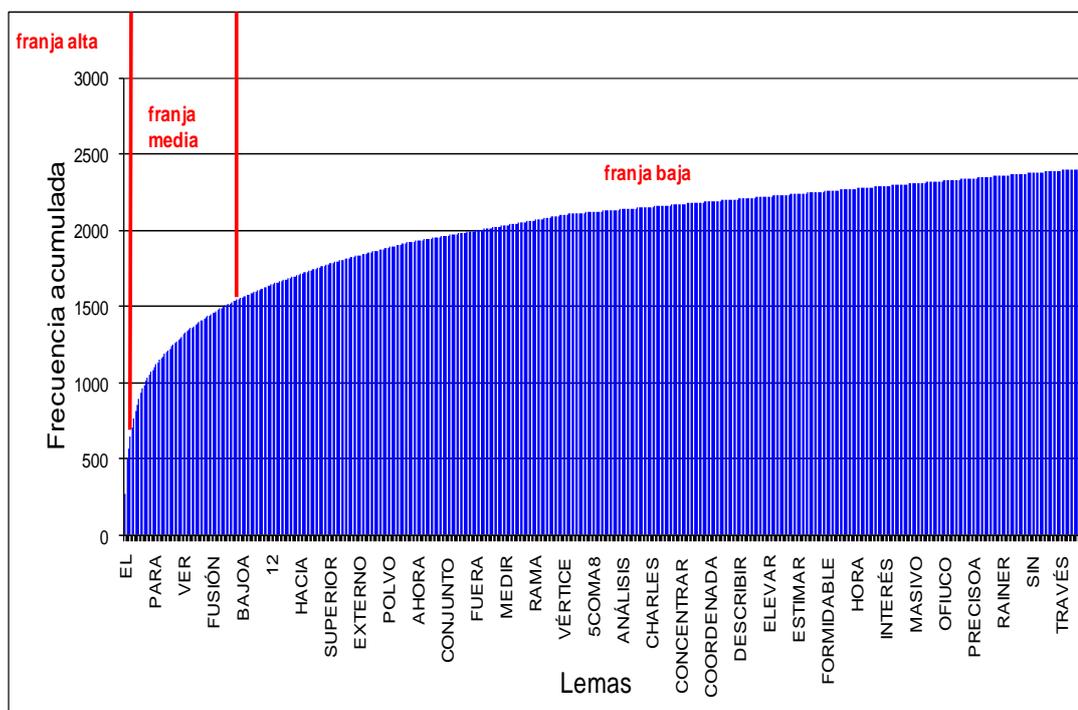


Figura 43.- Frecuencia acumulada del texto lematizado de FESTESE001V.

Debido a la eliminación de palabras funcionales y verbos auxiliares en los AF y AG (un 57-59% de las frecuencias), los límites de las franjas en frecuencia relativa han disminuido. Los porcentajes en número de lemas y frecuencias entre el AO y los AF y AG no son comparables por la misma razón. Por ello se ha añadido el dato de dichos porcentajes para los lemas léxicos de AO. El uso de la base

UN ESTUDIO DE CASO: UN TEXTO ESCOLAR DE ASTRONOMÍA

Eliminar produce una redistribución de los lemas léxicos donde parte de los lemas que aparecían en la franja media y baja del AO se han desplazado a la franja alta y media, respectivamente, en los AF y AG. La franja media ha aumentado hasta un 305-234%, mientras que la franja baja ha disminuido hasta un 80-86%. De la misma manera las frecuencias presentes en las franjas se han modificado: la franja media ha aumentado a un 115-108% y la franja baja ha disminuido hasta un 55-63%.

Tabla 42

Franjas en frecuencia acumulada (N) del texto lematizado de FESTESE001V

Análisis	Límites (f)	Vocablos límite	% N°	% $\sum n_i$
Franja alta ($\leq 15-20\%$ de N)				
AO ^a	113.561	el	0.18 (0.00)	11.36 (0.00)
AF	9.434-55.152	estrella-x	2.35	21.04
AG	10.631-56.697	estrella-superficial	1.81	19.56
Franja media ($\leq 60\%$ de N)				
AO	2.912-92.346	de-vida	10.25 (8.02)	50.96 (39.27)
AF	2.177-8.708	aproximadamente-vez	24.46	45.72
AG	2.835-9.922	masa-tipo	18.75	42.24
Franja baja (=100% de N)				
AO	0.416-2.496	cuando-útil	89.57 (91.98)	37.69 (60.23)
AF	0.726-1.451	0coma7-útil	73.19	33.24
AG	0.709-2.126	alguno-útil	79.44	38.20

Nota. N= frecuencia acumulada; f= frecuencia relativa; N°= número de vocablos; n_i = frecuencia absoluta del vocablo i.

^aLos porcentajes entre paréntesis del AO se refieren a los lemas léxicos del mismo.

Las diferencias entre el AF y AG son menores y se deben a la redistribución de lemas producida por la unificación de grafías. La redistribución de lemas y frecuencias produce un desplazamiento de ambas hacia las franjas de frecuencias menores (media y baja). En el AO la mayoría de las frecuencias de los lemas léxicos se encuentra en la franja baja, en los AF y AG esta mayoría se encuentra en la franja media. Los tres análisis coinciden en presentar el máximo y el mínimo número de lemas en la franja baja y alta, respectivamente.

UN ESTUDIO DE CASO: UN TEXTO ESCOLAR DE ASTRONOMÍA

Al comparar los tres análisis se ve que en términos de frecuencia se resalta el uso de los lemas léxicos entre sí en los AF y AG. Dado que esta investigación se centra en el análisis de contenido, resaltar el uso de los lemas léxicos proporcionará un análisis de contenido más rico.

El AO es el único que presenta lemas léxicos y funcionales en el texto lematizado, en una proporción en número de lemas de 91.91 y 8.09% y en frecuencias de 57.32 y 42.68% respectivamente. Otras características de los textos lematizado (AO) y transformados (AF y AG) se presentan en la Tabla 43. La franja alta, de frecuencias mayores, está formada por un grupo muy pequeño de vocablos que se repiten mucho y es de esperar que presenten gran número de relaciones (ver Tablas 42 y 43). En el caso del AF la franja baja está formada en su totalidad por lemas que aparecen 1 o 2 veces. En los AF y AG los hápax forman más de la mitad de la franja baja. Al igual que las franjas altas de los tres análisis son bastante distintas en cuanto al número de lemas, las franjas bajas lo son en cuanto a las frecuencias. En el AO las palabras funcionales (*el*) producen las frecuencias más altas, mientras que en los AF y AG las franjas en frecuencia están constituidas única y exclusivamente por lemas léxicos:

- AF: *estrella, ser, cúmulo, núcleo, más, globular, temperatura, masa, muy, principal, secuencia, x.*
- AG: *estrella, cúmulo, ser, globular, temperatura, luminosidad, núcleo, más, superficial.*

El AF y el AG tienen en común 7 lemas en la franja alta, que se han señalado subrayándose. La aparición de los vocablos *luminosidad* y *superficial* en el último se debe a la unificación de las grafías. Estos vocablos se encontraban antes de dos formas: como palabra y como símbolo matemático. En términos de

UN ESTUDIO DE CASO: UN TEXTO ESCOLAR DE ASTRONOMÍA

frecuencia la eliminación de palabras funcionales y verbos auxiliares resalta el uso de los lemas léxicos entre sí.

Tabla 43

Otras características del texto lematizado/transformado

Análisis	V _{alta}	V _{hápax}	% V _{n≤2}	% V _{hápax}	% $\sum n_{i, hápax}$
AO ^a	1 (0)	297 (290)	70.50 (67.27)	59.64 (61.70)	32.78 (34.94)
AF	12	290	73.19	77.54	63.32
AG	9	283	72.58	71.83	52.50

Nota. V_{alta}= número de vocablos en la franja alta; V_{hápax}= número de vocablos que son hápax; V_{n≤2}= número de vocablos con n≤2; % V_{hápax}= porcentaje de hápax en la franja baja; % $\sum n_{i, hápax}$ = porcentaje de frecuencias de los hápax en la franja baja.

^aLos valores entre paréntesis del AO se refieren a los lemas léxicos del mismo.

En el AO, al igual que pasaba con el texto preparado, se observa una gran cantidad de lemas procedentes de palabras léxicas (511) que proporcionalmente producen menos frecuencias que los 45 lemas funcionales. A partir de ahora se denominan lemas léxicos a los procedentes de palabras léxicas y lemas funcionales a los procedentes de palabras funcionales. De los 45 lemas funcionales, 7 son verbos auxiliares (un 1% de los lemas totales y un 16% de los lemas funcionales), de los cuales 2 se encuentran en la franja media (*poder, haber*) y el resto en la franja baja; de todos ellos, 2 tienen exclusivamente la función de verbo auxiliar (*poder, llegar*), 3 son verbos auxiliares la mayoría de las veces (*haber, comenzar, ir*), 1 lo es casi la mitad de las veces (*estar*) y el último apenas ejerce de verbo auxiliar (*ser*, 2 de 43 veces).

En el caso de los AF y AG todos los lemas del texto transformado son léxicos, lo cual permite centrar la atención en dicho tipo de lemas y resaltar el uso de los mismos entre sí.

Cuando se calcula el cociente tipo/caso y la densidad léxica para cada categoría gramatical se obtienen los distintos valores de la Tabla 44. Las cinco

últimas filas presentan los resultados de cada subcategoría léxica. El número de vocablos de las distintas subcategorías ha disminuido sólo para adjetivos, lenguaje matemático y sustantivos en el AG, puesto que fueron las subcategorías afectadas por la unificación de grafías, manteniéndose igual para el resto de las subcategorías en los tres análisis. La lematización ha hecho que tanto los lemas funcionales como los léxicos se repitan todavía más que en el texto preparado, por lo que su número de relaciones ha aumentado. En los textos transformados (AF y AG) han desaparecido los lemas funcionales, sin embargo los cocientes tipo/caso son similares para los tres análisis en las distintas categorías y subcategorías. La disminución del TTR en el AG se debe a la disminución de tipos debido a la unificación de grafías. Sólo se aprecian diferencias en los adverbios, cuyo TTR aumenta en el AG, y en adjetivos y sustantivos, cuyo TTR disminuye en el AG. En el AO y AF las subcategorías de adverbios y lenguaje matemático no varían su cociente tipo/caso respecto al texto preparado, ya que la lematización no les afecta, pues no existen variaciones gramaticales de ellos. El resto de los cocientes tipo/caso han disminuido, es decir, la lematización hace que disminuya el número de vocablos y que aumente su frecuencia. La disminución de vocablos es mayor en los lemas léxicos que en los funcionales, ya que la variación gramatical de los lemas léxicos es mayor. La disminución del cociente tipo/caso es mayor en verbos que en adjetivos y sustantivos, pues hay más formas verbales que formas de éstos últimos.

La densidad léxica, como indicador de la cantidad de información del texto, se reparte desigualmente por las subcategorías léxicas, y es igual en valores a la del texto preparado en el AO, puesto que la lematización no altera la cantidad de información. Debido a la supresión de palabras funcionales y verbos auxiliares la densidad léxica en el texto transformado del AF y AG es del 100%, es decir, estos

UN ESTUDIO DE CASO: UN TEXTO ESCOLAR DE ASTRONOMÍA

textos contienen exclusivamente toda la información del texto preparado. También han aumentado proporcionalmente las densidades léxicas de las subcategorías. Se puede ver como los sustantivos llevan la mayor parte de la información muy por encima de otras subcategorías, llegando casi al 50% de la información en los AF y AG, por lo que se espera que estén más representados tanto en los vocabularios frecuentes como en las redes semánticas.

Tabla 44

Características por categoría gramatical del texto lematizado o transformado

Categoría gramatical	AO		AF		AG	
	TTR	LD (%)	TTR	LD (%)	TTR	LD (%)
lemas funcionales	0.04					
lemas léxicos	0.37	57.32	0.37	100.00	0.35	100
verbos	0.43	8.65	0.43	15.09	0.43	14.74
adverbios	0.30	3.99	0.30	6.97	0.31	6.73
adjetivos	0.44	13.35	0.44	23.29	0.42	23.39
lenguaje matemático	0.35	5.70	0.35	9.94	0.35	8.58
sustantivos	0.33	25.62	0.33	44.70	0.30	46.56

Nota. TTR=cociente tipo/caso; LD=densidad léxica.

La influencia de las transformaciones que sufre el texto, debido a las variaciones en la metodología, se estudia en las tres franjas de frecuencia en función de las categorías gramaticales y subcategorías léxicas. En las siguientes figuras se puede observar la importancia de las categorías gramaticales tanto en número de ULL como en frecuencias producidas en el AO. Las distribuciones en número de lemas se presentan en las *Figuras 44, 46 y 48* y las de frecuencias en las *Figuras 45, 47 y 49*. En los datos que corresponden a la *Figura 44* hay menos verbos auxiliares en la franja media (un 4% de los lemas funcionales) que en la franja baja (un 11% de los lemas funcionales).

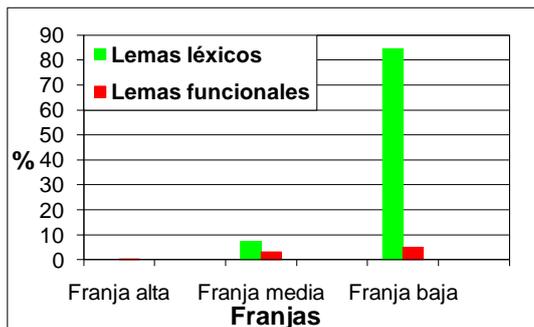


Figura 44.- Número de lemas por categoría gramatical en todo el texto de FESTESE001V.

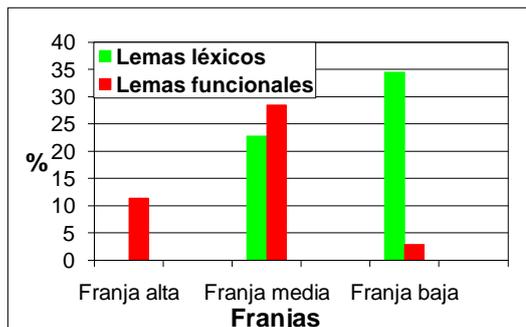


Figura 45.- Frecuencias por categoría gramatical en todo el texto de FESTESE001V.

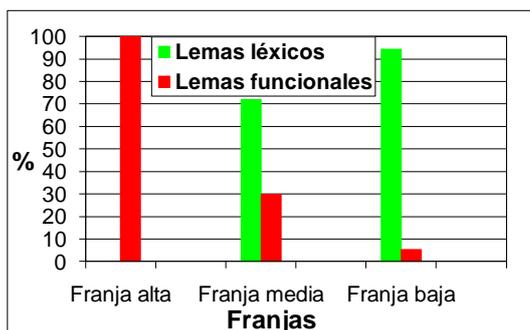


Figura 46.- Número de lemas por categoría gramatical en cada franja de FESTESE001V.

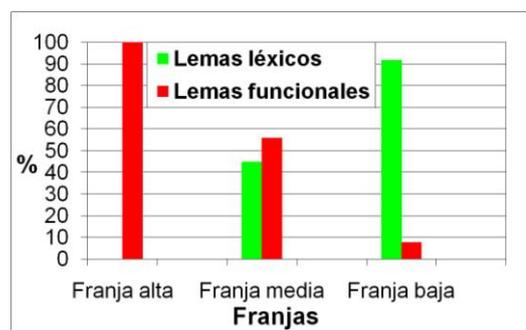


Figura 47.- Frecuencias por categoría gramatical en cada franja de FESTESE001V.

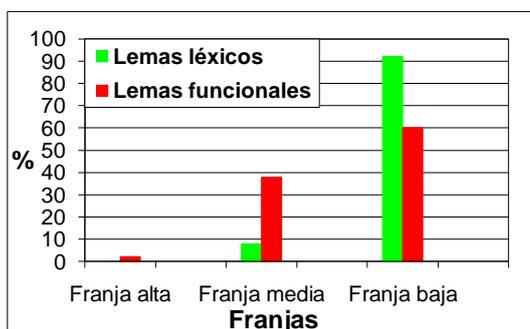


Figura 48.- Número de lemas de cada categoría gramatical en todo el texto de FESTESE001V.

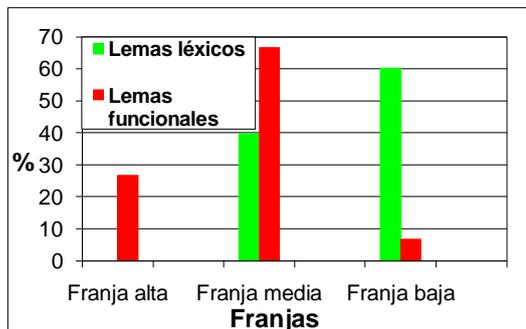


Figura 49.- Frecuencias de cada categoría gramatical en todo el texto de FESTESE001V.

Debido a la lematización, las distribuciones por franjas en cuanto a número de lemas y frecuencias son distintas de las del texto preparado (ver Figuras 28 y 44, 29 y 45, 32 y 48, 33 y 49). En el AO las palabras funcionales ocupan toda la franja alta con un solo lema (*el*). Las palabras léxicas han pasado a ocupar sólo las franjas media y baja, cosa que no ocurría en el texto preparado. En una metodología donde

UN ESTUDIO DE CASO: UN TEXTO ESCOLAR DE ASTRONOMÍA

se va a determinar un sistema óptimo para estudiar relaciones en función de la frecuencia, el mismo estará dominado por palabras funcionales (*Figura 47*).

Debido a la eliminación de palabras funcionales y verbos auxiliares los datos de los AF y AG sólo son comparables con los de las *Figuras 48 y 49* (Tabla 45). Las distribuciones en número de lemas y frecuencias sólo están correlacionadas en el AO para los lemas léxicos. En todos los análisis la categoría léxica aumenta en número de lemas hacia la franja baja, pero sus frecuencias en el AO presentan su máximo en la franja baja, y en los AF y AG en la franja media. La eliminación de palabras funcionales y verbos auxiliares produce una redistribución de la categoría léxica en número de lemas y sobre todo en las frecuencias. Las relaciones a estudiar en los AF y AG se darán exclusivamente entre lemas léxicos. Las diferencias entre el AF y AG son menores y se deben a la disminución de lemas y redistribución de frecuencias causados por la unificación de grafías.

Tabla 45

Distribución del número de lemas y frecuencias en todo el texto lematizado/transformado en los tres análisis

Análisis	% N ^o léxicos	% N ^o funcionales	% $\sum n_i$ léxicos	% $\sum n_i$ funcionales
Franja alta ($\leq 15-20\%$ de N)				
AO ^a	0.00	2.22	0.00	26.61
AF	2.35		21.04	
AG	1.81		19.56	
Franja media ($\leq 60\%$ de N)				
AO	8.02	37.78	39.77	66.67
AF	24.46		45.72	
AG	18.75		42.24	
Franja baja (=100% de N)				
AO	91.98	60.00	60.23	6.73
AF	73.19		33.24	
AG	79.44		38.20	

Nota. N^o léxicos= número de lemas léxicos; N^o funcionales= número de lemas funcionales; n_i léxicas= frecuencia absoluta del lema léxico i; n_i funcionales= frecuencia absoluta del lema funcional i.

^aLos datos del AO son los de cada categoría en todo el texto.

UN ESTUDIO DE CASO: UN TEXTO ESCOLAR DE ASTRONOMÍA

Dentro de los lemas léxicos hay cinco subcategorías bien diferenciadas: verbos, adverbios, adjetivos, sustantivos y lenguaje matemático. Los lemas léxicos se han clasificado en las subcategorías mencionadas consultando el Diccionario de la RAE online (2001) tal y como se ha explicado en el apartado anterior (4.2.). La clasificación de los lemas léxicos se puede ver en la quinta columna de las Tablas 96, 99 y 100 para los AO, AF y AG (Anexos I, P y Q, respectivamente). La importancia relativa de cada subcategoría en el AO se ve en las *Figuras 50 y 51*.

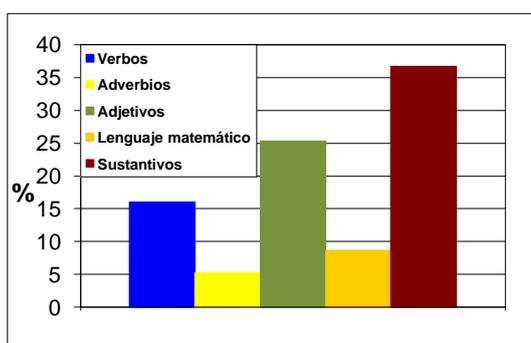


Figura 50.- Número de lemas por subcategoría léxica de FESTESE001V.

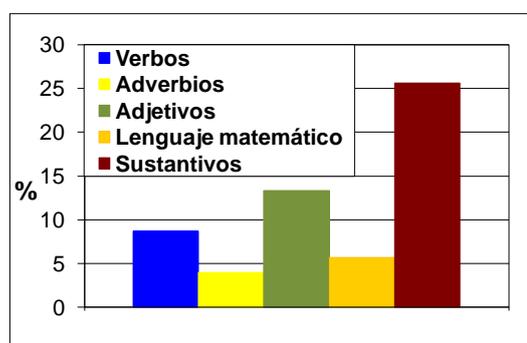


Figura 51.- Frecuencias por subcategoría léxica de FESTESE001V.

Las subcategorías léxicas presentan una menor diferencia entre el número de vocablos y las frecuencias producidas, comparado con los vocablos funcionales. Esto se debe a la escasa repetición de las palabras léxicas en comparación con las funcionales. A pesar de que el lenguaje matemático, verbos, adjetivos y sustantivos tienen mayor número de vocablo que los lemas funcionales, éstos producen más frecuencias (un 43%).

Las gráficas para los AF y AG son parecidas, por lo que se presentan los resultados en la Tabla 46 para facilitar la comparación. No hay grandes variaciones de valores para el número de vocablos y frecuencias entre los distintos análisis.

Las distribuciones en número de lemas en el AO se presentan en las *Figuras 52, 54 y 56*, y las de frecuencias en las *Figuras 53, 55 y 57*. El comportamiento de

UN ESTUDIO DE CASO: UN TEXTO ESCOLAR DE ASTRONOMÍA

cada subcategoría léxica está correlacionado con la tendencia general de su categoría, en cuanto al número de lemas y frecuencias, excepto adverbios y lenguaje matemático (ver *Figura 54*, donde los valores de este último en la franja media y baja son 8.77% y 8.63% respectivamente). Las tres subcategorías restantes constituyen un 85% del número de lemas y un 83% de las frecuencias de la categoría léxica (Tabla 46). El número de vocablos y las frecuencias del lenguaje matemático parecen estar correlacionados en el texto lematizado.

Tabla 46

Distribución de las subcategorías léxicas en los tres análisis en el texto lematizado/transformado

Subcategorías léxicas	AO ^a y AF		AG	
	% N ^o	% $\sum n_i$	% N ^o	% $\sum n_i$
verbos	17.42	15.09	17.94	14.74
adverbios	5.68	6.97	5.85	6.73
adjetivos	27.59	23.29	27.62	23.39
lenguaje matemático	9.39	9.94	8.47	8.58
sustantivos	39.92	44.70	40.12	46.56

Nota. N^o= número de vocablos; n_i= frecuencia absoluta del vocablo i.

^aEn el AO los datos son respecto a todos los lemas léxicos.

Las gráficas para los AF y AG son parecidas, por lo que se presentan los resultados en las Tablas 47, 48 y 49, para facilitar la comparación. La eliminación de palabras funcionales y verbos auxiliares en los AF y AG produce una redistribución de las subcategorías a lo largo de las tres franjas.

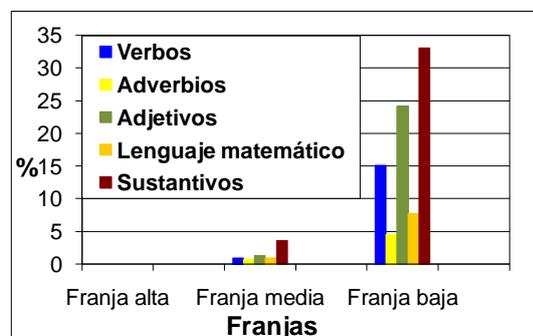


Figura 52.- Número de lemas por subcategoría léxica en todo el texto de FESTESE001V.

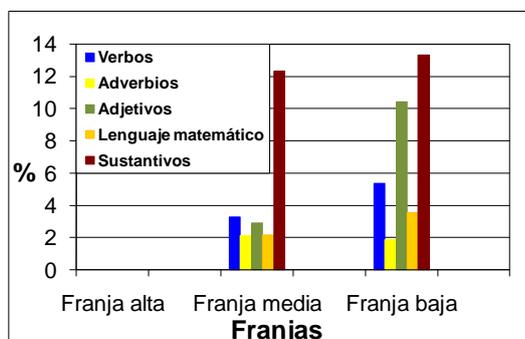


Figura 53.- Frecuencias por subcategoría léxica en todo el texto de FESTESE001V.

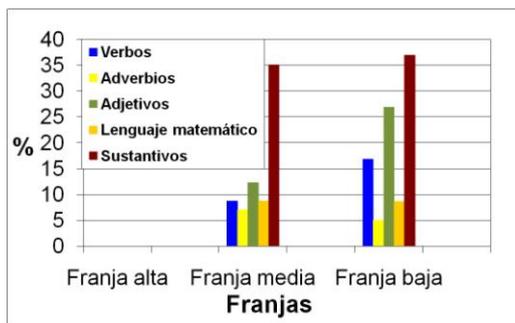


Figura 54.- Número de lemas por subcategoría léxica en cada franja de FESTEENSE001V.

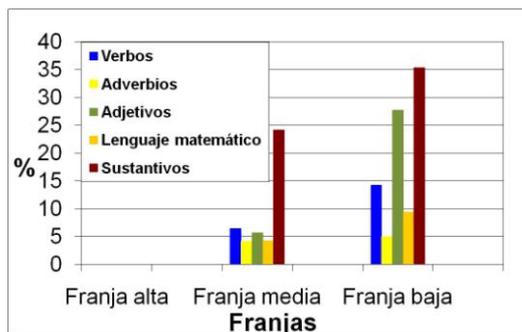


Figura 55.- Frecuencias por subcategoría léxica en cada franja de FESTEENSE001V.

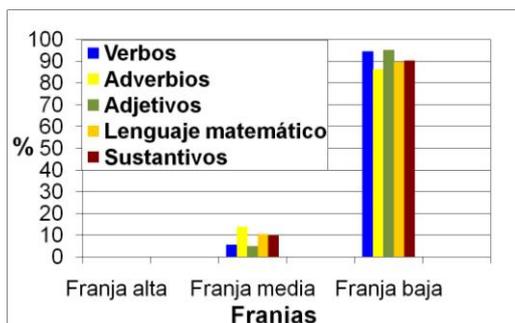


Figura 56.- Número de lemas de cada subcategoría léxica en todo el texto de FESTEENSE001V.

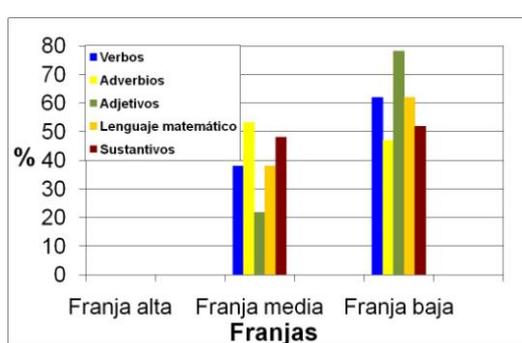


Figura 57.- Frecuencias de cada subcategoría léxica en todo el texto de FESTEENSE001V.

En el AF el comportamiento de cada subcategoría léxica está correlacionado con la tendencia general de su categoría, excepto verbos que se desvían de la tendencia general en cuanto a las frecuencias (Tabla 47). Las cuatro subcategorías restantes constituyen un 83% del número de lemas y un 86% de las frecuencias.

En el AG el comportamiento de cada subcategoría léxica no está correlacionado con la tendencia general de su categoría léxica, excepto en sustantivos, adverbios y lenguaje matemático, que en general lo están tanto en número de palabras como en frecuencias. Los adjetivos y verbos se desvían de la tendencia general en cuanto a las frecuencias (Tabla 47). Las tres subcategorías restantes constituyen un 54% del número de lemas y un 63% de las frecuencias.

UN ESTUDIO DE CASO: UN TEXTO ESCOLAR DE ASTRONOMÍA

Tabla 47

Distribución de las subcategorías léxicas en todo el texto lematizado/transformado en los tres análisis

Análisis	AO ^a		AF		AG	
	% N°	% $\sum n_i$	% N°	% $\sum n_i$	% N°	% $\sum n_i$
Fanja alta ($\leq 15-20\%$ de N)						
verbos	0.00	0.00	0.20	2.98	0.20	2.91
adverbios	0.00	0.00	0.39	2.25	0.20	1.20
adjetivos	0.00	0.00	0.39	2.18	0.40	2.69
lm	0.00	0.00	0.20	0.94	0.00	0.00
sustantivos	0.00	0.00	1.77	12.70	1.01	12.76
Fanja media ($\leq 60\%$ de N)						
verbos	0.90 (0.98)	3.29 (5.73)	2.94	5.81	2.02	4.61
adverbios	0.72 (0.78)	2.12 (3.70)	1.37	2.90	1.21	3.33
adjetivos	1.26 (1.37)	2.91 (5.08)	7.24	11.76	4.03	8.50
lm	0.90 (0.98)	2.16 (3.77)	2.94	5.88	3.02	6.17
sustantivos	3.60 (3.91)	12.31 (21.48)	9.98	19.38	8.47	19.63
Fanja baja (=100% de N)						
verbos	15.11 (16.44)	5.37 (9.36)	14.29	6.31	15.73	7.23
adverbios	4.50 (4.89)	1.87 (3.27)	3.91	1.81	4.44	2.20
adjetivos	24.10 (26.22)	10.44 (18.21)	19.96	9.36	23.19	12.19
lm	7.73 (8.41)	3.54 (6.17)	6.26	3.12	5.44	2.41
sustantivos	33.09 (36.01)	13.31 (23.22)	28.77	12.63	30.65	14.17

Nota. N°= número de vocablos; n_i = frecuencia absoluta del vocablo i; lm= lenguaje matemático.

^aLos porcentajes entre paréntesis del AO se refieren a los lemas léxicos del mismo.

Tabla 48

Distribución de las subcategorías léxicas en cada franja en los tres análisis en el texto lematizado/transformado

Análisis	AO		AF		AG	
	% N°	% $\sum n_i$	% N°	% $\sum n_i$	% N°	% $\sum n_i$
Fanja alta ($\leq 15-20\%$ de N)						
verbos	0.00	0.00	8.33	14.14	11.11	14.86
adverbios	0.00	0.00	16.67	10.69	11.11	6.16
adjetivos	0.00	0.00	16.67	10.34	22.22	13.77
lm	0.00	0.00	8.33	4.48	0.00	0.00
sustantivos	0.00	0.00	50.00	60.34	55.56	65.22
Fanja media ($\leq 60\%$ de N)						
verbos	8.77	6.45	12.00	12.70	10.75	10.91
adverbios	7.02	4.16	5.60	6.35	6.45	7.89
adjetivos	12.28	5.71	29.60	25.71	21.51	20.13
lm	8.77	4.24	12.00	12.86	16.13	14.60
sustantivos	35.09	24.16	40.80	42.38	45.16	46.48

UN ESTUDIO DE CASO: UN TEXTO ESCOLAR DE ASTRONOMÍA

Análisis	AO		AF		AG	
	% N°	% $\sum n_i$	% N°	% $\sum n_i$	% N°	% $\sum n_i$
Fanja baja (=100% de N)						
verbos	16.87	14.24	19.52	19.00	19.80	18.92
adverbios	5.02	4.97	5.35	5.46	5.58	5.75
adjetivos	26.91	27.70	27.27	28.17	29.19	31.91
lm	8.63	9.38	8.56	9.39	6.85	6.31
sustantivos	36.95	35.32	39.30	37.99	38.58	37.11

Nota. N°= número de vocablos; n_i = frecuencia absoluta del vocablo i; lm= lenguaje matemático.

Tabla 49

Distribución de cada subcategoría léxica en todo el texto lematizado/transformado en los tres análisis

Análisis	AO		AF		AG	
	% N°	% $\sum n_i$	% N°	% $\sum n_i$	% N°	% $\sum n_i$
Fanja alta ($\leq 15-20\%$ de N)						
verbos	0.00	0.00	1.12	19.71	1.12	19.71
adverbios	0.00	0.00	6.90	32.29	3.45	17.89
adjetivos	0.00	0.00	1.42	9.35	1.46	11.52
lm	0.00	0.00	2.08	9.49	0.00	0.00
sustantivos	0.00	0.00	2.94	28.41	2.51	27.40
Fanja media ($\leq 60\%$ de N)						
verbos	5.62	37.98	16.85	38.46	11.24	31.25
adverbios	13.79	53.13	24.14	41.67	20.69	49.47
adjetivos	4.96	21.81	26.24	50.47	14.60	36.36
lm	10.42	37.96	31.25	59.12	35.71	71.90
sustantivos	9.80	48.05	25.00	43.34	21.11	42.16
Fanja baja (=100% de N)						
verbos	94.38	62.02	82.02	41.83	87.64	49.04
adverbios	86.21	46.88	68.97	26.04	75.86	32.63
adjetivos	95.04	78.19	72.34	40.19	83.94	52.12
lm	89.58	62.04	66.67	31.39	64.29	28.10
sustantivos	90.20	51.95	72.06	28.25	76.38	30.44

Nota. N°= número de vocablos; n_i = frecuencia absoluta del vocablo i; lm= lenguaje matemático.

En todos los análisis hay 3 o 4 subcategorías que dominan el comportamiento de la categoría léxica. Las subcategorías léxicas aumentan en número de lemas hacia la franja baja mientras que sus frecuencias parecen presentar su máximo en la franja baja en el AO y en la franja media en los AF y AG. En los

UN ESTUDIO DE CASO: UN TEXTO ESCOLAR DE ASTRONOMÍA

AF y AG la franja alta está dominada por los sustantivos tanto en número de lemas como en frecuencias, por lo que se espera que dominen en el vocabulario frecuente y las redes semánticas.

Los porcentajes de hápax en cada subcategoría se presentan en la Tabla 50. Otros datos relativos a la importancia de cada subcategoría léxica respecto a su categoría, así como a la franja alta se presentan en la Tabla 51. En los tres análisis más de la mitad de los vocablos de las subcategorías léxicas son hápax tras la lematización, excepto los vocablos del lenguaje matemático en los AO y AF.

Tabla 50

Hápax en las subcategorías léxicas en el texto lematizado/transformado

Subcategorías léxicas	% N ^o hápax		% N ^o léxicas,hápax	
	AO y AF	AG	AO y AF	AG
verbos	66.29	66.29	11.55	11.90
adverbios	51.72	51.72	2.94	3.02
adjetivos	53.19	53.28	14.68	14.72
Im	43.75	50.00	4.11	4.23
sustantivos	58.82	57.79	23.48	23.19

Nota. N^o hápax= número de lemas de la subcategoría que son hápax en la franja baja; N^o léxicas,hápax= hápax respecto a la categoría léxica.

Tabla 51

Importancia de las subcategorías léxicas del texto transformado

Subcategorías léxicas	AF			AG		
	N ^o	Lemas en franja alta	% $\sum n_{i,subcategoría}$	N ^o	Lemas en franja alta	% $\sum n_{i,subcategoría}$
verbos	89	<i>ser</i>	19.71	89	<i>ser</i>	19.71
adverbios	29	<i>más, muy</i>	32.29	29	<i>más</i>	17.89
adjetivos	141	<i>globular, principal</i>	9.35	137	<i>globular, superficial</i>	11.52
Im	48	<i>x</i>	9.49	42		
sustantivos	204	<i>estrella, cúmulo, núcleo, temperatura, masa, secuencia</i>	28.41	199	<i>estrella, cúmulo, temperatura, luminosidad, núcleo</i>	27.40

Nota. N^o= número de lemas de la subcategoría; $\sum n_{i,subcategoría}$ = frecuencias en la franja alta respecto a la subcategoría.

La eliminación de las palabras funcionales y verbos auxiliares ha provocado una redistribución de los lemas léxicos, de manera que todas las subcategorías léxicas se hallan representadas en la franja alta del AF. La unificación de grafías vuelve a producir una redistribución donde el lenguaje matemático ha desaparecido de la franja alta. Ésta está dominada por los sustantivos en cuanto al número de lemas y frecuencias en los AF y AG.

4. 4.- Estudios de Entorno con PAFE: Relaciones

Una vez obtenidas las frecuencias de los textos preparado y lematizado/transformado, se pasa a analizar este último en el Estudio de Entorno del PAFE.

4. 4. 1.- Búsqueda de colocaciones

Para buscar las colocaciones se aplica el Estudio de Entorno del PAFE con las características de la Tabla 14 (ver apartado 3. 2. 2. 2.). El proceso se repitió tres veces en los AO y AG, y dos veces en el AF (Tablas 52, 53, 54 y 55).

UN ESTUDIO DE CASO: UN TEXTO ESCOLAR DE ASTRONOMÍA

Tabla 52

Número de colocaciones en los pares de lemas relacionados

Análisis	V	% N ^o _{ra≥2}	N ^o (ra _{ij} /n _i)≥0.5	N ^o _{consideradas}	N ^o _{añadidas}
Primera vez					
AO	556	21.94	77	45	9
AF	511	14.68	47	47	8
AG	496	15.12	50	50	5
Segunda vez					
AO	570	14.56	44	11	0
AF	535	4.86	8	8	0
AG	519	5.39	8	8	0
Tercera vez					
AO	572	12.59	36	3	0
AF	539	3.15	0	0	0
AG	522	3.83	1	1	0
Cuarta vez					
AO	573	12.04	0	0	0
AG	523	3.63	0	0	0

Nota. V= número de vocablos del texto lematizado/transformado; N^o= número de pares de lemas relacionados; ra_{ij}= relación absoluta de la ULL origen i con la ULL j; n_i= frecuencia absoluta de la ULL i.

Los datos de los tres análisis relativos a la primera vez que se intentan hallar colocaciones se presentan en la parte superior de la Tabla 53. El listado completo de los 122 lemas (21.94% del total) del AO, que en la primera vez tienen un mínimo de dos relaciones, se muestran en el Anexo J (Tabla 97). El criterio que indica una posible colocación es que el cociente entre el número total de relaciones (ra_i) de un lema origen i con otro lema dado y la frecuencia absoluta de i (n_i) sea igual o superior a 0.5 (ver tercera columna de la Tabla 52). En el caso del AO no todos los pares de lemas que cumplieron con este criterio fueron considerados como propuestas de unión entre lemas. Se tuvieron en cuenta todas aquellas relaciones donde el cociente era igual a 1, es decir, que siempre aparecen juntos. Muchos de estos casos se refieren a la unión de un lema léxico con otro funcional (ver el

ejemplo “**millón de**” señalado en negrita en la Tabla 53). Las 32 colocaciones encontradas mediante el cociente con un valor mayor que 0.5 pero inferior a 1, y que contenían un lema funcional fueron desechadas. Desde el punto de vista de la investigación son de más interés las posibles uniones entre lemas léxicos (ver el ejemplo “*temperatura superficial*” señalado en cursiva en la Tabla 53). Sin embargo hay uniones que han sido identificadas por el experto (un 17% del total) y no por el método del cociente (quinta columna en la Tabla 52). Es el caso de “*cúmulo globular*” señalado en cursiva y negrita en la Tabla 53, que a pesar de tener un cociente inferior a 0.5 define un tipo de objeto astronómico. En los AF y AG, todas las uniones halladas mediante el criterio del cociente son entre lemas léxicos, por lo que se consideran todas como colocaciones.

En la parte superior de las Tablas 53, 54 y 55 se presentan las uniones a considerar para seguir analizando el texto (ver Primera vez en la parte superior) en los AO, AF y AG, respectivamente. Los datos aparecen por orden léxicométrico del lema origen y de mayor a menor relación absoluta del lema relacionado con áquel.

Tabla 53

Listado de propuestas de uniones de lemas de FESTESE001V

ULL origen	ULL	ra_i	n_i	ra_i/n_i	Colocación
		Primera vez			
<i>cúmulo</i>	<i>globular</i>	17	39	0.44	
cúmulo	abiertoa	5	39	0.13	
<i>temperatura</i>	<i>superficial</i>	8	15	0.53	colocación
temperatura	central	2	15	0.13	
secuencia	principal	13	13	1.00	colocación
diagrama	hr	5	11	0.45	
l	solar	3	11	0.27	
año	luz	4	10	0.40	
millón	de	8	8	1.00	colocación
superficial	t	4	8	0.50	colocación
magnitud	visible	2	7	0.29	
gigante	rojo	6	6	1.00	colocación

UN ESTUDIO DE CASO: UN TEXTO ESCOLAR DE ASTRONOMÍA

ULL origen	ULL	ra_i	n_i	ra_i/n_i	Colocación
baja	masa	3	5	0.60	colocación
messier	12	2	5	0.40	
nuestro	galaxia	2	5	0.40	
vía	lácteo	5	5	1.00	colocación
ya	no	3	5	0.60	colocación
3	x	2	4	0.50	colocación
a2	x	2	4	0.50	colocación
a3	k	2	4	0.50	colocación
a4	k	2	4	0.50	colocación
cantidad	de	4	4	1.00	colocación
capa	externo	3	4	0.75	colocación
fig	6	2	4	0.50	colocación
helio	ser	2	4	0.50	colocación
herramienta	astronómico	4	4	1.00	colocación
k	tener	2	4	0.50	colocación
mucho	tiempo	2	4	0.50	colocación
nuevo	estrella	2	4	0.50	colocación
parte	superior	2	4	0.50	colocación
sección	herramienta	4	4	1.00	colocación
superior	derecho	2	4	0.50	colocación
cuatro	núcleo	3	3	1.00	colocación
deber	a	3	3	1.00	colocación
nebulosa	planetario	2	3	0.67	colocación
partir	de	3	3	1.00	colocación
0coma7	porciento	2	2	1.00	colocación
4	x	2	2	1.00	colocación
blanco	y	2	2	1.00	colocación
c	a2	2	2	1.00	colocación
ciento	de	2	2	1.00	colocación
combustión	del	2	2	1.00	colocación
comenzarx	a	2	2	1.00	colocación
diámetro	típico	2	2	1.00	colocación
ejercidoa	hacia	2	2	1.00	colocación
extender	aproximadamente	2	2	1.00	colocación
largo	de	2	2	1.00	colocación
manera	que	2	2	1.00	colocación
medir	el	2	2	1.00	colocación
mientras	el	2	2	1.00	colocación
nos	dar	2	2	1.00	colocación
orden	de	2	2	1.00	colocación
rama	de	2	2	1.00	colocación
residual	de	2	2	1.00	colocación
Segunda vez					
millónde	año	6	8	0.75	colocación
elemento	más	3	5	0.60	colocación
cantidadde	energía	2	4	0.50	colocación
yano	ser	2	3	0.67	colocación
0coma7porciento	residualde	2	2	1.00	colocación

UN ESTUDIO DE CASO: UN TEXTO ESCOLAR DE ASTRONOMÍA

ULL origen	ULL	ra _i	n _i	ra _i /n _i	Colocación
3x	10	2	2	1.00	colocación
combustión del	hidrógeno	2	2	1.00	colocación
diámetro típico	de	2	2	1.00	colocación
hacia	el	2	2	1.00	colocación
orden de	magnitud	2	2	1.00	colocación
residual de	el	2	2	1.00	colocación
Tercera vez					
elemento más	pesado a	2	3	0.67	colocación
0 coma 7 por ciento residual de	masa	2	2	1.00	colocación
diámetro típico de	uno	2	2	1.00	colocación

Nota. ra_i= relación absoluta; n_i= frecuencia absoluta.

Tabla 54

Listado de propuestas de uniones de lemas de FESTESE001V-F

ULL origen	ULL	ra _i	n _i	ra _i /n _i	Colocación
Primera vez					
cúmulo	globular	17	39	0.44	
cúmulo	abierto a	5	39	0.13	
temperatura	superficial	8	15	0.53	colocación
temperatura	central	2	15	0.13	
secuencia	principal	13	13	1.00	colocación
diagrama	hr	5	11	0.45	
l	solar	3	11	0.27	
año	luz	4	10	0.40	
proceso	fusión	5	9	0.56	colocación
millón	año	6	8	0.75	colocación
superficial	t	4	8	0.50	colocación
magnitud	visible	2	7	0.29	
vida	estrella	4	7	0.57	colocación
gigante	rojo	6	6	1.00	colocación
mayoría	estrella	3	6	0.50	colocación
mil	millón	5	6	0.83	colocación
abierto a	estrella	4	5	0.80	colocación
bajo a	masa	3	5	0.60	colocación
determinar	edad	3	5	0.60	colocación
elemento	más	3	5	0.60	colocación
gas	polvo	3	5	0.60	colocación
messier	12	2	5	0.40	
vía	lácteo	5	5	1.00	colocación
ya	no	3	5	0.60	colocación
3	x	2	4	0.50	colocación
a2	x	2	4	0.50	colocación
a3	k	2	4	0.50	colocación
a4	k	2	4	0.50	colocación

UN ESTUDIO DE CASO: UN TEXTO ESCOLAR DE ASTRONOMÍA

ULL origen	ULL	ra _i	n _i	ra _i /n _i	Colocación
cantidad	energía	2	4	0.50	colocación
capa	externo	3	4	0.75	colocación
fig	6	2	4	0.50	colocación
helio	ser	2	4	0.50	colocación
herramienta	astronómico	4	4	1.00	colocación
k	tener	2	4	0.50	colocación
mucho	tiempo	2	4	0.50	colocación
nuevo	estrella	2	4	0.50	colocación
parte	superior	2	4	0.50	colocación
sección	herramienta	4	4	1.00	colocación
superior	derecho	2	4	0.50	colocación
cuatro	núcleo	3	3	1.00	colocación
derecho	diagrama	2	3	0.67	colocación
equilibrio	presión	2	3	0.67	colocación
externo	estrella	2	3	0.67	colocación
fusionar	formar	3	3	1.00	colocación
mb	mv	2	3	0.67	colocación
nebulosa	planetario	2	3	0.67	colocación
0coma7	porciento	2	2	1.00	colocación
4	x	2	2	1.00	colocación
c	a2	2	2	1.00	colocación
combustión	hidrógeno	2	2	1.00	colocación
diámetro	típico	2	2	1.00	colocación
extender	aproximadamente	2	2	1.00	colocación
orden	magnitud	2	2	1.00	colocación
rama	gigante	2	2	1.00	colocación
residual	masa	2	2	1.00	colocación
Segunda vez					
cúmuloabiertoaestrella	ser	2	4	0.50	colocación
temperaturasuperficialt	estrella	2	4	0.50	colocación
temperaturasuperficialt	todo	2	4	0.50	colocación
cuatronúcleo	hidrógeno	2	3	0.67	colocación
elementomás	pesadoa	2	3	0.67	colocación
yano	ser	2	3	0.67	colocación
0coma7porciento	residualmasa	2	2	1.00	colocación
3x	10	2	2	1.00	colocación

Nota. ra_i= relación absoluta; n_i= frecuencia absoluta.

Tabla 55

Listado de propuestas de uniones de lemas de FESTESE001V-S

ULL origen	ULL	ra _i	n _i	ra _i /n _i	Colocación
Primera vez					
cúmulo	globular	23	41	0.56	colocación
cúmulo	abiertoa	7	41	0.17	
temperatura	superficial	15	21	0.71	colocación

UN ESTUDIO DE CASO: UN TEXTO ESCOLAR DE ASTRONOMÍA

ULL origen	ULL	ra_i	n_i	ra_i/n_i	Colocación
temperatura	central	2	21	0.10	
luminosidad	solar	3	19	0.16	
secuencia	principal	13	13	1.00	colocación
magnitud	aparente	6	12	0.50	colocación
magnitud	visible	2	12	0.17	
año	luz	4	10	0.40	
diagrama	hr	10	10	1.00	colocación
proceso	fusión	5	9	0.56	colocación
millón	año	6	8	0.75	colocación
abiertoa	estrella	6	7	0.86	colocación
índice	color	7	7	1.00	colocación
messier	12	4	7	0.57	colocación
vida	estrella	4	7	0.57	colocación
aparente	índice	6	6	1.00	colocación
gigante	rojo	6	6	1.00	colocación
mayoría	estrella	3	6	0.50	colocación
mil	millón	5	6	0.83	colocación
vía	lácteo	6	6	1.00	colocación
4	núcleo	3	5	0.60	colocación
bajoa	masa	3	5	0.60	colocación
determinar	edad	3	5	0.60	colocación
elemento	más	3	5	0.60	colocación
gas	polvo	3	5	0.60	colocación
ya	no	3	5	0.60	colocación
3	x	2	4	0.50	colocación
a2	x	2	4	0.50	colocación
a3	k	2	4	0.50	colocación
a4	k	2	4	0.50	colocación
cantidad	energía	2	4	0.50	colocación
capa	externo	3	4	0.75	colocación
disco	vía	2	4	0.50	colocación
helio	ser	2	4	0.50	colocación
herramienta	astronómico	4	4	1.00	colocación
k	tener	2	4	0.50	colocación
mucho	tiempo	2	4	0.50	colocación
parte	superior	2	4	0.50	colocación
sección	herramienta	4	4	1.00	colocación
superior	derecho	2	4	0.50	colocación
derecho	diagrama	2	3	0.67	colocación
equilibrio	presión	2	3	0.67	colocación
externo	estrella	2	3	0.67	colocación
fusionar	formar	3	3	1.00	colocación
nebulosa	planetario	2	3	0.67	colocación
nuevo	estrella	2	3	0.67	colocación
0coma7	porciento	2	2	1.00	colocación
c	a2	2	2	1.00	colocación
combustión	hidrógeno	2	2	1.00	colocación

UN ESTUDIO DE CASO: UN TEXTO ESCOLAR DE ASTRONOMÍA

ULL origen	ULL	ra _i	n _i	ra _i /n _i	Colocación
diámetro	típico	2	2	1.00	colocación
extender	aproximadamente	2	2	1.00	colocación
orden	magnitud	2	2	1.00	colocación
rama	gigante	2	2	1.00	colocación
residual	masa	2	2	1.00	colocación
Segunda vez					
magaparenteíndice color	b	3	6	0.50	colocación
magaparenteíndice color	v	3	6	0.50	colocación
4núcleo	hidrógeno	2	3	0.67	colocación
elementomás	pesadoa	2	3	0.67	colocación
yano	ser	2	3	0.67	colocación
0coma7porciento	residualmasa	2	2	1.00	colocación
3x	10	2	2	1.00	colocación
4	x	2	2	1.00	colocación
Tercera vez					
magaparenteíndice colorb	magaparenteíndice colorv	2	3	0.67	colocación

Nota. ra_i= relación absoluta; rr_i= relación relativa; n_i= frecuencia absoluta.

Las Tablas 56 y 57 muestran los porcentajes de uniones halladas con el criterio del cociente y las propuestas por el experto para las veces que se repitió el proceso en los AO y AG (Tabla 56) y AF (Tabla 57). En la primera vez que se hace el proceso el número de colocaciones a considerar es mayor para el valor 1 del cociente, excepto en el AO. En la segunda vez el máximo de colocaciones se encuentra para valores del cociente entre 0.5 y 1. En todos los análisis las propuestas del experto a considerar aparecen sólo en la primera vez.

Tabla 56

Porcentajes de uniones de FESTESE001V y FESTESE001V-S

Uniones	Cociente <0.5			0.5 ≤ cociente < 1			Cociente = 1		
	1 ^a	2 ^a	3 ^a	1 ^a	2 ^a	3 ^a	1 ^a	2 ^a	3 ^a
FESTESE001V									
Criterio del cociente	0	0	0	62	84	94	38	16	6
Consideradas	17	0	0	29	36	33	54	64	67
FESTESE001V-S									
Criterio del cociente	0	0	0	42	63	100	58	37	0
Consideradas	9	0	0	38	63	100	53	37	0

Nota. 1^a= primera vez; 2^a= segunda vez; 3^a= tercera vez.

Tabla 57

Porcentajes de uniones de FESTESE001V-F

Uniones	Cociente <0.5		0.5 ≤ cociente < 1		Cociente = 1	
	1 ^a	2 ^a	1 ^a	2 ^a	1 ^a	2 ^a
	FESTESE001V-F					
Criterio del cociente	0	0	38	75	62	25
Consideradas	14	0	33	75	53	25

Nota. 1^a= primera vez; 2^a= segunda vez.

Ya en la primera vez se pueden producir colocaciones de hasta 5 lemas a base de unirse lemas hacia la derecha que se encuentran fuertemente relacionados (ver parte superior de la Tabla 58). La mayoría de las uniones de dos lemas que forman parte de alguna de 3, 4 ó 5 lemas no desaparecen (Tabla 59). La conservación de parte de las colocaciones de 2 y 3 lemas se debe al hecho de que las uniones se han hallado considerando los lemas como entidades únicas, con lo que han aparecido algunos lemas que están unidos a la izquierda y a la derecha. Las colocaciones que son comunes a los tres análisis se muestran en negrita, las que están señaladas en cursiva se comportan de distinta manera en los AF y AG, y la que está señalada en negrita y cursiva no aparece en el AG debido a la unificación de grafías. La eliminación de palabras funcionales y verbos auxiliares produce más colocaciones entre lemas léxicos y con mayor número de lemas. En las colocaciones de los AF y AG hay más información sobre qué lemas léxicos aparecen próximos con mayor frecuencia. La unificación de grafías modifica ligeramente el modo en que se unen dichos lemas léxicos, y añade alguna colocación más. La formación de colocaciones de más de dos lemas junto con la desaparición de alguno de los pares hace que el número de colocaciones final se vea modificado, pasando de 53 a 51 en el AO y de 55 a 53 en los AF y AG. En los tres análisis la mayoría de las colocaciones son de 2 lemas (Tabla 58).

UN ESTUDIO DE CASO: UN TEXTO ESCOLAR DE ASTRONOMÍA

Tabla 58

Porcentaje del número de lemas en las colocaciones

Colocaciones	2 ULL	3 ULL	4 ULL	5 ULL	6 ULL	10 ULL
		Primera vez				
AO	90.57	9.43				
AF	81.13	16.98	1.89			
AG	79.25	16.98	1.89	1.89		
		Segunda vez				
AO	20.00	70.00		10.00		
AF		50.00	50.00			
AG	12.50	50.00	12.50	25.00		
		Tercera vez				
AO		33.33	33.33		33.33	
AG						100.00

La Tabla 60 presenta la proporción de las categorías gramaticales en las colocaciones de la primera vez para los tres análisis. Las subcategorías léxicas tienen la misma importancia relativa al formar colocaciones en todos los análisis, excepto los sustantivos, ya que del AO a los AF y AG, el porcentaje ocupado por los lemas funcionales pasa a ser prácticamente de los sustantivos. La subcategoría que más entra a formar parte de colocaciones es la de sustantivos, seguida por adjetivos, lenguaje matemático, verbos y adverbios. Las diferencias entre el AF y el AG resaltan la desaparición de lenguaje matemático a favor de sustantivos.

En el AO (AF y AG) un 24% (44%) de los lemas forman colocación con otro de su misma categoría gramatical; la mitad de este tipo de colocaciones corresponde respectivamente al lenguaje matemático y sustantivos. La proporción con que las distintas categorías gramaticales se colocan entre sí se presenta en la Tabla 61. Por cada columna y análisis se señala con qué categoría se coloca más la categoría señalada en la columna. Para los tres análisis los sustantivos se colocan más con los adjetivos, los adjetivos con los sustantivos y el lenguaje matemático

UN ESTUDIO DE CASO: UN TEXTO ESCOLAR DE ASTRONOMÍA

consigo mismo. Estos resultados coinciden con los de los estudios de franjas, en cuanto que los sustantivos y adjetivos presentan mayor número de relaciones, siendo las colocaciones un tipo de ellas.

Tabla 59

Formación de colocaciones de más de 2 lemas en la Primera vez

2 lemas	3 lemas	4 lemas	5 lemas
AO			
a3k	a3ktener		
a4k	a4ktener		
	secherrastronómico		
<i>temperaturasuperfici</i>	<i>temperaturasuperficia</i>		
<i>al</i>	<i>lt</i>		
	<i>partesuperiorderecho</i>		
AF			
cúmuloabiertoa	cúmuloabiertoaestrell		
	a		
<i>temperaturasuperfici</i>	<i>temperaturasuperficia</i>		
<i>al</i>	<i>lt</i>		
millónaño	millónaño		
capaexterno	capaexternoestrella		
	secherrastronómico		
<i>derechodiagrama</i>	<i>partesuperiorderecho</i>	<i>partesuperiorderechodiag</i>	
		<i>rama</i>	
	<i>ramagiganterojo</i>		
a3k	a3ktener		
a4k	a4ktener		
AG			
cúmuloabiertoa	cúmuloabiertoaestrell		
	a		
índicecolor		magaparenteíndicecolor	
millónaño	millónaño		
vialácteo	discovialácteo		
a4k	a4ktener		
a3k	a3ktener		
	secherrastronómico		
	<i>derdiaghr</i>		<i>partesuperiorderdiag</i>
			<i>hr</i>
	<i>partesuperiorderecho</i>		<i>partesuperiorderdiag</i>
			<i>hr</i>
capaexterno	capaexternoestrella		
<i>giganterojo</i>	<i>ramagiganterojo</i>		

UN ESTUDIO DE CASO: UN TEXTO ESCOLAR DE ASTRONOMÍA

Tabla 60

Porcentajes de las categorías gramaticales en las colocaciones de la Primera vez

Colocaciones	funcionales	verbos	adverbios	adjetivos	lm	sustantivos
AO	17.59	6.48	4.63	24.07	17.59	29.63
AF		5.45	3.64	25.45	19.09	46.36
AG		5.45	3.64	25.45	13.64	51.82

Tabla 61

Porcentaje de colocaciones entre las categorías gramaticales en la Primera vez

Colocaciones	funcionales	verbos	adverbios	adjetivos	lm	sustantivos
AO						
funcionales	5.56	66.67	33.33	20.00	0.00	24.14
verbos	22.22	0.00	0.00	0.00	8.33	3.45
adverbios	5.56	0.00	33.33	0.00	0.00	3.45
adjetivos	27.78	0.00	0.00	8.00	16.67	55.17
lm	0.00	16.67	0.00	8.00	58.33	6.90
sustantivos	38.89	16.67	33.33	64.00	16.67	6.90
AF						
verbos		20.00	50.00	0.00	7.69	5.13
adverbios		20.00	50.00	0.00	0.00	2.56
adjetivos		0.00	0.00	7.69	15.38	56.41
lm		20.00	0.00	7.69	61.54	5.13
sustantivos		40.00	0.00	84.62	15.38	30.77
AG						
verbos		20.00	33.33	0.00	11.11	4.65
adverbios		20.00	33.33	0.00	0.00	2.33
adjetivos		0.00	0.00	7.69	0.00	55.81
lm		20.00	0.00	0.00	66.67	4.65
sustantivos		40.00	33.33	92.31	22.22	32.56

Tras introducir los cambios en el texto lematizado/transformado, es decir, unir los lemas, se han vuelto a estudiar posibles uniones. Las nuevas características de los textos transformados y las colocaciones halladas se muestran en la Segunda vez de las Tablas 52, 53, 54 y 55. Todas las colocaciones a considerar fueron encontradas con el criterio del cociente (Tabla 52). En el AO aparece una colocación entre palabras funcionales (*hacia el*) para la que el criterio del cociente tiene valor 1, por lo que se ha considerado. En los AF y AG el máximo de

colocaciones se encuentra para valores del cociente entre 0.5 y 1 (Tabla 57). Para los tres análisis la mayoría de las colocaciones de la Segunda vez son de 3 lemas (Tabla 58).

Tras reiterar el anterior proceso de uniones se vuelven a estudiar posibles colocaciones obteniéndose las que aparecen en la parte inferior de las Tablas 53, 54 y 55(ver Tercera vuelta). En el AF en ningún caso el cociente entre el número de relaciones y la frecuencia absoluta es igual o superior a 0.5, por lo que en esta tercera vez no se han propuesto nuevas uniones. Para los AO y AG todas las colocaciones a considerar fueron encontradas con el criterio del cociente (Tabla 52) estando la mayoría para valores del cociente comprendidos entre 0.5 y 1 (Tabla 56). La única unión a considerar en el AG forma una ULL de 10 lemas (ver parte inferior de Tabla 55). Esta unión proviene de dos símbolos de magnitudes que en el AG pasaron a expresarse por varios lemas.

Se repite el proceso una vez más en los AO y AG estudiándose posibles uniones (Tabla 52). En el AO todas las uniones halladas con el criterio del cociente son de una ULL léxica con una funcional, por lo que en esta cuarta vez no se han propuesto nuevas uniones. En el AG en ningún caso el cociente entre el número de relaciones y la frecuencia absoluta es igual o superior a 0.5, por lo que en esta cuarta vez tampoco hay nuevas uniones.

El criterio del cociente entre las relaciones y la frecuencia absoluta tuvo un éxito superior al 85% en los tres análisis (Tabla 62). La producción de colocaciones de más de 2 lemas provocó a veces la desaparición de otras, con lo que el número final de colocaciones (ver quinta columna de Tabla 62) es más reducido al hallado inicialmente (suma de la segunda y tercera columna de la misma tabla). La mayoría de las colocaciones son de 2 lemas en los tres análisis, seguida por un 22% de

UN ESTUDIO DE CASO: UN TEXTO ESCOLAR DE ASTRONOMÍA

uniones formadas por 3 lemas (Tabla 63). El lenguaje matemático forma parte de un 20 a un 24% de las colocaciones, siendo un 9% uniones consigo mismo en todos los análisis. Un 80% de las colocaciones que contienen lenguaje matemático están formadas por 2 lemas en los tres análisis.

Tabla 62

Colocaciones totales en los tres análisis

Análisis	Nº ($ra_{ij}/n_i \geq 0.5$)	Nº añadidas	% éxito ($ra_{ij}/n_i \geq 0.5$)	Total colocaciones
AO	59	9	87	56
AF	55	8	87	55
AG	59	5	92	58

Nota. Nº= número de colocaciones; ra_i = relación absoluta; n_i = frecuencia absoluta.

Tabla 63

Porcentaje de colocaciones por número de lemas

Análisis	2 lemas	3 lemas	4 lemas	5 lemas	6 lemas	10 lemas
AO	75.00	21.43	1.79		1.79	
AF	69.09	21.82	9.09			
AG	67.24	24.14	1.72	5.17		1.72

Hay colocaciones de 3 lemas en el AO que se convierten en colocaciones de 2 lemas en los AF y AG al desaparecer la palabra funcional que separaba dichos lemas. Es el caso de “*millóndeño*” (ver Segunda vez en Tabla 53) que se convierte en “*millónaño*” (ver Primera vez en Tablas 54 y 55). Los tres análisis tienen en común 36 colocaciones que aparecen ya en el AO (un 62-65% de las uniones de cada análisis).

Para poder estudiar las relaciones entre ULL se usará el último texto transformado que incluye todas las colocaciones. El ejemplo de este texto para el caso del AO, llamado FESTESE001VD, se presenta en el Anexo I. Las características de los textos transformados definitivos para los tres análisis se muestran en la Tabla 64. La diferencia en número de ULL entre el AO y los AF y

AG se deben a la supresión de las ULL funcionales. Las pequeñas diferencias entre el AF y el AG se deben a la unificación de grafías.

Tabla 64

Características de los textos transformados definitivos (medidos con la aplicación Análisis de Frecuencias del PAFE)

Análisis	N _{ULL}	V _{ULL}	TTR	LD (%)
AO	2191	573	0.26	55.82
AF	1175	539	0.46	100.00
AG	1168	523	0.45	100.00

Nota. N_{ULL}= número de ULL; V_{ULL}= número de lemas de ULL; TTR= cociente tipo/caso; LD= densidad léxica.

En el AO todas las uniones, excepto una (*hacia el*), contienen al menos un lema léxico y han sido consideradas como lemas léxicos. El valor de TTR sigue siendo más cercano al de los textos científicos que a la de textos dedicados a la educación, mientras que la densidad léxica es próxima a la de libros de divulgación. Las propiedades del texto en cuanto a los valores de TTR y LD no han variado con la unión de lemas. Los AF y AG presentan valores más altos de TTR, indicando una menor repetición de sus ULL, y poseen toda la información del texto tal y como muestra la LD.

Las diferencias en los gráficos de frecuencias de los textos transformados definitivos entre el AO y los AF y AG se deben a la supresión de las palabras funcionales y verbos auxiliares en los AF y AG. Las diferencias entre estos dos últimos se producen por la unificación de grafías. Los gráficos de frecuencias del texto transformado definitivo en el AO se pueden ver en las *Figuras 58 y 59*. La *Figura 58* muestra la frecuencia absoluta tal y como aparece en el listado del Anexo J. Tanto en la *Figura 58* como en la *59* las ULL se encuentran en orden lexicométrico. En la *Figura 58* la frecuencia absoluta cae bruscamente en un número reducido de ULL, mientras que la distribución hacia la derecha parece

UN ESTUDIO DE CASO: UN TEXTO ESCOLAR DE ASTRONOMÍA

tender hacia una asíntota a lo largo de un gran número de ULL. En la *Figura 59* la frecuencia acumulada tiene en el eje de las ordenadas un comportamiento análogo pero invertido. De acuerdo con la metodología (apartado 3. 2. 2. 2. 2.) se pueden distinguir tres franjas de ULL atendiendo a la frecuencia acumulada (ver *Figura 59* y *Tabla 65*), que se representan también en el gráfico de la frecuencia absoluta (*Figura 58*). Nótese cómo cambian de lugar los lemas en la lista del texto transformado en comparación con la lista de frecuencias de los lemas del texto lematizado (ver Anexos G y K) debido a las colocaciones. La eliminación de palabras funcionales y verbos auxiliares hace que haya una redistribución de las ULL en los AF y AG, las cuales son sólo léxicas. Esto se puede ver en el cambio que hay en los límites de las franjas, así como en el número de ULL y frecuencias que forman cada una de ellas.

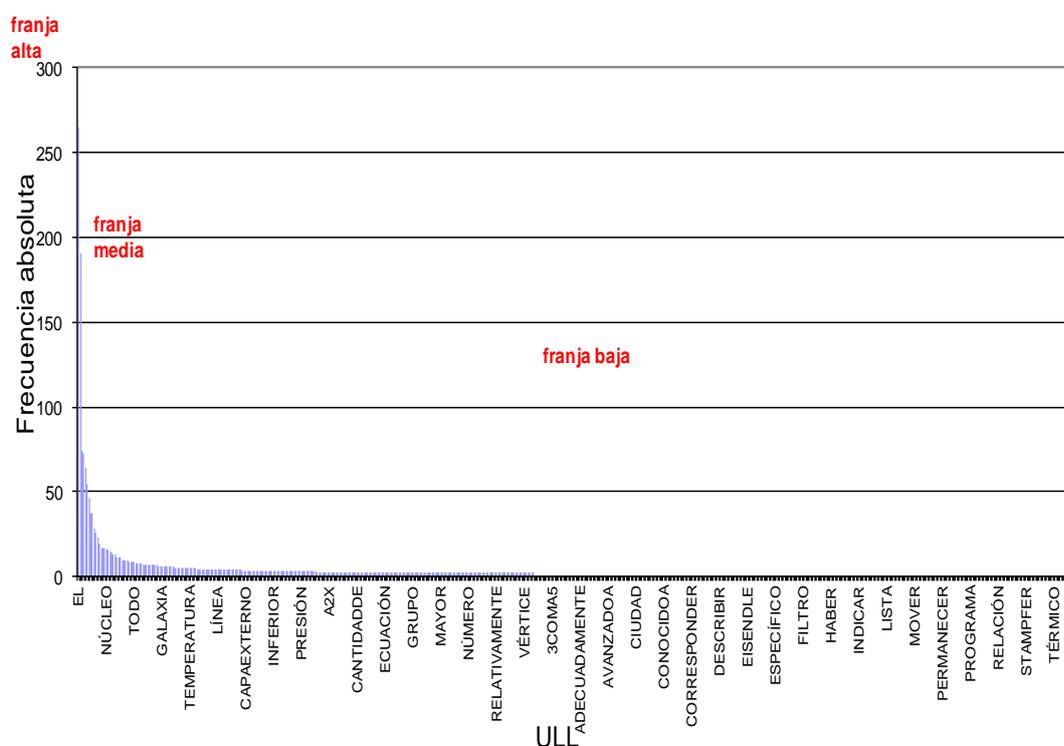


Figura 58.- Frecuencia absoluta de FESTESE001VD.

UN ESTUDIO DE CASO: UN TEXTO ESCOLAR DE ASTRONOMÍA

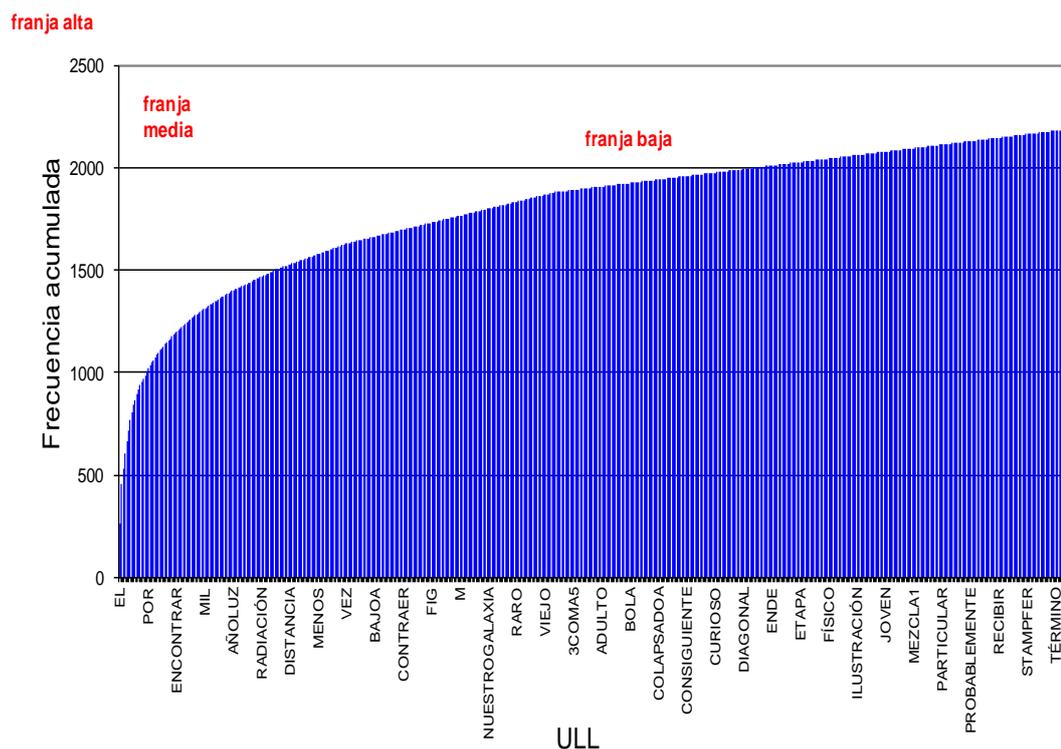


Figura 59.- Frecuencia acumulada de FESTEESE001VD.

Tabla 65

Franjas en frecuencia acumulada (N) de los textos transformados definitivos

Análisis	Límites (f)	ULL límite	% N°	% $\sum n_i$
Franja alta ($\leq 15-20\%$ de N)				
AO ^a	0		0.00 (0.00)	0.00 (0.00)
AF	8.511-50.213	estrella-tener	2.41	20.17
AG	9.418-53.938	estrella-ver	2.10	18.92
Franja media ($\leq 60\%$ de N)				
AO	2.738-120.949	el-t	9.60 (6.62)	60.79 (33.93)
AF	1.702-7.660	10-área	39.33	53.11
AG	1.712-8.562	aproximadamente-área	38.81	54.62
Franja baja (=100% de N)				
AO	0.456-2.282	=-útil	90.40 (93.38)	39.21 (66.07)
AF	0.851	100-útil	58.26	26.72
AG	0.856	100-útil	59.08	26.46

Nota. N= frecuencia acumulada; f= frecuencia relativa; N°= número de ULL; n_i = frecuencia absoluta de la ULL i.

^aLos porcentajes entre paréntesis del AO se refieren a las ULL léxicas del mismo.

UN ESTUDIO DE CASO: UN TEXTO ESCOLAR DE ASTRONOMÍA

Desde el texto preparado a este texto transformado definitivo se han producido cambios que han afectado a la franja alta y media. En el caso del AO ya no hay ULL en la franja alta, mientras que en los AF y AG las colocaciones son importantes en dicha franja (Tabla 66). Para los AF y AG la franja baja sólo tiene hápax y los datos relativos a las frecuencias más bajas son similares.

Tabla 66

Otras características de los textos transformados definitivos

Análisis	V _{alta}	% C _{alta}	V _{hápax}	% V _{n≤2}	% V _{hápax}	% ∑n _{i, hápax}
AO ^a	0 (0)	0.00	311 (304)	76.09 (72.95)	55.44 (57.14)	36.20(37.62)
AF	13	15.38	314	80.71	100.00	100.00
AG	10	30.00	309	80.88	100.00	100.00

Nota. V_{alta}= número de ULL en la franja alta; C_{alta}= número de colocaciones en la franja alta; V_{hápax}= número de ULL que son hápax; V_{n≤2}= número de ULL con n≤2; % V_{hápax}= porcentaje de hápax en la franja baja; % ∑n_{i, hápax}= porcentaje de frecuencias de los hápax en la franja baja.

^aLos valores entre paréntesis del AO se refieren a las ULL léxicas del mismo.

El AO es el único que presenta ULL léxicas y funcionales en el texto transformado, en una proporción en número de ULL de 92.32 y 7.68% y en frecuencias de 55.82 y 44.18%, respectivamente. A partir de ahora se denominan ULL léxicas a los procedentes de lemas léxicos y ULL funcionales a los procedentes de lemas funcionales. De las 44 ULL funcionales 7 son verbos auxiliares (un 1% de las ULL totales y un 16% de las ULL funcionales). Dos se encuentran en la franja media (*poder, haber*) y el resto en la franja baja. De los 7 verbos 2 tienen exclusivamente la función de verbo auxiliar (*poder, llegar*), 3 son verbos auxiliares la mayoría de las veces (*haber, comenzar, ir*), 1 lo es casi la mitad de las veces (*estar*) y 1 apenas ejerce de verbo auxiliar (*ser*, 2 de 37 veces). Otras 6 formas del verbo ser que no ejercen de verbo auxiliar se han unido a otros lemas del texto.

Cuando se calcula el cociente tipo/caso y la densidad léxica para cada categoría gramatical se obtienen los distintos valores de la Tabla 67. Las cinco últimas filas de cada análisis presentan los resultados de cada subcategoría léxica. La unión de lemas ha hecho que tanto los lemas funcionales como los léxicos se repitan todavía menos que en el texto lematizado, por lo que su número de relaciones ha disminuido (ver Tablas 44 y 67). En los tres análisis las subcategorías que más han aumentado su TTR respecto al texto lematizado o transformado sin colocaciones son las de adjetivos, lenguaje matemático y sustantivos, que a su vez son las subcategorías léxicas que más formaron parte de uniones, generando muchas veces nuevas ULL. En total se generaron en torno a un 10% de nuevas ULL léxicas que no estaban en el anterior texto lematizado o transformado. La densidad léxica disminuyó ligeramente respecto al texto lematizado en el AO. Al igual que ocurría con el TTR las mayores variaciones de LD respecto al texto lematizado corresponden a las subcategorías de adjetivos, sustantivos y lenguaje matemático en los tres análisis. Aún así los sustantivos llevan la mayor parte de la información muy por encima de otras subcategorías, por lo que se espera que estén más representados tanto en los vocabularios frecuentes como en las redes semánticas. Debido a la formación de ULL léxicas mixtas, algunas de éstas han quedado sin clasificar en las subcategorías. Esto produce errores variables en el recuento de las frecuencias de cada subcategoría (ver 5ª columna en Tabla 67).

La influencia de las transformaciones que sufre el texto, debido a las variaciones en la metodología, se estudia en las tres franjas de frecuencia en función de las categorías gramaticales y subcategorías léxicas. En las siguientes figuras se puede observar la importancia de las categorías gramaticales tanto en número de ULL como en frecuencias producidas en el AO.

UN ESTUDIO DE CASO: UN TEXTO ESCOLAR DE ASTRONOMÍA

Tabla 67

Características por categoría gramatical de los textos transformados definitivos

Categoría gramatical	V	TTR	LD (%)	$\delta\sum n_i/\sum n_i$ (%)
AO				
lemas funcionales	44	0.05		
lemas léxicos	529	0.43	55.82	
verbos	89	0.45	9.13	0.65
adverbios	29	0.35	3.83	0.49
adjetivos	130	0.57	10.50	6.70
lenguaje matemático	49	0.46	4.84	1.23
sustantivos	201	0.40	22.87	6.95
AF				
lemas léxicos	539	0.46	100.00	
verbos	88	0.46	16.34	1.11
adverbios	30	0.35	7.32	0.56
adjetivos	130	0.59	18.81	5.94
lenguaje matemático	49	0.47	8.85	0.93
sustantivos	204	0.44	39.66	6.12
AG				
lemas léxicos	523	0.45	100.00	
verbos	88	0.45	16.61	1.15
adverbios	30	0.35	7.28	0.76
adjetivos	126	0.59	18.24	5.54
lenguaje matemático	44	0.49	7.62	1.34
sustantivos	198	0.43	39.55	6.31

Nota. V= número de vocablos; TTR=cociente tipo/caso; LD=densidad léxica; $\delta\sum n_i/\sum n_i$ = error relativo en frecuencias absolutas.

Las distribuciones en número de ULL se presentan en las *Figuras 60, 62 y 64* y las de frecuencias en las *Figuras 61, 63 y 65*. En los datos que corresponden a la *Figura 60* hay menos verbos auxiliares en la franja media (un 5% de las ULL funcionales) que en la franja baja (un 11% de las ULL funcionales).

En el AO las distribuciones son análogas a las del texto lematizado con la salvedad de que la franja alta está vacía. Las ULL han pasado a ocupar sólo las franjas media y baja, cosa que no ocurría en el texto preparado. En una metodología

donde se va a precisar un sistema óptimo en función de la frecuencia para estudiar relaciones, el mismo estará dominado por palabras funcionales.

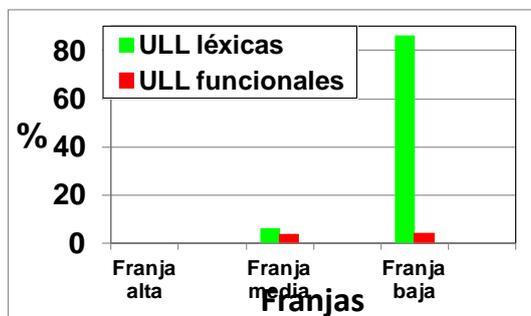


Figura 60.- Número de ULL por categoría gramatical en todo el texto de FESTESE001VD.

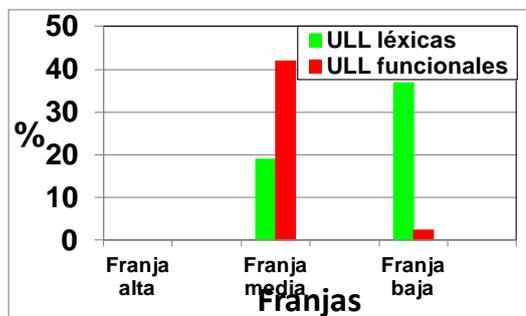


Figura 61.- Frecuencias por categoría gramatical en todo el texto de FESTESE001VD.

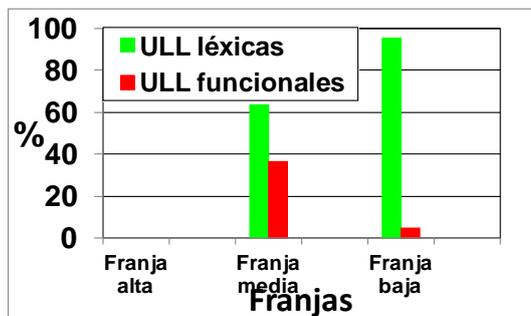


Figura 62.- Número de ULL por categoría gramatical en cada franja de FESTESE001VD.

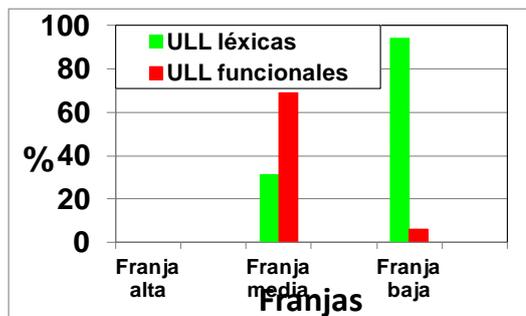


Figura 63.- Frecuencias por categoría gramatical en cada franja de FESTESE001VD.

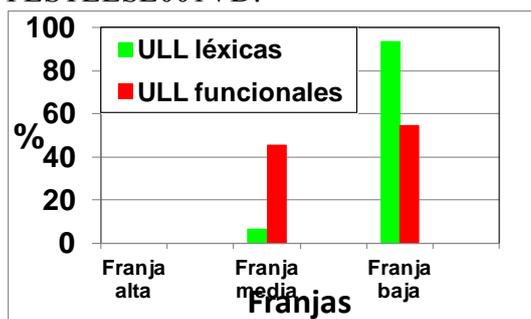


Figura 64.- Número de ULL de cada categoría gramatical en todo el texto de FESTESE001VD.

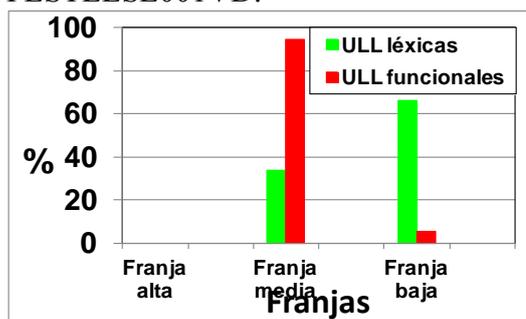


Figura 65.- Frecuencias de cada categoría gramatical en todo el texto de FESTESE001VD.

Debido a la eliminación de palabras funcionales y verbos auxiliares los datos de los AF y AG sólo son comparables con los de las Figuras 64 y 65 (Tabla 68). Las distribuciones en número de ULL y frecuencias sólo están correlacionadas en el AO para los lemas léxicos. En todos los análisis la categoría léxica aumenta en

UN ESTUDIO DE CASO: UN TEXTO ESCOLAR DE ASTRONOMÍA

número de ULL hacia la franja baja, pero sus frecuencias en el AO presentan su máximo en la franja baja, y en los AF y AG en la franja media. La eliminación de palabras funcionales y verbos auxiliares produce una redistribución de la categoría léxica en número de ULL y frecuencias. Las relaciones a estudiar en los AF y AG se darán exclusivamente entre ULL léxicas. Las diferencias entre el AF y AG son menores y se deben a la disminución de ULL y redistribución de frecuencias causados por la unificación de grafías. Estos comportamientos son semejantes a los del texto lematizado/transformado previo a las colocaciones (Tabla 45). Los resultados para las ULL léxicas en la franja alta son análogos en cuanto a valores a los del texto lematizado/transformado en los tres análisis.

Tabla 68

Distribución del número de ULL y frecuencias en todo el texto transformado definitivo en los tres análisis

Análisis	% N ^o léxicas	% N ^o funcionales	% $\sum n_i$, léxicas	% $\sum n_i$, funcionales
Franja alta ($\leq 15-20\%$ de N)				
AO ^a	0.00	0.00	0.00	0.00
AF	2.41		20.17	
AG	2.10		18.92	
Franja media ($\leq 60\%$ de N)				
AO	6.62	45.45	33.93	94.73
AF	39.33		53.11	
AG	38.81		54.62	
Franja baja (=100% de N)				
AO	93.38	54.55	66.07	5.27
AF	58.26		26.72	
AG	59.08		26.46	

Nota. N^o léxicas = número de ULL léxicas; N^o funcionales = número de vocablos funcionales; n_i léxicas = frecuencia absoluta de la ULL léxica i ; n_i , funcionales = frecuencia absoluta de la ULL funcional i .

^aLos datos del AO son los de cada categoría en todo el texto.

Dentro de las ULL hay cinco subcategorías bien diferenciadas: verbos, adverbios, adjetivos, sustantivos y lenguaje matemático. Las ULL se han clasificado

en las subcategorías mencionadas consultando el Diccionario de la RAE online (2001) tal y como se ha explicado en un apartado anterior (4. 2.). La clasificación de las ULL para el AO se puede ver en la quinta columna de la Tabla 101 (Anexo R). La importancia relativa de cada subcategoría en el AO se observa en las *Figuras 66 y 67*.

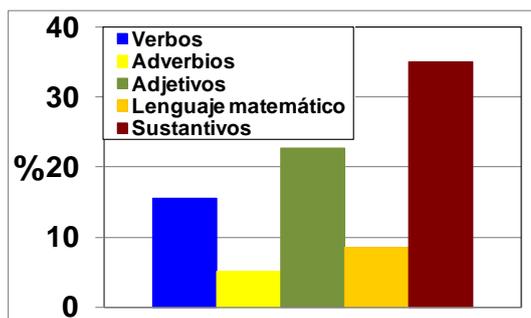


Figura 66.- Número de ULL por subcategoría léxica de FESTESE001VD.

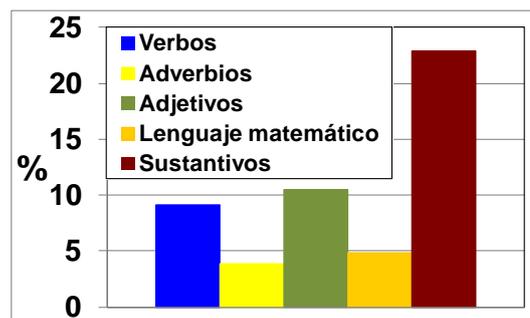


Figura 67.- Frecuencias por subcategoría léxica de FESTESE001VD.

Las subcategorías léxicas presentan una menor diferencia entre el número de ULL y las frecuencias producidas, comparado con las unidades funcionales. Esto se debe a la escasa repetición de ULL en comparación con las funcionales. A pesar de que el lenguaje matemático, verbos, adjetivos y sustantivos tienen mayor número de ULL que las funcionales, éstas producen más frecuencias (un 44%).

Las gráficas para los AF y AG son parecidas, por lo que se presentan los resultados en la Tabla 69 para facilitar la comparación. Debido a las uniones entre lemas de distintas subcategorías no todas las ULL han sido clasificadas dentro de las subcategorías mostradas en la Tabla 69. No hay grandes variaciones de valores para el número de ULL y frecuencias entre los distintos análisis.

La distribución en número de ULL en el AO se presenta en la *Figuras 68, 70 y 72*, y las de frecuencias en las *Figuras 69, 70 y 72*. Las ULL que estaban formadas por más de una subcategoría no forman parte de estas distribuciones.

UN ESTUDIO DE CASO: UN TEXTO ESCOLAR DE ASTRONOMÍA

Tabla 69

Distribución de las subcategorías léxicas en los tres análisis en el texto transformado definitivo

Subcategorías léxicas	AO ^a		AF		AG	
	% N ^o	% $\sum n_i$	% N ^o	% $\sum n_i$	% N ^o	% $\sum n_i$
verbos	16.82	16.35	16.33	16.34	16.83	16.61
adverbios	5.48	6.87	5.57	7.32	5.74	7.28
adjetivos	24.57	18.81	24.12	18.81	24.09	18.24
lenguaje matemático	9.26	8.67	9.09	8.85	8.41	7.62
sustantivos	38.00	40.96	37.85	39.66	37.86	39.55

Nota. N^o= número de vocablos; n_i= frecuencia absoluta del vocablo i.

^aEn el AO los datos son respecto a todos los lemas léxicos.

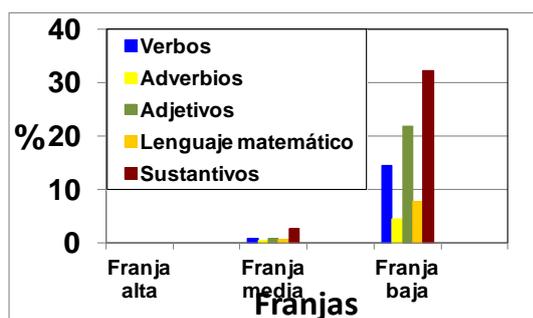


Figura 68.- Número de ULL por subcategoría léxica en todo el texto de FESTESE001VD.

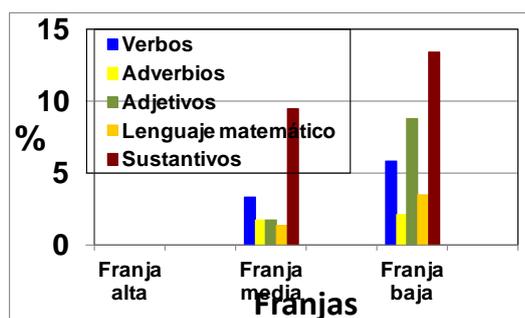


Figura 69.- Frecuencias por subcategoría léxica en todo el texto de FESTESE001VD.

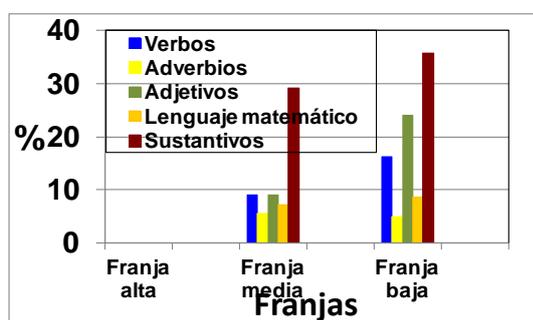


Figura 70.- Número de ULL por subcategoría léxica en cada franja de FESTESE001VD.

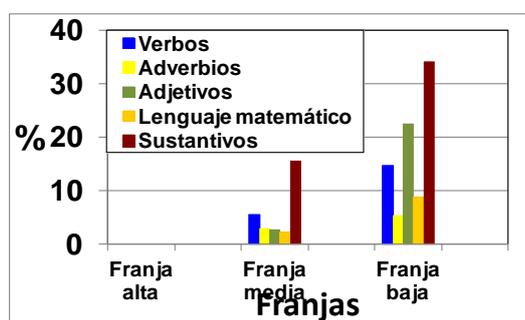


Figura 71.- Frecuencias por subcategoría léxica en cada franja de FESTESE001VD.

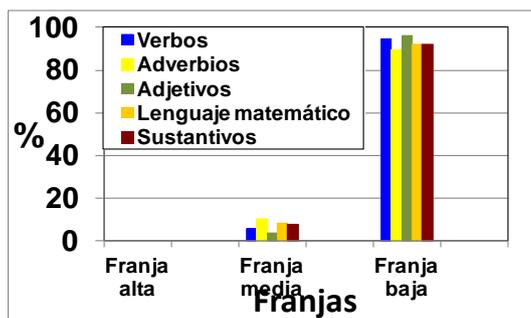


Figura 72.- Número de ULL de cada subcategoría léxica en todo el texto de FESTESE001VD.

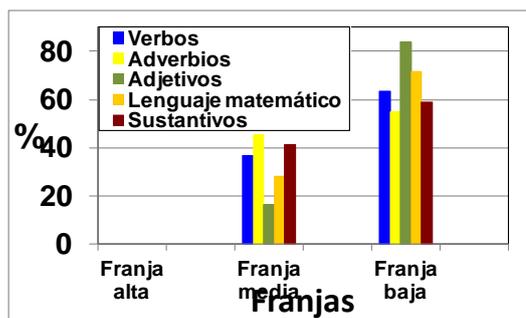


Figura 73.- Frecuencias de cada subcategoría léxica en todo el texto de FESTESE001VD.

Sólo el comportamiento de los adverbios en número de ULL no está correlacionado con la tendencia general de su categoría gramatical. Sustantivos, adjetivos, verbos y lenguaje matemático constituyen un 94% del número de ULL y un 93% de las frecuencias de la categoría léxica (Tabla 69). El número de ULL y las frecuencias de estas cuatro subcategorías parecen estar correlacionados en el texto transformado.

Las gráficas para los AF y AG son parecidas, por lo que se presentan los resultados en las Tablas 70, 71 y 72, para facilitar la comparación. La eliminación de palabras funcionales y verbos auxiliares en los AF y AG produce una redistribución de las subcategorías a lo largo de las tres franjas.

Respecto al texto transformado sin colocaciones ya no hay adjetivos ni lenguaje matemático en la franja alta en los AF y AG (ver Tablas 47, 48 y 49). Los resultados en los AF y AG son similares en las Tablas 70 y 71, siendo más diferentes para verbos y adverbios en las franjas alta y media de la Tabla 72. La eliminación de palabras funcionales y verbos auxiliares produce una redistribución de las subcategorías léxicas de los AF y AG respecto al AO. El comportamiento de cada subcategoría léxica está correlacionado con la tendencia general de la categoría léxica en los AF y AG. Las subcategorías léxicas aumentan en número de

UN ESTUDIO DE CASO: UN TEXTO ESCOLAR DE ASTRONOMÍA

lemas hacia la franja baja mientras que sus frecuencias parecen presentar su máximo en la franja baja en el AO y en la franja media en los AF y AG. En los AF y AG la franja alta está dominada por los sustantivos tanto en número de lemas como en frecuencias (casi un 50%), por lo que se espera que dominen en el vocabulario frecuente y las redes semánticas.

Tabla 70

Distribución de las subcategorías léxicas en todo el texto transformado definitivo en los tres análisis

Análisis	AO ^a		AF		AG	
	% N ^o	% $\sum n_i$	% N ^o	% $\sum n_i$	% N ^o	% $\sum n_i$
Franja alta ($\leq 15-20\%$ de N)						
verbos	0.00	0.00	0.56	4.77	0.38	4.11
adverbios	0.00	0.00	0.56	3.23	0.38	2.31
adjetivos	0.00	0.00	0.00	0.00	0.00	0.00
lm	0.00	0.00	0.00	0.00	0.00	0.00
sustantivos	0.00	0.00	0.93	9.62	0.76	9.08
Franja media ($\leq 60\%$ de N)						
verbos	0.87 (0.95)	3.33 (5.97)	4.82	6.55	5.16	7.45
adverbios	0.52 (0.57)	1.73 (3.11)	2.04	2.72	2.29	3.60
adjetivos	0.87 (0.95)	1.69 (3.03)	9.28	12.00	8.99	11.47
lm	0.70 (0.76)	1.37 (2.45)	4.45	6.72	3.82	5.57
sustantivos	2.79 (3.02)	9.45 (16.93)	13.73	19.40	13.96	20.12
Franja baja (=100% de N)						
verbos	14.66 (15.88)	5.80 (10.38)	10.95	5.02	11.28	5.05
adverbios	4.54 (4.91)	2.10 (3.76)	2.97	1.36	3.06	1.37
adjetivos	21.82 (23.63)	8.81 (15.78)	14.84	6.81	15.11	6.76
lm	7.85 (8.51)	3.47 (6.21))	4.64	2.13	4.59	2.05
sustantivos	32.29 (34.97)	13.42 (24.04)	23.19	10.64	23.14	10.36

Nota. N^o= número de vocablos; n_i= frecuencia absoluta del vocablo i; lm= lenguaje matemático.

^aLos porcentajes entre paréntesis del AO se refieren a los lemas léxicos del mismo.

Tabla 71

Distribución de las subcategorías léxicas en cada franja en los tres análisis en el texto transformado definitivo

Análisis	AO		AF		AG	
	% N°	% $\sum n_i$	% N°	% $\sum n_i$	% N°	% $\sum n_i$
Franja alta ($\leq 15-20\%$ de N)						
verbos	0.00	0.00	23.08	23.63	18.18	21.72
adverbios	0.00	0.00	23.08	16.03	18.18	12.22
adjetivos	0.00	0.00	0.00	0.00	0.00	0.00
lm	0.00	0.00	0.00	0.00	0.00	0.00
sustantivos	0.00	0.00	38.46	47.68	36.36	47.96
Franja media ($\leq 60\%$ de N)						
verbos	9.09	5.48	12.26	12.34	13.30	13.64
adverbios	5.45	2.85	5.19	5.13	5.91	6.58
adjetivos	9.09	2.78	23.58	22.60	23.15	21.00
lm	7.27	2.25	11.32	12.66	9.85	10.19
sustantivos	29.09	15.54	34.91	36.54	35.96	36.83
Franja baja (=100% de N)						
verbos	16.22	14.78	18.79	18.79	19.09	19.09
adverbios	5.02	5.36	5.10	5.10	5.18	5.18
adjetivos	24.13	22.47	25.48	25.48	25.57	25.57
lm	8.69	8.85	7.96	7.96	7.77	7.77
sustantivos	35.71	34.23	39.81	39.81	39.16	39.16

Nota. N°= número de vocablos; n_i = frecuencia absoluta del vocablo i ; lm= lenguaje matemático.

Tabla 72

Distribución de cada subcategoría léxica en todo el texto transformado definitivo en los tres análisis

Análisis	AO		AF		AG	
	% N°	% $\sum n_i$	% N°	% $\sum n_i$	% N°	% $\sum n_i$
Franja alta ($\leq 15-20\%$ de N)						
verbos	0.00	0.00	3.41	29.17	2.27	24.74
adverbios	0.00	0.00	10.00	44.19	6.67	31.76
adjetivos	0.00	0.00	0.00	0.00	0.00	0.00
lm	0.00	0.00	0.00	0.00	0.00	0.00
sustantivos	0.00	0.00	2.45	24.25	2.02	22.94
Franja media ($\leq 60\%$ de N)						
verbos	5.62	36.50	29.55	40.10	30.68	44.85
adverbios	10.34	45.24	36.67	37.21	40.00	49.41
adjetivos	3.85	16.09	38.46	63.80	37.30	62.91
lm	8.16	28.30	48.98	75.96	45.45	73.03
sustantivos	7.96	41.32	36.27	48.93	36.87	50.87

UN ESTUDIO DE CASO: UN TEXTO ESCOLAR DE ASTRONOMÍA

Análisis	AO		AF		AG	
	% N°	% $\sum n_i$	% N°	% $\sum n_i$	% N°	% $\sum n_i$
Franja baja (=100% de N)						
verbos	94.38	63.50	67.05	30.73	67.05	30.41
adverbios	89.66	54.76	53.33	18.60	53.33	18.82
adjetivos	96.15	83.91	61.54	36.20	62.70	37.09
lm	91.84	71.70	51.02	24.04	54.55	26.97
sustantivos	92.04	58.68	61.27	26.82	61.11	26.19

Nota. N°= número de vocablos; n_i = frecuencia absoluta del vocablo i ; lm= lenguaje matemático.

Los porcentajes de hápax en cada subcategoría se presentan en la Tabla 73.

Otros datos relativos a la importancia de cada subcategoría léxica respecto a su categoría, así como a la franja alta se presentan en la Tabla 74. En los tres análisis más de la mitad de los vocablos de las subcategorías léxicas son hápax tras la lematización. Las ULL que estaban formadas por más de una subcategoría no forman parte de los datos de las tablas.

Tabla 73

Hápax en las subcategorías léxicas en el texto transformado definitivo

Subcategorías léxicas	% N° _{hápax}			% N° _{léxicas,hápax}		
	AO	AF	AG	AO	AF	AG
verbos	70.24	67.05	67.05	19.41	10.95	11.28
adverbios	61.54	53.33	53.33	5.26	2.97	3.06
adjetivos	63.20	61.54	62.70	25.99	14.84	15.11
lm	53.33	51.02	54.55	7.89	4.64	4.59
sustantivos	65.95	61.27	61.11	40.13	23.19	23.14

Nota. N°_{hápax}= número de lemas de la subcategoría que son hápax; N°_{léxicas,hápax}= hápax respecto a la categoría léxica.

La eliminación de las palabras funcionales y verbos auxiliares ha provocado una redistribución de los lemas léxicos, de manera que sólo 3 subcategorías se hallan representadas en la franja alta en los AF y AG. La unificación de grafías vuelve a producir una redistribución de ULL en el AG. La franja alta está dominada por los sustantivos en cuanto al número de lemas y frecuencias en los AF y AG. En

la franja alta aparecen colocaciones que no se han clasificado en la Tabla 74: *cúmulo globular* y *secuencia principal* en el AF (12.66% de las frecuencias de la franja alta), *cúmulo globular*, *temperatura superficial* y *secuencia principal* en el AF (18.10% de las frecuencias de la franja alta). Dentro de estas colocaciones hay adjetivos, por lo que se podría decir que en la franja alta sólo falta la subcategoría de lenguaje matemático tras las colocaciones.

Tabla 74

Importancia de las subcategorías léxicas del texto transformado definitivo

Subcategorías léxicas	AF			AG		
	Nº	Lemas en franja alta	% $\sum n_{i,subcategoría}$	Nº	Lemas en franja alta	% $\sum n_{i,subcategoría}$
verbos	88	<i>ser, ver, tener</i>	29.17	88	<i>ser, ver</i>	24.74
adverbios	30	<i>más, muy, aproximadamente</i>	44.19	30	<i>más, muy</i>	31.76
adjetivos	130		0.00	126		0.00
lm	49		0.00	44		0.00
sustantivos	204	<i>estrella, cúmulo, núcleo, luminosidad, formal secuencia</i>	24.25	198	<i>estrella, luminosidad, núcleo, cúmulo</i>	22.94

Nota. Nº= número de lemas de la subcategoría; $\sum n_{i,subcategoría}$ = frecuencias en la franja alta respecto a la subcategoría.

4. 4. 2.- Vocabulario más frecuente y vocabulario específico

El análisis de la distribución de frecuencias absolutas y frecuencias acumuladas para el texto transformado definitivo del AO se presenta en las Figuras 74 y 75. En ambas figuras el mejor ajuste a la distribución viene dado por la curva marrón, cuyos parámetros se ven en la parte inferior derecha.

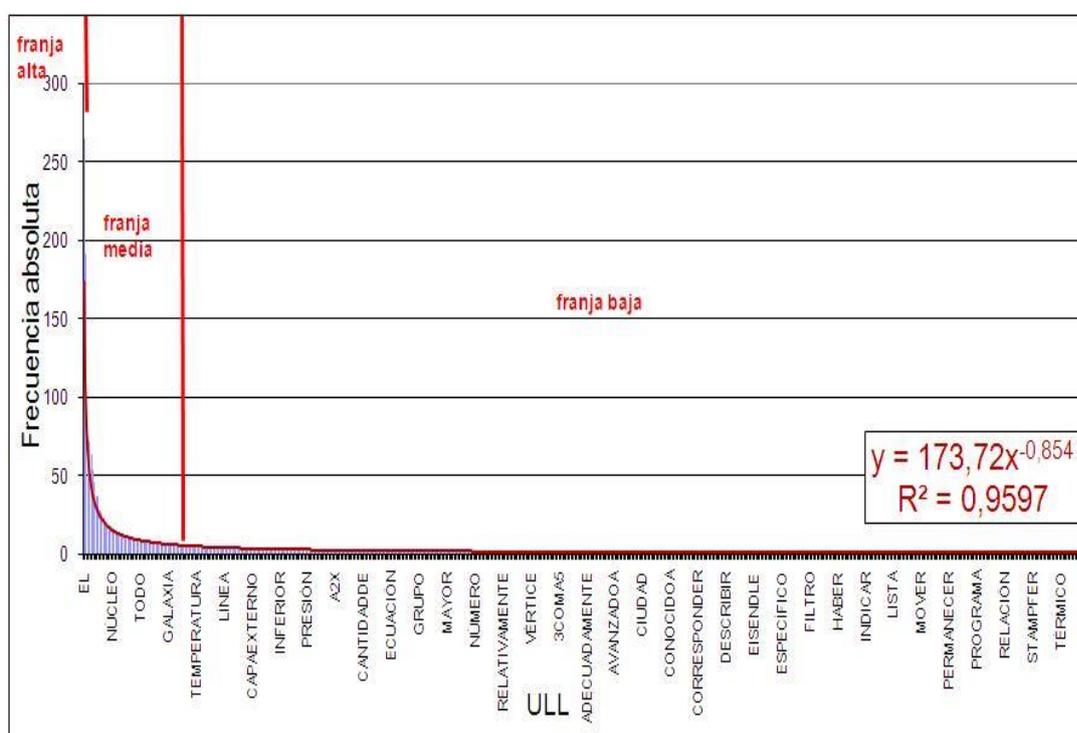


Figura 74.- Ajuste de la distribución de frecuencias absolutas del texto lematizado FESTESE001VD.

En los AF y AG las distribuciones y los ajustes son similares. Las distribuciones de frecuencia absoluta presentan los ajustes $y=173.72x^{-0.854}$ ($R^2=0.9597$) e $y=52.462x^{-0.665}$ ($R^2=0.9418$), respectivamente, mientras que las distribuciones de frecuencia acumulada presentan los ajustes $y=333.22\ln(x)+19.69$ ($R^2=0.9909$) e $y=244.34\ln(x)-441.04$ ($R^2=0.9525$). Los valores del coeficiente de determinación, R^2 , indican que el ajuste explica en más del 94% la variación de los datos de los tres análisis.

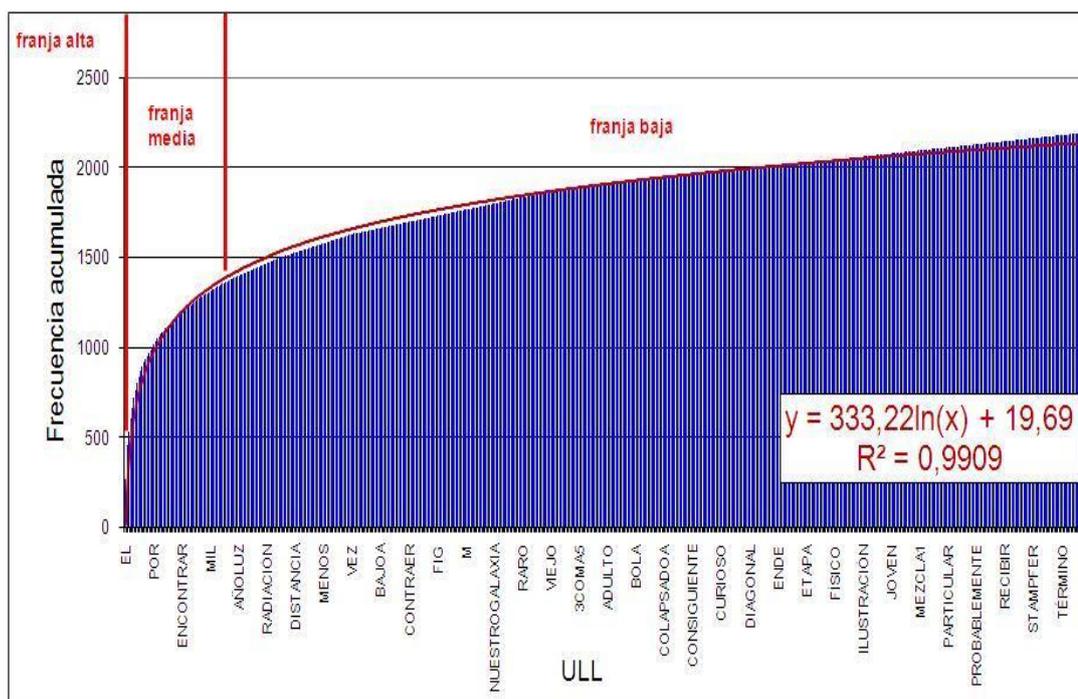


Figura 75.- Ajuste de la distribución de frecuencias acumuladas del texto lematizado FESTESE001VD.

Para los dos tipos de distribuciones se halla la derivada de la función y en qué punto la pendiente de la tangente es -1 o +1. Esto define la franja de vocabulario más frecuente, cuyas características vienen dadas en la Tabla 75.

Tabla 75

Franja de vocabulario más frecuente analizando las frecuencias de los textos transformados definitivos

Distribución de frecuencias	Función derivada	Tangente	% N°	%($\sum n_i$)
AO				
Absoluta	$y' = -173,72x^{-1,854}$	$x=14.83$ (m= -1)	2.27	42.72
Acumulada	$y' = 333,22/x$	$x=333.22$ (m= +1)	100	100
AF				
Absoluta	$y' = -31.691x^{-1.648}$	$x=8.13$ (m= -1)	1.48	15.74
Acumulada	$y' = 250.76/x$	$x=250.76$ (m= +1)	100	100
AG				
Absoluta	$y' = -34.887x^{-1,665}$	$x=8,62$ (m= -1)	1.72	17.98
Acumulada	$y' = 244.34/x$	$x=244.34$ (m= +1)	100	100

Nota. m= pendiente de la recta tangente; N°= número de ULL; n= frecuencia absoluta.

UN ESTUDIO DE CASO: UN TEXTO ESCOLAR DE ASTRONOMÍA

En los tres análisis la distribución de frecuencias absolutas se ajusta con una función de potencias del tipo $(a/x^b) \pm c$ (*Figura 74*), mientras que la distribución de frecuencias acumuladas se ajusta con una función logarítmica del tipo $(a \cdot \ln x) \pm b$ (*Figura 75*). La función logarítmica no tiene ninguna asíntota, por lo que presenta un cambio de curvatura más suave que la función de potencias. En ambas representaciones en el eje de ordenadas se representan datos numéricos, mientras que en el eje de abscisas se representan las ULL por orden lexicométrico (números naturales). La apariencia de las curvas puede ser engañosa. Visualmente el ajuste de la *Figura 75* parece tener una tangente con pendiente de valor +1 cerca del eje de ordenadas, pero al hacer el análisis matemático se encuentra que dicho punto está muy alejado del eje para los tres análisis (*Tabla 75*). El estudio de la distribución en frecuencias acumuladas proporciona una franja que contiene todo el vocabulario del texto analizado, mientras que el análisis de la distribución de frecuencias absolutas proporciona una franja de pocas ULL que producen bastantes frecuencias. Se podría considerar como vocabulario más frecuente el situado a la izquierda del punto donde la pendiente es igual a -1 en el análisis de la distribución de frecuencias absolutas (parte superior de la *Tabla 76*). En el caso del AO se escoge como límite del rango de frecuencia absoluta 19, puesto que no hay un rango con frecuencia absoluta igual a 18.

Juilland & Chang-Rodríguez (1964) muestran las distribuciones de frecuencias en rangos y con los datos ajustados a una función (ver *Graph1*, p. XLII). En el eje de las ordenadas se representa la frecuencia absoluta promedio de las 500 palabras que forman cada rango, y en el eje de las abscisas el rango de frecuencias, cuyas palabras están por orden lexicométrico. Si se dibujara una tangente de pendiente -1 a dicha función se observaría que es una buena aproximación al

vocabulario más frecuente de su corpus, ya que estaría un poco más a la derecha del primer rango.

Tabla 76

Características de la franja determinada por $m=-1$ en las distribuciones de frecuencias absolutas

Análisis	$m=-1$	$n \geq$	Nº	ULL límite	% Nº	%($\sum n_i$)
Distribución frente a las ULL ^a						
AO	(14.83, 17.37)	19	13	<i>el-su</i>	2.27	42.72
AF	(8.13, 12.58)	13	8	<i>estrella-</i> <i>secuenciaprincipal</i>	1.48	15.74
AG	(8.62, 12.52)	13	9	<i>estrella-</i> <i>secuenciaprincipal</i>	1.72	17.98
Distribución frente a las frecuencias acumuladas ^b						
AO	(522.95, 144.85)	191	2	<i>el-de</i>	0.35	20.81
AF	(78.24, 51.75)	59	1	<i>estrella</i>	0.19	5.02
AG	(88.12, 54.63)	63	1	<i>estrella</i>	0.19	5.39

Nota. m = pendiente de la recta tangente; n = frecuencia absoluta; Nº= número de ULL.

^a Para $m= -1$ se escribe el punto (x, y) donde se encuentra dicha tangente, siendo x una ULL e $y=n$.

^b Para $m= -1$ se escribe el punto (N, n) donde se encuentra dicha tangente, siendo N la frecuencia acumulada.

Otra manera de determinar de forma matemática la franja de vocabulario más frecuente sería en una gráfica de frecuencia absoluta frente a la frecuencia acumulada. La *Figura 76* presenta dicha gráfica para el AO. Las gráficas de los AF y AG son similares y el ajuste para las distribuciones en los tres análisis es una función de potencias del tipo $(a/x^b) \pm c$. Las características de la franja definida mediante la tangente con pendiente -1 (parte inferior de la Tabla 76) contiene aún menos ULL que en el caso descrito para la *Figura 74*.

A la hora de decidir cuál es el mejor método para determinar el vocabulario más frecuente se tendrá en cuenta que la cantidad de frecuencias producida por el mismo sea máxima. Por tanto se elegirá la franja definida por el análisis del ajuste de la *Figura 74*. A partir de la franja de vocabulario más frecuente (Tabla 77), y

UN ESTUDIO DE CASO: UN TEXTO ESCOLAR DE ASTRONOMÍA

aplicando el criterio descrito en Chung (2003), se obtiene el vocabulario específico (señalado en negrita en dicha tabla).

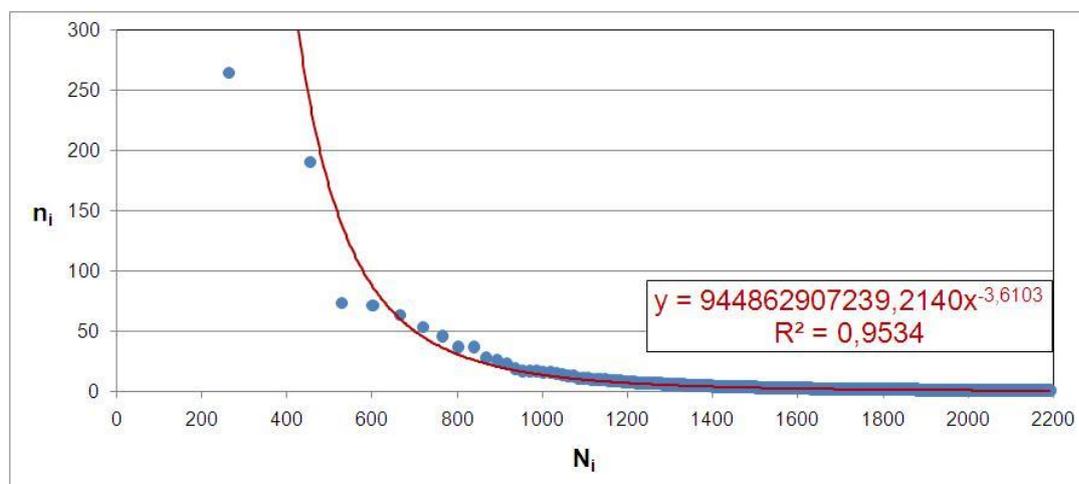


Figura 76.- Determinación del vocabulario más frecuente mediante las frecuencias absoluta y acumulada.

Tabla 77

Vocabulario más frecuente y específico del texto transformado definitivo

OL	ULL	f_i (a1000)	f_{CREA} (a 1000)	Cociente (f/f_{CREA})
AO				
1	EL	120.949	104.538	1.157
2	DE	87.175	64.994	1.341
3	ESTRELLA	33.775	0.097	348.663
4	UNO	32.862	21.258	1.546
5	EN	29.21	27.501	1.062
6	SE	24.646	13.115	1.879
7	Y	20.995	28.073	0.748
8	SER	16.887	11.366	1.486
9	QUE	16.887	30.341	0.557
10	DEL	12.78	12.057	1.060
11	ESTO	11.867	4.964	2.391
12	A	10.497	21.561	0.487
13	SU	8.672	7.162	1.211
AF				
1	ESTRELLA	50.213	0.097	518.355
2	SER	29.787	11.366	2.621
3	CÚMULO	14.468	0.008	1758.951
4	CÚMULOGLOBULAR	14.468	0.000	33819.838
5	NÚCLEO	13.617	0.045	303.586
6	MÁS	12.766	4.291	2.975
7	MUY	11.064	1.351	8.191
8	SECUENCIAPRINCIPAL	11.064	0.000	68277.736

UN ESTUDIO DE CASO: UN TEXTO ESCOLAR DE ASTRONOMÍA

OL	ULL	f_i (a1000)	f_{CREA} (a 1000)	Cociente (f/f_{CREA})
		AG		
1	ESTRELLA	53.938	0.099	542.697
2	SER	31.678	12.382	2.558
3	CÚMULOGLOBULAR	19.692	0.000	46011.450
4	LUMINOSIDAD	13.699	0.006	2233.149
5	NÚCLEO	13.699	0.046	298.637
6	TEMPERATURASUPERFICIAL	12.842	0.000	30006.045
7	MÁS	11.986	4.550	2.634
8	MUY	11.13	1.531	7.271
9	SECUENCIAPRINCIPAL	11.13	0.000	68655.477

Nota. OL= orden lexicométrico; f_i = frecuencia relativa en el texto FESTESE001VD; f_{CREA} =frecuencia relativa en el corpus CREA.

Las características del vocabulario más frecuente y específico para los tres análisis se muestran en la Tabla 78. Los porcentajes de ULL y frecuencias producidas por las distintas categorías gramaticales que forman parte del vocabulario más frecuente se muestran en la Tabla 79 y las del vocabulario específico en la Tabla 80. Los resultados para el AO reflejan el hecho de que en un listado de frecuencias las palabras de mayor frecuencia son principalmente palabras funcionales (Renouf, 1992; Sinclair, 1991). En los AF y AG los sustantivos son los que más frecuencias producen en el vocabulario más frecuente. En los tres análisis todas las ULL del vocabulario específico contienen un sustantivo (Tabla 80), siendo estos los que más frecuencias producen. En los AF y AG hay además adjetivos colocados con sustantivos.

Tabla 78

Características del vocabulario más frecuente y específico

Análisis	% N° franja	% $\sum n_i$ franja	% N° vocabulario específico	% $\sum n_i$ vocabulario específico
AO ^a	23.64	70.27	7.69	7.50
AF	61.54	78.06	62.50	65.95
AG	81.82	95.02	66.67	69.52

Nota. N° franja= número de ULL respecto a la franja alta; $\sum n_i$ franja= frecuencias respecto a la franja alta; N° vocabulario específico= número de ULL del vocabulario más frecuente que es específico; $\sum n_i$ vocabulario específico= frecuencias del vocabulario más frecuente que es específico.

^aLos valores del AO se refieren a la franja media.

UN ESTUDIO DE CASO: UN TEXTO ESCOLAR DE ASTRONOMÍA

Tabla 79

Categorías gramaticales del vocabulario más frecuente

Categorías gramaticales	AO		AF		AG	
	% N°	% Σn _i	% N°	% Σn _i	% N°	% Σn _i
ULL funcionales	84.62	88.14				
ULL léxicas	15.38	11.86	100.00	100.00	100.0	100.00
Sustantivos	7.69	7.91	37.50	49.73	33.33	45.24
Verbos	7.69	3.95	12.50	18.92	11.11	17.62
Adverbios			25.00	15.14	22.22	12.86
Sustantivos-adjetivos			25.00	16.22	33.33	24.29

Nota. N°= número de ULL; n= frecuencia absoluta.

Tabla 80

Categorías gramaticales del vocabulario específico

Categorías léxicas	AO		AF		AG	
	% N°	% Σn _i	% N°	% Σn _i	% N°	% Σn _i
Sustantivos	100.00	100.00	60.00	75.41	50.00	65.07
Sustantivos-adjetivos			40.00	24.59	50.00	34.93

Nota. N°= número de UL; n= frecuencia absoluta.

Al eliminar las palabras funcionales aparecen colocaciones y adverbios en el vocabulario más frecuente. La mayor diferencia entre el AF y el AG es el aumento de colocaciones. La única ULL específica común para los tres análisis es *estrella*, debido a que los lemas funcionales, que no se han eliminado en el AO, tienen frecuencias muy altas. Los AF y AG tienen en común 7 ULL, que no ocupan el mismo orden lexicométrico. Del AF al AG desaparece *cúmulo* y aparece una nueva colocación, *temperaturasuperficial*, como consecuencia de la unificación de grafías.

4. 4. 3.- Determinación del sistema a estudiar

Para determinar el sistema a estudiar se aplica el Estudio de Entorno del PAFE con las características de la Tabla 16 (apartado 3. 2. 2. 2. 3.). Las características de los sistemas a analizar vienen dados en la Tabla 81.

Tabla 81

Sistemas a analizar

Sistema (f_{\min})	ULL límite	% N°	% (Σn_i)	% (Σra_i)	% ($\Sigma ra_i / \Sigma n_i * 2$)
AO					
0.456 ^a	el-útil	100.00	100.00	100.00	66.00
0.913	el-él	45.72	85.81	79.94	61.49
1.369	el-vez	23.91	74.40	78.63	69.75
1.826	el-unoa	16.40	68.51	66.25	63.82
2.738 ^b	<i>el-t</i>	9.60	60.79	52.14	56.61
5.933	el-secuenciaprincipal	3.84	49.02	30.22	40.69
8.672	el-su	2.27	42.72	21.44	33.12
16.887	el-que	1.57	38.34	18.33	31.55
29.210	el-en	0.87	30.40	12.86	27.93
AF					
0.851 ^a	estrella-útil	100.00	100.00	100.00	91.40
1.702 ^b	<i>estrella-área</i>	41.93	73.28	53.35	66.55
2.553	estrella-vida	19.29	52.68	27.37	47.50
3.404	estrella-vidaestrella	12.06	42.72	17.88	38.25
5.106	estrella-t	4.64	28.00	8.75	28.57
8.511 ^c	estrella-tener	2.41	20.17	4.93	22.36
11.064	estrella-secuenciaprincipal	1.48	15.74	3.26	18.9
AG					
0.856 ^a	estrella-útil	100.00	100.00	100.00	91.35
1.712 ^b	<i>estrella-área</i>	40.92	73.54	53.89	66.94
2.553	estrella-vida	19.12	54.02	28.77	48.65
3.404	estrella-víalácteo	12.43	45.03	19.87	40.30
5.106	estrella-radio	5.35	31.34	10.68	31.15
9.418 ^c	estrella-ver	1.91	18.92	4.87	23.53
11.130	estrella-secuenciaprincipal	1.72	17.98	3.84	19.52

Nota. f_{\min} = frecuencia relativa mínima; N°=número de ULL; n= frecuencia absoluta; ra= relación absoluta.

^aTexto transformado definitivo completo. ^bFranjas alta y media. ^cFranja alta.

UN ESTUDIO DE CASO: UN TEXTO ESCOLAR DE ASTRONOMÍA

En la Tabla 81 se muestran selecciones de sistemas basados en los límites de las franjas alta, media y baja, la franja del vocabulario más frecuente (en negrita) y otros cortes que proporcionan valores intermedios o menores en cuanto a la proporción de frecuencias absolutas que contienen. La proporción de ULL no varía de igual modo que el de las frecuencias absolutas, pues siempre es un pequeño número de ULL el que produce un gran número de frecuencias; véase por ejemplo el sistema definido por $f_{\min}=29.210$ en el AO. La suma de frecuencias absolutas tampoco varía de igual modo que la suma de relaciones absolutas. Las relaciones disminuyen mucho más rápidamente en los AF y AG que en el AO, como se ve en la columna $\%(\Sigma r_i)$ de la Tabla 81.

Los sistemas definidos por $f_{\min}>0.456$, 0.851 y 0.856 para los AO, AF y AG, respectivamente, no contienen ninguna relación con hápax o entre ellos, es decir, la diferencia en las relaciones de los sistemas de la primera a la segunda fila se debe a la supresión de los hápax. Eso no quiere decir que no existan más relaciones de valor 1, puesto que se pueden dar entre ULL que no sean hápax.

A la hora de estudiar las ULL que etiquetan a los conceptos, no sólo interesa la frecuencia de las mismas sino también las relaciones absolutas, ya que dicha ULL puede ser importante no sólo por tener una frecuencia alta sino también por tener relaciones absolutas altas con otra ULL. Podría darse el caso de que una ULL con una frecuencia no muy alta presentara unas relaciones fuertes con otras ULL (ver la relación entre *pérdida* y *masa* en las Figuras 18 y 19 que se ha perdido con el corte de frecuencias). Por lo tanto la selección del sistema óptimo no se puede basar exclusivamente en el criterio de las frecuencias sino que hay que tener en cuenta las relaciones absolutas contenidas en el sistema. Para intentar encontrar un

indicador de cuál podría ser el sistema óptimo se ha hallado el cociente entre las relaciones absolutas y las frecuencias absolutas (ver sexta columna de la Tabla 81). Supongamos una ULL i con frecuencia absoluta n_i . En un caso ideal en el que su entorno 1 está lleno de otras ULL las relaciones totales, rat_i , son el doble que su frecuencia: $rat_i = 2 n_i$. Por tanto una medida de las relaciones que se conservan para una ULL en un sistema dado será el cociente $rat_i / n_i * 2$. Cuando este cociente sea igual o mayor que el 80% se considerará que la ULL ha conservado sus relaciones. Para todas las ULL de un sistema se puede considerar el cociente $\Sigma rat_i / \Sigma n_i * 2$, siendo $\Sigma rat_i = \Sigma ra_i$, ya que Σra_i se encuentra duplicado. Cuando el cociente sea del 100%, significa que el entorno de todas las ULL está completamente lleno (caso ideal). Debido a los puntos y seguidos y puntos y apartes habrá ULL que se alejen del caso ideal. En el texto transformado definitivo hay 65 puntos y seguidos y 35 puntos y aparte, que suponen un vacío de 200 ULL en entornos de 1. Es decir hay una desviación global del 6.47% del caso ideal en el AO (8.52 y 8.57% en los AF y AG, respectivamente). El sistema del vocabulario más frecuente ha perdido muchas relaciones (ver porcentajes de Σra_i y $\Sigma ra_i * 2 / \Sigma n_i$). El sistema definido por la franja media, señalado en cursiva en la Tabla 81, conserva bastantes relaciones y contiene un elevado número de ULL (55, 226 y 214 para los AO, AF y AG, respectivamente). De elegir este último sistema hay que recordar que en los análisis de redes semánticas no aparecen todas las ULL, igual que en las *Figuras 18, 19 y 20* no aparecieron todas las ULL que constituían los sistemas.

La *Figura 77* muestra en escala logarítmica el número de ULL (representados con símbolos) y el número de relaciones (representados con barras) de dicho valor de relación para los nueve sistemas definidos mediante una frecuencia relativa mínima en el AO. Las distribuciones para los AF y AG son

UN ESTUDIO DE CASO: UN TEXTO ESCOLAR DE ASTRONOMÍA

parecidas, sólo que hay menos valores de relación (de 1 a 6) y un número menor (7) de sistemas definidos para estudiar. Todos los sistemas contienen el mismo número de ULL y producen el mismo número de relaciones para valores de relación igual o superior a 14 (4 y 5 para los AF y AG, respectivamente). Debido al corte en la frecuencia relativa mínima de los sistemas, a veces se pierden ULL que contribuyen a relaciones. A mayor frecuencia relativa mínima, menos relaciones de valor bajo se tienen. La desaparición de algunas ULL por el corte en frecuencia relativa puede provocar la pérdida de valor de relaciones altos. Es el caso de los sistemas con $f_{\min} \geq 5.933$ en el AO. Todos los sistemas con $f_{\min} \leq 2.738$ (1.702 y 1.712) tienen el mismo número de ULL y relaciones a partir del valor de relación 5 (4 y 5).

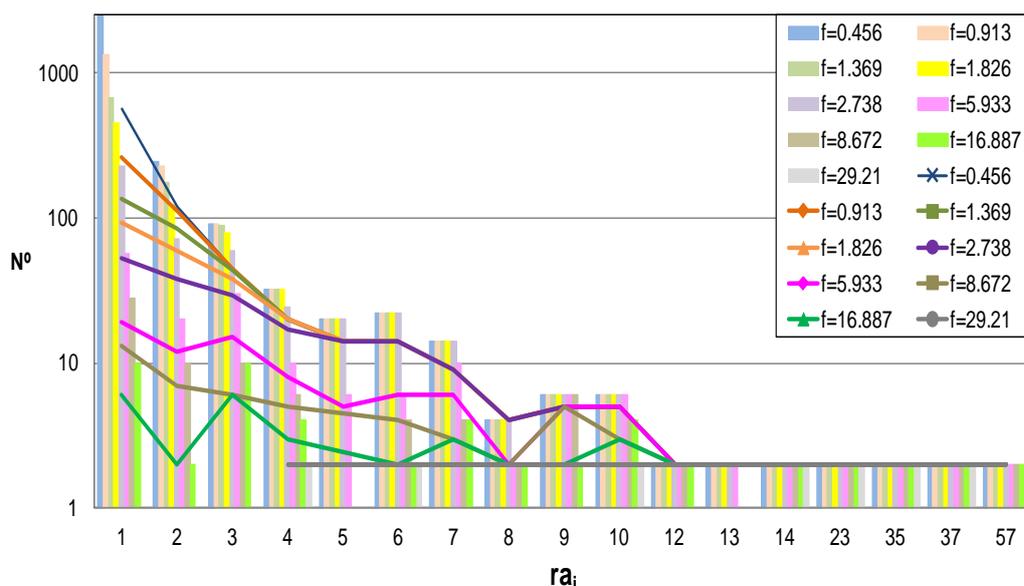


Figura 77.- Distribución del número de ULL (símbolos) y relaciones (barras) respecto al valor de relación en FESTESE001VD. ra_i =valor de relaciones absoluto, N° = número.

La Figura 78 muestra en escala logarítmica el número de ULL (representados con símbolos) y el número de relaciones (representados con barras) del valor de relación acumulado para los nueve sistemas definidos mediante una frecuencia absoluta mínima en el AO. Todos los sistemas contienen el mismo

número de ULL y producen el mismo número de relaciones para valores de relación acumulados igual o superior a 14 (4 y 5 para los AF y AG, respectivamente).

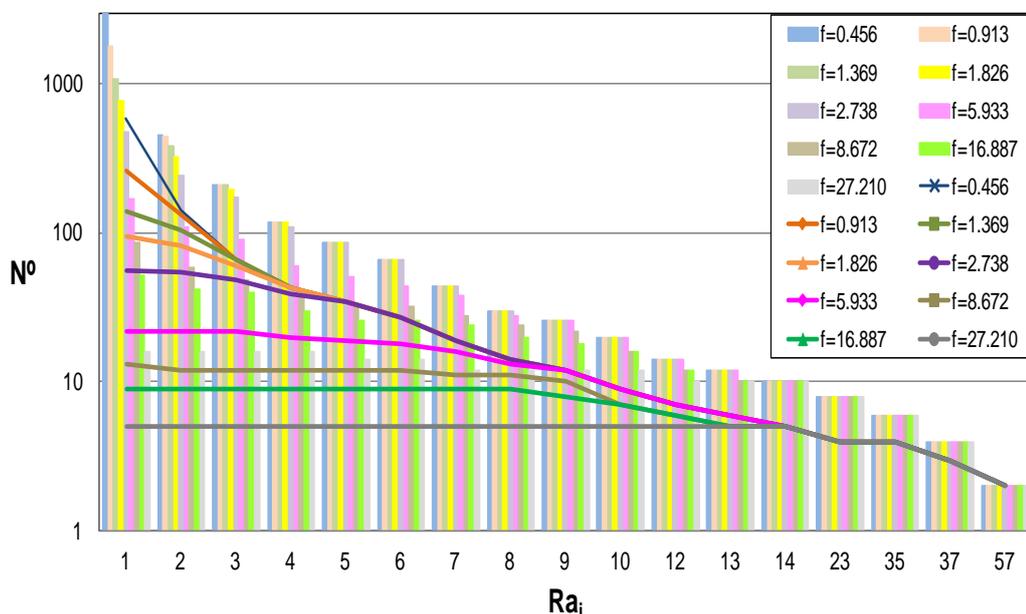


Figura 78.- Distribución del número de ULL (símbolos) y relaciones (barras) respecto al valor de relación acumulado en FESTESE001VD. Ra_i=valor de relaciones acumulado, N°= número.

En el AO, el sistema definido por $f_{\min}=2.738$ (1.702 y 1.712 en el AF y AG, respectivamente) contiene todas las relaciones de valor 5 (3 para los AF y AG) del sistema total. El valor 6 representa el corte que se va a dar en relaciones para estudiar las redes semánticas. Para tener una idea de las relaciones que conserva cada ULL de dicho sistema se calcula el cociente ra_i/n_i*2 para algunos de los sistemas (ver 4^a a 8^a columna en Tablas 82, 83 y 84). Teniendo en cuenta que la desviación máxima del cociente, en porcentaje, se ha calculado en 6.47% y que se considera que las ULL han conservado sus relaciones cuando su cociente sea igual o superior a 80%, se establece el 73.53% como límite del cociente para definir la conservación de las relaciones en el AO (71.48 y 71.43% para los AF y AG). Las ULL que cumplen con este último criterio para los sistemas definidos por $f_{\min} =$

UN ESTUDIO DE CASO: UN TEXTO ESCOLAR DE ASTRONOMÍA

2.738, 1.702 y 1.712 en los AO, AF y AG, se han señalado en negrita en las Tablas 82, 83 y 84. Además se encuentran señaladas en cursiva aquellas ULL que de acuerdo con el criterio del cociente (Chung, 2003) son específicas. Debido a que el corpus CREA (Real Academia Española: Banco de datos (CREA) [en línea], 2008) no permite la consulta de ciertos símbolos matemáticos se han señalado dichas ULL específicas subrayándolas. Ya en la Tabla 82 aparecen ULL que desde el punto de vista del experto no deberían ser consideradas como específicas, como por ejemplo *aproximadamente*. Esto se debe al hecho de que el criterio para hallar ULL específicas es un criterio basado en frecuencias con lo que puede ocurrir que haya ULL que se usen más en un corpus específico que en un corpus general.

Tabla 82

Conservación de relaciones en cada ULL para distintos sistemas definidos en FESTESE001VD

OL	ULL	n_i	$f=0.456$	$f=1.369$	$f=2.738$	$f=8.672$	$f=29.210$
1	EL	265	92.26	71.51	55.85	31.51	24.34
2	DE	191	100.00	67.80	51.31	29.58	24.61
3	ESTRELLA	74	94.59	81.08	75.68	68.24	48.65
4	UNO	72	97.22	61.81	50.69	28.47	20.83
5	EN	64	94.53	59.38	50.00	40.63	36.72
6	SE	54	96.30	62.96	49.07	16.67	
7	Y	46	100.00	71.74	46.74	25.00	
8	QUE	37	100.00	66.22	51.35	35.14	
9	SER	37	97.30	82.43	72.97	44.59	
10	DEL	28	100.00	60.71	42.86	1.79	
11	ESTO	26	82.69	61.54	46.15	34.62	
12	A	23	95.65	60.87	43.48	30.43	
13	SU	19	100.00	73.68	57.89	34.21	
14	<i>CÚMULO</i>	17	<i>91.18</i>	<i>76.47</i>	<i>67.65</i>		
15	<i>CÚMULOGLOBULAR</i>	17	<i>100.00</i>	<i>76.47</i>	<i>67.65</i>		
16	PARA	17	94.12	55.88	35.29		
17	<i>NÚCLEO</i>	16	<i>96.88</i>	<i>81.25</i>	<i>71.88</i>		
18	POR	16	84.38	53.13	46.88		
19	MÁS	15	100.00	60.00	43.33		
20	PODERX	14	100.00	53.57	32.14		
21	MUY	13	100.00	50.00	26.92		
22	SECUENCIAPRINCIPAL	13	92.31	88.46	84.62		
23	COMO	11	90.91	59.09	50.00		
24	LUMINOSIDAD	11	95.45	90.91	90.91		
25	VER	11	100.00	86.36	72.73		
26	<i>APROXIMADAMENTE</i>	10	<i>100.00</i>	<i>80.00</i>	<i>70.00</i>		
27	FORMA1	10	100.00	90.00	85.00		
28	HIDRÓGENO	10	90.00	80.00	75.00		
29	TENER	10	95.00	90.00	75.00		

UN ESTUDIO DE CASO: UN TEXTO ESCOLAR DE ASTRONOMÍA

OL	ULL	n _i	f=0.456	f=1.369	f=2.738	f=8.672	f=29.210
30	10	9	100.00	33.33	16.67		
31	POCO	9	100.00	77.78	72.22		
32	PROCESO	9	100.00	83.33	83.33		
33	TODO	9	83.33	77.78	61.11		
34	CON	8	100.00	81.25	62.50		
35	ENCONTRAR	8	100.00	93.75	87.50		
36	L	8	100.00	75.00	62.50		
37	MASA	8	87.50	87.50	75.00		
38	DIFERENTE	7	92.86	71.43	50.00		
39	EDAD	7	100.00	100.00	100.00		
40	ENTRE	7	100.00	64.29	50.00		
41	FORMAR	7	100.00	78.57	71.43		
42	FUSIÓN	7	100.00	85.71	78.57		
43	HABERX	7	100.00	50.00	35.71		
44	SOL	7	78.57	78.57	64.29		
45	VIDA	7	100.00	100.00	100.00		
46	X	7	100.00	35.71	14.29		
47	CUANDO	6	83.33	41.67	41.67		
48	DIAGRAMA	6	83.33	58.33	58.33		
49	GALAXIA	6	83.33	75.00	58.33		
50	GIGANTEROJO	6	75.00	66.67	66.67		
51	MAYORÍA	6	100.00	100.00	100.00		
52	MIL	6	100.00	83.33	75.00		
53	MILLÓNDEAÑO	6	75.00	66.67	66.67		
54	O	6	100.00	83.33	41.67		
55	T	6	100.00	50.00	50.00		

Nota. OL= orden lexicométrico; f= frecuencia relativa mínima; n= frecuencia absoluta.

Tabla 83

Conservación de relaciones en cada ULL para distintos sistemas definidos en FESTESE001VD-F

OL	ULL	n _i	f=0.851	f=1.702	f=2.553	f=3.404	f=5.106
1	ESTRELLA	59	86.44	65.25	48.31	39.83	27.97
2	SER	35	91.43	70.00	57.14	51.43	41.43
3	CÚMULO	17	79.41	58.82	44.12	29.41	26.47
4	CÚMULOGLOBULAR	17	91.18	64.71	44.12	32.35	20.59
5	NÚCLEO	16	93.75	65.63	37.50	34.38	15.63
6	MÁS	15	100.00	76.67	50.00	33.33	26.67
7	MUY	13	100.00	76.92	50.00	34.62	19.23
8	SECUENCIAPRINCIPAL	13	92.31	80.77	53.85	46.15	30.77
9	LUMINOSIDAD	11	90.91	63.64	59.09	50.00	40.91
10	VER	11	95.45	59.09	50.00	50.00	31.82
11	APROXIMADAMENTE	10	100.00	80.00	70.00	65.00	40.00
12	FORMA1	10	100.00	70.00	50.00	30.00	25.00
13	TENER	10	95.00	85.00	65.00	60.00	30.00
14	10	9	100.00	66.67	33.33	27.78	16.67
15	POCO	9	100.00	77.78	55.56	50.00	33.33
16	ENCONTRAR	8	100.00	93.75	81.25	43.75	31.25
17	GALAXIA	8	81.25	62.50	56.25	31.25	12.50
18	HIDRÓGENO	8	87.50	56.25	37.50	31.25	25.00
19	L	8	100.00	87.50	68.75	56.25	50.00
20	MASA	8	81.25	50.00	50.00	31.25	6.25

UN ESTUDIO DE CASO: UN TEXTO ESCOLAR DE ASTRONOMÍA

OL	ULL	n _i	f=0.851	f=1.702	f=2.553	f=3.404	f=5.106
21	DIFERENTE	7	92.86	50.00	50.00	42.86	21.43
22	SOL	7	78.57	78.57	71.43	64.29	50.00
23	TODO	7	78.57	57.14	50.00	50.00	21.43
24	X	7	100.00	78.57	35.71	21.43	14.29
25	T	6	100.00	91.67	50.00	41.67	41.67
26	=	5	100.00	70.00	50.00	50.00	
27	DIAGRAMAHR	5	90.00	60.00	30.00	30.00	
28	EJERCICIO	5	80.00	60.00	20.00	20.00	
29	HACER	5	90.00	70.00	30.00	20.00	
30	MILMILLÓNANO	5	70.00	50.00	20.00	10.00	
31	NO	5	100.00	40.00	30.00	20.00	
32	PORCIENTO	5	100.00	50.00	30.00	30.00	
33	PROCESOFUSIÓN	5	80.00	40.00	30.00	20.00	
34	SÓLO	5	100.00	50.00	30.00	30.00	
35	TEMPERATURA	5	100.00	90.00	70.00	50.00	
36	USAR	5	100.00	90.00	40.00	20.00	
37	VALOR	5	90.00	50.00	10.00	10.00	
38	VÍALÁCTEO	5	70.00	50.00	40.00	40.00	
39	AÑOLUZ	4	87.50	25.00	25.00	25.00	
40	COMPONER	4	87.50	87.50	75.00	75.00	
41	DAR	4	87.50	25.00	25.00	12.50	
42	DIAGRAMA	4	87.50	50.00	25.00	0.00	
43	DISCO	4	87.50	62.50	50.00	50.00	
44	EDAD	4	100.00	50.00	37.50	37.50	
45	ENERGÍA	4	100.00	37.50	25.00	25.00	
46	ESTABLE	4	75.00	62.50	50.00	50.00	
47	ESTELAR	4	87.50	87.50	50.00	37.50	
48	FASE	4	87.50	50.00	50.00	50.00	
49	FORMAR	4	100.00	87.50	62.50	37.50	
50	GIGANTEROJO	4	75.00	62.50	37.50	25.00	
51	INCLUSO	4	100.00	75.00	62.50	37.50	
52	LUGAR	4	100.00	62.50	62.50	62.50	
53	LÍNEA	4	87.50	75.00	37.50	37.50	
54	MOSTRAR	4	100.00	87.50	50.00	50.00	
55	PEQUEÑO	4	87.50	62.50	25.00	25.00	
56	PROCESO	4	87.50	75.00			
57	R	4	87.50	87.50	37.50	25.00	
58	RADIACIÓN	4	100.00	37.50	12.50	12.50	
59	REACCIÓN	4	87.50	37.50	25.00	12.50	
60	SECHERRASTRONÓMICO	4	75.00	50.00	50.00	37.50	
61	TANTO	4	75.00	50.00	50.00	37.50	
62	TEMPERATURASUPERFICIAL	4	100.00	75.00	62.50	50.00	
63	TIPO	4	100.00	100.00	62.50	62.50	
64	UNOA	4	100.00	62.50	50.00	25.00	
65	VIDAESTRELLA	4	100.00	87.50	50.00	37.50	
66	ALGUNO	3	100.00	83.33	33.33		
67	BAJOAMASA	3	100.00	83.33	83.33		
68	COLOR	3	100.00	66.67	33.33		
69	CONVERTIR	3	100.00	66.67	50.00		
70	DEBER	3	83.33	66.67	66.67		
71	DETERMINAREDA	3	100.00	83.33	83.33		
72	DISTANCIA	3	100.00	50.00	50.00		
73	ESFÉRICO	3	100.00	83.33	33.33		
74	ESTAR	3	100.00	66.67	66.67		

UN ESTUDIO DE CASO: UN TEXTO ESCOLAR DE ASTRONOMÍA

OL	ULL	n _i	f=0.851	f=1.702	f=2.553	f=3.404	f=5.106
75	<i>FUSIONARFORMAR</i>	3	100.00	66.67	33.33		
76	<i>GASPOLVO</i>	3	83.33	66.67	33.33		
77	GENERAL	3	100.00	33.33	16.67		
78	GRANDE	3	100.00	100.00	83.33		
79	<u>HASTAM</u>	3	100.00	83.33	16.67		
80	IMAGEN	3	100.00	83.33	33.33		
81	IMPORTANTE	3	100.00	50.00	33.33		
82	<i>INFERIOR</i>	3	100.00	66.67	50.00		
83	LLAMADOA	3	83.33	83.33	33.33		
84	<i>LSOLAR</i>	3	66.67	50.00	50.00		
85	LUZ	3	66.67	50.00	50.00		
86	<i>M12</i>	3	66.67	66.67	50.00		
87	<i>MAGNITUD</i>	3	100.00	66.67	66.67		
88	MAYORÍA	3	83.33	50.00	50.00		
89	<i>MAYORÍAESTRELLA</i>	3	83.33	33.33	33.33		
90	MENOS	3	100.00	66.67	16.67		
91	<i>MESSIER</i>	3	83.33	33.33	33.33		
92	OBJETO	3	83.33	50.00	50.00		
93	OTRO	3	83.33	66.67	33.33		
94	PARTIR	3	100.00	66.67	33.33		
95	PESADOA	3	100.00	100.00	83.33		
96	PROPIO	3	100.00	83.33	50.00		
97	PUNTO	3	83.33	50.00	33.33		
98	RADIO	3	83.33	50.00	50.00		
99	SISTEMA	3	100.00	66.67	50.00		
100	TAL	3	83.33	66.67	50.00		
101	TAMBIÉN	3	83.33	83.33	83.33		
102	TOTAL	3	100.00	100.00	66.67		
103	VEZ	3	100.00	83.33	66.67		
104	VIDA	3	83.33	66.67	50.00		
105	0COMA7PCIENTORESIDLNASA	2	75.00	75.00			
106	1	2	100.00	25.00			
107	12	2	75.00	75.00			
108	2	2	75.00	50.00			
109	3	2	100.00	75.00			
110	3X10	2	100.00	75.00			
111	4X	2	100.00	75.00			
112	<u>A2X</u>	2	100.00	100.00			
113	<u>A3</u>	2	100.00	100.00			
114	<u>A4</u>	2	100.00	100.00			
115	<u>A6</u>	2	100.00	50.00			
116	AHORA	2	100.00	50.00			
117	APARECER	2	75.00	50.00			
118	APROXIMADOA	2	100.00	50.00			
119	B	2	100.00	100.00			
120	BAJOA	2	100.00	75.00			
121	BLANCO	2	100.00	75.00			
122	BRILLANTE	2	100.00	100.00			
123	CA2	2	75.00	75.00			
124	CADA	2	75.00	75.00			
125	CALCULAR	2	75.00	75.00			
126	CALIENTE	2	100.00	75.00			
127	CANTIDAD	2	100.00	75.00			
128	CANTIDADENERGÍA	2	50.00	25.00			

UN ESTUDIO DE CASO: UN TEXTO ESCOLAR DE ASTRONOMÍA

OL	ULL	n _i	f=0.851	f=1.702	f=2.553	f=3.404	f=5.106
129	<i>CAPAEXTERNOESTRELLA</i>	2	75.00	25.00			
130	CIENTO	2	100.00	75.00			
131	COMBUSTIÓNHIDRÓGENO	2	100.00	75.00			
132	COMENZAR	2	100.00	50.00			
133	COMPACTO	2	100.00	100.00			
134	COMÚN	2	100.00	75.00			
135	CONJUNTO	2	75.00	75.00			
136	CONTENER	2	100.00	50.00			
137	<i>CONTRAER</i>	2	50.00	50.00			
138	CUATRONÚCLEOHIDRÓGENO	2	100.00	75.00			
139	<i>CÚMULOABESTRELLASER</i>	2	75.00	50.00			
140	CÚMULOABIERTOAEESTRELLA	2	100.00	100.00			
141	DETERMINAR	2	100.00	75.00			
142	DISTRIBUCIÓN	2	75.00	50.00			
143	<i>DIÁMETROTÍPICO</i>	2	75.00	50.00			
144	DOS	2	100.00	75.00			
145	<i>ECUACIÓN</i>	2	100.00	50.00			
146	EJEMPLO	2	100.00	100.00			
147	<i>EJERCIDOA</i>	2	100.00	50.00			
148	ELEMENTO	2	100.00	50.00			
149	<i>ELEMENTOMÁSPESADOA</i>	2	50.00	25.00			
150	EQUILIBRIOPRESIÓN	2	100.00	75.00			
151	ESTRUCTURA	2	100.00	75.00			
152	EVOLUCIONAR	2	100.00	75.00			
153	EVOLUCIÓN	2	100.00	75.00			
154	EXTENDERAPROXIMADAMENTE	2	100.00	75.00			
155	FAMOSO	2	100.00	100.00			
156	FIG	2	100.00	50.00			
157	<i>FIG6</i>	2	50.00	50.00			
158	FORMACIÓN	2	100.00	75.00			
159	FUERA	2	100.00	75.00			
160	<i>FUSIÓN</i>	2	75.00	50.00			
161	GAS	2	100.00	50.00			
162	<i>GRAVITATORIO</i>	2	100.00	50.00			
163	GRUPO	2	100.00	100.00			
164	<i>HALO</i>	2	75.00	50.00			
165	HELIO	2	75.00	75.00			
166	<i>HELIOSER</i>	2	100.00	50.00			
167	HISTORIA	2	100.00	50.00			
168	LARGO	2	100.00	75.00			
169	LENTAMENTE	2	75.00	50.00			
170	LLAMAR	2	75.00	75.00			
171	<i>LUMÍNICO</i>	2	50.00	0.00			
172	M	2	100.00	50.00			
173	MAGNITUDVISIBLE	2	75.00	75.00			
174	MANERA	2	100.00	100.00			
175	MANTENER	2	100.00	50.00			
176	MAYOR	2	100.00	50.00			
177	<i>MBMV</i>	2	100.00	25.00			
178	MEDIDA1	2	100.00	75.00			
179	MEDIDOA	2	100.00	75.00			
180	<i>MEDIR</i>	2	100.00	50.00			
181	<i>MESSIER12</i>	2	75.00	50.00			
182	MIENTRAS	2	75.00	50.00			

UN ESTUDIO DE CASO: UN TEXTO ESCOLAR DE ASTRONOMÍA

OL	ULL	n _i	f=0.851	f=1.702	f=2.553	f=3.404	f=5.106
183	MILLÓN	2	100.00	75.00			
184	MUCHO	2	50.00	50.00			
185	MUCHOTIEMPO	2	75.00	50.00			
186	NEBULOSAPLANETARIO	2	100.00	75.00			
187	NGC	2	75.00	50.00			
188	NORMAL	2	100.00	100.00			
189	NUEVO	2	75.00	25.00			
190	NUEVOESTRELLA	2	75.00	50.00			
191	NÚMERO	2	100.00	75.00			
192	OBSERVACIÓN	2	75.00	75.00			
193	ORDENMAGNITUD	2	100.00	75.00			
194	PARECER	2	100.00	75.00			
195	PARTE	2	100.00	100.00			
196	PASO1	2	100.00	75.00			
197	PESAR	2	75.00	25.00			
198	PLÉYADES	2	75.00	75.00			
199	PRIMERO	2	75.00	50.00			
200	PROPIEDAD	2	100.00	75.00			
201	PRUEBA1	2	100.00	75.00			
202	QUEMAR	2	75.00	50.00			
203	RAMAGIGANTEROJO	2	75.00	50.00			
204	RARO	2	100.00	75.00			
205	RECTO	2	100.00	100.00			
206	RELATIVAMENTE	2	100.00	50.00			
207	SIGMA	2	100.00	100.00			
208	SIGUIENTE	2	75.00	75.00			
209	SOLAR	2	75.00	75.00			
210	STEFANBOLTZMANN	2	75.00	50.00			
211	SUCEDER	2	100.00	100.00			
212	SUPERIOR	2	100.00	50.00			
213	TEMPERATURACENTRAL	2	100.00	50.00			
214	TEMPSUPERFICIALTESTRE	2	50.00	0.00			
215	TEMPSUPERFICIALTTODO	2	100.00	100.00			
216	TIEMPO	2	75.00	75.00			
217	V	2	50.00	50.00			
218	VALIOSO	2	100.00	75.00			
219	VARIO	2	100.00	75.00			
220	VIEJO	2	100.00	75.00			
221	VÉRTICE	2	100.00	75.00			
222	YA	2	75.00	50.00			
223	YANOSER	2	100.00	75.00			
224	≈	2	100.00	100.00			
225	ÁREA	2	100.00	25.00			

Nota. OL= orden lexicométrico; f= frecuencia relativa mínima; n= frecuencia absoluta.

UN ESTUDIO DE CASO: UN TEXTO ESCOLAR DE ASTRONOMÍA

Tabla 84

Conservación de relaciones en cada ULL para distintos sistemas definidos en FESTESE001VD-S

OL	ULL	n _i	f=0.856	f=1.712	f=2.553	f=3.404	f=5.106
1	ESTRELLA	63	85.71	64.29	50.00	41.27	30.16
2	SER	37	93.24	71.62	59.46	52.70	44.59
3	CÚMULOGLOBULAR	23	89.13	63.04	50.00	36.96	26.09
4	LUMINOSIDAD	16	93.75	68.75	56.25	50.00	40.63
5	NÚCLEO	16	93.75	65.63	37.50	34.38	18.75
6	TEMPERATURASUPERFICIAL	15	96.67	83.33	53.33	46.67	46.67
7	MÁS	14	100.00	78.57	46.43	32.14	25.00
8	MUY	13	100.00	76.92	50.00	34.62	19.23
9	SECUENCIAPRINCIPAL	13	92.31	80.77	53.85	46.15	34.62
10	CÚMULO	11	81.82	54.55	45.45	27.27	22.73
11	VER	11	95.45	59.09	59.09	54.55	36.36
12	APROXIMADAMENTE	10	100.00	80.00	75.00	70.00	50.00
13	TENER	10	100.00	90.00	70.00	65.00	35.00
14	10	9	100.00	66.67	33.33	27.78	16.67
15	FORMA1	9	100.00	72.22	50.00	27.78	27.78
16	MASA	9	83.33	55.56	55.56	38.89	16.67
17	POCO	9	100.00	77.78	55.56	50.00	33.33
18	TODO	9	83.33	66.67	61.11	61.11	33.33
19	DIAGRAMAHR	8	87.50	62.50	37.50	25.00	6.25
20	ENCONTRAR	8	100.00	93.75	62.50	37.50	31.25
21	HIDRÓGENO	8	87.50	56.25	37.50	31.25	25.00
22	DIFERENTE	7	92.86	50.00	50.00	42.86	21.43
23	SOL	7	78.57	78.57	71.43	64.29	50.00
24	X	7	100.00	78.57	42.86	28.57	28.57
25	1	6	100.00	50.00	41.67	25.00	16.67
26	CÚMULOABIERTOAEESTRELLA	6	75.00	75.00	58.33	58.33	58.33
27	GALAXIA	6	83.33	66.67	50.00	25.00	16.67
28	RADIO	6	83.33	66.67	33.33	33.33	25.00
29	=	5	100.00	80.00	70.00	70.00	
30	EJERCICIO	5	80.00	60.00	20.00	20.00	
31	ENERGÍA	5	100.00	40.00	30.00	30.00	
32	FIGURA	5	80.00	50.00	30.00	30.00	
33	HACER	5	90.00	70.00	30.00	20.00	
34	MILMILLÓNANO	5	70.00	50.00	20.00	10.00	
35	NO	5	100.00	40.00	30.00	20.00	
36	PORCIENTO	5	100.00	50.00	40.00	40.00	
37	PROCESOFUSIÓN	5	80.00	40.00	30.00	20.00	
38	SÓLO	5	100.00	50.00	30.00	30.00	
39	USAR	5	100.00	90.00	40.00	30.00	
40	VALOR	5	90.00	50.00	10.00	10.00	
41	2	4	87.50	62.50	62.50	62.50	
42	AÑOLUZ	4	87.50	25.00	25.00	25.00	
43	COMPONER	4	87.50	87.50	75.00	75.00	
44	DAR	4	87.50	25.00	25.00	12.50	

UN ESTUDIO DE CASO: UN TEXTO ESCOLAR DE ASTRONOMÍA

OL	ULL	n _i	f=0.856	f=1.712	f=2.553	f=3.404	f=5.106
45	EDAD	4	100.00	50.00	37.50	37.50	
46	ESTABLE	4	75.00	62.50	50.00	50.00	
47	ESTELAR	4	87.50	87.50	50.00	37.50	
48	FASE	4	87.50	50.00	50.00	50.00	
49	FORMAR	4	100.00	87.50	62.50	37.50	
50	GIGANTEROJO	4	75.00	62.50	50.00	37.50	
51	INCLUSO	4	100.00	75.00	62.50	37.50	
52	LUGAR	4	100.00	62.50	62.50	62.50	
53	LÍNEA	4	87.50	75.00	37.50	37.50	
54	MESSIER12	4	62.50	50.00	50.00	25.00	
55	MOSTRAR	4	100.00	87.50	50.00	50.00	
56	PEQUEÑO	4	87.50	62.50	25.00	25.00	
57	PROCESO	4	87.50	75.00	12.50	12.50	
58	RADIACIÓN	4	100.00	37.50	12.50	12.50	
59	REACCIÓN	4	87.50	37.50	25.00	12.50	
60	SECHERRASTRONÓMICO	4	75.00	50.00	37.50	37.50	
61	TANTO	4	75.00	50.00	50.00	37.50	
62	TEMPERATURA	4	100.00	87.50	87.50	75.00	
63	TIPO	4	100.00	100.00	75.00	75.00	
64	VIDAESTRELLA	4	100.00	87.50	50.00	37.50	
65	VÍALÁCTEO	4	50.00	37.50	25.00	25.00	
66	ALGUNO	3	100.00	83.33	33.33		
67	BAJOAMASA	3	100.00	83.33	83.33		
68	CONVERTIR	3	100.00	66.67	50.00		
69	DEBER	3	83.33	66.67	66.67		
70	DETERMINAREDA	3	100.00	83.33	83.33		
71	DISTANCIA	3	100.00	50.00	50.00		
72	ESFÉRICO	3	100.00	83.33	33.33		
73	ESTAR	3	100.00	83.33	83.33		
74	FUSIONARFORMAR	3	100.00	66.67	33.33		
75	GASPOLVO	3	83.33	66.67	33.33		
76	GRANDE	3	100.00	100.00	83.33		
77	HASTAM	3	100.00	83.33	16.67		
78	IMAGEN	3	100.00	83.33	33.33		
79	IMPORTANTE	3	100.00	50.00	33.33		
80	INFERIOR	3	100.00	66.67	50.00		
81	LLAMADOA	3	83.33	83.33	33.33		
82	LUMINOSIDAD SOLAR	3	66.67	50.00	50.00		
83	LUZ	3	66.67	50.00	50.00		
84	MAYORÍA	3	83.33	50.00	50.00		
85	MAYORÍAESTRELLA	3	83.33	33.33	33.33		
86	MENOS	3	100.00	66.67	16.67		
87	MESSIER	3	83.33	33.33	33.33		
88	OBJETO	3	83.33	50.00	50.00		
89	OTRO	3	83.33	66.67	33.33		
90	PARTIR	3	100.00	66.67	33.33		
91	PESADOA	3	100.00	100.00	83.33		
92	PLÉYADES	3	66.67	66.67	50.00		
93	PROPIO	3	100.00	83.33	50.00		

UN ESTUDIO DE CASO: UN TEXTO ESCOLAR DE ASTRONOMÍA

OL	ULL	n _i	f=0.856	f=1.712	f=2.553	f=3.404	f=5.106
94	PUNTO	3	83.33	50.00	33.33		
95	SISTEMA	3	100.00	66.67	33.33		
96	TAL	3	83.33	66.67	50.00		
97	TAMBIÉN	3	83.33	83.33	83.33		
98	TOTAL	3	100.00	100.00	66.67		
99	VEZ	3	100.00	83.33	66.67		
100	VIDA	3	83.33	66.67	50.00		
101	0COMA7PCIENTORESIDLNASA	2	75.00	75.00			
102	12	2	75.00	75.00			
103	3	2	100.00	75.00			
104	3X10	2	100.00	75.00			
105	4NÚCLEOHIDRÓGENO	2	100.00	75.00			
106	4X	2	100.00	75.00			
107	6	2	75.00	75.00			
108	A2X	2	100.00	100.00			
109	A3	2	100.00	100.00			
110	A4	2	100.00	100.00			
111	A6	2	100.00	50.00			
112	AHORA	2	100.00	50.00			
113	APARECER	2	75.00	50.00			
114	APROXIMADOA	2	100.00	50.00			
115	B	2	100.00	75.00			
116	BAJOA	2	100.00	75.00			
117	BLANCO	2	100.00	75.00			
118	BRILLANTE	2	100.00	100.00			
119	CA2	2	75.00	75.00			
120	CADA	2	75.00	75.00			
121	CALCULAR	2	75.00	75.00			
122	CALIENTE	2	100.00	75.00			
123	CANTIDAD	2	100.00	75.00			
124	CANTIDADENERGÍA	2	50.00	25.00			
125	CAPAEXTERNOESTRELLA	2	75.00	25.00			
126	CIENTO	2	100.00	75.00			
127	COLOR	2	100.00	75.00			
128	COMBUSTIÓNHIDRÓGENO	2	100.00	75.00			
129	COMENZAR	2	100.00	50.00			
130	COMPACTO	2	100.00	100.00			
131	COMÚN	2	100.00	75.00			
132	CONJUNTO	2	75.00	75.00			
133	CONTENER	2	100.00	50.00			
134	CONTRAER	2	50.00	50.00			
135	DETERMINAR	2	100.00	75.00			
136	DISCO	2	75.00	25.00			
137	DISCOVÍALÁCTEO	2	100.00	50.00			
138	DISTRIBUCIÓN	2	75.00	50.00			
139	DIÁMETROTÍPICO	2	75.00	50.00			
140	ECUACIÓN	2	100.00	50.00			
141	EJEMPLO	2	100.00	100.00			
142	EJERCIDOA	2	100.00	50.00			

UN ESTUDIO DE CASO: UN TEXTO ESCOLAR DE ASTRONOMÍA

OL	ULL	n _i	f=0.856	f=1.712	f=2.553	f=3.404	f=5.106
143	ELEMENTO	2	100.00	50.00			
144	ELEMENTOMÁSPESADOA	2	50.00	25.00			
145	EQUILIBRIOPRESIÓN	2	100.00	75.00			
146	ESTRUCTURA	2	100.00	75.00			
147	EVOLUCIONAR	2	100.00	75.00			
148	EVOLUCIÓN	2	100.00	75.00			
149	EXTENDERAPROXIMADAMENTE	2	100.00	75.00			
150	FAMOSO	2	100.00	100.00			
151	FORMACIÓN	2	100.00	75.00			
152	FUERA	2	100.00	75.00			
153	FUSIÓN	2	75.00	50.00			
154	GAS	2	100.00	50.00			
155	GENERAL	2	100.00	50.00			
156	GRAVITATORIO	2	100.00	50.00			
157	GRUPO	2	100.00	100.00			
158	HALO	2	75.00	50.00			
159	HELIO	2	75.00	75.00			
160	HELIOSE	2	100.00	50.00			
161	HISTORIA	2	100.00	50.00			
162	LARGO	2	100.00	75.00			
163	LENTAMENTE	2	75.00	50.00			
164	LLAMAR	2	75.00	75.00			
165	LUMÍNICO	2	50.00	0.00			
166	MAGAPARENTEÍNDICECOLORBV	2	100.00	50.00			
167	MAGNITUD	2	100.00	75.00			
168	MAGNITUDVISIBLE	2	75.00	75.00			
169	MANERA	2	100.00	100.00			
170	MANTENER	2	100.00	50.00			
171	MAYOR	2	100.00	50.00			
172	MEDIDA1	2	100.00	75.00			
173	MEDIDOA	2	100.00	75.00			
174	MEDIR	2	100.00	50.00			
175	MIENTRAS	2	75.00	50.00			
176	MILLÓN	2	100.00	75.00			
177	MUCHO	2	50.00	50.00			
178	MUCHOTIEMPO	2	75.00	50.00			
179	NEBULOSAPLANETARIO	2	100.00	75.00			
180	NGC	2	75.00	25.00			
181	NORMAL	2	100.00	100.00			
182	NUEVOESTRELLA	2	75.00	50.00			
183	NÚMERO	2	100.00	75.00			
184	OBSERVACIÓN	2	75.00	75.00			
185	ORDENMAGNITUD	2	100.00	75.00			
186	PARECER	2	100.00	75.00			
187	PARTE	2	100.00	100.00			
188	PASO1	2	100.00	75.00			
189	PESAR	2	75.00	25.00			
190	PRIMERO	2	75.00	50.00			
191	PROPIEDAD	2	100.00	75.00			

UN ESTUDIO DE CASO: UN TEXTO ESCOLAR DE ASTRONOMÍA

OL	ULL	n_i	$f=0.856$	$f=1.712$	$f=2.553$	$f=3.404$	$f=5.106$
192	PRUEBA1	2	100.00	75.00			
193	QUEMAR	2	100.00	75.00			
194	<i>RAMAGIGANTEROJO</i>	2	75.00	50.00			
195	RARO	2	100.00	75.00			
196	RECTO	2	100.00	100.00			
197	<i>RELATIVAMENTE</i>	2	100.00	50.00			
198	SIGMA	2	100.00	100.00			
199	SIGUIENTE	2	75.00	75.00			
200	SOLAR	2	75.00	75.00			
201	<i>STEFANBOLTZMANN</i>	2	75.00	50.00			
202	SUCEDER	2	100.00	100.00			
203	SUPERIOR	2	100.00	50.00			
204	<i>TEMPERATURACENTRAL</i>	2	100.00	50.00			
205	TIEMPO	2	75.00	75.00			
206	V	2	50.00	50.00			
207	VALIOSO	2	100.00	75.00			
208	VARIO	2	100.00	75.00			
209	VIEJO	2	100.00	75.00			
210	VÉRTICE	2	100.00	75.00			
211	YA	2	75.00	50.00			
212	YANOSER	2	100.00	75.00			
213	≈	2	100.00	100.00			
214	ÁREA	2	100.00	25.00			

Nota. OL= orden lexicométrico; f = frecuencia relativa mínima; n = frecuencia absoluta.

Las ULL que aparecen en las Tablas 82, 83 y 84 corresponden a las que constituyen los sistemas definidos por $f_{\min}=2.738$, 1.702 y 1.712 en los AO, AF y AG, respectivamente. Las ULL de menor frecuencia que forman parte de sistemas con menor f_{\min} no aparecen tabuladas. Nótese que para algunas ULL, en sistemas con f_{\min} mayor, no aparecen valores del cociente, puesto que no existen esas ULL en dichos sistemas.

La selección de un sistema en función de la frecuencia no es suficiente a la hora de estudiar los conceptos, puesto que en éstos son también importantes las relaciones. En las Tablas 82, 83 y 84 se aprecia que los sistemas donde las ULL conservan mayor número de relaciones corresponden a los definidos por $f_{\min}=2.738$, 1.702 y 1.712 en los AO, AF y AG. Para comparar lo que sucede en los tres

análisis, se presentan en la Tabla 85 los porcentajes de ULL que cumplen distintas condiciones respecto a la especificidad y la conservación de relaciones. En el AO todas estas ULL son léxicas. Los valores del AO son menores que los del AF y AG debido a la gran cantidad de ULL funcionales. Los valores de los AF y AG son similares y cercanos al 50% tanto en número como en frecuencias de ULL específicas o que conservan relaciones. Sin embargo estos valores se reducen a menos de la mitad cuando se exige la condición de ser específico y conservar relaciones. Estas últimas ULL son a las que más atención hay que prestar a la hora de interpretar las redes semánticas.

Tabla 85

Comparación de la conservación de relaciones

Análisis	N°	%					
		N° _r	%(∑n _i) _r	N° _{esp}	%(∑n _i) _{esp}	N° _{r,esp}	%(∑n _i) _{r,esp}
AO	55	25.45	13.96	25.45	15.47	9.09	8.63
AF	225	47.11	39.14	44.89	48.55	20.44	16.61
AG	214	50.47	45.17	44.39	49.24	20.44	17.89

Nota. N°=número de ULL; N°_r= número de ULL que conservan más del 80% de sus relaciones; n_r= frecuencia absoluta de las ULL que conservan más del 80% de sus relaciones; N°_{esp}= número de ULL que son específicas; n_{esp}= frecuencia absoluta de las ULL que son específicas; N°_{r,esp}= número de ULL que conservan más del 80% de sus relaciones y que además son específicas; n_{r,esp}= frecuencia absoluta de las ULL que conservan más del 80% de sus relaciones y que además son específicas.

La importancia relativa de las categorías gramaticales para los sistemas definidos por $f_{\min} = 2.738, 1.702$ y 1.712 en los AO, AF y AG se pueden ver en la Tabla 86. En esta tabla se ha añadido como categoría adicional la de colocaciones formadas por uniones de palabras de distintas categorías, ya que el resto de colocaciones se han clasificado en la categoría correspondiente. Los datos del AO son relativos a la categoría léxica para poder compararlos con los otros dos análisis. De nuevo los resultados para los AF y AG son similares y distintos del AO debido a

UN ESTUDIO DE CASO: UN TEXTO ESCOLAR DE ASTRONOMÍA

la eliminación de palabras funcionales y verbos auxiliares. La única diferencia notable entre los AF y AG son las frecuencias de los verbos, mucho mayores en el AG debido a que en éste *ser* conserva un 80% de las relaciones. Se observa que la categoría más representada en número, tanto en el grupo de ULL que conserva relaciones como en las ULL específicas, es la de sustantivos. Dado que los sustantivos son los que más forman parte de las colocaciones, estos resultados muestran la importancia que tendrá dicha categoría a la hora de interpretar las redes semánticas. Cuando se exige a la vez la condición de que la ULL conserve relaciones y sea específica la categoría de lenguaje matemático llega a alcanzar la misma importancia que la de sustantivos.

En los tres análisis el sistema escogido para analizar con redes semánticas corresponde al que contiene la franja alta y la franja media. En técnicas de construcción de tesauros se considera que las palabras representativas de un texto son aquéllas que se encuentran en la franja media de la distribución de frecuencias (Velasco, Díaz, Lloréns, de Amescua, & Martínez, 1999).

Hay 33 ULL en común entre los AO y AF, y 210 entre los AF y AG. Los sistemas óptimos del AF y AG son parecidos en cuanto al número y frecuencias de las ULL que los forman, tanto en términos generales y de categorías gramaticales como en términos particulares de características de las ULL. La diferencia entre los AF y AG y el AO sigue siendo la ausencia de palabras funcionales y verbos auxiliares. Esta ausencia resalta la importancia de las ULL léxicas, que etiquetan y conforman los conceptos.

Tabla 86

Comparación de categorías gramaticales en la conservación de relaciones

Categoría gramatical	%N°	% N° _r	%(∑n _i) _r	% N° _{esp}	%(∑n _i) _{esp}	% N° _{r,esp}	%(∑n _i) _{r,esp}
AO							
verbos	14.29	14.29	9.68	0.00	0.00	0.00	0.00
adverbios	8.57	0.00	0.00	7.14	4.85	0.00	0.00
adjetivos	14.29	7.14	3.23	7.14	2.91	0.00	0.00
lm	11.43	0.00	0.00	7.14	3.40	0.00	0.00
sustantivos	45.71	71.43	80.11	64.29	74.27	80.00	88.70
colocaciones	5.71	7.14	6.99	14.29	14.56	20.00	11.30
AF							
verbos	12.89	11.32	13.95	8.91	6.22	10.87	11.89
adverbios	6.22	5.66	13.95	1.98	2.87	2.17	6.99
adjetivos	22.22	27.36	21.96	14.85	9.33	17.39	13.99
lm	10.67	14.15	14.24	12.87	9.81	21.74	22.38
sustantivos	35.11	30.19	24.33	33.66	49.04	21.74	24.48
colocaciones	12.89	11.32	11.57	27.72	22.73	26.09	20.28
AG							
verbos	13.55	13.89	22.94	9.47	6.15	13.04	12.34
adverbios	6.54	5.56	11.86	2.11	2.84	2.17	6.49
adjetivos	21.96	25.93	18.56	15.79	9.22	17.39	12.99
lm	9.35	12.96	9.54	11.58	7.33	21.74	18.83
sustantivos	35.98	31.48	23.71	33.68	47.75	21.74	15.58
colocaciones	12.62	10.19	13.40	27.37	26.71	23.91	33.77

Nota. lm= lenguaje matemático; N°= número de ULL; N°_r= número de ULL que conservan más del 80% de sus relaciones; n_r= frecuencia absoluta de las ULL que conservan más del 80% de sus relaciones; N°_{esp}= número de ULL que son específicas; n_{esp}= frecuencia absoluta de las ULL que son específicas; N°_{r,esp}= número de ULL que conservan más del 80% de sus relaciones y que además son específicas; n_{r,esp}= frecuencia absoluta de las ULL que conservan más del 80% de sus relaciones y que además son específicas.

Cuando se comparan las Tablas 82, 83 y 84 se observan diferencias en la especificidad y conservación de relaciones de algunas ULL. Del AO al AF las ULL *L* y *masa* pasan a ser específicas (un 6% de las 33 ULL en común). Este cambio se debe a la variación de la frecuencia relativa de las ULL entre los dos tipos de análisis. La variación en frecuencia relativa está causada por la supresión de palabras funcionales y verbos auxiliares en el texto transformado (ver Tablas 96 y 99 en los Anexos I y P, respectivamente). Además las ULL *estrella*, *luminosidad*,

UN ESTUDIO DE CASO: UN TEXTO ESCOLAR DE ASTRONOMÍA

formal, hidrógeno, masa, edad, fusión, vida, mayoría pasan a no conservar un 80% de relaciones. También hay ULL que conservan un 80% de relaciones en el AF pero no en el AO: *más, muy, aproximadamente, poco, formar, sol, x*. En total un 48% de las ULL comunes en los sistemas elegidos para el AO y el AF han cambiado la conservación cuantitativa de sus relaciones. Esto se debe a la eliminación de las palabras funcionales y verbos auxiliares.

La unificación de grafías provoca diferencias en la frecuencia relativa de las ULL entre los AF y AG: *disco* es específica sólo en el AF y *fase* lo es sólo en el AG. Estas ULL representan un 0.95% de las 210 ULL que hay en común entre los sistemas elegidos del AF y el AG. De igual modo se producen cambios en la conservación de relaciones de ciertas ULL. *Pléyades* conserva un 80% de relaciones en el AF pero no en el AG. Las ULL que conservan un 80% de relaciones en el AG pero no en el AF son =, *estar, color* y *quemar*.

4. 4. 4.- Redes semánticas

Los sistemas definidos por $f_{\min}=2.738$, 1.702 y 1.712 para los AO, AF y AG, respectivamente, se analizan con un entorno 3 y las características de la Tabla 21. Con la ayuda de la gráfica elaborada en el PAFE y las tablas generadas con las relaciones de las ULL se realizan las gráficas en Excel 2007 correspondientes: *Figuras 79, 80, 81 y 82* para el AO; *Figuras 83, 84, 85 y 86* para el AF; *Figuras 87, 88, 89 y 90* para el AG. En todas las gráficas las ULL se centran en torno a la ULL más frecuente: *el* en el AO, *estrella* en el AF y AG. El rango de frecuencias se

divide en cuatro intervalos de igual tamaño, que se ordenan de mayor a menor frecuencia mediante los colores verde, rojo, azul y amarillo. De igual manera el rango de relaciones entre ULL y el de ULL respecto a una ULL origen se dividen en otros cuatro intervalos de igual tamaño, codificados de la misma manera que las frecuencias. Los rangos de frecuencias y relaciones representados se pueden ver en la Tabla 87, siendo éstos parecidos en los AF y AG. La diferencia en los rangos respecto al AF se debe a la supresión de palabras funcionales y verbos auxiliares. En el listado de relaciones relativas a la ULL origen aparecieron relaciones de una ULL consigo misma que han sido obviadas en la representación gráfica y que no son tenidas en cuenta a la hora de calcular la relación total de una ULL origen. Las ULL específicas se representan en cursiva y las que conservan un 80% de relaciones o más se encuentran rodeadas de un marco negro.

Tabla 87

Rangos de frecuencias y relaciones en las gráficas

Análisis	f_{\min}	f_r	rr_{pareja}	rr_{origen}
AO	2.738	[0, 120.949]	[0, 49.832]	[0, 588.235]
AF	1.702	[0, 50.213]	[0, 25.510]	[0, 750.000]
AG	1.712	[0, 53.938]	[0, 22.727]	[0, 750.000]

Nota. f_{\min} = frecuencia relativa mínima; f_r = frecuencia relativa; rr_{pareja} = relación relativa entre parejas de ULL; rr_{origen} = relación relativa a la ULL origen.

Las características de las gráficas de los tres análisis se pueden resumir en las Tablas 88 y 89. Los datos entre el AF y el AG son más parecidos que entre éstos y el AO. Los datos referidos a la especificidad y conservación de relaciones en el AO se refieren sólo a las ULL léxicas. Es decir, los AF y AG muestran más del doble de ULL específicas y que conservan un 80% de relaciones o más, que el AO. Sólo hay 23 ULL en común entre las gráficas del AO y el AF, mientras que hay 52 ULL en común entre las del AF y el AG. Los datos de la conservación de relaciones en el AG son ligeramente superiores a los del AF.

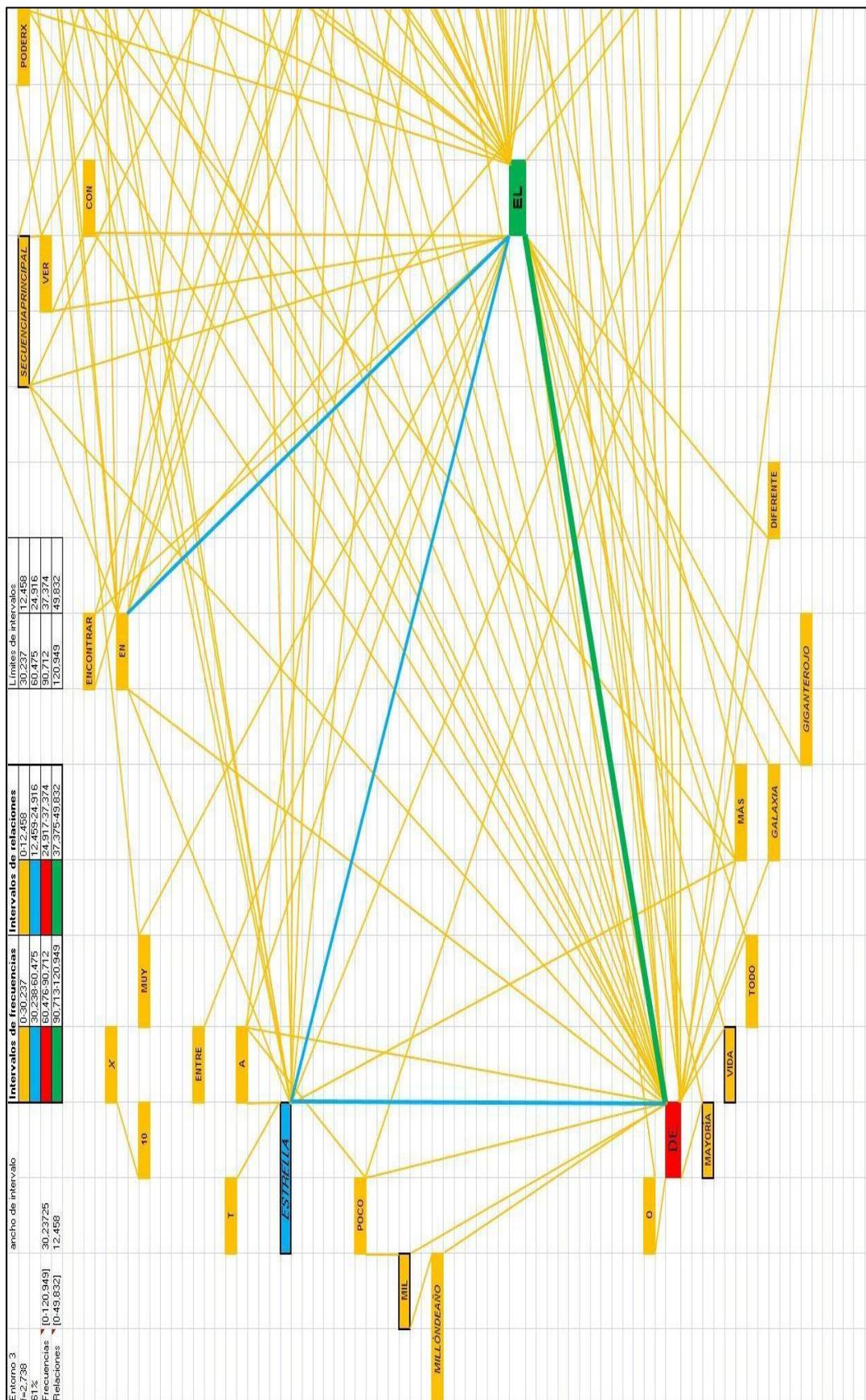


Figura 79.- Parte izquierda de la gráfica con relaciones relativas al listado completo de FESTESE001VD.

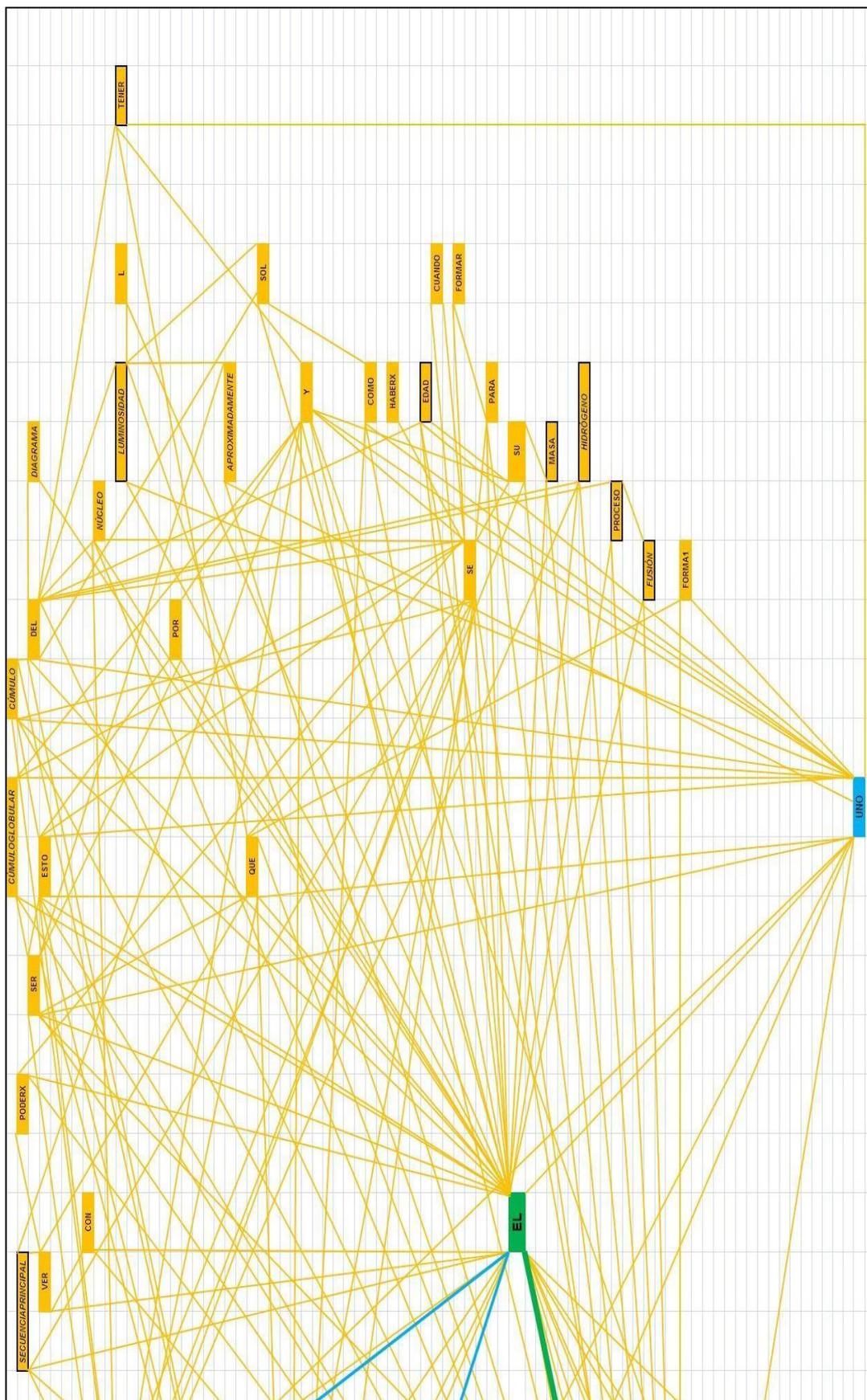


Figura 80.- Parte derecha de la gráfica con relaciones relativas al listado completo de FESTESE001VD.



Figura 81.- Parte izquierda de la gráfica con relaciones relativas a la ULL origen de FESTESE001VD.

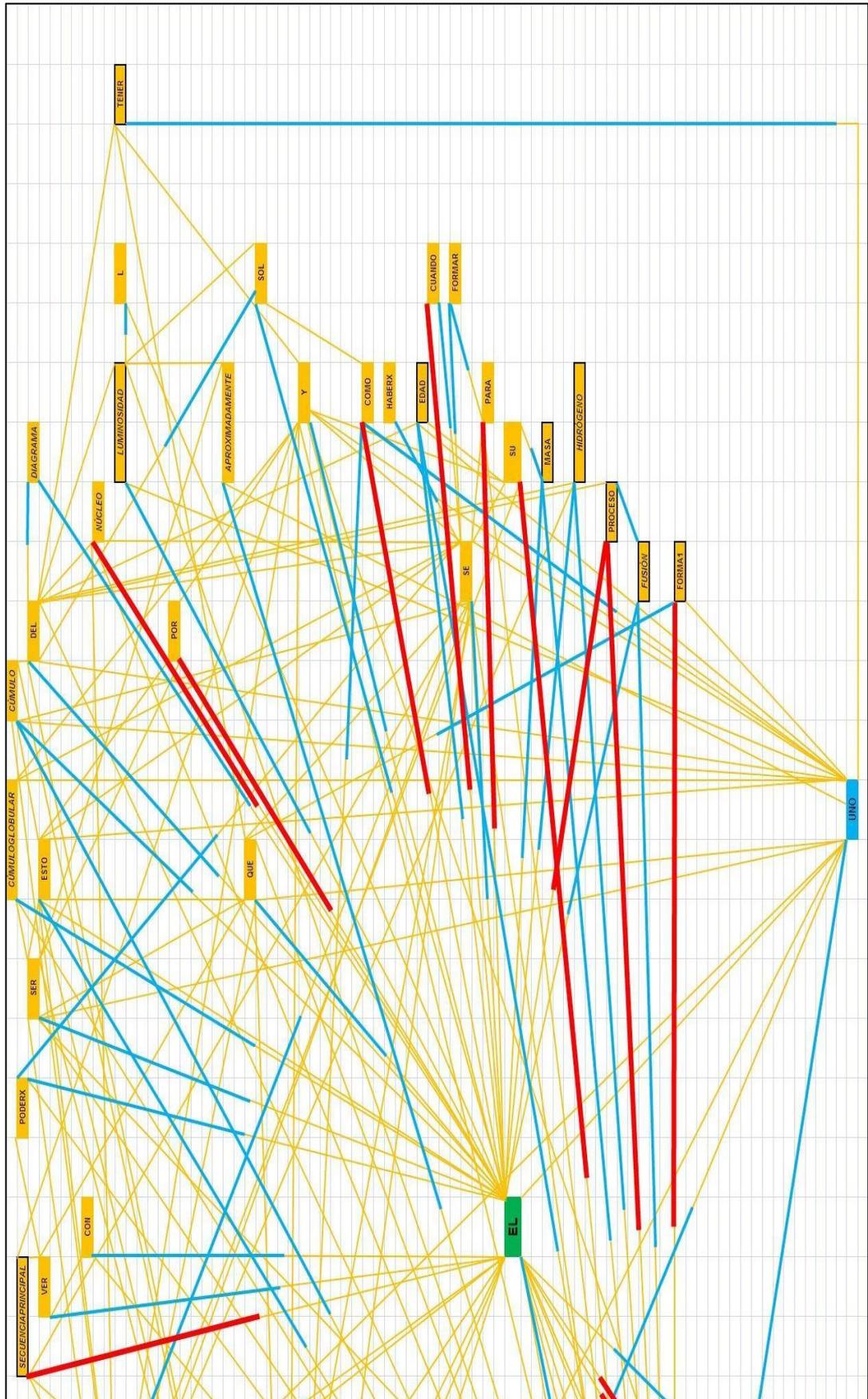


Figura 82.- Parte derecha de la gráfica con relaciones relativas a la ULL origen de FESTEESE001VD

UN ESTUDIO DE CASO: UN TEXTO ESCOLAR DE ASTRONOMÍA

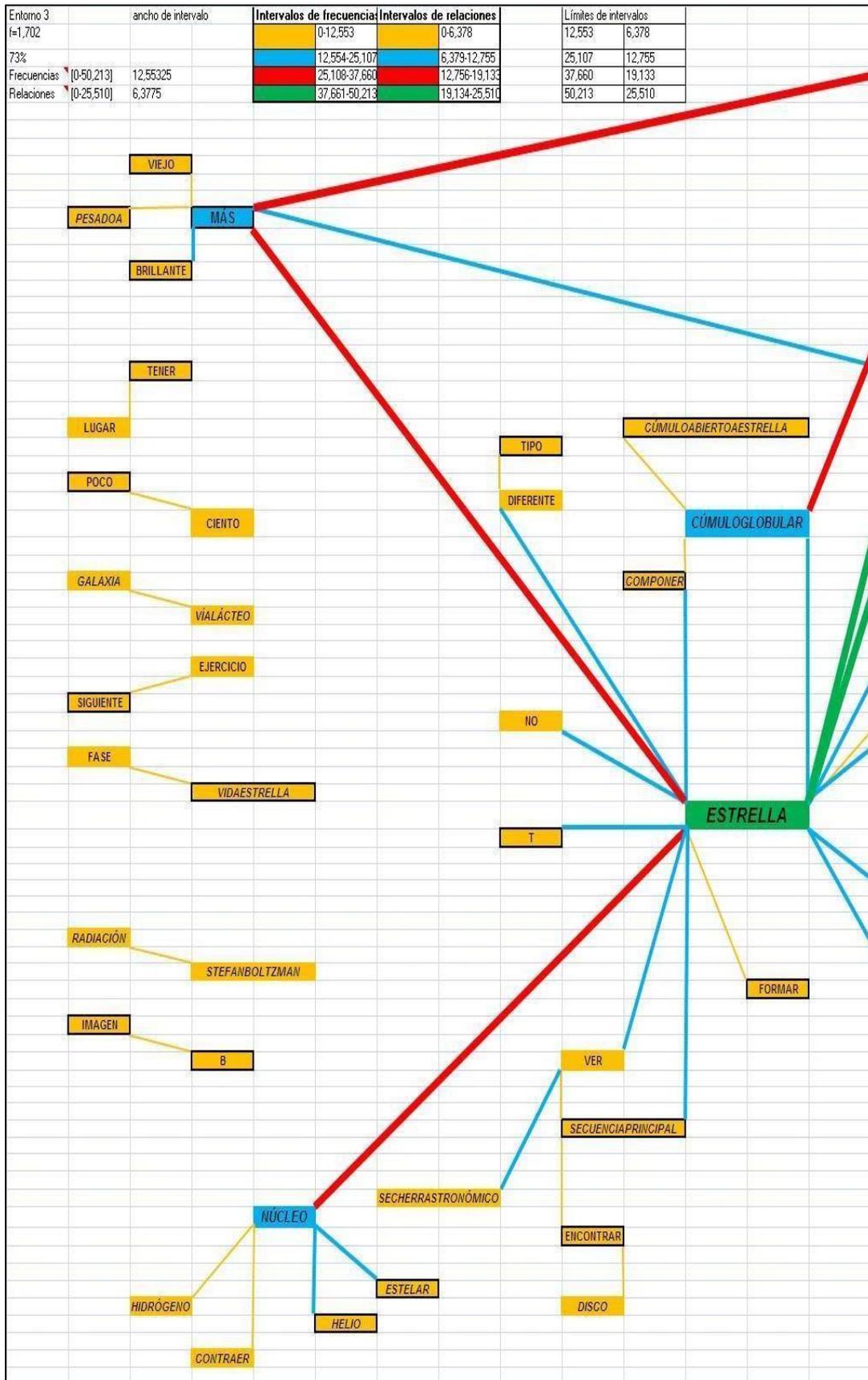


Figura 83.- Parte izquierda de la gráfica con relaciones relativas al listado completo de FESTESE001VD -F.

UN ESTUDIO DE CASO: UN TEXTO ESCOLAR DE ASTRONOMÍA

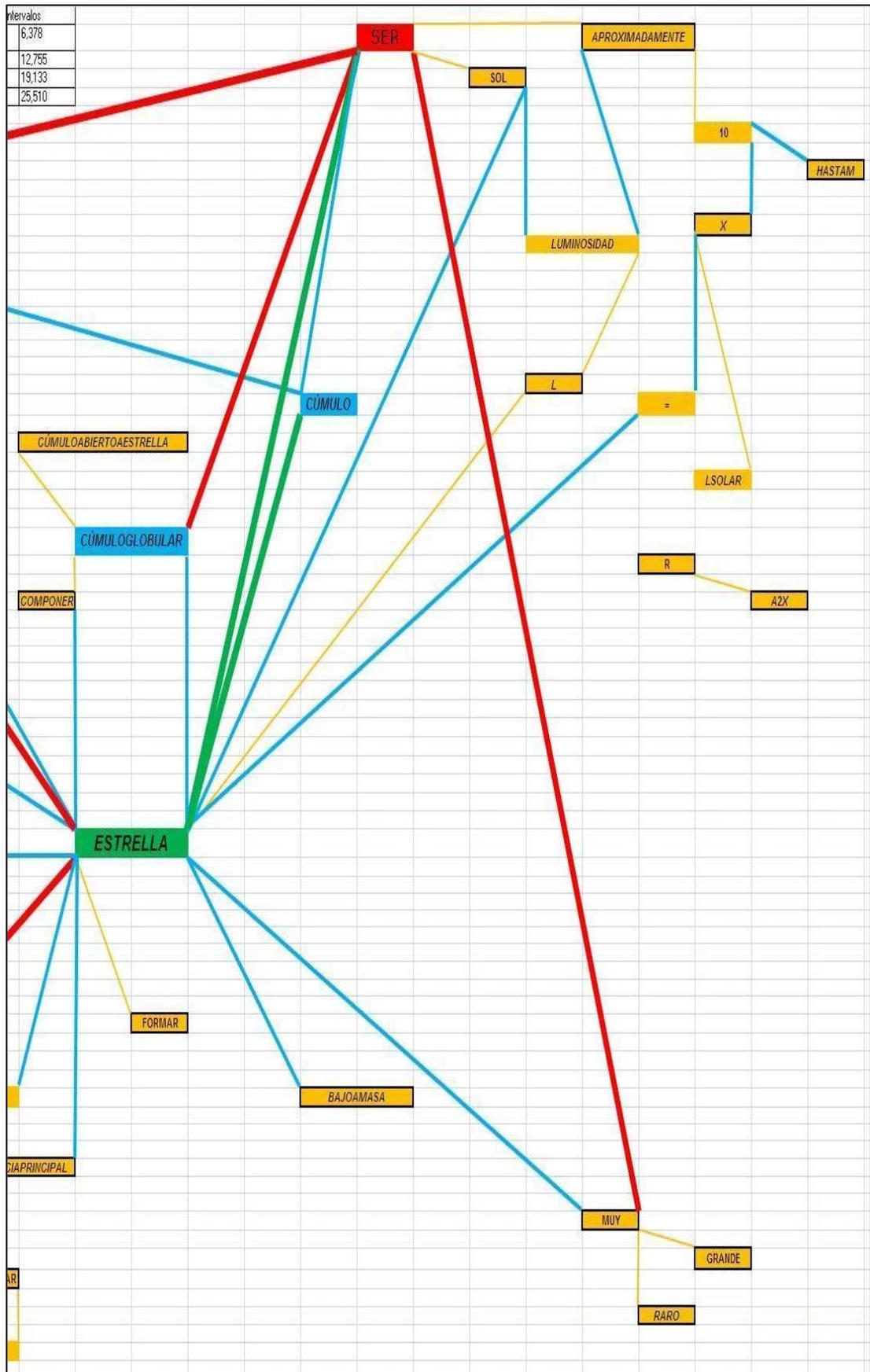


Figura 84.- Parte derecha de la gráfica con relaciones relativas al listado completo de FESTESE001VD -F.

UN ESTUDIO DE CASO: UN TEXTO ESCOLAR DE ASTRONOMÍA

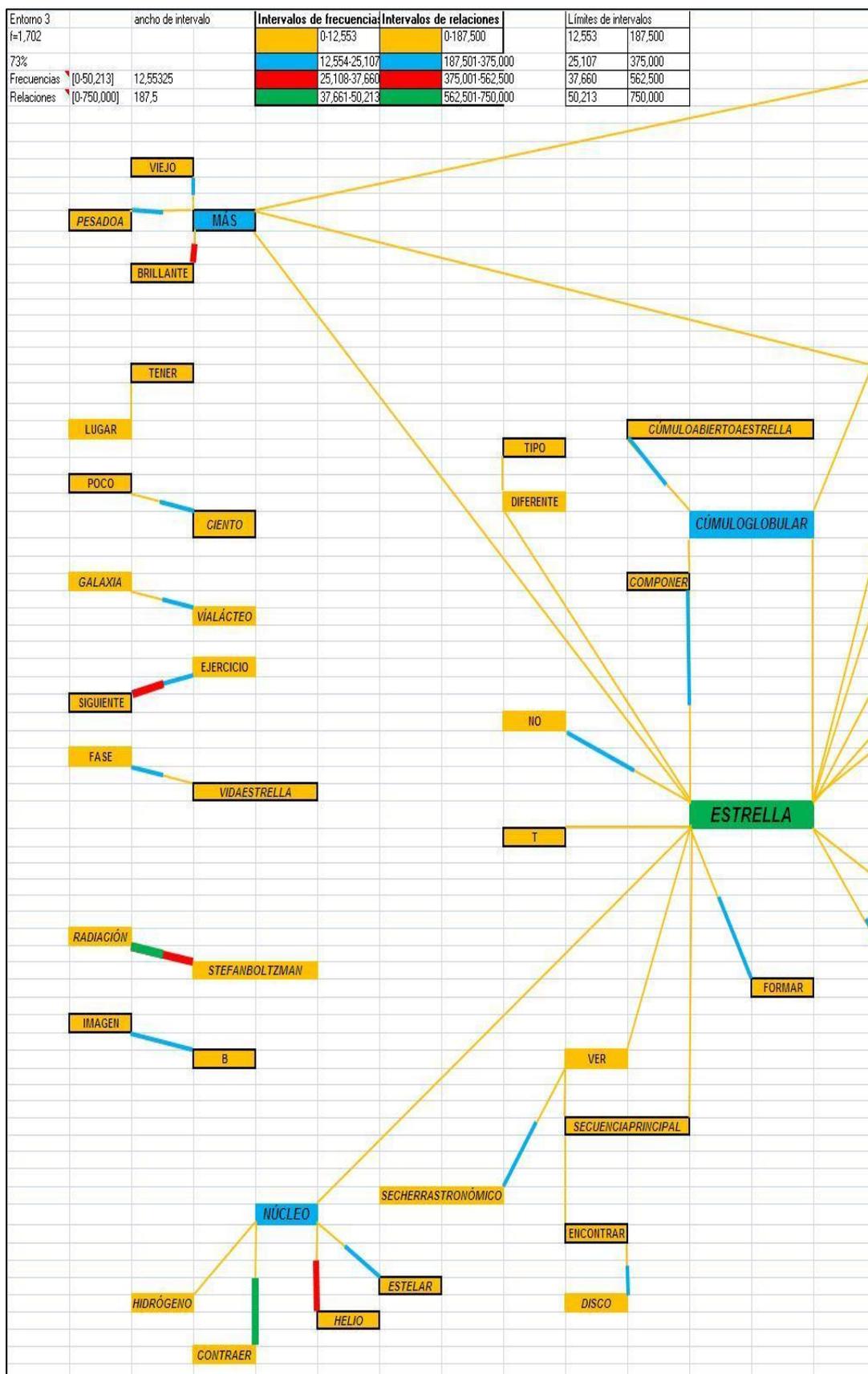


Figura 85.- Parte izquierda de la gráfica con relaciones relativas a la ULL origen de FESTESE001VD-F.

UN ESTUDIO DE CASO: UN TEXTO ESCOLAR DE ASTRONOMÍA

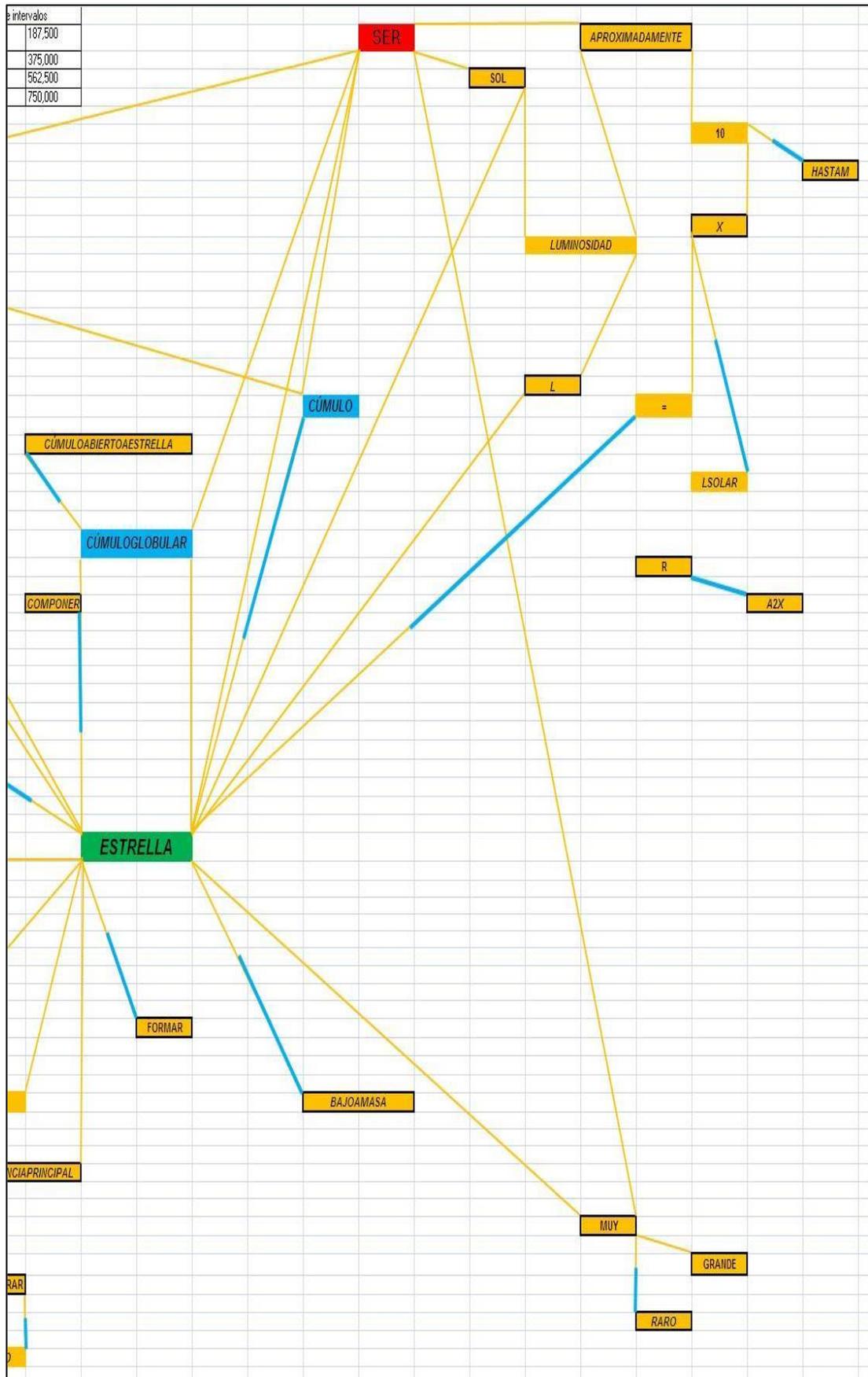


Figura 86.- Parte derecha de la gráfica con relaciones relativas a la ULL origen de FESTESE001VD-F.

UN ESTUDIO DE CASO: UN TEXTO ESCOLAR DE ASTRONOMÍA

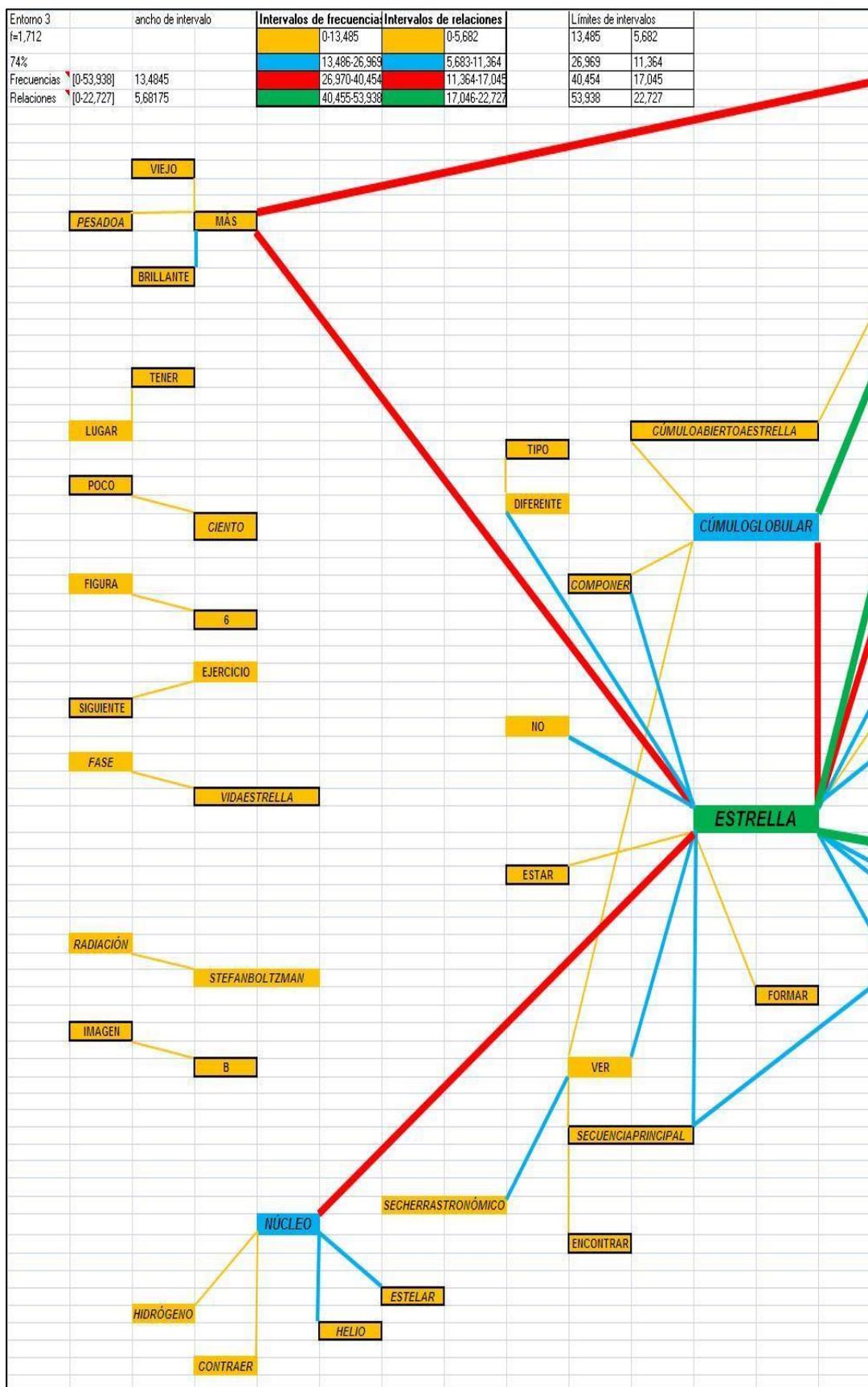


Figura 87.- Parte izquierda de la gráfica con relaciones relativas al listado completo de FESTESE001VD -S.

UN ESTUDIO DE CASO: UN TEXTO ESCOLAR DE ASTRONOMÍA

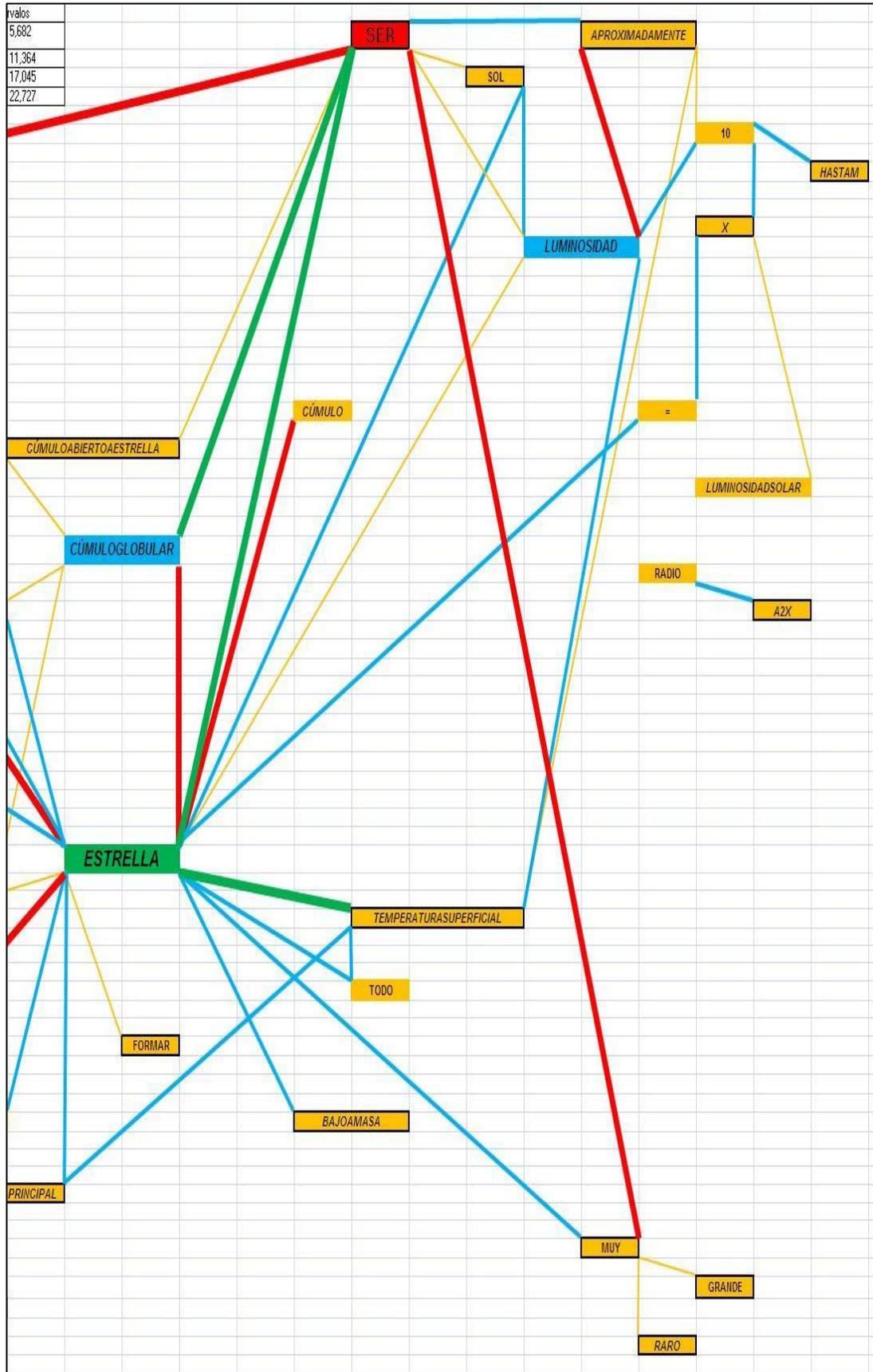


Figura 88.- Parte derecha de la gráfica con relaciones relativas al listado completo de FESTESE001VD -S.

UN ESTUDIO DE CASO: UN TEXTO ESCOLAR DE ASTRONOMÍA

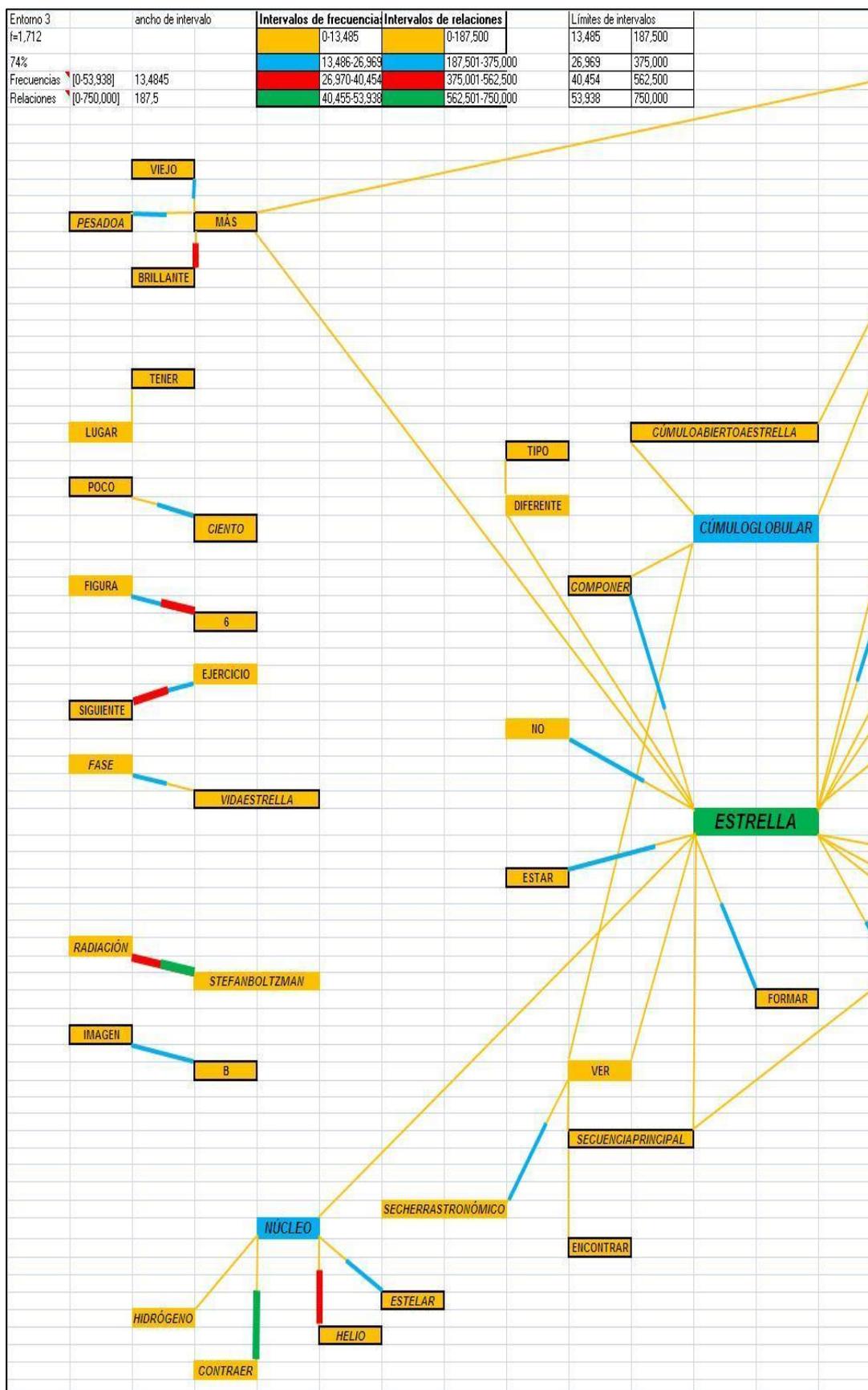


Figura 89.- Parte izquierda de la gráfica con relaciones relativas a la ULL origen de FESTESE001VD-S.

UN ESTUDIO DE CASO: UN TEXTO ESCOLAR DE ASTRONOMÍA

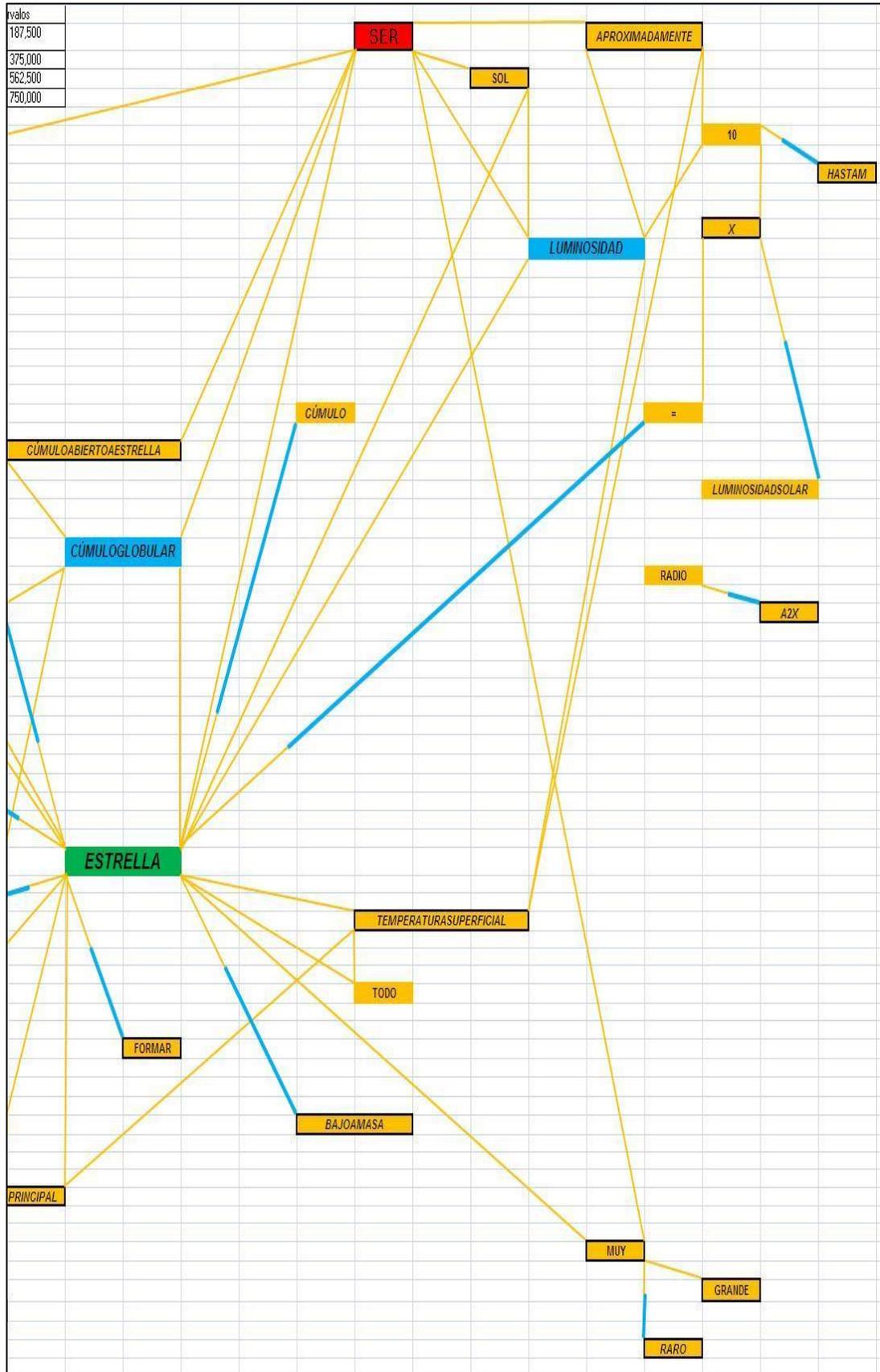


Figura 90.- Parte derecha de la gráfica con relaciones relativas a la ULL origen de FESTESE001VD-S.

UN ESTUDIO DE CASO: UN TEXTO ESCOLAR DE ASTRONOMÍA

La importancia relativa de las categorías gramaticales para las representaciones gráficas de las redes semánticas se puede ver en las Tablas 90 y 91. En estas tablas se ha añadido como categoría adicional la de colocaciones formadas por uniones de palabras de distintas categorías, ya que el resto de colocaciones se han clasificado en la categoría correspondiente. Los datos en el AO son relativos a la categoría léxica para poder compararlos con los otros dos análisis. De nuevo los resultados para los AF y AG son similares y distintos del AO debido a la eliminación de palabras funcionales y verbos auxiliares. La categoría léxica más representada en las gráficas es la de sustantivos, incluso cuando se mira la especificidad, lo cual es coherente con el alto grado de nominalización que se espera de los textos de ciencias (Shanahan & Shanahan, 2008; Yore, Bisanz, & Hand, 2003) y con los resultados obtenidos en los distintos pasos de la metodología. Sin embargo cuando se considera la conservación de relaciones de cada ULL son más importantes los adjetivos, el lenguaje matemático y las colocaciones. Esto se debe a que los sustantivos perderán más fácilmente las numerosas relaciones que puedan tener con los cortes en frecuencia. Ocurría lo mismo con las ULL funcionales en el AO.

Tabla 88

Características de las ULL representadas en las gráficas

Análisis	f_{\min}	Nº	% N ^o _{sistema}	% ($\sum n_i$) _{sistema}
AO	2.738	55	100.00	100.00
AF	1.702	54	24.00	47.39
AG	1.712	54	25.23	49.71

Nota. f_{\min} = frecuencia relativa mínima; N^o=número de ULL representadas en las gráficas; N^o_{sistema}= número de ULL de la gráfica que forman parte del sistema elegido; n_{sistema} = frecuencia absoluta de las ULL de la gráfica que forman parte del sistema elegido.

Tabla 89

Especificidad y relaciones de las ULL representadas en las gráficas

Análisis	% N ^o _r	%(∑n _i) _r	% N ^o _{esp}	%(∑n _i) _{esp}	% N ^o _{r,esp}	%(∑n _i) _{r,esp}
AO	25.45	13.96	25.45	15.47	9.09	8.63
AF	57.41	39.95	55.56	57.35	27.78	16.91
AG	59.26	40.52	51.85	56.44	29.63	19.91

Nota. N^o_r= número de ULL de la gráfica que conservan más del 80% de sus relaciones; n_r= frecuencia absoluta de las ULL de la gráfica que conservan más del 80% de sus relaciones; N^o_{esp}= número de ULL de la gráfica que son específicas; n_{esp}= frecuencia absoluta de las ULL de la gráfica que son específicas; N^o_{r,esp}= número de ULL de la gráfica que conservan más del 80% de sus relaciones y que además son específicas; n_{r,esp}= frecuencia absoluta de las ULL de la gráfica que conservan más del 80% de sus relaciones y que además son específicas.

Tabla 90

Comparación de categorías gramaticales representadas en las gráficas

Categoría gramatical	AO		AF		AG	
	% N ^o	%(∑n _i)	% N ^o	%(∑n _i)	% N ^o	%(∑n _i)
verbos	14.29	17.59	12.96	18.14	14.81	18.50
adverbios	8.57	9.16	7.41	10.54	7.41	9.84
adjetivos	14.29	8.92	18.52	8.82	20.37	10.54
lm	11.43	7.23	16.67	11.27	12.96	7.03
sustantivos	45.71	49.88	31.48	39.71	31.48	38.41
colocaciones	5.71	7.23	12.96	11.52	12.96	15.69

Nota. lm= lenguaje matemático; N^o=número de ULL; n= frecuencia absoluta.

En las gráficas de las redes semánticas se ha obviado la representación numérica que aparecía en la *Figura 17* dado que sólo se van a comparar rangos de frecuencias y rangos de relaciones. En el AO sólo aparece un grupo aislado de la red principal, mientras que en el AF y AG hay 8 grupos aislados de la misma. El grupo aislado en el AO (parte superior izquierda de la *Figura 79*) aparece incorporado a la red principal en los AF y AG (parte superior derecha de las *Figuras 84 y 88*).

UN ESTUDIO DE CASO: UN TEXTO ESCOLAR DE ASTRONOMÍA

Tabla 91

Categorías gramaticales respecto a su especificidad y conservación de relaciones

Categoría gramatical	% N ^o _r	%(∑n _i) _r	% N ^o _{esp}	%(∑n _i) _{esp}	% N ^o _{r,esp}	%(∑n _i) _{r,esp}
AO						
verbos	14.29	9.68	0.00	0.00	0.00	0.00
adverbios	0.00	0.00	7.14	4.85	0.00	0.00
adjetivos	7.14	3.23	7.14	2.91	0.00	0.00
lm	0.00	0.00	7.14	3.40	0.00	0.00
sustantivos	71.43	80.11	64.29	74.27	80.00	88.70
colocaciones	7.14	6.99	14.29	14.56	20.00	11.30
AF						
verbos	12.90	15.95	6.67	2.56	6.67	5.80
adverbios	9.68	23.31	3.33	4.27	6.67	14.49
adjetivos	29.03	17.79	13.33	4.70	26.67	15.94
lm	22.58	19.63	16.67	10.68	26.67	28.99
sustantivos	16.13	12.27	36.67	57.69	13.33	8.70
colocaciones	9.68	11.04	23.33	20.09	20.00	26.09
AG						
verbos	15.63	16.76	7.14	2.49	6.25	4.17
adverbios	9.38	21.39	3.57	4.15	6.25	11.76
adjetivos	28.13	16.76	14.29	4.56	25.00	12.94
lm	18.75	12.14	14.29	7.05	25.00	20.00
sustantivos	15.63	11.56	35.71	53.94	12.50	7.06
colocaciones	12.50	21.39	25.00	27.80	25.00	43.53

Nota. lm= lenguaje matemático; N^o=número de ULL; N^o_r= número de ULL que conservan más del 80% de sus relaciones; n_r= frecuencia absoluta de las ULL que conservan más del 80% de sus relaciones; N^o_{esp}= número de ULL que son específicas; n_{esp}= frecuencia absoluta de las ULL que son específicas; N^o_{r,esp}= número de ULL que conservan más del 80% de sus relaciones y que además son específicas; n_{r,esp}= frecuencia absoluta de las ULL que conservan más del 80% de sus relaciones y que además son específicas.

Existen relaciones fuertes de ULL de alta frecuencia con ULL de baja frecuencia, como por ejemplo *en-el* (AO) y *estrella-sol* (AF y AG). En el AO las diferencias entre las *Figuras 79 y 80* y las *Figuras 81 y 82* se refieren a la forma en que se ha calculado la relación relativa. En las *Figuras 79 y 80* las relaciones relativas lo son respecto a la suma de relaciones absolutas entre todas las ULL, que se obtienen en el listado del botón inferior del PAFE *Lista de relaciones*. En las *Figuras 81 y 82* las relaciones relativas lo son respecto a la suma de relaciones

absolutas de la ULL origen. Por ello, en estas últimas figuras, algunas de las relaciones no son simétricas. Por ejemplo, *estrella* está fuertemente relacionada con *el*, pero ésta no está fuertemente relacionada con *estrella* (Figura 81). Es decir, hay una alta probabilidad de que *el* aparezca en la cercanía de *estrella*, pero no de que *estrella* aparezca cerca de *el*. Algo parecido ocurre en las gráficas de los AF y AG.

En el AO la mayoría de las ULL se encuentran fuertemente relacionadas con una ULL funcional, como por ejemplo *estrella-el*. Sin embargo, dichas ULL funcionales no presentan una fuerte relación respecto a las otras ULL. Las ULL que más fuertemente se relacionan con otras son *el* y *de*, que a la vez son las de mayor frecuencia. En los AF y AG la ULL que más fuertemente se relaciona con otras es *estrella*, pero la cantidad y fuerza de estas relaciones es claramente inferior a las observadas en el AO. Ya se había predicho que la escasa repetición de las ULL léxicas produciría un menor número de relaciones entre las mismas.

Se puede ver como aparecen relaciones muy fuertes entre ULL de baja frecuencia, como por ejemplo *10-x* en la parte superior izquierda de la Figura 81 (AO). En este ejemplo la relación se calcula respecto a una ULL origen que no ha conservado bien sus relaciones, por lo que no se puede generalizar el resultado al que tenía en el texto completo. En la parte inferior derecha de la Figura 82 se observa una relación fuerte entre dos ULL que conservan un 80% de sus relaciones o más (*proceso-fusión*). Estas mismas consideraciones se pueden hacer en las gráficas de los AF y AG.

Desde el punto de vista del análisis de los conceptos las gráficas del AO presentan una información muy pobre, dado que las relaciones se concentran en torno a ULL funcionales (ver el caso de *estrella*). Sin embargo las gráficas del AF y el AG proporcionan una red rica en relaciones entre ULL léxicas y resalta la

UN ESTUDIO DE CASO: UN TEXTO ESCOLAR DE ASTRONOMÍA

importancia de las mismas. En los AF y AG hay varias ULL en las que detenerse, dada su importancia en frecuencia y relaciones (*estrella, ser, cúmulo globular, núcleo, cúmulo, luminosidad y más*). La Tabla 92 muestra las ULL relacionadas con *estrella* en los tres tipos de análisis. En ésta aparecen señaladas en cursiva las ULL específicas, en negrita aquellas que conservan más del 80% de sus relaciones y subrayadas aquellas que están fuertemente relacionadas con la ULL origen *estrella*. Además el fondo de la casilla indica el color de la relación entre la ULL origen y dicha ULL.

Tabla 92

ULL relacionadas con estrella

AO	AF	AG
EL	SER	SER
DE	<i>CÚMULO</i>	<i>TEMPERATURASUPERFICIAL</i>
<u>T</u>	MÁS	MÁS
	<i>NÚCLEO</i>	<i>CÚMULOGLOBULAR</i>
	<i>CÚMULOGLOBULAR</i>	<i>NÚCLEO</i>
	<u>COMPONER</u>	<u>COMPONER</u>
	DIFERENTE	DIFERENTE
	SOL	SOL
	<u>NO</u>	<u>NO</u>
	T	NO
	VER	VER
	<i>SECUENCIAPRINCIPAL</i>	<i>SECUENCIAPRINCIPAL</i>
	<u>BAJOAMASA</u>	<u>BAJOAMASA</u>
	MUY	MUY
	≡	≡
	<u>FORMAR</u>	<u>FORMAR</u>
		TODO
		<u>ESTAR</u>

Algunas de las ULL no ofrecen ninguna información desde el punto de vista de la didáctica, como son *más* o *no*. El análisis resalta no sólo la importancia de ULL específicas y con alto número de relaciones, sino también de ULL que tienen un uso normal en el lenguaje, el cual es el vehículo de comunicación analizado. Por

tanto, se hace necesario recurrir una vez más al experto para discernir entre lenguaje común o no. Los resultados del análisis muestran por un lado vocabulario frecuente en torno a *estrella* (*no,...*) y por otro conceptos relacionados con ésta (*cúmulo,...*). En el AF los conceptos relacionados con *estrella* serían: *cúmulo, núcleo, cúmulo globular, sol, T, secuencia principal y baja masa*. Mientras que en el AG serían: *temperatura superficial, cúmulo globular, núcleo, cúmulo, sol, secuencia principal y baja masa*. La única diferencia entre estos dos análisis lo constituye la variación de la frecuencia de las ULL al unificar grafías.

La aplicación PAFE permite hallar las redes semánticas usando un fragmento de “punto y aparte” en un entorno 3 y con las características de la Tabla 21. Así se obtienen las gráficas presentadas en las *Figuras 91, 92, 93 y 94* para el AO; *Figuras 95, 96, 97 y 98* para el AF; *Figuras 99, 100, 101 y 102* para el AG.

Las diferencias de estas gráficas respecto al punto y seguido (*Figuras 79 a 90*) están en las relaciones entre las ULL, ya que la frecuencia de éstas no ha variado. Aparecen algunas relaciones nuevas entre parejas de ULL, por ejemplo *el-aproximadamente* (*Figuras 80 y 92*) en el AO, *ejercicio-estrella* (*Figuras 83 y 95, Figuras 87 y 99*) en los AF y AG. En estos últimos han aparecido además grupos nuevos aislados, como *ya-ca2* (*Figuras 95 y 99*); ULL nuevas que se unen a la red principal, como *distancia* (*Figuras 96 y 100*) o a otras redes como *V* (*Figuras 95 y 99*), redes aisladas que se unen a la red principal como *siguiente-ejercicio a estrella* (*Figuras 95 y 99*) y relaciones asimétricas respecto a la ULL origen que se encuentran por debajo del umbral de relaciones escogido y se representan con una fina línea negra, como *estrella* respecto a *distancia* (*Figuras 98 y 102*).

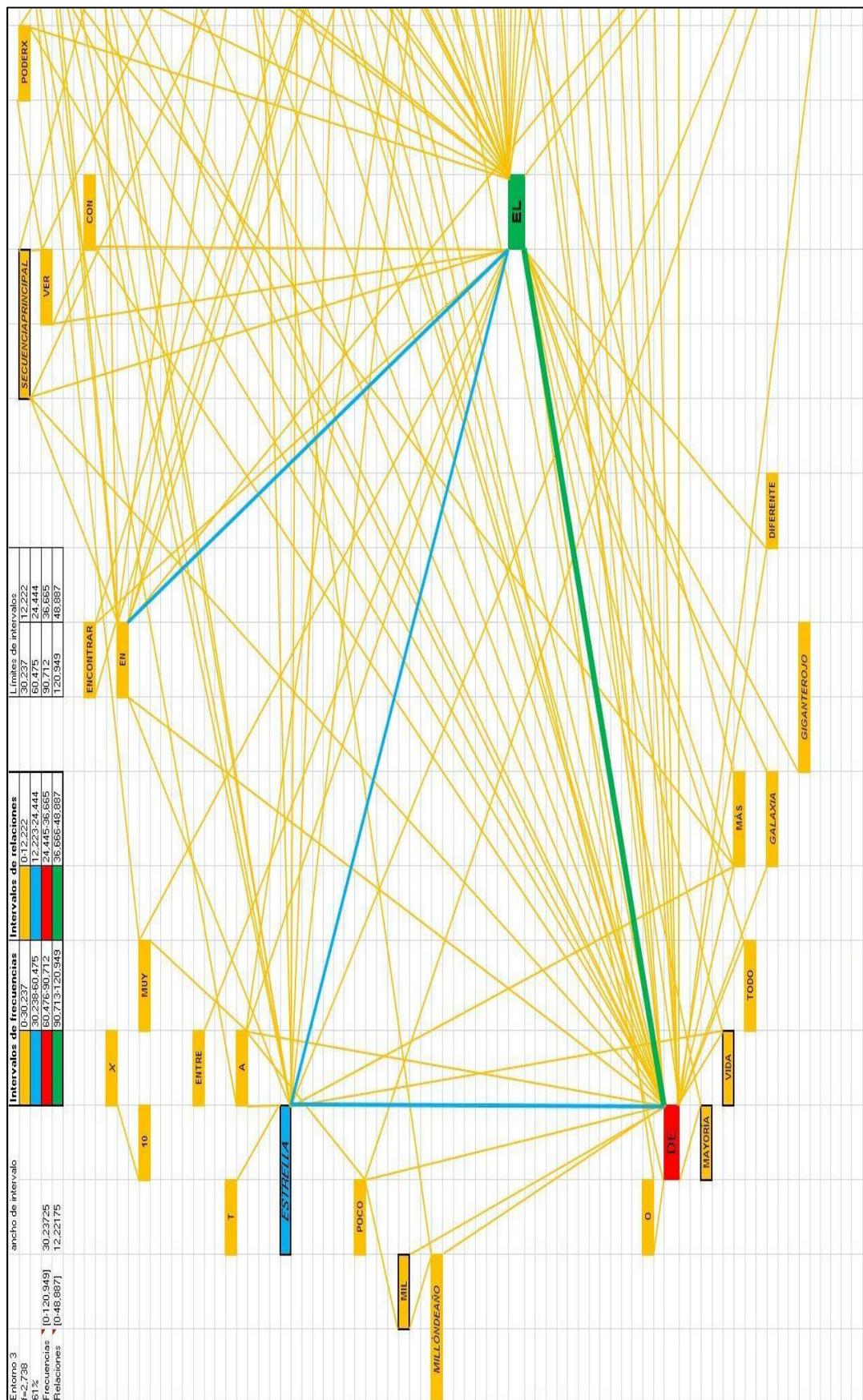


Figura 91.- Parte izquierda de la gráfica con relaciones relativas al listado completo, fragmento="punto y aparte", de FESTESE001VD.



Figura 92.- Parte derecha de la gráfica con relaciones relativas al listado completo, fragmento="punto y aparte", de FESTESE001VD.

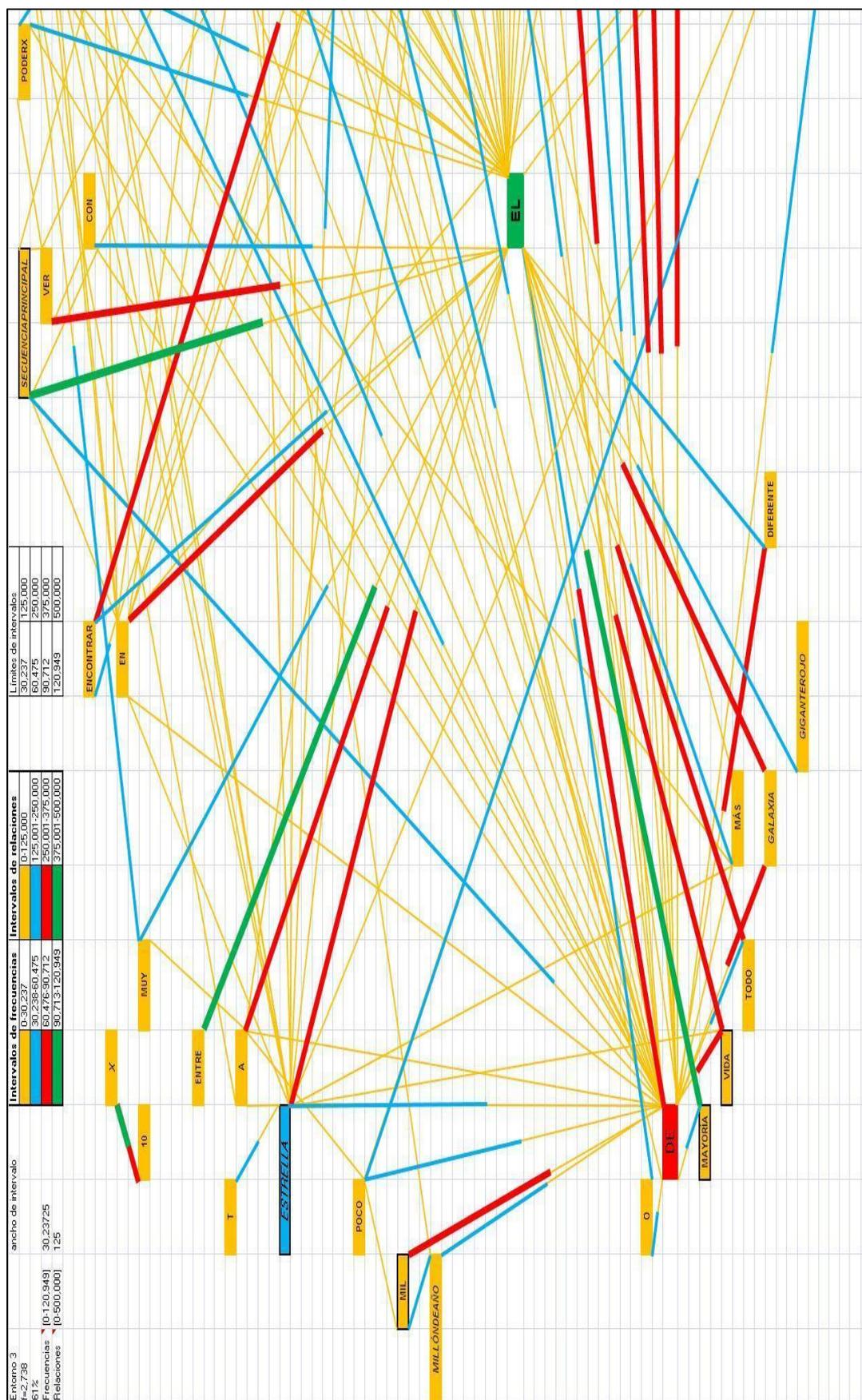


Figura 93.- Parte izquierda de la gráfica con relaciones relativas a la ULL origen, fragmento="punto y aparte", de FESTEES001VD.

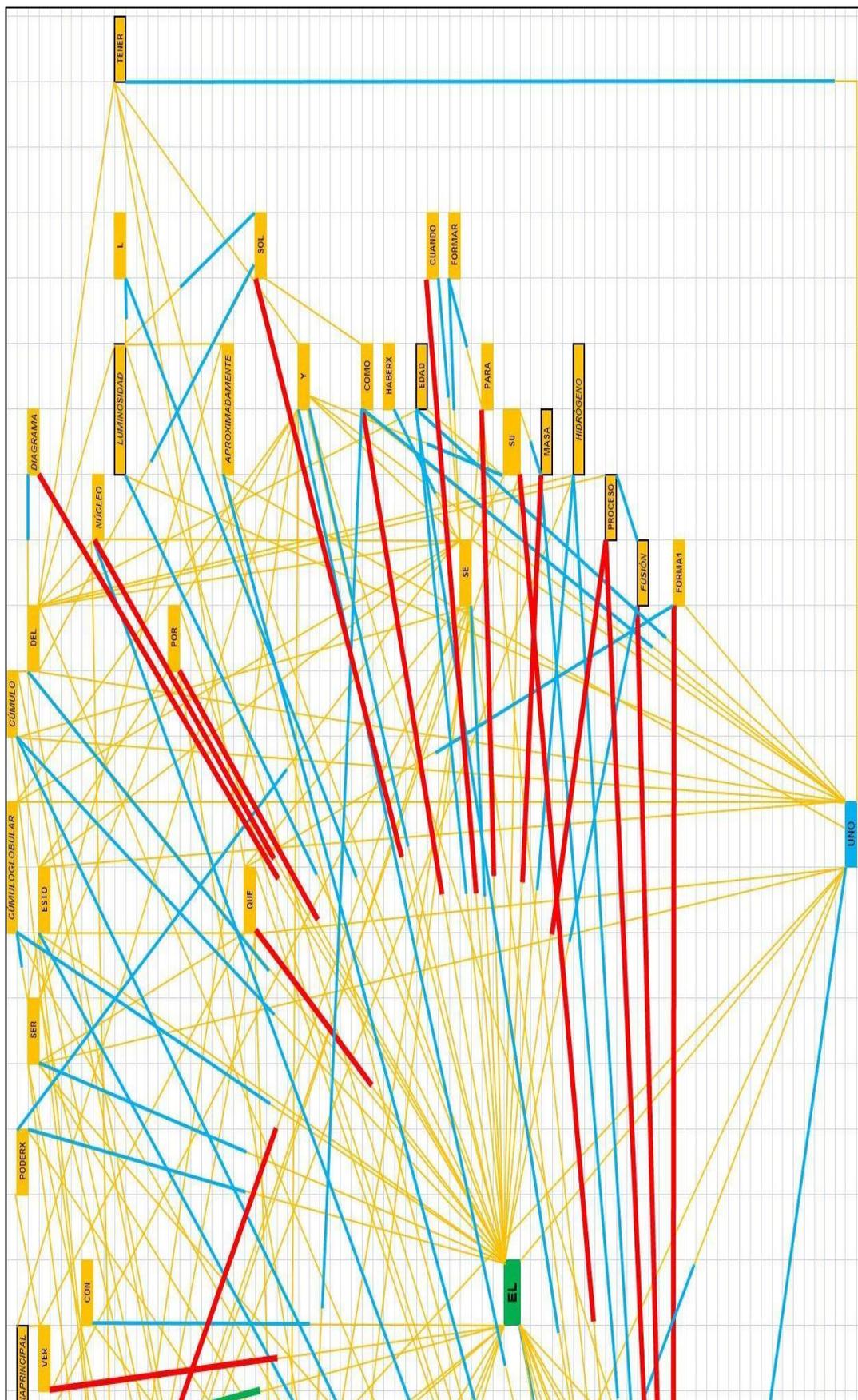


Figura 94.- Parte derecha de la gráfica con relaciones relativas a la ULL origen, fragmento="punto y aparte", de FESTESE001VD.

UN ESTUDIO DE CASO: UN TEXTO ESCOLAR DE ASTRONOMÍA

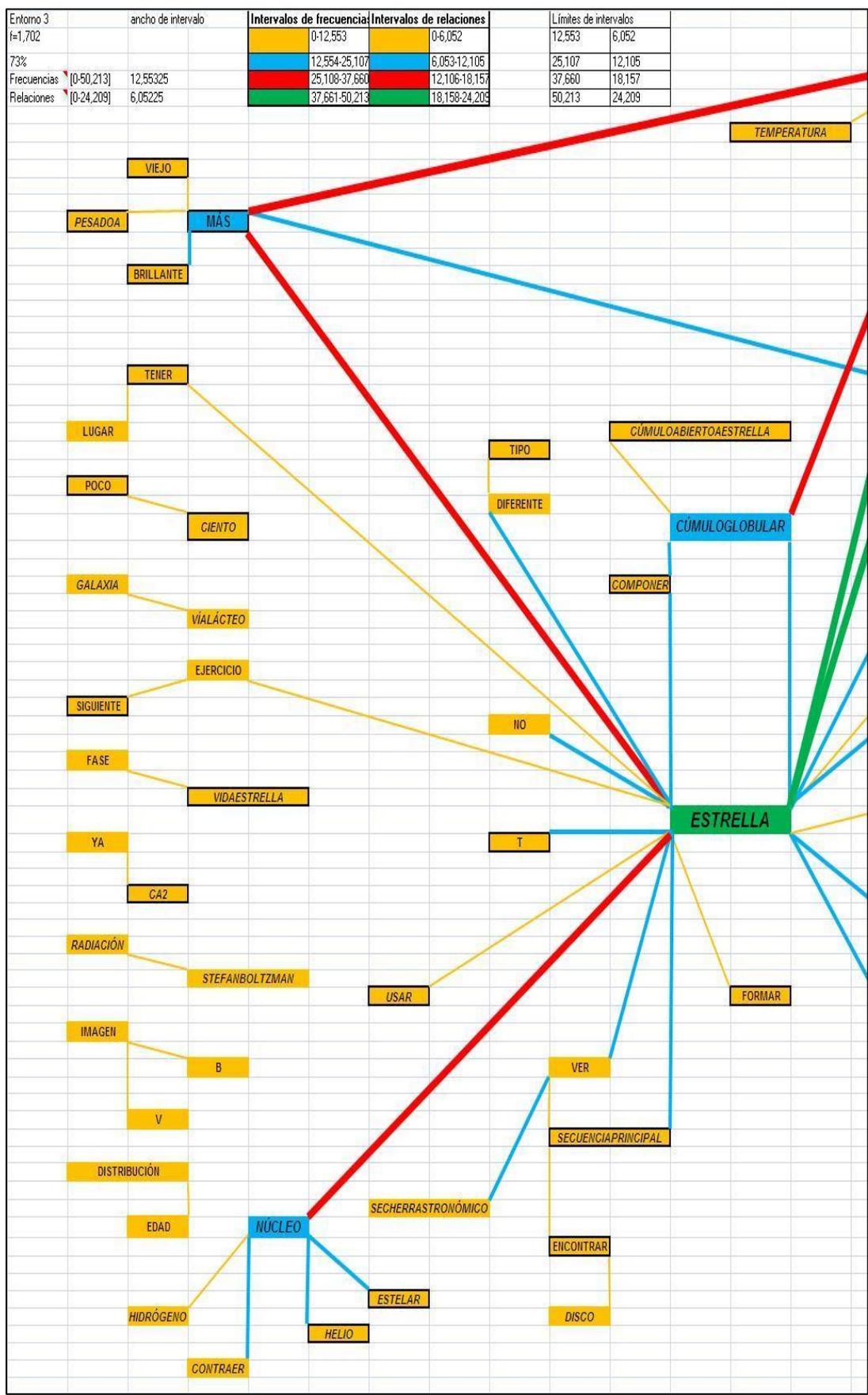


Figura 95.- Parte izquierda de la gráfica con relaciones relativas al listado completo, fragmento="punto y aparte", de FESTESE001VD-F.

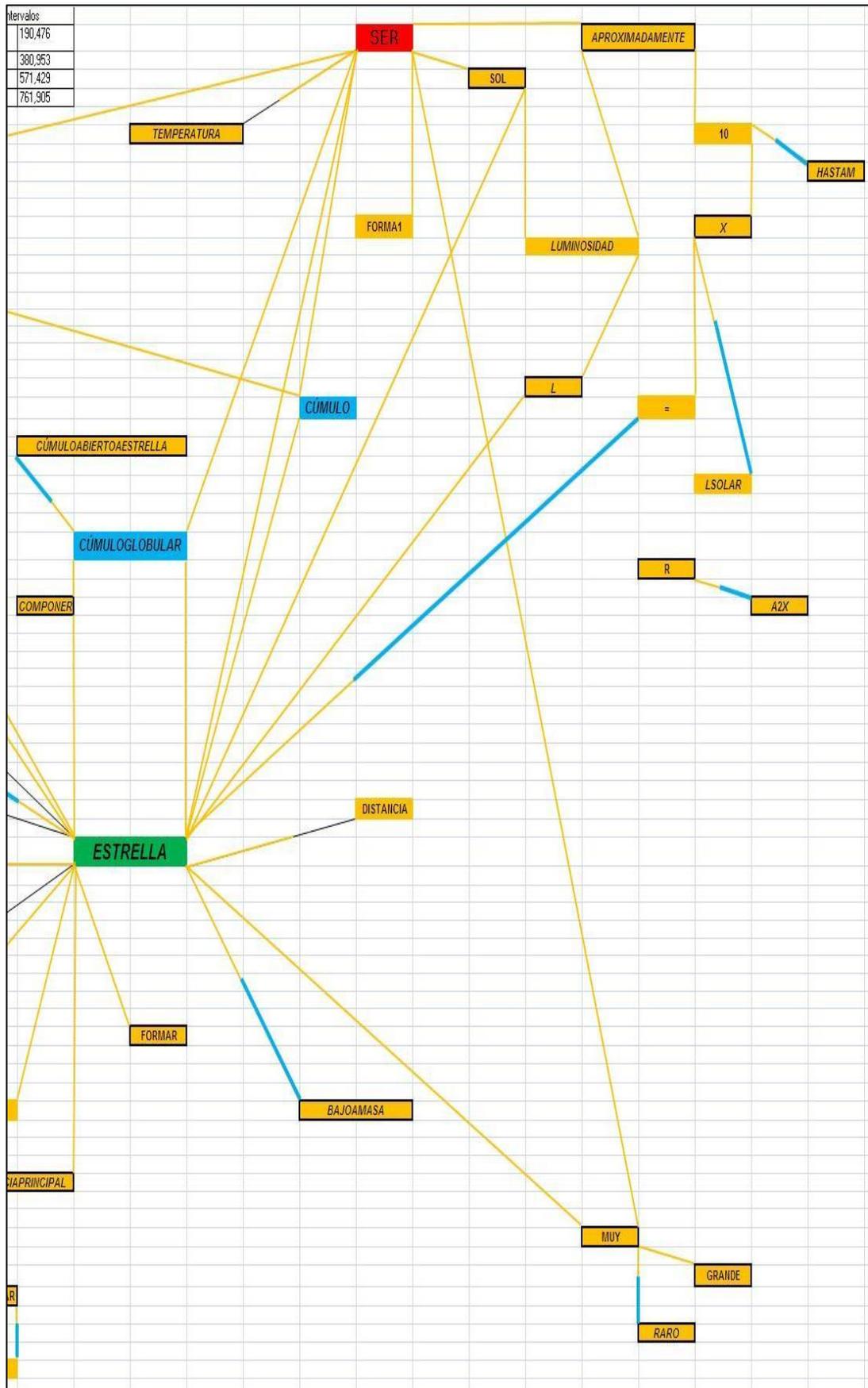


Figura 96.- Parte derecha de la gráfica con relaciones relativas al listado completo, fragmento="punto y aparte", de FESTESE001VD-F.

UN ESTUDIO DE CASO: UN TEXTO ESCOLAR DE ASTRONOMÍA

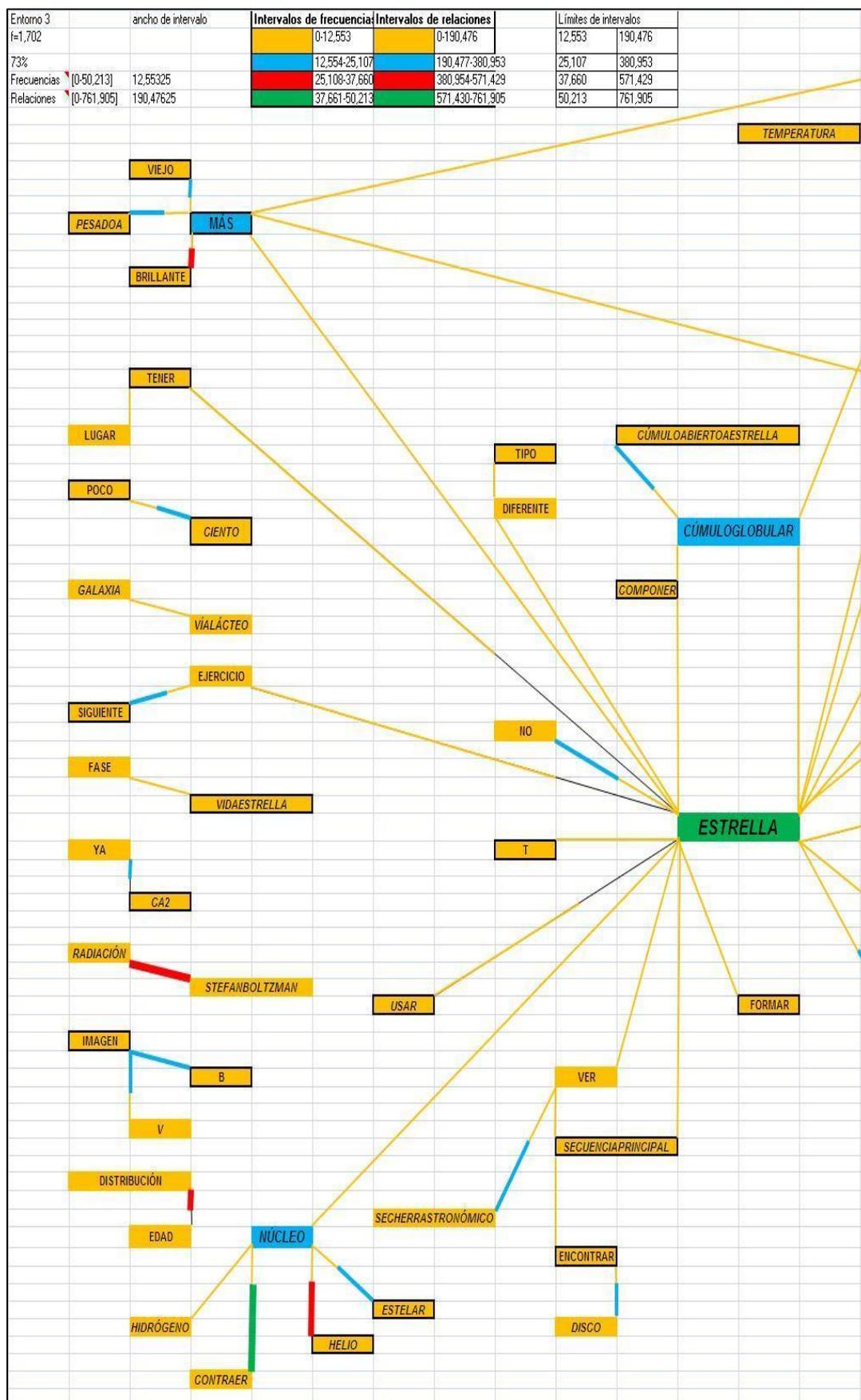


Figura 97.- Parte izquierda de la gráfica con relaciones relativas a la ULL origen, fragmento="punto y aparte", de FESTESE001VD-F.

UN ESTUDIO DE CASO: UN TEXTO ESCOLAR DE ASTRONOMÍA

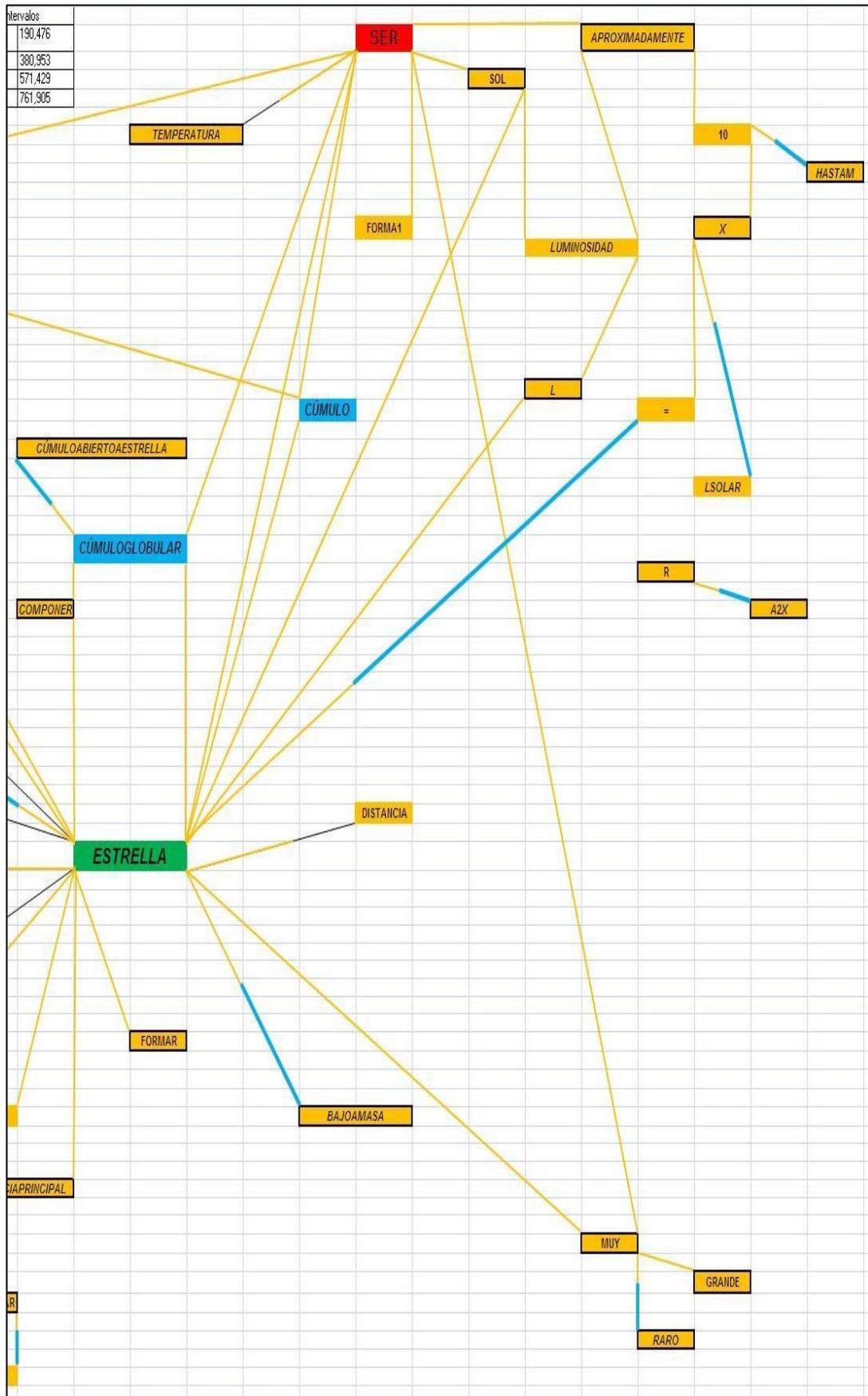


Figura 98.- Parte derecha de la gráfica con relaciones relativas a la ULL origen, fragmento="punto y aparte", de FESTESE001VD-F.

UN ESTUDIO DE CASO: UN TEXTO ESCOLAR DE ASTRONOMÍA

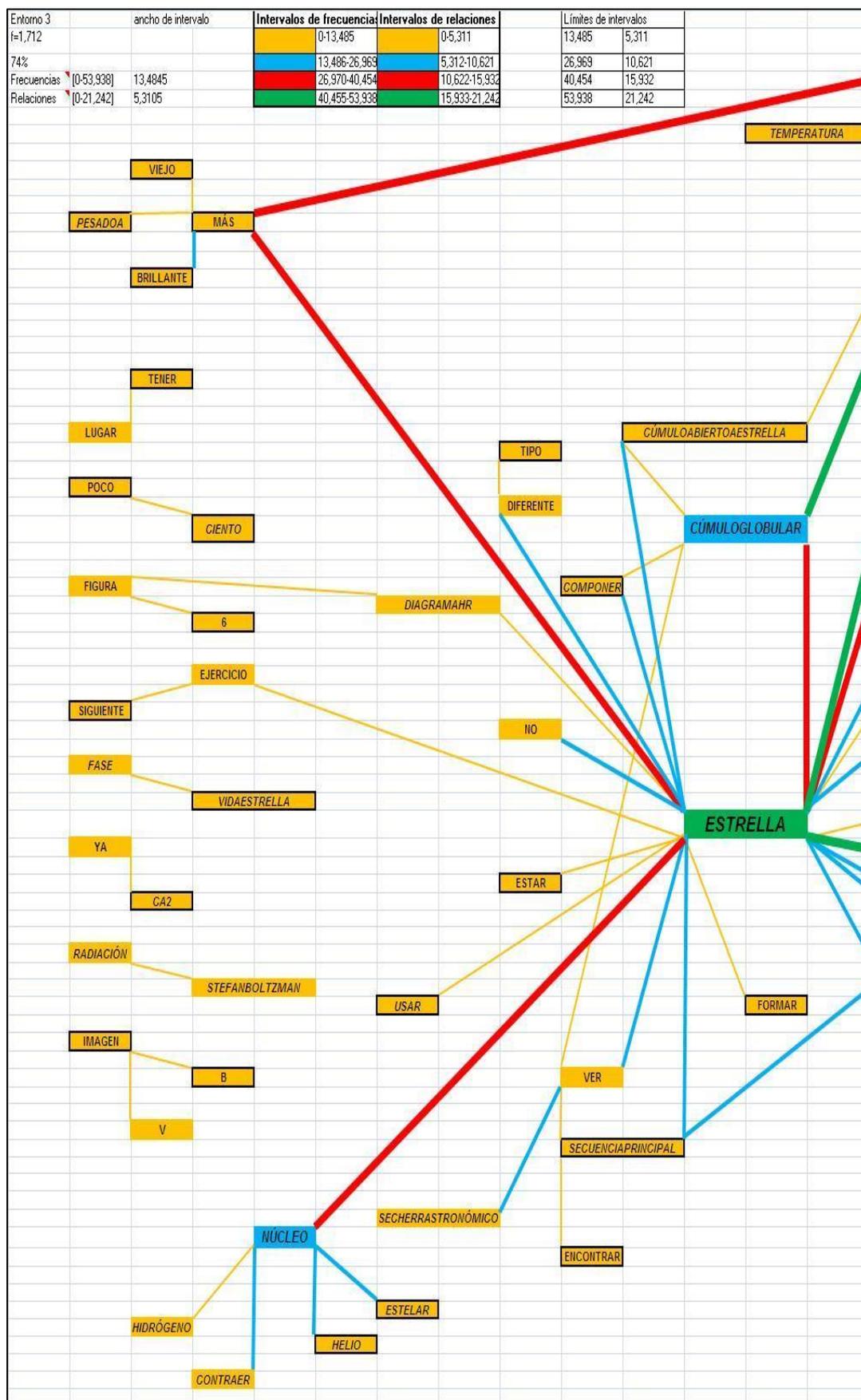


Figura 99.- Parte izquierda de la gráfica con relaciones relativas al listado completo, fragmento="punto y aparte", de FESTESE001VD-S.

UN ESTUDIO DE CASO: UN TEXTO ESCOLAR DE ASTRONOMÍA

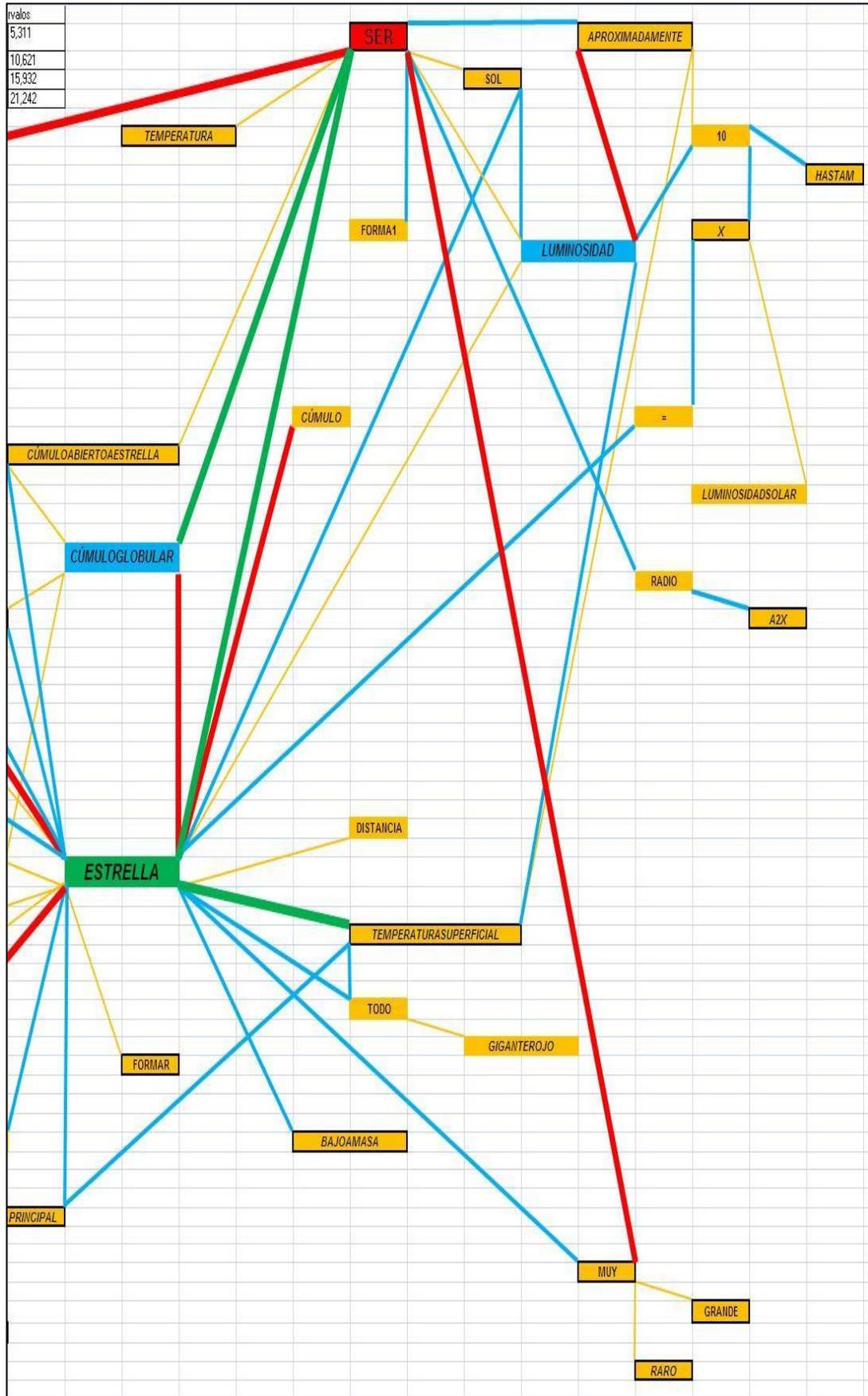


Figura 100.- Parte derecha de la gráfica con relaciones relativas al listado completo, fragmento="punto y aparte", de FESTESE001VD-S.

UN ESTUDIO DE CASO: UN TEXTO ESCOLAR DE ASTRONOMÍA

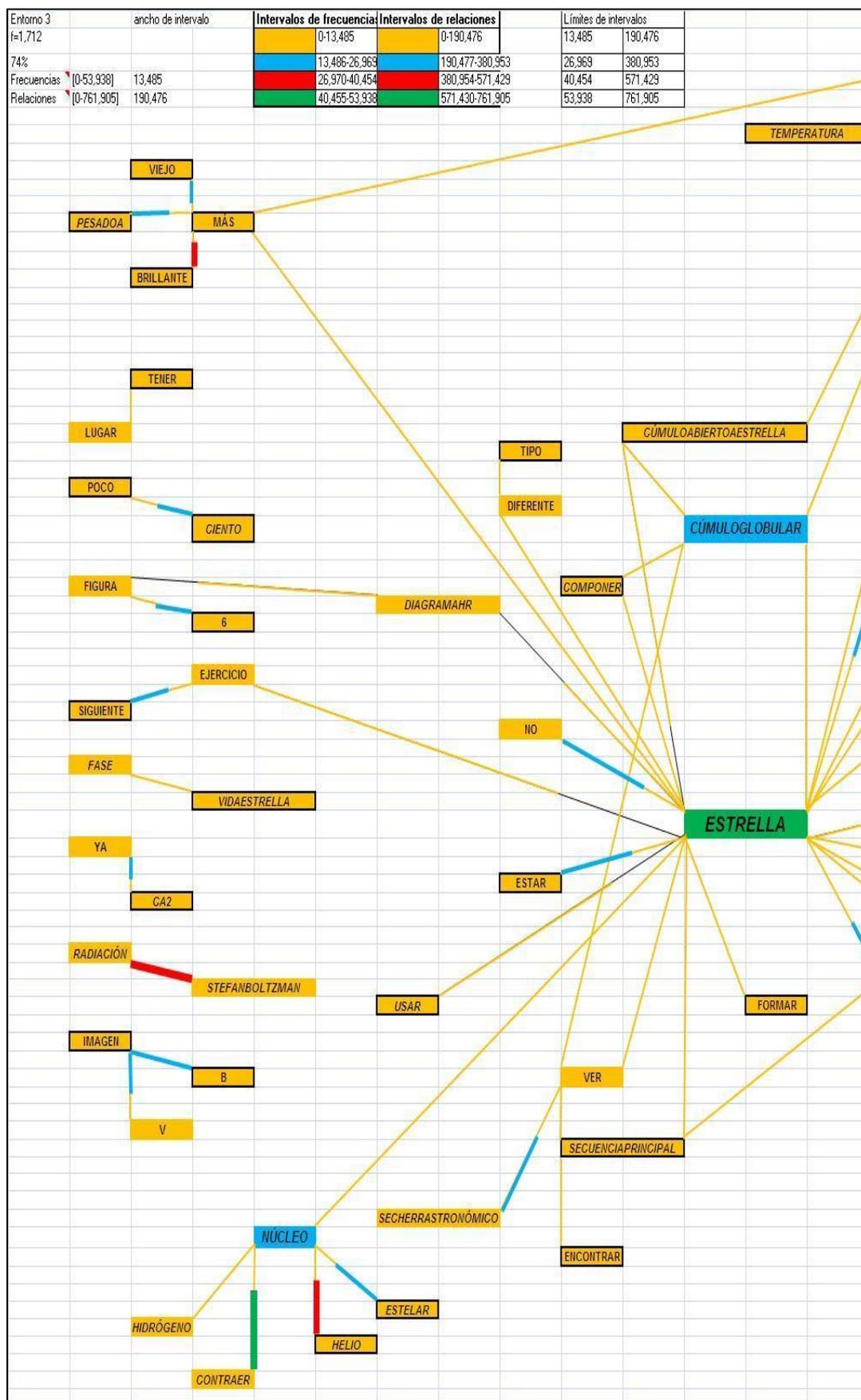


Figura 101.- Parte izquierda de la gráfica con relaciones relativas a la ULL origen, fragmento="punto y aparte", de FESTESE001VD-S.

UN ESTUDIO DE CASO: UN TEXTO ESCOLAR DE ASTRONOMÍA

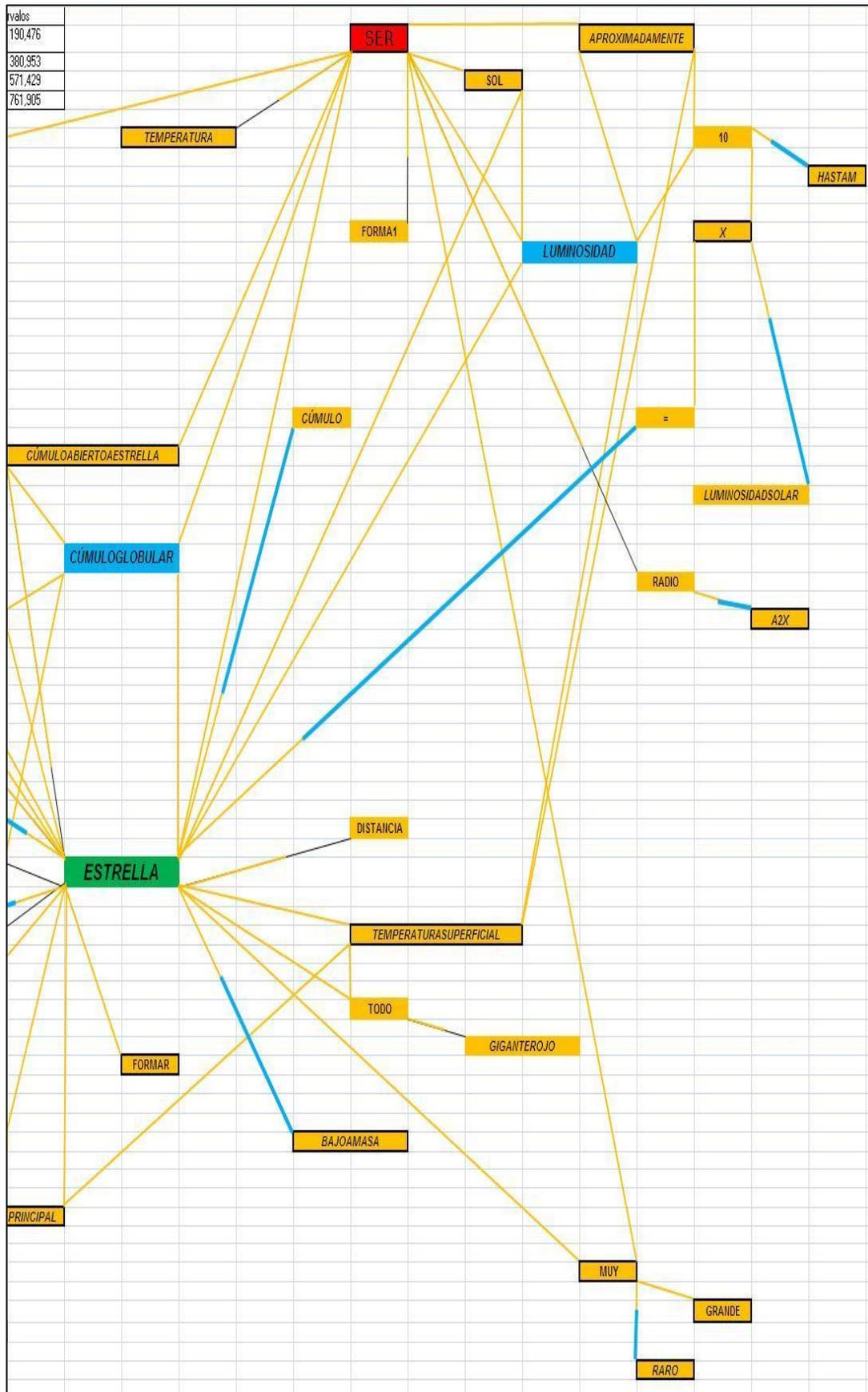


Figura 102.- Parte derecha de la gráfica con relaciones relativas a la ULL origen, fragmento="punto y aparte", de FESTEESE001VD-S.

UN ESTUDIO DE CASO: UN TEXTO ESCOLAR DE ASTRONOMÍA

En la representación en cuatro colores la mayoría de las relaciones entre ULL se vuelven a representar en el mismo color tanto para el fragmento terminado en punto y seguido como para aquel terminado en punto y aparte. Sólo algunas relaciones entre parejas de ULL han pasado a un intervalo superior, como por ejemplo *núcleo-contraer* en los AF y AG (ver *Figuras 95 y 99*). Las mayores diferencias respecto a las relaciones se observan en aquellas que son respecto a una ULL origen, como por ejemplo *entre-el* en el AO donde la relación de *entre* respecto a *el* ha subido de rango o *componer* respecto a *estrella* en los AF y AG que ha bajado de rango. Al haber mayor número de relaciones para una ULL origen, sus relaciones totales han aumentado y las relaciones relativas de otra ULL respecto a aquella disminuyen del análisis con punto seguido al análisis con punto y aparte. La supresión del punto y seguido como final de entorno puede hacer que las ULL cercanas a los mismos aumenten sus relaciones. Para estas ULL puede ocurrir que sus relaciones relativas a la ULL origen no disminuyan tanto. La representación con punto y aparte resalta más las diferencias entre relaciones relativas a la ULL origen en el AO, sin embargo no lo hace tanto en los AF y AG, siendo más interesante en estos últimos los resultados respecto a las apariciones de nuevas ULL y nuevas relaciones entre ULL pre-existentes.

Desde el punto de vista del análisis de los conceptos, el análisis con un fragmento de “punto y aparte” no cambia las ULL que son de interés, de acuerdo a su importancia en frecuencia y relaciones, en los AF y AG (*estrella, ser, cúmulo globular, núcleo, cúmulo, luminosidad y más*). Sin embargo se producen ligeras variaciones respecto a la fuerza de la relación de las ULL respecto a una ULL origen como *estrella* (Tabla 93). En la Tabla 93 aparecen señaladas en cursiva las ULL específicas, en negrita aquellas que conservan más del 73.53% de sus

UN ESTUDIO DE CASO: UN TEXTO ESCOLAR DE ASTRONOMÍA

relaciones y subrayadas aquellas que están fuertemente relacionadas con la ULL origen *estrella*. Además el fondo de la casilla indica el color de la relación entre la ULL origen y dicha ULL.

Tabla 93

ULL relacionadas con estrella con el fragmento “punto y aparte”

AO	AF	AG
EL	SER	SER
DE	CÚMULO	TEMPERATURASUPERFICIAL
T	MÁS	MÁS
	NÚCLEO	CÚMULOGLOBULAR
	CÚMULOGLOBULAR	NÚCLEO
	COMPONER	CÚMULO
	DIFERENTE	COMPONER
	SOL	DIFERENTE
	NO	SOL
	T	NO
	VER	VER
	SECUENCIAPRINCIPAL	SECUENCIAPRINCIPAL
	<u>BAJOAMASA</u>	<u>BAJOAMASA</u>
	MUY	MUY
	≡	≡
		TODO
		CÚMULOABIERTOAESTRELLA
		<u>ESTAR</u>

El cambio más importante es la aparición en el AG de una ULL específica: *cúmuloabiertoaestrella*. En el AF los conceptos relacionados con *estrella* son los mismos que los encontrados con el fragmento de “punto y seguido”: *cúmulo*, *núcleo*, *cúmulo globular*, *sol*, *T*, *secuencia principal* y *baja masa*. Mientras que en el AG aparece un nuevo concepto: *temperatura superficial*, *cúmulo globular*, *núcleo*, *cúmulo*, *sol*, *secuencia principal*, *baja masa* y *cúmulo abierto de estrellas*.

5.- DETERMINACIÓN DE

ERRORES

20	CUMULO	16	6.656	993	sustantivo
21	POR	16	6.656	1009	funcional
22	TEMPERATURA	15	6.24	1024	sustantivo
23	NÚCLEO	14	5.824	1038	sustantivo
24	MUY	13	5.408	1051	adverbio
25	PRINCIPAL	13	5.408	1064	adjetivo
26	SECUENCIA	13	5.408	1077	sustantivo
27	X	13	5.408	1090	lm(v)
28	APROXIMADAMENTE	12	4.992	1102	adverbio
29	HIDRÓGENO	12	4.992	1114	sustantivo
30	SU	12	4.992	1120	funcional
31	10	11	4.576	1137	lm(adj)
32	COMO	11	4.576	1146	funcional
33	DIAGRAMA	11	4.576	1159	sustantivo
34	GLOBULARES	11	4.576	1170	adjetivo
35	L	11	4.576	1181	lm(s)
36	AÑOS	10	4.16	1191	sustantivo
37	LUMINOSIDAD	10	4.16	1201	sustantivo
38	MASA	10	4.16	1211	sustantivo
39	T	10	4.16	1221	lm(s)

se	llamar	2	1.854	54	0.04
y	el	10	9.268	48	0.21
y	polvo	3	2.780	48	0.06
y	se	3	2.780	48	0.06
y	cúmulo	2	1.854	48	0.04
y	mv	2	1.854	48	0.04
y	por	2	1.854	48	0.04
y	su	2	1.854	48	0.04
y	tener	2	1.854	48	0.04
y	uno	2	1.854	48	0.04
ser	el	8	7.414	41	0.20
ser	uno	8	7.414	41	0.20
ser	muy	4	3.707	41	0.10
ser	de	3	2.780	41	0.07
ser	aproximadamente	2	1.854	41	0.05
ser	estable	2	1.854	41	0.05
ser	por	2	1.854	41	0.05
cúmulo	globular	17	15.755	39	0.44
cúmulo	abiertoa	5	4.634	39	0.13
cúmulo	de	5	4.634	39	0.13
que	el	10	9.268	39	0.26
que	se	3	2.780	39	0.08
que	ser	3	2.780	39	0.08

5.- DETERMINACIÓN DE ERRORES

Los programas informáticos usados durante la investigación (PAFE, One View, Excel 2007) no producen errores por sí mismos. Es el uso que de ellos hace el investigador lo que puede llevar a errores. La metodología aplicada contiene distintos pasos que requieren la intervención del experto-investigador en mayor o menor medida. Los pasos donde más interviene el investigador son la preparación del texto, la clasificación de palabras y la unión de lemas. Es ahí donde se puede generar la mayor fuente de errores. A continuación se detallan las limitaciones y posibles errores en cada uno de los pasos de la metodología.

Las estimaciones de error se han hecho con los resultados de los análisis de entorno producidos con un fragmento de “punto y seguido”. No se espera que difieran si se hacen con un fragmento de “punto y aparte”.

5. 1.- Errores en la Preparación del Texto

En la preparación del texto se pueden encontrar dos fuentes de error: no distinguir distintos significados expresados con la misma grafía o no distinguir

DETERMINACIÓN DE ERRORES

sinónimos. Ambas fuentes de error provienen de la capacidad del experto-investigador de hacer semejantes distinciones.

En el estudio de caso del capítulo 4 se constató la no separación de significados de *E* como energía y *e* como conjunción y de *final* como adjetivo o como sustantivo. El error fue subsanado durante la investigación. Al tratarse de palabras con una frecuencia absoluta baja (3 y 2 respectivamente) no hubiera producido ningún error en el AO dado que todas las ULL con una frecuencia absoluta menor a 6 no formaron parte del sistema óptimo. Sin embargo en los AF y AG sí hubieran formado parte del sistema óptimo puesto que la frecuencia absoluta mínima era 2. No todas las ULL tienen relaciones como para aparecer en la gráfica de las redes semánticas. Sólo un 5 a 6% de las ULL específicas con frecuencia absoluta 2 aparecen en las redes, y sucede lo mismo con un 10 a 11% de las de frecuencia absoluta 3. Esto lleva a un error de un 1 a 3% en la aparición de dicha ULL en la red semántica.

Estas palabras no hubieran formado parte del vocabulario más frecuente en ninguno de los tres tipos de análisis llevados a cabo. Dado el sistema óptimo en los AF y AG, sólo una (*E*) sería específica, dando lugar a un error de un 1% en cuanto al vocabulario específico en dicho sistema.

Ninguna de estas palabras entró a formar parte de las colocaciones.

De las ULL listadas en la Tabla 92 que se relacionan con *estrella* dos tienen $n=3$: *estar* y *bajoamasa* en el AG. Y de éstas sólo la última se considera a la hora de interpretar los resultados desde el punto de vista de la didáctica, con lo que se volvería a incurrir en un error del 1%.

5. 2.- Errores en el Análisis del Texto Preparado

La Tabla 94 muestra los errores producidos en la clasificación de vocablos en categorías gramaticales en las tablas de Excel, que posteriormente se subsanaron en la investigación.

Tabla 94

Errores producidos por el investigador en FESTESE001V

Vocablos	n_i	Categoría gramatical errónea	Categoría gramatical
sí	1	adverbio	funcional
e	3	funcional	lm 1, funcional 2
galaxia	7	funcional	sustantivo
final	2	adjetivo	sustantivo 1, adjetivo 1
sólo	5	funcional	adverbio
mucho	2	adverbio	adjetivo
muchos	2	adverbio	adjetivo

Nota: n_i = frecuencia absoluta del vocablo i ; lm=lenguaje matemático.

Hay en total 22 frecuencias mal clasificadas, ya sea porque no se diferenciaron en significados, como es el caso de *e*, o porque se cometió un error en la categorización gramatical. En el caso de *e*, no se distinguió cuando era el símbolo de energía de cuando funcionaba como conjunción. La mayoría de estos errores ocurrieron en la franja media de frecuencias, no habiendo ninguno en la franja alta. El error producido en la franja media está en torno a un 2% y en la franja baja en un 1%. Sin embargo el error producido en las categorías gramaticales está entre un 1 y un 5%. Cabe destacar que este error se debe a la categorización gramatical hecha por el experto-investigador con el objeto de estudiar la influencia de la metodología en las categorías gramaticales. En una investigación donde se aplica la metodología

DETERMINACIÓN DE ERRORES

sobre corpus no se hacen estudios sobre categorías gramaticales con lo que este error no va a afectar a los resultados.

Para la aplicación PAFE se preparó un archivo con palabras funcionales que se introdujeron en la base *Eliminar*, con lo que *galaxia* no habría sido suprimida.

5. 3.- Errores en el Análisis del Texto Lematizado/Transformado

En este paso cabe esperar que el experto-investigador incurra en errores durante la clasificación en categorías gramaticales de los lemas o ULL, al igual que ocurría en el texto preparado. Para intentar solventar la aparición de estos errores se procedió a copiar la categoría gramatical de las palabras correspondientes del texto preparado. Es decir, los errores cometidos en el texto preparado se propagaron al texto lematizado/transformado. La revisión de la copia de la categoría gramatical del texto preparado permitió detectar los errores señalados en la Tabla 94. Como ya se ha dicho, en un análisis de corpus este error no aparecería.

La lematización implica la unificación de variantes gráficas por motivos gramaticales en un mismo lema. Dado que un 19% de los vocablos eran variantes gráficas de un lema, éste es el error que cabría esperar en un estudio de relaciones entre palabras donde no se lematizara el texto. Por tanto, se puede considerar la lematización como un proceso metodológico obligatorio.

El experto-investigador no detectó más errores de clasificación gramatical en esta fase.

5. 4.- Errores en el Estudio de Entorno

5. 4. 1- Errores en la búsqueda de colocaciones

En el caso de la unión de lemas se necesita que el investigador sea experto en el tema estudiado, pues en caso contrario se puede cometer un error de hasta un 14% (ver 3. 2. 2. 2. 2.). El resto de los pasos se hacen sin la intervención directa del investigador, siendo las distintas aplicaciones informáticas las encargadas de trabajar con los datos.

La primera vez que se buscan colocaciones mediante el criterio del cociente, el investigador-experto añade otras que suponen entre un 9 y un 17% del total de colocaciones finalmente consideradas. De las colocaciones representadas en las redes semánticas un 25%, 33% y 22% de las mismas correspondían a las halladas por el experto en los AO, AF y AG. Una de las colocaciones presentes en las ULL-conceptos relacionados con *estrella* es de este tipo (*cúmulo globular*), estando también en el vocabulario más frecuente. El 100% de las colocaciones son específicas.

La importancia que las colocaciones tienen en los sistemas óptimos y redes semánticas se ha discutido ya (Tablas 86 y 90). En los AF y AG las colocaciones representan un mínimo de un 13% de las ULL de los sistemas óptimos y un 17% de las ULL representadas en las redes semánticas.

DETERMINACIÓN DE ERRORES

Se puede decir que no hallar colocaciones, ni hacer uso de un experto para ello, puede producir un error de hasta un 20% en las redes semánticas y en la interpretación de los conceptos que se relacionan con uno dado.

5. 4. 2.- Errores en el vocabulario más frecuente y el vocabulario específico

En esta tesis se estima por primera vez cuál es el vocabulario más frecuente mediante ajustes a una función de potencias de las frecuencias absolutas, con un coeficiente R cercano a 1. En función del valor de R^2 (apartado 4. 4. 2.) el error del ajuste está entre un 4 (AO y AF) y un 6% (AG).

El método del criterio del cociente para determinar si un vocablo es específico presenta un error del 14% (Chung, 2003). En el estudio de caso no se han revisado dichos resultados desde el punto de vista del experto. Posteriormente se han detectado ULL que deben ser consideradas como específicas (ver T en Tabla 83) así como ULL que debido a la frecuencia de uso en el texto son específicas por el criterio del cociente pero no deben ser consideradas como tales para los objetivos de la investigación.

En las Tablas 92 y 93 hay 6 términos que deberían haber sido revisados lo que supone un error de un 11, 13 y 33% en los AG, AF y AO, respectivamente. Las variaciones observadas se pueden deber a la selección de ULL hecha en dichas tablas, que no es representativa del texto completo. Para evitar este error en la

metodología, un experto debe hacer una revisión de la especificidad en los listados de los sistemas óptimos (Tablas 82, 83 y 84).

5. 4. 3.- Errores la determinación del sistema óptimo y las redes semánticas

El sistema óptimo se ha elegido mediante criterios. En los AF y AG, este sistema ha coincidido con el definido por la franja alta y la franja media, es decir, aquellas ULL con $n \geq 2$. Las ULL que no formaban parte del sistema eran hápax y por lo tanto sólo se relacionan una o dos veces con otra ULL dada. Por tanto, los sistemas óptimos contienen todas las relaciones del texto completo que son de nuestro interés.

Las redes semánticas se hallan a partir de los datos del PAFE, los cuales no tienen errores. Los errores de las redes semánticas provienen de los pasos que previamente conducen hasta las mismas. El mayor error es el de no hallar colocaciones (hasta un 20%), seguido por el de la especificidad de las ULL que Chung (2003) fija en un 14%. La señalización de la especificidad y conservación de relaciones en las redes facilita una mejor interpretación de las mismas.

6.- CONCLUSIONES



6.- CONCLUSIONES

6. 1.- Preparación del Texto

- Del texto original un 2% de las palabras se encuentran en los títulos y un 12% se encontraron en los pies de figura descriptivos. En total se eliminó un 14% de las palabras en la preparación para su uso en la aplicación PAFE.
- Un 7% de los vocablos que aparecen en el texto preparado sufrieron una modificación previa al análisis con el PAFE. La mayoría de las modificaciones fueron debidas a una posible confusión de funciones gramaticales o semánticas (un 82% en total). Dentro de éstas la mayor parte se deben a distinguir una palabra de una forma verbal (un 71%). Un 9% de los adjetivos provienen de participios de verbos. Un 18% de las modificaciones se deben a caracteres no admitidos por la aplicación PAFE. Estos caracteres representan el 15% de los símbolos matemáticos del texto.
- El verbo *ser* sólo es usado como auxiliar un 5% de las veces. La señalización de sus formas auxiliares no es necesaria en la preparación del texto puesto que los resultados en cuanto a vocabularios frecuentes, específicos e interpretación de redes no se verían modificados.
- Dentro de las palabras funcionales los verbos auxiliares son poco importantes representando un 2% del total de las palabras; se encuentran en las franjas

CONCLUSIONES

media y baja. Tras la lematización hay más verbos auxiliares en la franja baja que en la franja media. Un error en la señalización de los verbos auxiliares no tiene importancia en los resultados de vocabulario más frecuente, específico e interpretación de redes. Por tanto, la señalización de los verbos auxiliares no es necesaria en la preparación del texto.

- Una de cada diez palabras léxicas es una palabra del lenguaje matemático. No es correcto omitir el uso del lenguaje matemático en el estudio y análisis de vocablos, colocaciones y estudios de entorno. Es decir, el uso del lenguaje matemático en el lenguaje de las ciencias (en particular en astronomía) debe ser considerado a la hora de estudiar los conceptos y las construcciones lingüísticas de los textos escolares. Esto ha sido previamente confirmado por otros estudios preliminares en otros ámbitos de la ciencia cercanos a la Astronomía, como es la Física (Lemke, 1998).
- La unificación de grafías sólo es importante para lemas situados en la franja alta o en la parte superior de la franja media. En el estudio de caso la importancia de la unificación de grafías fue mayor para los conceptos etiquetados de dos formas diferentes, con palabras y símbolos matemáticos, influyendo en las redes semánticas en los rangos superiores de frecuencias de las ULL representadas. La unificación de grafías, como paso de la metodología, sólo es necesaria para las palabras situadas en las franjas alta y media.
- La separación de significados que tienen la misma grafía requiere la manipulación de un experto.

6. 2.- Análisis de Frecuencias del Texto Preparado y Lematizado

- Un 19% de los vocablos del texto preparado eran variantes gráficas de un mismo lema, ya sea en número, género o conjugación verbal. Es decir, la no lematización del texto produciría un error de hasta un 20% en el estudio de relaciones y frecuencias.
- En el texto preparado un 10 % de palabras (funcionales) producen un 43% de las frecuencias, mientras que un 90% (léxicas) producen el restante 57%. Estos resultados indican la gran importancia que tienen las palabras funcionales en la construcción del texto. Las palabras funcionales, vacías de contenido semántico, tienen pues una gran importancia a la hora de comprender un texto de ciencias. Será difícil abordar la comprensión de un texto sin el conocimiento previo de la lengua en la que se expresa.
- Las palabras funcionales predominan en la franja alta debido al alto número de frecuencias que produce un pequeño número de las mismas; las palabras léxicas muestran una tendencia a concentrarse en las franjas media y baja tanto en número como en frecuencias. Por tanto en el estudio de entorno se obtienen sistemas óptimos y redes semánticas dominados por palabras funcionales. Dado que el objetivo de la investigación son las palabras léxicas y sus relaciones, se hace necesaria la supresión de las palabras funcionales mediante la base de datos *Eliminar*.
- En los AF y AG, donde se han eliminado las palabras funcionales y verbos auxiliares, los lemas léxicos presentan más frecuencias en la franja media tanto en los textos preparados como en los transformados. En el AO, cuando las

CONCLUSIONES

palabras léxicas se toman en el contexto gramatical, estas producen más frecuencias en la franja baja. La franja donde se concentran mayor número de vocablos/lemas léxicos es la franja baja en todos los análisis y prácticamente para todas las subcategorías léxicas, excepto el lenguaje matemático que tiene el máximo en la franja media.

- El número de palabras y frecuencias no está correlacionado entre palabras funcionales y palabras léxicas, es decir, un pequeño número de palabras funcionales produce más frecuencias que un gran número de palabras léxicas. En el AO, en el texto preparado, dentro de la categoría léxica, las distintas subcategorías no siguen la tendencia de la misma en cuanto a número de palabras y frecuencias. La categoría léxica está dominada en cuanto a número de palabras y frecuencias por sustantivos y adjetivos, los cuales sí parecen estar correlacionados con el comportamiento general de la categoría. La lematización hace que los verbos sigan también la tendencia general de la categoría.
- En los tres análisis más de la mitad de los vocablos de las subcategorías léxicas son hápax tras la lematización, es decir, se encuentran en la franja baja.
- Desde el punto de vista teórico la lematización debería influir más en los verbos, ya que debido a la conjugación presentan un mayor número de grafías distintas que otras subcategorías. Sin embargo el mayor número de vocablos de sustantivos hace que sean éstos a los que la lematización afecta más.
- La eliminación de las palabras funcionales en los AF y AG provoca una redistribución de los lemas léxicos, de manera que todas las subcategorías léxicas, con la excepción del lenguaje matemático en el AG, se hallan representadas en la franja alta, estando ésta dominada por los sustantivos en cuanto al número de lemas y frecuencias.

- La categoría léxica dominante es la de sustantivos, lo cual es coherente con la alta nominalización de un texto de ciencias.
- El texto lematizado del AF presenta menos repeticiones que cuando las palabras funcionales están incluidas, y por tanto tiene menor número de relaciones. No hay apenas diferencia en los cocientes tipo/caso de los textos preparados, es decir, la riqueza léxica del texto donde se han unificado las grafías es prácticamente la misma. El cociente tipo/caso del caso de estudio está más cercano a los valores de un texto científico que a otro tipo de textos.
- La densidad léxica del estudio de caso es elevada (un 57%) y prácticamente no cambia al unificar grafías, es decir, la cantidad de información no se ve afectada por la unificación de grafías. La eliminación de las palabras funcionales y verbos auxiliares produce un texto transformado que contiene toda la información del texto preparado. La densidad léxica está más cercana a los valores de textos de divulgación que a otro tipo de textos para el caso de estudio. La mayor parte de la información se debe a los sustantivos.
- El texto educativo del estudio de caso, propuesto por una institución científica, muestra características de la riqueza léxica usada por los científicos y de la cantidad de información usada por los mismos cuando se divulga, pero no tiene valores cercanos a los textos educativos.
- Para elaborar un corpus habrá que tener en cuenta la autoría de los textos, el cociente tipo/caso y la densidad léxica, como indicadores de la semejanza de los textos entre sí.
- El análisis por categoría gramatical del estudio de caso, que contiene más de 1000 palabras, es suficiente para hacerse una idea de cómo afecta la

CONCLUSIONES

metodología a los resultados finales. Este análisis no es necesario a la hora de obtener vocabularios frecuentes, específicos ni redes semánticas en un corpus.

6. 3.- Estudio de Entorno

6. 3. 1.- Búsqueda de colocaciones

- El criterio del cociente $ra_{ij}/n_i \geq 0.5$ para hallar colocaciones resultó efectivo por encima de un 85% de los casos en los tres tipos de análisis. Esto quiere decir que de no intervenir un experto se obtendría un 15% de error en la estimación de las colocaciones. Parte de las colocaciones halladas por el experto se representan en las redes semánticas, constituyendo un 17% de las colocaciones representadas y hasta un 20% de las encontradas en las ULL-conceptos relacionadas con *estrella* en el estudio de caso. Por tanto es necesaria la intervención de un experto en este paso de la metodología.
- La categoría léxica que más forma parte de las colocaciones es la de sustantivos. Y la que más se coloca consigo misma es la del lenguaje matemático. El estudio de las colocaciones por categorías gramaticales no es necesario en la metodología a aplicar en un corpus.
- Unir lemas en forma de colocaciones no cambia las características generales de los textos, en cuanto a distribuciones de ULL y frecuencias. Pero sí influye en la

representación gráfica de las redes semánticas e interpretación de los resultados.

Por tanto, la búsqueda de colocaciones es un paso necesario en la metodología.

6. 3. 2.- Vocabulario más frecuente y vocabulario específico

- Para determinar el vocabulario más frecuente se debe usar el análisis del ajuste de la distribución de frecuencias absolutas. Es decir, la determinación de la pendiente con valor igual a -1 determina la franja de frecuencias absolutas límite de la franja de vocabulario más frecuente.
- La alta frecuencia de las palabras funcionales, así como su alta capacidad de relación, hace que los vocabularios de palabras más frecuentes se llenen de esta categoría gramatical y que las redes semánticas de sistemas con vocabulario frecuente no reflejen relaciones con contenido léxico. Por esta razón la metodología debe incluir la eliminación de dichas palabras introduciéndolas en la base *Eliminar*.
- El criterio del cociente de frecuencias de una ULL entre un corpus específico y un corpus general (Chung, 2003) no es suficiente para distinguir qué ULL son específicas. Se necesita la intervención de un experto para subsanar los errores en los resultados obtenidos aplicando dicho criterio.

6. 3. 3.- Determinación del sistema óptimo

- El sistema óptimo se define estudiando la conservación de relaciones total y de cada ULL.
- Cuando el porcentaje del cociente $rat_i / n_i * 2$ de una ULL i sea igual o mayor que 80% se considerará que dicha ULL ha conservado bien sus relaciones. Debido a la presencia de puntos y seguidos y puntos y aparte se produce una desviación del número de relaciones ideal que puede tener una ULL dada. Por tanto es necesario calcular dicha desviación para cada corpus/texto o tipo de análisis.
- En el caso de estudio, el sistema óptimo de los tres análisis quedó siempre definido por el límite inferior de la franja media.
- La especificidad y la conservación de relaciones del sistema óptimo son mucho mayores en los AF y AG que en el AO. La conservación de relaciones es ligeramente mayor en el AG que en el AF por lo que es más adecuado realizar la investigación unificando grafías en un corpus o texto.
- Los datos de la conservación de relaciones de las ULL aportan información para interpretar las redes semánticas.

6. 3. 4.- Redes semánticas

- Los AF y AG proporcionan vocabularios frecuentes, específicos y redes semánticas ricas en contenido léxico, que permiten observar la importancia relativa y las relaciones entre los lemas léxicos. El AO produce redes

semánticas dominadas por palabras funcionales. Debido a los objetivos que persigue la investigación es necesario la eliminación de palabras funcionales mediante el uso de la base *Eliminar* en el *PAFE*.

- Aunque las redes semánticas del AF y el AG son casi iguales, es conveniente unificar grafías en las franjas media y alta, dado que lo que se busca son las relaciones entre significados (conceptos).
- El análisis con fragmento de “punto y aparte” proporciona modificaciones hacia rangos de relaciones superiores de algunas de las relaciones pre-existentes en el análisis con fragmento de “punto y seguido”. Teniendo en cuenta que un párrafo es un conjunto de oraciones unidas para formar una idea, parece natural escoger este tipo de análisis para estudiar las relaciones entre ULL.
- A la hora de interpretar las redes semánticas es necesaria la representación de las relaciones respecto a una ULL origen, la señalización de la especificidad y de la conservación de relaciones de cada ULL que aparece en la gráfica, así como la intervención de un experto.
- Para el estudio de caso, las ULL-conceptos relacionados con *estrella* son: *temperatura superficial, cúmulo globular, núcleo, cúmulo, sol, secuencia principal, baja masa y cúmulo abierto de estrellas*. Todas estas ULL-conceptos son específicas.

6. 3. 5.- Análisis de errores

- La mayor fuente de error en la metodología se debe a la necesidad de que intervenga un experto en distintos pasos: separación de significados, hallar colocaciones e interpretación de los resultados.
- Los errores provenientes de no distinguir verbos auxiliares son despreciables.
- Los errores de clasificación de palabras introducidos por el propio experto se generaron en la franja baja y media. Estos errores no influyen en los resultados en cuanto a la determinación de vocabularios frecuentes y específicos, sistema óptimo y redes semánticas.
- Previamente al proceso de clasificación, es conveniente introducir en la base *Eliminar* todas las palabras funcionales y aquellas que se deseen eliminar.
- Deben unificarse grafías en la franja alta y media, pues de no hacerse se producen variaciones importantes en las frecuencias de los significados, vocabularios frecuentes y redes semánticas.
- El método de hallar colocaciones de forma automatizada produce un error de hasta un 15% de las mismas. De ahí que sea necesaria la intervención de un experto para subsanar dicho error.

6. 4.- Esquema de la Nueva Metodología

Las conclusiones de la investigación sobre el estudio de caso implican cambios en la metodología a usar en un análisis de corpus. Estos cambios se muestran en las Figuras 103 y 104.

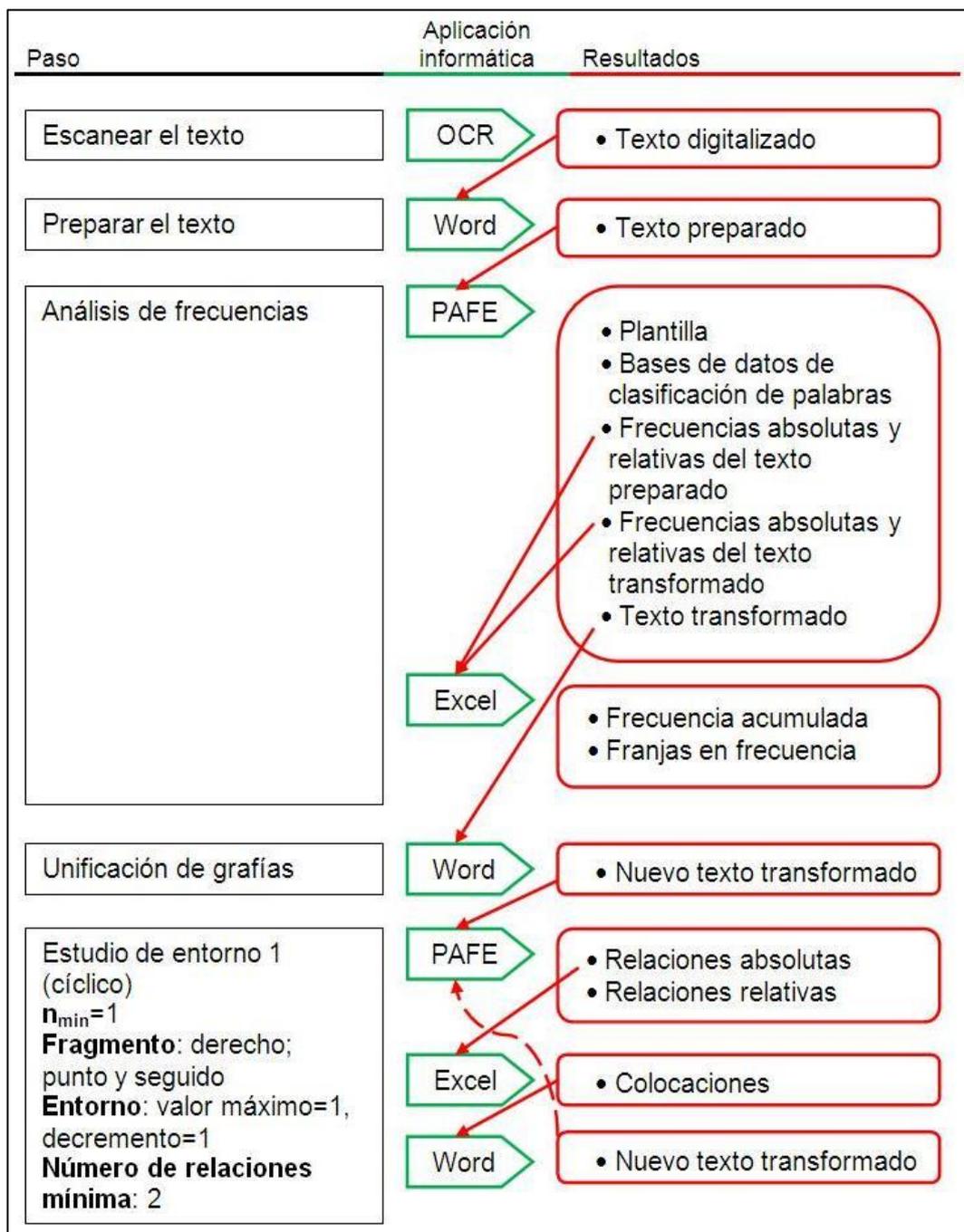


Figura 103.- Primera parte del esquema de la metodología.

CONCLUSIONES



Figura 104.- Segunda parte del esquema de la metodología.

En la preparación del texto se puede obviar el señalar los verbos auxiliares, pues el error cometido es mínimo. La unificación de grafías se hará tras el primer análisis del PAFE. Para ello se buscan sinónimos que tengan un número elevado de

apariciones en las franjas alta y la franja media. Sólo se unificarán las grafías de dichos sinónimos. El lenguaje matemático lleva hasta un 10% de la información léxica e influye en los resultados al unificar grafías. Por tanto, el lenguaje matemático debe formar parte del texto analizado.

Se usará la base *Eliminar* donde se clasificarán las palabras funcionales. Se ha obviado el análisis de las categorías gramaticales, puesto que no van a aportar nada a la hora de determinar el vocabulario más frecuente, el sistema óptimo ni el análisis de relaciones. Es necesario el estudio de la conservación de las relaciones para cada ULL para la interpretación de las redes semánticas.

En la primera parte del esquema de la metodología (ver *Figura 103*) se parte del texto original y se llega finalmente a un texto transformado que contiene todas las colocaciones halladas. En la segunda parte (*Figura 104*) se determinan el vocabulario más frecuente, el vocabulario específico y el sistema óptimo de ULL; se analizan las relaciones dentro del fragmento del “punto y aparte” y se interpretan los resultados de las redes semánticas.

7.- CUESTIONES ABIERTAS



7.- CUESTIONES ABIERTAS

7. 1.- Preparación del Texto

1. Hacer un estudio sobre la importancia de los títulos que se eliminan en el caso particular de los niveles escolares más bajos. En los primeros niveles escolares la información a estudiar se reduce a una oración.
2. ¿Qué influencia tendría sobre los resultados el no señalar los participios que actúan como adjetivos?
3. ¿La influencia de la unificación de grafías en un texto es la misma que se produce en un corpus?
4. Automatizar el hallazgo de sinónimos. ¿Qué influencia tendría el separar las oraciones coordinadas y subordinadas para no omitir sujetos? ¿Qué importancia tienen las oraciones subordinadas en cuanto a la cantidad de sinónimos?

7. 2.- Análisis de Frecuencias del Texto Preparado y Lematizado

5. Hacer un estudio sobre la importancia del lenguaje matemático en los corpus educativos, divulgativo y académico.
6. Además de la eliminación de las palabras funcionales, se podría considerar eliminar locuciones en las que intervienen palabras léxicas que pierden su significado original como *a medida que*, *a continuación*, etc. y de ciertos adverbios como *ahí*, *generalmente*, etc. Muchas de estas palabras y locuciones han sido detectadas en otros estudios no publicados. Un listado de palabras y locuciones a eliminar donde también están incluidas las palabras funcionales está en el anexo Q. Estas locuciones pueden parecer en principio de poca importancia a la hora de obtener frecuencias; sin embargo al permanecer la parte léxica en los textos transformados influyen en el entorno entre lemas, es decir, la locución queda incluida en el estudio de entornos.
7. Otro tipo de palabras a eliminar son los adverbios y los verbos, estos últimos haciendo el estudio sin la base *Verbos*. En el estudio de caso estas tipos de palabras aparecen en los vocabularios frecuentes, vocabularios específicos y las redes semánticas, pero no han sido interpretadas por el experto como relevantes en torno al concepto *estrella*. ¿Hasta qué punto son importantes a la hora de aportar datos para la Enseñanza de las Ciencias? Existen verbos específicos que sí interesarían en la investigación, pero ¿vale la pena considerarlos?
8. Si aproximadamente un 75% de un texto está formado por palabras que sólo aparecen una (hápax) o dos veces ¿qué influencia tiene la riqueza de vocabulario en la comprensión lectora? Dado que en el análisis se obtienen los

datos de esta parte del vocabulario, convendría indagar en la bibliografía estudios sobre esta cuestión para enriquecer la interpretación de los resultados del PAFE.

9. Existen otro tipo de medidas procedentes del análisis de contenido para caracterizar los textos que se podrían obtener de nuestros datos (Calderero, 2003):
 - **Longitud de frases:** la media de número de palabras por frase, considerándose como legible una media de 15 palabras por frase (Bardin, 1986) citado en Calderero (2003).
 - **Coefficiente de lexicalidad:** formas plenas de significado divididas por las formas vacías o gramaticales. Calderero (2003) toma la clasificación de Bardin (1986, citado en Calderero, 2003) donde se consideran como palabras con significado a sustantivos, adjetivos y verbos. Las palabras vacías serían el resto de categorías gramaticales.
 - **Índice de conceptualización:** número de términos específicos de un tema concreto que hay en un texto dividido entre el número de términos específicos que se encuentran en todos los textos.
 - **Densidad conceptual:** definido como el número de términos específicos que aparecen en el texto dividido por el total de palabras del mismo.
10. La densidad léxica indica la cantidad de información de un texto. En el caso de que la investigación requiriera clasificar palabras no funcionales en la base de datos *Eliminar*, la densidad léxica sería un parámetro a tener en cuenta para medir la pérdida de información y ver hasta qué punto es importante. Otro indicador sería el cociente tipo/caso en cuanto a la riqueza léxica.

CUESTIONES ABIERTAS

11. ¿Es correcto considerar a las palabras unidas por guiones como ULL? Es decir, si no se las considerara ULL ¿se encontraría a posteriori que sí lo son?
12. Detrás de distintas categorías gramaticales (por ejemplo brillo, brillante, brillar) puede estar el mismo concepto, lo cual explicaría cómo los niños son capaces de usar la categoría gramatical para interpretar el significado de una palabra (Hall & Bélanger, 2005). O expresado de otra manera, el concepto cambia de palabra según sea el contexto gramatical. ¿Cómo cambian los resultados de la metodología planteada en la tesis si además todas las categorías gramaticales se reducen a una sola (concepto)?

7. 3.- Estudio de Entorno

7. 3. 1.- Búsqueda de colocaciones

13. ¿Se puede mejorar cómo hallar colocaciones para que no se requiera la intervención del experto?
14. En el caso de estudio se hallaron todas las colocaciones posibles. ¿Qué error se cometería si sólo se aplica la metodología de hallar colocaciones una vez?

7.3.2.- Vocabulario más frecuente y vocabulario específico

15. Las estimaciones tradicionales de vocabularios frecuentes se hacen en función de un número determinado de palabras (Juilland & Chang-Rodríguez, 1964), que relamente no es representativo de un corpus o texto. ¿Se podría determinar un método mejor que el propuesto en esta tesis? ¿Depende del tamaño del texto o corpus? ¿Se puede determinar a priori como una franja determinada por un porcentaje?
16. Encontrar un método para obtener el vocabulario específico con un error menor del 20%.
17. Estudiar la influencia del investigador-experto en la determinación del vocabulario específico.
18. Dado un vocabulario frecuente donde hay lemas no específicos, se podría considerar la posibilidad de volver a hacer el análisis para que en el texto transformado no aparecieran estos lemas. Suponiendo que estuviera libre de ellos habría que ver cómo influye en las redes semánticas. Por ejemplo si estrella se relaciona con más y más con cúmulo, pero cúmulo no estaba relacionado con estrella: ¿aparecerá ahora una relación entre estrella y cúmulo al haber suprimido más?

7. 3. 3.- Determinación del sistema óptimo

19. Confirmar que las distribuciones del número de relaciones y número de ULL de los mismos sistemas en distintos entornos son análogas en corpus. ¿Hasta qué punto el análisis con dos entornos diferentes produce resultados distintos? ¿Qué influencia tiene la posición de una ULL j en torno a otro origen i cuando los entornos son distintos? Algunas de estas preguntas han sido consideradas en estudios no publicados del GICEC.
20. Estudiar las diferencias en las redes semánticas en corpus al realizar el análisis usando como fragmento “punto y seguido” o “punto y aparte”.
21. El caso de estudio dio como resultado para el sistema óptimo el definido por el límite inferior de la franja media. ¿Ocurre lo mismo para cualquier corpus?
22. Dado el sistema óptimo en un corpus ¿cuál sería el valor de relación mínima a usar para poder visualizar las relaciones más importantes?

7. 3. 4.- Redes semánticas

23. Buscar aplicaciones informáticas que grafiquen las redes semánticas de forma automática, de acuerdo con las necesidades de la investigación.
24. Estudiar la influencia en las relaciones de las redes semánticas del vocabulario más frecuente que no es específico dentro de un corpus.

7. 4.- Otros

25. El texto que se ha escogido, aunque escrito para alumnos de secundaria, ha sido escrito desde una institución que se dedica a la investigación científica. Esto podría explicar que su cociente tipo/caso sea más cercano al de los textos científicos que al de textos dedicados a la educación. ¿Hasta qué punto los materiales escolares desarrollados desde las instituciones científicas son adecuados como materiales que se pueden usar en el aula? ¿Se puede usar la metodología como evaluación de los textos desarrollados desde otro tipo de instituciones que no son las meramente educativas?
26. Usar la metodología en estudios históricos sobre la formación y evolución de un concepto.
27. Rediseñar el programa PAFE para obtener más rápidamente los resultados.
28. De acuerdo con Weist, Lyytinen, Wysocka, & Atanassova (1997), en experiencias con niños, cuando el desarrollo del conocimiento se basa en los conceptos, apenas se encuentra influencia del lenguaje sobre el mismo. En el caso de los textos escritos por un mismo experto para públicos distintos (educativo, divulgativo, científico): ¿influye el lenguaje en los resultados de vocabularios frecuentes, específicos y redes semánticas? ¿Se obtienen resultados análogos usando el mismo texto escrito en dos idiomas distintos?
29. Diseño, compilación, caracterización y representatividad de corpus científicos, divulgativos y educativos en los que aplicar la metodología.
30. Aplicación de la metodología en corpus científicos, divulgativos y educativos con el objetivo de hallar vocabularios más frecuentes y específicos, sistemas óptimos y redes semánticas. La comparación de los resultados proporciona

CUESTIONES ABIERTAS

datos sobre cómo se conforma un concepto. ¿Con los datos proporcionados por los corpus educativos y divulgativos se puede determinar si los libros de texto escolares responden a las necesidades de conocimiento que debe tener un ciudadano medio? ¿Con los datos proporcionados por los corpus educativos y científicos se puede determinar si los libros de texto responden a las necesidades de formación de un profesional de ciencias?

31. A partir de los resultados obtenidos en las comparaciones de los corpus educativo, divulgativo y científico, investigar cuál es la mejor forma de enseñar un concepto dado. Para ello habrá que comparar distintos modelos de enseñanza, tener en cuenta las dificultades de aprendizaje del concepto, los errores conceptuales estudiados en torno al concepto, los medios disponibles para el profesor y las condiciones psicoevolutivas del alumnado. Diseñar en caso de que no los hubiera recursos de enseñanza de un concepto dado.

GLOSARIO

Vocabulario más frecuente y específico del texto transformado definitivo

OL	ULL	f_i a1000)	f_{TEXT} (a 1000)	Cociente (f_i/f_{TEXT})
		AO		
1	EL	120.949	104.538	1.157
2	DE	87.175	64.994	1.341
3	ESTRELLA	33.775	0.097	348.663
4	UNO	32.862	21.258	1.546
5	EN	29.21	27.501	1.062
6	SE	24.646	13.115	1.879
7	Y	20.995	28.073	0.748
8	SER	16.887	11.366	1.486
9	QUE	16.887	30.341	0.557
10	DEL	12.78	12.057	1.060
11	ESTO	11.867	4.964	2.391
12	A	10.497	21.561	0.487
13	SU	8.072	7.162	1.211
		AF		
1	ESTRELLA	50.213	0.097	518.355
2	SER	29.787	11.366	2.621
3	CÚMULO	14.468	0.008	1758.951
4	CÚMULO GLOBULAR	14.468	0.009	33819.838
5	NÚCLEO	13.617	0.045	303.586
6	MÁS	12.766	4.291	2.975
7	MUY	11.064	1.351	8.191
8	SECUENCIA PRINCIPAL	11.064	0.000	68277.736
		AG		
1	ESTRELLA	53.938	0.099	542.697
2	SER	31.078	12.382	2.518
3	CÚMULO GLOBULAR	19.692	0.000	46011.450
4	LUMINOSIDAD	13.699	0.006	2233.149
5	NÚCLEO	13.699	0.046	298.637
6	TEMPERATURA SUPERFICIAL	12.842	0.000	30006.045
7	MÁS	11.986	4.550	2.634
8	MUY	11.113	1.531	7.271
9	SECUENCIA PRINCIPAL	11.113	0.000	68655.477

GLOSARIO

AF: Análisis sin palabras funcionales, en el que se clasifican todas las palabras funcionales y verbos auxiliares en la base de datos *Eliminar*. El texto de partida se denomina FESTESE001V-F.

AG: Análisis con unificación de grafías, en el que la preparación del texto incluye la unificación de sinónimos y se explicitan las sustituciones de palabras. Además se clasifican todas las palabras funcionales y verbos auxiliares en la base de datos *Eliminar*. El texto de partida se denomina FESTESE001V-S.

AO: Análisis ortodoxo, en el que no se ha clasificado ninguna palabra en la base de datos *Eliminar*. El texto de partida se denomina FESTESE001V.

Cociente tipo/caso (TTR): Cociente entre el número de vocablos (tipos) y el número de palabras (casos) $TTR=V/N$, siendo V los vocablos y N las palabras. Cuando el cociente tipo/caso es 1 indica que no se repite ninguna palabra, mientras que cuanto más cerca está de 0 indica que se repiten más palabras. Es también una medida de la variabilidad léxica. Es decir, da idea del dominio del lenguaje que se debe tener a la hora de entender un texto y

GLOSARIO

por tanto de la dificultad de lectura que puede tener (Baker, Hardie, & McEnery, 2006).

Colocación: ULL formada por dos o más palabras que con frecuencia aparecen juntas (Álvarez Cavanillas & Chacón Beltrán, 2003). En esta tesis se consideran también aquellas conocidas a priori por el experto. Estas colocaciones producen un significado distinto o más concreto que las palabras separadas. Por ejemplo: *gigante roja* se refiere a un tipo de estrella concreto, que se distingue de una estrella *gigante* o de una estrella *roja*.

Corpus: Colección de textos almacenados de forma digital, que contienen miles o millones de palabras (Baker, Hardie, & McEnery, 2006). Un corpus es una colección de trozos del lenguaje, seleccionados de acuerdo a criterios externos para representar, tanto como se pueda, el lenguaje como fuente de datos en una investigación lingüística (Sinclair, 2005).

Densidad léxica (LD): Porcentaje de palabras léxicas respecto al total de palabras (Ure, 1971; Johansson, 2008). La densidad léxica da idea de la cantidad de información que contiene un texto, ya que cuantifica el empaquetamiento de las palabras léxicas en la estructura gramatical. En esta tesis se usará el siguiente cociente para calcular la densidad léxica: $LD = N_{\text{léxicas}}/N$, siendo $N_{\text{léxicas}}$ el número de palabras léxicas y N el número de palabras totales. La densidad léxica se representa en porcentaje, por lo que el resultado del cociente se multiplica por 100. Los textos hablados tienen una densidad

léxica de menos del 40%, variando entre un 24% y un 43%. Los textos escritos tienen una densidad léxica mayor del 40%, variando entre un 36% hasta un 57%.

Entorno: El entorno de una ULL dada está integrado por n ULL a cada lado de la misma, pudiéndosele atribuir a estas últimas un valor diferente de acuerdo a su proximidad.

Extensión del entorno: Está determinada por la unidad de contexto y la simetría del entorno respecto a una ULL origen. La unidad de contexto se refiere a si el entorno acaba en un punto y seguido o en un punto y aparte. La finalización del entorno en un punto y seguido, donde la unidad de contexto es la frase, permite estudiar las relaciones entre ULL que suelen aparecer juntas, como es hallar colocaciones en frases (ver apartado 3. 2. 2. 2.). Mientras que la finalización del entorno en un punto y aparte, donde la unidad de contexto es el párrafo, permite hallar relaciones entre ULL en frases contiguas dentro de un mismo párrafo. En el PAFE la extensión de entorno se determina en el *Fragmento*. En esta tesis se determina una extensión marcando hacia la derecha y punto y seguido para la determinación de las colocaciones; y una extensión simétrica con punto y seguido o punto y aparte para estudiar las relaciones entre ULL.

Forma: Es el conjunto de sonidos o letras que forman una palabra (Martínez Celdrán, 2002).

GLOSARIO

Frecuencia absoluta: Número de veces que una palabra, lema o ULL aparece u ocurre en un texto. Para una palabra con orden lexicométrico i la frecuencia absoluta es n_i .

Frecuencia acumulada: Suma de las frecuencias absolutas por orden lexicométrico. Para una palabra con orden lexicométrico i la frecuencia acumulada es $N_i = n_i + \sum n_{i-1}$.

Frecuencia relativa: Frecuencia absoluta dividida por el número total de palabras, lemas o ULL de un texto. En esta tesis la frecuencia relativa se normaliza a la de un texto de 1000 palabras. Para una palabra con orden lexicométrico i la frecuencia relativa es $f_i = n_i * 1000 / \sum n_i$.

Género: Clasificación del uso del lenguaje para determinados propósitos (Martin, 1984 citado en Eggins, 2002). Describe la influencia que el contexto de la cultura tiene sobre el lenguaje.

Hápax (hápax legomena): Palabra que sólo aparece una vez en el texto o corpus (Baker, Hardie, & McEnery, 2006).

Índice de frecuencias: “Listado de palabras al que se le adjunta una determinada frecuencia de aparición en un corpus considerado” (Ávila Muñoz, 1999 p.62).

LD: ver Densidad léxica.

Lema: Representación mediante un vocablo de la “reunión de los vocablos que cuentan con una misma raíz y con un significado equivalente para los fines del recuento” (Jordá Lliteras & Pàmies, 1996, p. 154, citado en López Pérez 2008). En esta tesis el vocablo elegido es la palabra en singular, y en masculino si existiera esa forma. El lema de *estrella* y *estrellas* es *estrella*, mientras que el lema de *roja*, *rojas*, *rojo* y *rojos* es *rojo*. Las formas conjugadas de los verbos regulares e irregulares se reunirán bajo un mismo lema: su infinitivo. Es decir, el lema de *es*, *será* y *fue* es *ser*.

Lematización: “Proceso mediante el cual representamos bajo una forma léxica única todas las posibilidades paradigmáticas de un término” (Ávila Muñoz, 1999, p. 56).

Lenguaje matemático: Signos matemáticos, incluidos los números, y símbolos de magnitudes y de sistemas de unidades de medidas.

Léxico: Lista de palabras (Baker, Hardie, & McEnery, 2006).

Lingüística computacional: Campo de la lingüística que estudia científicamente el lenguaje mediante ordenadores (Baker, Hardie, & McEnery, 2006).

GLOSARIO

Número de relaciones: Número de relaciones con un valor de relación dado. Es decir, dado el valor de relación 6 que aparece sólo entre dos ULL *i* y *j* el número de relaciones de valor 6 será 1. Si hubiera 3 parejas distintas de ULL que tuvieran el valor de relación 6 el número de relaciones es 3. En esta tesis, debido a que los listados provenientes del PAFE duplican la información relativa a las parejas de ULL, puesto que cada una de ellas aparece como ULL origen, el número de relaciones aparece multiplicado siempre por 2.

Orden lexicométrico: Las palabras, vocablos o ULL se ordenan “según frecuencias y recurriendo al orden alfabético en los casos de igual frecuencia” (Jordá Lliteras & Pàmies, 1996, p. 155 citado en López Pérez 2008).

PAFE: Programa informático de Análisis de Frecuencias y Estudio de Entornos desarrollado por el grupo GICEC para el entorno Windows (Ceballos, et al., 2000; Pérez García & Pérez Ceballos, 1997).

Palabra: “Segmento del discurso unificado habitualmente por el acento, el significado y pausas potenciales inicial y final” (Diccionario de la RAE online, 2001). Es decir, *estrella* y *estrellas* serían dos palabras distintas.

Palabras funcionales: Palabras que carecen de contenido descriptivo y actúan como instrumentos (relación, determinantes, cuantificadores,...) (Morera, 2007; Escandell Vidal & Leonetti, 2000).

Palabras léxicas: Palabras que poseen significación denotativa o conceptual (Morera, 2007; Escandell Vidal & Leonetti, 2000).

Red semántica: Grafo donde los nodos están ocupados por ULL y las aristas indican las relaciones entre las mismas. La caracterización de los nodos da una idea de la frecuencia de las ULL y la de las aristas indica el número de conexiones o relaciones entre las ULL. En esta tesis, dicha caracterización se hace en función de cuatro colores: amarillo, azul, rojo y verde. Estos indican intervalos de frecuencias y relaciones relativas que aumentan del amarillo hacia el verde.

Registro: Clasificación del lenguaje que describe la influencia que tiene el contexto inmediato de un acto verbal sobre el lenguaje utilizado. Se define con tres dimensiones: campo (sobre lo que se trata al hacer uso del lenguaje), modo (oficio que el lenguaje lleva a cabo en la interacción) y tenor (papel de las relaciones entre los interlocutores) (Halliday (1978, 1985b) citado en Eggins 2002).

Texto escaneado: Texto original que se ha escaneado usando un programa de reconocimiento óptico de caracteres (OCR) que transforma el formato de imagen del escaneado a formato de texto. En esta tesis el texto se revisa de modo que sea fiel al original y se guarda en los formatos *.doc y *.txt.

GLOSARIO

Texto lematizado: Texto en el que se han eliminado todas las variaciones gramaticales de un mismo significado. Es decir, todos los sustantivos y adjetivos están en masculino singular, y todas las formas conjugadas de los verbos están en infinitivo. Si en el texto preparado no se han unido palabras (colocaciones) y con la aplicación informática PAFE no se clasifican palabras en la base *Eliminar*, el texto transformado resultante es sólo un texto lematizado.

Texto original: Texto en formato impreso o digital que se va a analizar en la investigación de esta tesis.

Texto preparado: Texto escaneado o digital en el que se han eliminado títulos y pies de figuras descriptivos, se han distinguido grafías que eran iguales para distintos conceptos, se ha transformado la simbología matemática y se ha adaptado el texto para ser procesado por la aplicación informática PAFE. Pueden aparecer además palabras unidas (colocaciones) y unificarse grafías que tienen un mismo significado. Este texto se encuentra en formato digital con las extensiones *.doc y *.txt con saltos de líneas. Este último formato es el que utiliza la aplicación informática PAFE.

Texto transformado: Texto lematizado del que se han eliminado palabras del texto preparado y en el que también pueden aparecer palabras unidas (colocaciones). Es el texto que resulta tras usar la aplicación informática PAFE clasificando palabras del texto preparado en las distintas bases.

TTR: ver Cociente tipo/caso.

Unidad de contexto: Grupos de una o más frases dentro del cual se hace el análisis textual (Brier & Hopp, 2011). En esta tesis se usan dos unidades de contexto: frase y párrafo.

UL: ver Unidad léxica.

ULL: Unidad léxica lematizada.

Unidad léxica: Palabra/lema o conjunto de palabras/lemas con un significado léxico. Por ejemplo *gigante roja*.

Valor de relación: Número de relaciones absolutas entre dos ULL. Dadas dos ULL *i* y *j* que aparecen juntas 6 veces en un análisis de entorno 1, su valor de relación es 6.

Valores de entorno: Indican cuántas ULL se van a considerar alrededor de la ULL a estudiar (*Valor máximo*), así como el valor que se le va a dar por cercanía a dichas ULL (*Decremento*). Ejemplos de ellos se pueden ver en la Tabla 13 (ver apartado 3.2. 2. 2. 2). En el *PAFE* los valores de entorno se

GLOSARIO

predeterminan en el *Entorno*. En esta tesis se usan los valores máximos 1 y 3 para hallar colocaciones y relaciones entre ULL, respectivamente, y un decremento de 1.

Vocablo: Cada una de las palabras diferentes que constituyen un texto o corpus (Grasset, 2007). Es decir, en todo este párrafo hay 34 palabras y sólo 30 vocablos, puesto que se repiten: *palabras, que*.

Vocabulario: Conjunto cerrado de palabras o unidades léxicas de un autor, especialidad, etc. (Alcaraz Varó, 2000).

Vocabulario más frecuente: formado por aquellos vocablos que aparecen más veces en un texto o corpus. En esta tesis se determina el vocabulario más frecuente usando la distribución de frecuencias de los vocablos o ULL. Para ello se halla el punto donde la pendiente de la tangente a la curva de la distribución de frecuencias toma el valor de -1 (ver apartado 3. 2. 2. 2. 3.).

Vocabulario específico: el que hace referencia a aquellas ULL usadas únicamente en un dominio del conocimiento, peculiares en virtud de su especialidad, con un significado altamente específico y un único referente conceptual (Gómez González-Jover, 2007). El vocabulario específico o técnico se diferencia del general en que sus términos son monosémicos, es decir, tienen un solo significado (Alcaraz Varó, 2000). El vocabulario específico se puede

determinar comparando las frecuencias de las ULL con un corpus general de referencia (Chung, 2003).

BIBLIOGRAFÍA



BIBLIOGRAFÍA

- Akhtar, N. (2004). Contexts of early word learning. In D. G. Hall, & S. R. Waxman (Eds.), *Weaving a lexicon* (pp. 485-507). Cambridge, MA, EEUU: MIT Press.
- Alameda, J. R., & Cuetos, F. (1995). *Diccionario de frecuencias de las unidades lingüísticas del castellano*. Servicio de Publicaciones de la Universidad de Oviedo.
- Alarcos Llorach, E. (2009). *Gramática de la Lengua Española* (18ª reimpresión ed.). Madrid: Editorial Espasa Calpe, S.A.
- Alcaraz Varó, E. (2000). *El inglés profesional y académico*. Madrid, España: Alianza Editorial.
- Almela, R., Cantos, P., Sánchez, A., Sarmiento, R., & Almela, M. (2005). *Frecuencias del español. Diccionario y estudios léxicos y morfológicos*. (E. Universitas, Editor). Recuperado de <http://www.um.es/lacell/proyectos>
- Álvarez Cavanillas, J. L. & Chacón Beltrán, R. (2003). La enseñanza de “colocaciones” en español como L2: una propuesta didáctica. *ELIA: Estudios de lingüística inglesa aplicada*, 4, 237-253. Recuperado de <http://dialnet.unirioja.es/>

BIBLIOGRAFÍA

Anglin, J. M. (1993). Vocabulary Development: A Morphological Analysis.

Monographs of the Society for Research in Child Development , 58 (10), 1-186. Recuperado de <http://www.jstor.org>

Arunachalam, S., & Waxman, S. R. (2010). Language and conceptual development.

Wiley Interdisciplinary Reviews: Cognitive Science , 1 (4), 548-558. doi: 10.1002/wcs.37

Ash, D. (2004). Reflective scientific sense-making dialogue in two languages: The

Science in the dialogue and the dialogue in the Science. *Science Education* , 88 (6), 855-884. doi: 10.1002/sce.20002

Ausubel, D. P. (1968). *Psicología educativa*. Mexico: Trillas.

Ávila Muñoz, M. (1999). *Léxico de frecuencia del español hablado en la ciudad de*

Málaga. Málaga: Universidad de Málaga.

Baker, P., Hardie, A., & McEnery, T. (2006). *A Glossary of Corpus Linguistics*.

Edinburgh: Edinburgh University Press.

Baldwin, D. A. (1995). Understanding the link between joint attention and

language. In C. Moore, & P. Dunham (Eds.), *Joint attention: its origins and role in development* (pp. 131-158). Hillsdale, NJ: Earlbaum. Recuperado de <http://books.google.com/>

Baldwin, D. A., Markman, E. M., Bill, B., Desjardins, R. N., Irwin, J. M., &

Tidball, G. (1996). Infants' Reliance on a Social Criterion for Establishing Word-Object Relations. *Child Development* , 67 (6), 3135-3153.

Recuperado de <http://www.jstor.org/>

Barberá, O., & Valdés, P. (1996). El trabajo práctico en la enseñanza de las

ciencias: una revisión. *Enseñanza de las Ciencias* , 14, 365-379.

- Bécue, M. (1991). *Análisis Estadístico de datos textuales: Métodos de análisis y algoritmos*. París: Cisia.
- Bécue-Bertaut, M., Lebart, L., & Rajadell, N. (1992). El análisis estadístico de datos textuales: la lectura según los escolares de enseñanza primaria. *Anuario de psicología* , 55, 7-22. Recuperado de <http://www.raco.cat/index.php/anuariopsicologia/>
- Behl-Chadha, G. (1996). Basic-level and superordinate-like categorical representations in early infancy. *Cognition* , 60, 105-141. doi:10.1016/0010-0277(96)00706-8
- Benzécri, J. P. (1973). *L'Analyse des données: La Taxinomie* (4ª, 1984 ed., Vol. I). París: Dunod.
- Benzécri, J. P. (1976). *L'Analyse des données: L'Analyse des correspondances* (Vol. II). París: Dunod.
- Biber, D. (1990). Methodological issues regarding corpus-based analyses of linguistic variation. *Literary and Linguistic Computing* , 5, 257-269.
- Biber, D., Conrad, S., & Reppen, R. (1998). *Corpus linguistics : investigating language structure and use*. Cambridge: Cambridge University Press.
- Brier, A., & Hopp, B. (2011). Computer assisted text analysis in the social sciences. *Quality & Quantity* , 45 (1), 103-128. doi 10.1007/s11135-010-9350-8

BIBLIOGRAFÍA

- Butler, F., Bailey, A. L., Stevens, R., Huang, B., & Lord, C. (2004). *Academic English in Fifth-grade Mathematics, Science, and Social Studies Textbooks*. CSE Report 642 Los Angeles: CRESST/University of California. Los Angeles: Center for the Study of Evaluation (CSE)/National Center for Research on Evaluation, Standards, and Student Testing (CRESST). Recuperado de <http://www.eric.ed.gov/>
- Buzan, T., & Buzan, B. (1996). *El libro de los mapas mentales*. (E. editorial, Trans.) Barcelona: Ediciones Urano S.A.
- Caamaño, A. (1992). Los trabajos prácticos en ciencias experimentales. *Aula de innovación educativa*, 9, 61-68.
- Calderero, J. (2003). *Estudio de libros de texto de ciencias de la naturaleza mediante el análisis cuantitativo basado en teoría de grafos*. Madrid: Tesis doctoral Universidad Complutense de Madrid. Recuperado de <http://eprints.ucm.es/>
- Campbell, A. L., & Namy, L. L. (2003). The role of social-referential context in verbal and nonverbal symbol learning. *Child Development*, 74 (2), 549-563. doi: 10.1111/1467-8624.7402015
- Ceballos, J. P., Galotti, A., González, B., & Varela, C. (1997). El Sonido: aproximación al Lenguaje en textos de Educación Secundaria. *Resúmenes de las Comunicaciones del 7º Encuentro Ibérico para la Enseñanza de la Física* (pp. 31-32). Universidad de Las Palmas de Gran Canaria.
- Ceballos, J. P., Galotti, A., Pérez, D., López, C., Álvarez, J. M., Vadillo, A., López, A. (2000). PAFE – Programa de Análisis de Frecuencias y Entornos. *Actas XII Congreso Nacional y I Iberoamericano de Pedagogía*, 2, pp. 51-52.

- Ceballos, J. P., Galotti, A., Santana, F., & Varela, C. (2000a). Waves in Secondary Education. *International Conference on Physics Education. Physics Teacher Education Beyond 2000 & PTTIS - Abstracts* (p. 111). España: GIREP - Universidad Autónoma de Barcelona.
- Ceballos, J. P., Galotti, A., Santana, F., & Varela, C. (2000b). Waves in Secondary Education. *International Conference on Physics Education. Physics Teacher Education Beyond 2000 & PTTIS – CD-ROM version Proceedings* . España: GIREP - Universidad Autónoma de Barcelona.
- Ceballos, J. P., Galotti, A., Santana, F., & Varela, C. (2000c). Ondas en textos de Educación Secundaria Obligatoria. *XIX Encuentros de Didáctica de las Ciencias Experimentales, Resúmenes* (pp. 102-103). España: Universidad Complutense de Madrid.
- Ceballos, J., Galotti, A., Santana, F., & Varela, C. (2001). Ondas en textos de educación secundaria obligatoria. In *Reflexiones sobre la didáctica de las ciencias experimentales*. Madrid: Nivola.
- Ceballos, J. P., Galotti, A., & Varela, C. (1998). El análisis de textos y los contenidos en Ciencias. *Actas 8º Encontro Ibérico para o Ensino da Física* (pp. 176-177). Sociedade Portuguesa de Física.
- Ceballos, J. P., Galotti, A., Varela, C., & Leal, P. (1998). Análisis de textos de Enseñanza Secundaria italianos y españoles: el Sonido. *Livro de Resumos do 8º Encontro Ibérico para o Ensino da Física* (pp. 174-175). Portugal: Sociedade Portuguesa de Física - Universidade do Porto.

BIBLIOGRAFÍA

- Ceballos, J. P., Galotti, A., Varela, C., & Talavera, J. A. (1998). El Sonido en Textos de la ESO. *XVIII Encuentros de Didáctica de las Ciencias Experimentales* (pp. 223-225). Facultade de Ciencias da Educación - Universidade da Coruña.
- Ceballos, J. P., Galotti, A., Varela Calvo, C., & Talavera Sosa, J. A. (1999). El Sonido en Textos de la Educación Secundaria Obligatoria. In C. Martínez Losada, & S. García Barros, *La Didáctica de las Ciencias. Tendencias Actuales* (pp. 605-614). La Coruña: Servicio de Publicaicones - Universidade da Coruña.
- Ceballos, J. P., Hansen, C., & Stengler, E. (1999). Las estrellas en las revistas de divulgación científica. In B. León, A. Izaguirre, & A. Martínez (Ed.), *Divulgar la ciencia. Actas XIV Jornadas Internacionales de Comunicación* (pp. 293-303). Pamplona, España: Ediciones Eunate.
- Ceballos, J. P., & Hansen-Ruiz, C. S. (2002). La formación del profesorado y el currículo legislado para la ESO respecto a los contenidos de Astronomía. In Elortegui, Medina, Fernández, Varela, & Jarabo (Ed.), *XX Encuentros de Didáctica de las Ciencias Experimentales, "Secundaria y Universidad"*. La Laguna.
- Ceballos, J. P., & Hansen-Ruiz, C. S. (2005). Astronomy in Popular Science Magazines Published in Spain. In T. J. Mahoney (Ed.), *Communicating Astronomy* (pp. 213-216). La Laguna, España: Instituto de Astrofísica de Canarias.

- Ceballos, J. P., Hansen-Ruiz, C. S., Galotti, A., & Stengler, E. (2005). What Science Concepts Appear in Popular Science Magazines? How are They Related? The Case of Stars. In T. J. Mahoney (Ed.), *Communicating Astronomy* (pp. 199-203). La Laguna, España: Instituto de Astrofísica de Canarias.
- Ceballos, J., Santos, A., Varela, C., & Galotti, A. (2000). Dinámica de poblaciones en textos universitarios de Ecología. *XIX Encuentros de Didáctica de las Ciencias Experimentales, Resúmenes* (p. 78). España: Universidad Complutense de Madrid.
- Ceballos, J., Santos, A., Varela, C., & Galotti, A. (2001). Dinámica de poblaciones en textos universitarios de ecología. In *Reflexiones sobre la didáctica de las ciencias experimentales*. Madrid: Nivola.
- Ceballos, J. P., Varela, C., & Galotti, A. (1999). Diseño de una línea de investigación en Didáctica de las Ciencias. In C. Martínez Losada, S. García Barros, & C. M. García (Ed.), *La Didáctica de las Ciencias. Tendencias Actuales* (pp. 439-448). La Coruña: Servicio de Publicaciones - Universidade da Coruña.
- Chung, T. M. (2003). A Corpus Comparison Approach for Terminology Extraction. *Terminology*, 9 (2), 221-246. doi: 10.1075/term.9.2.05chu
- Clark, E. V. (2004). How language acquisition builds on cognitive development. *TRENDS in Cognitive Sciences*, 8 (10), 472-478.
doi:10.1016/j.tics.2004.08.012

BIBLIOGRAFÍA

Conrad, S. M. (1996). Investigating academic texts with corpus-based techniques:

An example from biology. *Linguistics and Education* , 8 (3), 299-326.

doi:10.1016/S0898-5898(96)90025-X

Davies, M. (2002-). *Corpus del Español (100 million words, 1200s-1900s)*.

Recuperado de <http://www.corpusdelespanol.org>

De Rosa, A. S. (2002). The associative network: A Technique for detecting

structure, contents, polarity and stereotyping indexes of the semantic fields.

European Review of Applied Psychology , 52 (3-4), 181-200. Recuperado de

<http://www.euophd.eu/>

Delgado, J. M., & Gutiérrez, J. (1999). *Métodos y técnicas cualitativas de*

investigación en Ciencias Sociales. Madrid: Síntesis.

Devetak, I., Vogrinc, J., & Glazar, S. A. (2010). States of Matter Explanations in

Slovenian Textbooks for Students Aged 6 to 14. *International Journal of*

Environmental and Science Education , 5 (2), 217-235. Recuperado de

<http://www.ijese.com/>

Domínguez, M. C. (2006). *Análisis de conceptos en el estudio de textos de*

enseñanza básica sobre el Sistema Solar. Contribuciones a la Didáctica de

las Ciencias. La Laguna: Tesis doctoral. Universidad de La Laguna.

Domínguez, M. C., & Galotti, A. (2003). Modelo de Análisis de lenguaje utilizado

en textos: Aplicación al tema el Sistema Solar. *Actas XXIX de Reunión*

Bienal de la Real Sociedad Española de Física, (pp. 115-116).

- Domínguez, M. C., & Varela, C. (2001). Ejemplo de análisis comparativo de textos divulgativos de astronomía. *Actas VI Congreso Internacional sobre investigación en la Didáctica de las Ciencias. extra (2)*, p. 163. Enseñanza de las Ciencias.
- Domínguez, M. C., & Varela, C. (2005a). Análisis del lenguaje utilizado en textos sobre el Sistema Solar durante el cambio de etapa. *Actas VII Congreso Internacional sobre la investigación en la Didáctica de la Enseñanza de las Ciencias, formato CD-ROM*.
- Domínguez, M. C., & Varela, C. (2005b). Un ejemplo de análisis de textos de Enseñanza Primaria: El concepto del día y la noche. *Actas XXX Reunión Bienal de la Real Sociedad Española de Física*, (pp. 256-257).
- Domínguez Burgos, A. (2002). Lingüística computacional: un esbozo. *Boletín de Lingüística* , 18, 104-119. Recuperado de <http://redalyc.uaemex.mx/>
- Domínguez Herrera, M. C. (2006). *Análisis de conceptos en el estudio de textos de enseñanza básica sobre el sistema solar : contribuciones a la didáctica de las ciencias*. La Laguna, Santa Cruz de Tenerife, España: La Laguna.
- Domínguez Herrera, M. C. (2007). Análisis conceptual en textos de enseñanza básica: El día y la noche. *Qurrriculum* , 20, 133-146.
- Domínguez Herrera, M. C. (2009). La visión egocéntrica del universo en textos de enseñanza básica sobre el Sistema Solar. *Qurrriculum* , 22, 151-164. Recuperado de <http://webpages.ull.es/users/revistaq/>

BIBLIOGRAFÍA

- Domínguez Herrera, M. C., Galotti, A., & Varela Calvo, C. (2002). El Sistema Solar: Análisis del lenguaje utilizado en los textos. El paso de la Educación Primaria y Secundaria. *Actas XX Encuentros de Didáctica de las Ciencias Experimentales*, 2, pp. 641-648.
- Domínguez Herrera, M. C., & Varela Calvo, C. (2008). Aplicación de una técnica de análisis textual a textos escolares sobre el Sistema Solar. *Revista Electrónica de Enseñanza de las Ciencias*, 7 (1), 261-274.
- Domínguez Herrera, M. C., & Varela Calvo, C. (2009). Muestra de una metodología de investigación en didáctica de las ciencias. *Enseñanza de las Ciencias, Número Extra VIII Congreso Internacional sobre Investigación en Didáctica de las Ciencias*, pp. 1967-1971. Barcelona. Recuperado de <http://ensciencias.uab.es/>
- Donovan, C. A., & Smolkin, L. B. (2001). Genre and Other Factors Influencing Teachers' Book Selections for Science Instruction. *Reading Research Quarterly*, 36 (4), 412-440. Recuperado de <http://www.jstor.org/stable/748058>
- Doyle, P. (2005). Replicating Corpus-Based Linguistics: Investigating Lexical Networks in Text. *Proceedings from the Corpus Linguistics Conference Series, 1, 1*, pp. 1-19. Birmingham. Recuperado de <http://www.birmingham.ac.uk/research/activity/corpus/index.aspx>
- Drouin, P. (2004). Detection of Domain Specific Terminology Using Corpora Comparison. *Proceedings of the Fourth International Conference on Language Resources and Evaluation (LREC)*, (pp. 79-82). Lisbon, Portugal. Recuperado de <http://olst.ling.umontreal.ca/>

- Dugac, Z., Kern, J., & Majdancic, Z. (2005). Health-educative publications (1872-1938)-content analysis by Spad-T application. *Drustvena istrazivanja* , 14 (4-5), 867-883. Recuperado de <http://hrcak.srce.hr/>
- Echeverría, M. S., Vargas, R., Urzúa, P., & Ferreira, R. (2008). DispoGrafo una nueva herramienta computacional para el análisis de relaciones semánticas en el léxico disponible. *Revista de lingüística teórica y aplicada* , 46 (1), 81-92. doi: 10.4067/S0718-48832008000100005
- Eggins, S. (2002). *Introducción a la lingüística sistémica* (Traducción de F. Alcántara). Logroño: Universidad de La Rioja, Servicio de Publicaciones. (Original publicado en 1994).
- Ellis, H. C. (1972). *Fundamentos del aprendizaje y procesos cognoscitivos del hombre*. México, México: Trillas.
- Escandell Vidal, M. V., & Leonetti, M. (2000). Categorías funcionales y gramática procedimental. In M. Martínez Hernández, D. García Padrón, D. Corbella, C. Corrales, F. Cortés, J. Gómez Soliño, et al. (Eds.), *Cien años de investigación semántica: De Michel Bréal a la actualidad* (Vol. 1, pp. 363-378). Madrid: Ed. Clásicas. Recuperado de <http://www.uned.es/>
- Etxeberria, J., García, E., Gil, J., & Rodríguez, G. (1995). *Análisis de datos y textos*. Madrid: Ra-ma.
- Fajardo Rodríguez, M. B., & Varela Calvo, C. (1999). El tratamiento de los conceptos medioambientales en los libros de texto y su asimilación por parte del alumnado. Estudio de un caso en Canarias. *Actas del VIII Congreso Diálogo Fe-Cultura: "Y ahora Europa"* (pp. 285-300). La Laguna (Tenerife): Centro de Estudios Teológicos.

BIBLIOGRAFÍA

- Fang, Z. (2005). Scientific Literacy: A Systemic Functional Linguistics Perspective. (I. Wiley Periodicals, Ed.) *Science Education* , 2, 335–347. doi
10.1002/sce.20050
- Fang, Z. (2006). The Language Demands of Science Reading in Middle School. *International Journal of Science Education* , 28 (5), 491-520 . doi:
10.1080/09500690500339092
- Freeman, G., & Taylor, V. (2006). *Integrating Science and Literacy Instruction: A Framework for Bridging the Gap*. Lanham: Rowman and Littlefield Education.
- Furió, C., & Guisasola, J. (1998). Dificultades de aprendizaje de los conceptos de carga y de campo eléctrico en estudiantes de bachillerato y universidad. *Enseñanza de las Ciencias* , 16 (1), 131-146.
- Galagovsky, L. R. (1993). Redes conceptuales: base teórica e implicaciones para el proceso de enseñanza-aprendizaje. *Enseñanza de las ciencias* , 11 (3), 301-307.
- Galán Rodríguez, C., & Montero Melchor, J. (2002). *El discurso tecnocientífico: la caja de herramientas del lenguaje* (Vols. 75, Cuadernos de Lengua Española). Madrid, Madrid, España: Arco Libros S.L.
- Galindo, C. (1998). *Técnicas de investigación en Sociedad, Cultura y comunicación*. México: Addison Wesley Longman.
- Galotti, A., & Ceballos, J. (2001). Conceptos básicos de Física Cuántica. *Resúmenes de las comunicaciones del 11º Encuentro Ibérico para la Enseñanza de la Física*, 2, pp. 291-292.

- Galotti, A., Ceballos, J. P., & Matus, M. L. (2003). Física Cuántica: análisis en textos de Bachillerato con el PAFE. *Actas XXIX de Reunión Bienal de la Real Sociedad Española de Física, 1*, pp. 105-106.
- García Barros, S., Martínez Losada, C., Mondelo Alonso, M., & Vega Marcote, P. (1997). La Astronomía en textos escolares de Educación Primaria. *Enseñanza de las Ciencias, 15* (2), 225-232.
- García Hoz, V. (1976). *El vocabulario general de orientación científica y sus estratos. (Bases para una enseñanza vertebrada)*. Madrid: Instituto de Pedagogía-CSIC.
- Gattino, S., & Miglietta, A. (2004). Entre la marginación y la integración. Un estudio sobre el prejuicio étnico, orientación política y el empleo del lenguaje. *Boletín de Psicología, 80*, 37-57. Recuperado de <http://www.uv.es/seoane/boletin/boletin.html>
- Gil, D., & Solbes, J. (1993). The introduction of modern physics: overcoming a deformed vision of science. *International Journal of Science Education, 1* (3), 255-260.
- Glenn, W. H. (1990). Treatment of Selected Concepts of Organic Evolution and the History of Life on Earth in Three Series of High School Earth Science Textbooks, 1960-1989. *Science Education, 4* (1), 37-52.
- Gnostice Information Technologies Private Limited. (2002). Oneview. *software, versión 1.1*. Disponible en <http://www.gnostice.com/>
- Gómez González-Jover, A. (2007). Léxico especializado y traducción. In E. Alcaraz Varó, J. Mateo Martínez, & F. Yus Ramos (Eds.), *Las lenguas profesionales y académicas* (pp. 27-40). Barcelona, Barcelona, España: Ariel.

BIBLIOGRAFÍA

Gómez Tórrego, L. (2007). *Gramática didáctica del español* (Novena edición ed.).

Madrid: Ediciones SM.

Graham, S. A., Kilbreath, C. S., & Welder, A. N. (2004). Thirteen-Month-Olds

Rely on Shared Labels and Shape Similarity for Inductive Inferences. *Child*

Development, 75 (2), 409-427. Recuperado de <http://www.jstor.org/>

Grasset, E. (2007). ¿ Ruptura poética y nuevo repertorio ? : el repertorio léxico

como posible síntoma crítico en la poesía castellana de Pere Gimferrer.

Pandora: revue d'etudes hispaniques (7), 233-248. Recuperado de

<http://dialnet.unirioja.es/>

Hall, D. G., & Bélanger, J. (2005). Young children's use of range-of-reference

information in word learning. (B. P. Ltd., Ed.) *Developmental Science*, 8

(1), 8-15. doi: 10.1111/j.1467-7687.2005.00388.x

Halliday, M. A. (1993). Some grammatical problems in scientific English. In M. A.

Halliday, & J. R. Martin (Eds.), *Writing science: Literacy and discursive*

power (pp. 69 – 85). Pittsburgh: University of Pittsburgh Press.

Hansen-Ruiz, C. S., Ceballos, J. P., & Stengler, E. (2005a). Astronomy in the

Spanish Pre-University educational System: The Case of the Canary Islands.

In T. J. Mahoney (Ed.), *Communicating Astronomy* (pp. 147-152). La

Laguna, España: Instituto de Astrofísica de Canarias.

Hansen-Ruiz, C. S., Ceballos, J. P., & Stengler, E. (2005b). La Astronomía en las

revistas de la NSTA. *XXX Reunión Bienal de la Real Sociedad Española de*

Física y 15º Encuentro Ibérico para la Enseñanza de la Física (pp. 252-

253). Orense, España: Real Sociedad Española de Física.

- Heimlich, J. & Pittelman, S. (1990). *Los mapas semánticos. Estrategias de aplicación en el aula*. Madrid. Visor.
- Henderson, A. M., & Graham, S. A. (2005). Two-Year-Olds' Appreciation of the Shared Nature of Novel Object Labels. *Journal of Cognition & Development* , 6 (3), 381-402. doi:10.1207/s15327647jcd0603_4
- Hills, T. T., Maouene, J., Riordan, B., & Smith, L. B. (2010). The associative structure of language: Contextual diversity in early word learning. *Journal of Memory and Language* , 63, 259-273. doi:10.1016/j.jml.2010.06.002
- Hubber, P., Tytler, R., & Haslam, F. (2010). Teaching and Learning about Force with a Representational Focus: Pedagogy and Teacher Change. *Research in Science Education* , 40 (1), 5-28. doi 10.1007/s11165-009-9154-9
- Janicki, K. (2006). *Language Misconceived: Arguing for Applied Cognitive Sociolinguistics*. New Jersey, EEUU: Lawrence Erlbaum Associates .
Recuperado de <http://books.google.es/>
- Johansson, V. (2008). Lexical diversity and lexical density in speech and writing: a developmental perspective. *Lund Working Papers in Linguistics* , 53, 61-79.
Recuperado de <http://www.sciecom.org/>
- Juilland, A., & Chang-Rodríguez, E. (1964). *Frequency dictionary of Spanish words*. La Haya: Mouton.
- Justicia Justicia, F. (1995). *El desarrollo del vocabulario. Diccionario de frecuencias*. Granada: Servicio de Publicaciones de la Universidad de Granada. Campus Universitario de Cartuja.

BIBLIOGRAFÍA

- Kirk, M. K., & Layman, J. W. (1996). A Pre-Lab Guide for General Chemistry: Improving Student Understanding of Chemical Concepts and Processes. *Annual Meeting of the National Association for Research in Science* . St. Louis. Recuperado de <http://www.eric.ed.gov/>
- Klausmeier, H. J. (1990). Conceptualizing. In B. J. (Eds.), *Dimensions of thinking and cognitive instruction* (pp. 93-138). Hillsdale, NJ: Erlbaum.
- Klein, P. D., & Kirkpatrick, L. C. (2010). Multimodal Literacies in Science: Currency, Coherence and Focus. *Research in Science Education* , 40, 87-92. doi 10.1007/s11165-009-9159-4
- Koenig, M. A., & Woodward, A. (2007). Word learning. In M. G. Gaskell (Ed.), *Oxford Handbook of Psycholinguistics* (pp. 617-626). Oxford: Oxford University Press. Recuperado de www.bsos.umd.edu/
- Kohl, P., & Finkelstein. (2005). Representational format, student choice, and problem solving in physics. *AIP Conference Proceedings*, 790, 121-124. doi:10.1063/1.2084716
- Küçüközer, H., Korkusuz, M. E., Küçüközer, H. A., & Yürümezoğlu, K. (2009). The Effect of 3D Computer Modeling and Observation-Based Instruction on the Conceptual Change Regarding Basic Concepts of Astronomy in Elementary School Students. *Astronomy Education Review* , 8 (1). doi:10.3847/AER2009006
- Lasry, N., Finkelstein, N., & Mazur, E. (2009). Are most people too dumb for Physics? *The Physics teacher* , 47, 418-422. Recuperado de <http://tpt.aapt.org/>

- Lebart, L., & Salem, A. (1994). *Statistique Textuelle*. París: Dunod. Recuperado de <http://lexicometrica.univ-paris3.fr/livre/st94/st94-tdm.html>
- Lebart, L., Morineau, A., & Bécue, M. (1989). *Système Portable pour l'Analyse des Données Textuelles. SPAD-T. Manuel de l'utilisateur*. París: Cisia.
- Lebart, L., Morineau, A., Bécue, M., & Haeusler, L. (1994). *Spad-T. Système Portable pour l'Analyse des Données Textuelles*. París: Cisia.
- Lebart, L., Salem, A., & Bécue, M. (2000). *Análisis Estadístico de Textos*. Madrid: Milenio.
- Leech, G. (2001). The Role of Frequency in ELT: New Corpus Evidence Brings a Re-appraisal. In H. Wenzhong (Ed.), *ELT in China 2001: Papers presented at the 3rd International Symposium on ELT in China* (pp. 1-23). Beijing: Foreign Language Teaching and Research Press. Recuperado de <http://www.lancaster.ac.uk/>
- Lemay, C., L'Homme, M. C., & Drouin, P. (2005). Two methods for extracting “specific” single-word terms from specialized corpora. *International Journal of Corpus Linguistics*, 10 (2), 227-255.
- Lemke, J. (1998). Multiplying meaning: visual and verbal semiotics in scientific text. In J. R. Martin, & R. Veel (Eds.), *Reading science: critical and functional perspectives of discourses of science* (pp. 87–111). Nueva York: Routledge.
- Lillo Beviá, J. (1994). Los trabajos prácticos de Ciencias Naturales como actividad reflexiva, crítica y creativa. *Alambique*, 2, 47-56.

BIBLIOGRAFÍA

- Linebarger, D. L., & Vaala, S. E. (2010). Screen media and language development in infants and toddlers: An ecological perspective. *Developmental Review*, 30, 176-202. doi:10.1016/j.dr.2010.03.006
- Llisterri, J. (2007). *Herramientas para el análisis textual*. (U. A. Barcelona, Editor). Consultado en http://liceu.uab.es/~joaquim/language_resources/lang_res/Herram_TecnTex.html
- López Pérez, M. V. (2008). Estudios sobre la lengua de instrucción (LI): un índice de frecuencias léxicas lematizado a partir de un manual escolar. *RESLA*, 21, 201-229. Recuperado de <http://dialnet.unirioja.es/>
- Lovell, K. (1977). *Desarrollo de los conceptos básicos matemáticos y científicos en los niños*. Madrid, España: Morata.
- Mandler, J. M. (2000). Perceptual and Conceptual Processes in Infancy. *Journal of Cognition and Development*, 1, 3-36. doi: 10.1207/S15327647JCD0101N_2
- Marimón Llorca, C., & Santamaría Pérez, M. I. (2007). Los géneros y las lenguas de especialidad (II): el contexto científico-técnico. In E. Alcaraz Varó, J. Mateo Martínez, & F. Yus Ramos (Eds.), *Las lenguas profesionales y académicas* (pp. 127-140). Barcelona, España: Ariel.
- Martínez Celdrán, E. (2002). *Lingüística. Teoría y aplicaciones*. Barcelona: Masson.
- Morera, M. (2007). *La gramática del léxico español*. Badajoz: Editorial @becedario.
- Morrison, C. M., & Conway, M. A. (2010). First words and first memories. *Cognition* (32), 23-32. doi:10.1016/j.cognition.2010.03.011

- Ozdemir, O. F. (2004). *The coexistence of alternative and scientific conceptions in physics. Unpublished doctoral dissertation, The Ohio State University.* Columbus, Ohio, EEUU. Recuperado de <http://etd.ohiolink.edu/>
- Peñas, A., Verdejo, F., & Gonzalo, J. (2001). Corpus-based terminology extraction applied to information access. *13th Special Issue. Proceedings of the Corpus Linguistics 2001* (pp. 458-465). UCREL Technical Paper Lancaster University. Recuperado de <http://citeseerx.ist.psu.edu/>
- Pérez García, D., & Pérez Ceballos, J. G. (1997). PAFE, Programa de Análisis de Frecuencias y Entornos. *versión 2.0*. La Laguna, España.
- Phillips, M. (1985). *Aspects of Text Structure: An Investigation of the Lexical Organisation of Text*. Amsterdam: North Holland.
- Phillips, M. (1989). *Lexical Structure of Text* (Vols. Discourse Analysis Monograph, nº12). Birmingham: University of Birmingham.
- Pozo Muncio, J. I., & Gómez Crespo, M. A. (1998). *Aprender y enseñar ciencia: del conocimiento cotidiano al conocimiento científico* (1ª ed.). Madrid, España: Morata.
- Prater, M. A. (1993). Teaching concepts: Procedures for the design and delivery of instruction. *Remedial & Special Education* , 14 (5), 51-62.
- Real Academia Española. (2001). *Diccionario de la Lengua Española*. (22ª edición) Consultado en <http://www.rae.es/rae.html>
- Real Academia Española: Banco de datos (CREA) [en línea]. (2008). *Corpus de referencia del español actual*. Consultado en <http://www.corpus.rae.es/creanet.html>

BIBLIOGRAFÍA

- Renouf, A. (1992). What do you think of that: a pilot study of the phraseology of the core words of English. In G. Leitner, *New Direction in English Language Corpora*. Mouton de Gruyter. Recuperado de <http://books.google.es>
- Roth, W. M. (1995). *Authentic school science*. Dordrecht: Kluwer. Recuperado de <http://books.google.com/>
- Sánchez Gómez, P. J., Cervelló, J., & Martín, M.,. (2001). Una síntesis entre los enfoques cualitativo y cuantitativo en el estudio de materiales escritos: la codificación en red. En *Retos de la Enseñanza de las Ciencias en el Siglo XXI. Actas del VI Congreso Internacional sobre Investigación en la Didáctica de las Ciencias. Enseñanza de las Ciencias*. Número extra 2001, tomo I, pp. 235-236.
- Sánchez Gómez, P., Morcillo, J., Martín, M., & Silván, E. (2002). Análisis estructural del contenido de textos de alumnos universitarios sobre estructura atómica. En N. Elortegui, M. Medina, J. Fernández Calvo, C. Varela, & F. Jarabo, *Relación Secundaria Universidad. Actas de los XX Encuentros de Didáctica de las Ciencias Experimentales* (pp. 455-463). La Laguna.
- Schachner, D. A., Shaver, P. R., & Mikulincer, M. (2005). Patterns of Nonverbal Behavior and Sensivity in the Context of Attachment Relations. *Journal of Nonverbal Behavior* , 29 (3), 141-169. doi: 10.1007/s10919-005-4847-x
- Schöttler, P. (1989). Historians and discourse Analysis. *History Workshop Journal* , 37-65. doi:10.1093/hwj/27.1.37

- Scott, M. (2010). *WordSmith Tools. Version 5.0*. Liverpool: Lexical Analysis Software. Consultado en <http://www.lexically.net/wordsmith/index.html>
- Sebastián Gallés, N., Martí Antonín, M., Carreiras Valiña, M., & Cuetos Vega, F. (2000). LEXESP, léxico informatizado del español. Barcelona, España: Edicions de la Universitat de Barcelona.
- Sevilla Muñoz, M., & Sevilla Muñoz, J. (2003). Una clasificación del texto científico-técnico desde un enfoque multidireccional. *Language Design* , 5, 93-98. Recuperado de <http://elies.rediris.es/>
- Shanahan, T., & Shanahan, C. (2008). Teaching Disciplinary Literacy to Adolescents: Rethinking Content-Area Literacy. *Harvard Educational Review* , 78 (1), 40-59. Recuperado de <http://www.metapress.com>
- Sinclair, J. M. (1991). *Corpus, concordance and collocation*. Oxford University Press. Recuperado de <http://pocketknowledge.tc.columbia.edu/>
- Sinclair, J. (2005). Corpus and Text - Basic Principles. In M. Wynne (Ed.), *Developing Linguistic Corpora: a Guide to Good Practice* (pp. 1-16). Oxford: Oxford Books.
- Snow, C. E. (2010). Academic Language and the Challenge of Reading for Learning About Science. *Science* , 328, 450-452. doi: 10.1126/science.1182597
- Spelke, E. S. (2000). Core Knowledge. *American Psychologist* , 55, 1233-1243.
- Stuart, K., & Botella, A. (2009). Corpus Linguistics, Network Analysis and Co-occurrence Matrices. *International Journal of English Studies* , 9 (Special Issue), 1-20. Recuperado de <http://revistas.um.es/ijes/article/view/99481>

BIBLIOGRAFÍA

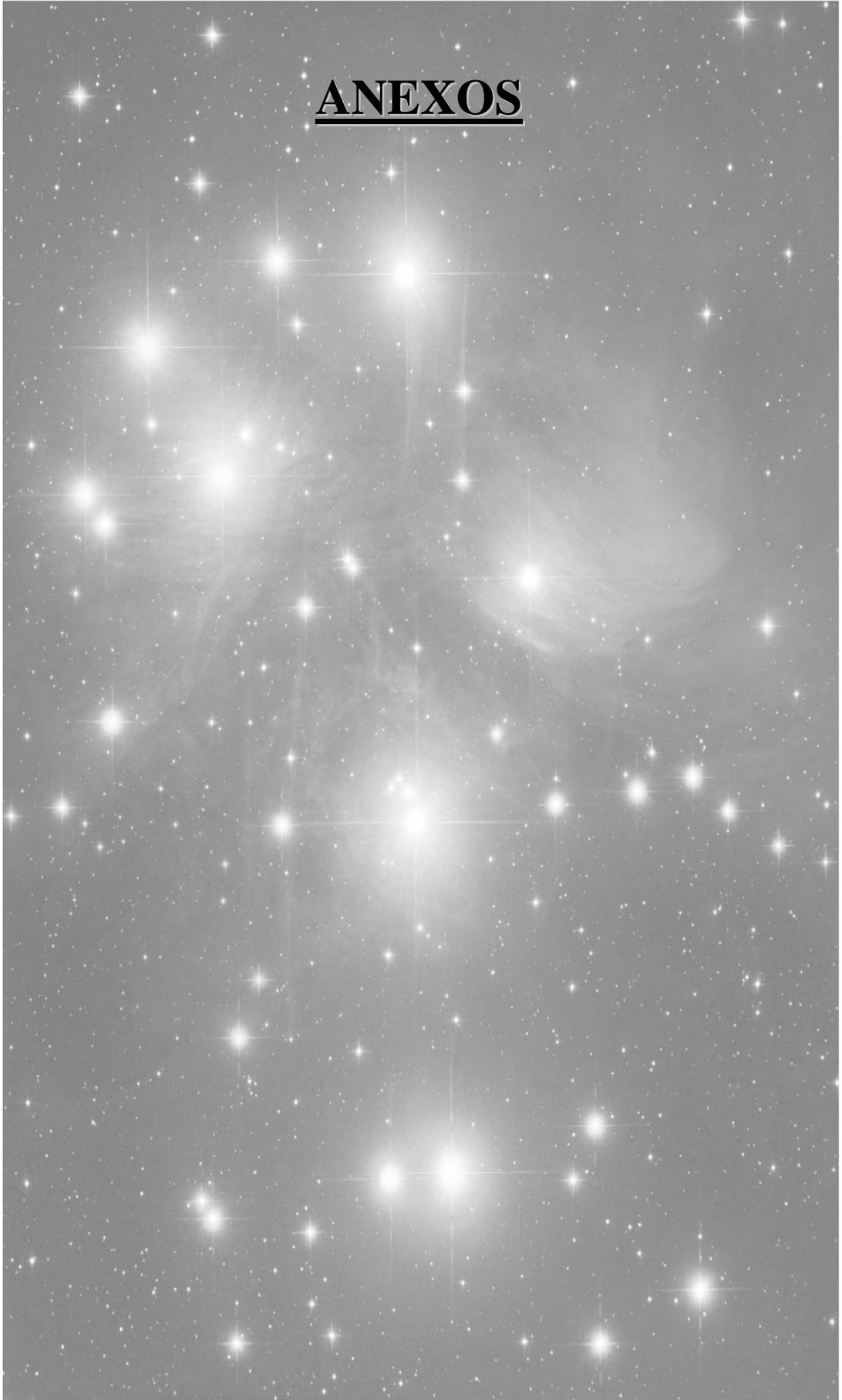
- Tagliacozzo, R. (1976). Levels of Technicality in Scientific Communication. *Information Processing and Management*, 12 (2), 95-110. Recuperado de <http://deepblue.lib.umich.edu/>
- Teich, E., & Fankhauser, P. (2010). Exploring a corpus of scientific texts using data mining. In S. T. Gries, S. Wulff, & M. Davies (Eds.), *Corpus-linguistic applications. Current studies, new directions*. (Vol. 71 (1), pp. 233-247). Amsterdam: Rodopi. Recuperado de <http://www.ingentaconnect.com/>
- Tools and Deliverables Dascitex*. (n.d.). Consultado el 27-03-2011. Recuperado de http://www.linglit.tu-darmstadt.de/index.php?id=lingpro_deliverable
- Ure, J. (1971). Lexical density and register differentiation. In G. E. Perren, & J. L. Trim (Eds.), *Applications of linguistics* (pp. 443-452). Cambridge: Cambridge University Press.
- Varela Calvo, C., & Fajardo Rodríguez, M. B. (2001). ¿Influyen los campos electromagnéticos en nuestra salud?: Una propuesta didáctica. *Las Didácticas de las Áreas Curriculares en el siglo XXI (Volumen I), I Congreso Nacional de Didácticas Específicas* (pp. 545-555). Granada, España: Grupo Editorial Universitario.
- Velasco, I., Díaz, Lloréns, J., de Amescua, A., & Martínez, V. (1999). Algoritmo de filtrado multi-término para la obtención de relaciones jerárquicas en la construcción automática de un tesoro. *Revista Española de Documentación Científica*, 22 (1), 34-49. doi:10.3989/redc.1999.v22.i1.333
- Vogel, D. (2003). Using Generic Corpora to Learn Domain-Specific Terminology. *Workshop on Link Analysis for Detecting Complex Behavior, Washington DC, USA, August 27, 2003*. Recuperado de <http://www.cs.cmu.edu>

- Vosniadou, S. (2002). Mental Models in Conceptual Development. In L. Magnani, & N. J. Nersessian (Eds.), *Model-Based Reasoning: Science, Technology, Values* (pp. 353-368). Nueva York: Kluwer Academic Press.
- Vosniadou, S. (2007). The Cognitive-Situative Divide and the Problem of Conceptual Change. *Educational Psychologist* , 42 (1), 55-66.
- Vosniadou, S., Pagondiotis, C., & Deliyi, M. (2005). Review: From the Pragmatics of Classification Systems to the Metaphysics of Concepts. *Journal of the Learning Sciences* , 14 (1), 115-125. Recuperado de <http://www.jstor.org/>
- Vygotsky, L. (1986). *Thought and language*. Cambridge: MIT Press.
- Wang, X. (2006). Grammatical Concepts and their Application in Foreign Language Teaching. *AARE Conference Papers and Abstracts on the WWW 2006*. Recuperado de <http://www.aare.edu.au/>
- Waxman, S. R., & Gelman, S. A. (2009). Early word-learning entails reference, not merely associations. *Trends in Cognitive Sciences* , 13 (6), 258-263.
doi:10.1016/j.tics.2009.03.006
- Waxman, S. R., & Lidz, J. (2006). Early Word Learning. In D. Kuhn, & R. Siegler (Eds.), *Handbook of Child Psychology: Cognition, Perception and Language (6th Edition)* (pp. 299-335). NY, EEUU: John Wiley & Sons.
- Weist, R. M., Lyytinen, P., Wysocka, J., & Atanassova, M. (1997). The interaction of language and thought in children's language acquisition: a crosslinguistic study. *Journal of Child Language* , 24 (1), 81-121 . Recuperado de <http://journals.cambridge.org>
- Williams, G. (1999). *French discourse analysis : the method of post-structuralism*. London: Routledge. Recuperado de <http://books.google.es>

BIBLIOGRAFÍA

- Woodward, A. L., & Hoyne, K. L. (1999). Infants' learning about words and sound in relation to objects. *Child Development* , 70 (1), 65-77. doi: 10.1111/1467-8624.00006
- Yore, L. D., Bisanz, G. L., & Hand, B. M. (2003). Examining the literacy component of science literacy: 25 years of language arts and science research. *International Journal of Science Education* , 25 (6), 689–725. doi: 10.1080/0950069032000076661
- Yore, L. D., Hand, B., Goldman, S. R., Hildebrand, G. M., Osborne, J. F., Treagust, D. F., & Wallace, C. S. (2004). New directions in language and science education research. *Reading Research Quarterly* , 39 (3), 347-355 .
- Yore, L. D., Hand, B. M., & Prain, V. (2002). Scientists as writers. *Science Education* , 86 (5), 672-692. doi: 10.1002/sce.10042
- Young, E. (2005). The language of Science, the language of students. *Classroom Projects and Curriculum Ideas* , 42 (2), 12.

ANEXOS



ANEXO A: Instalación y Administración del Programa PAFE

La aplicación informática PAFE está grabada en 2 CDs: la última versión de instalación es de marzo de 2006 y las carpetas que corrigen fallos de dicha instalación son de septiembre de 2006. En ambos CDs la estructura de carpetas se muestra en la *Figura 105*.

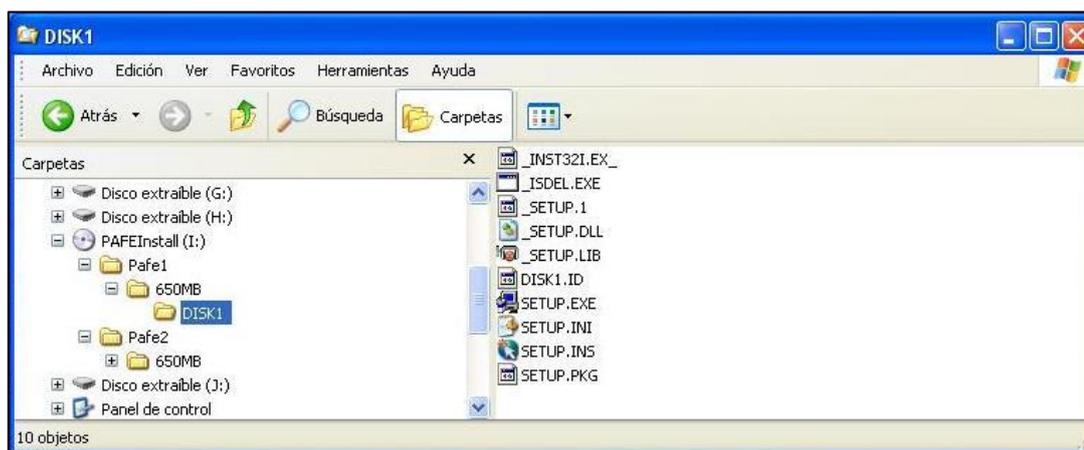


Figura 105.- Estructura de las carpetas de instalación.

La aplicación informática PAFE tiene dos partes independientes, el Análisis de Frecuencias (AF) y el Estudio de Entornos (EE). Cada una de ellas tiene su propia carpeta de instalación: Pafe1 y Pafe2, respectivamente. Para instalar el programa, primero se instala el Análisis de Frecuencias y luego se instala el Estudio de Entornos. La instalación se hace mediante el programa setup.exe que se encuentra en el subdirectorio Pafe1/650MB/DISK1/. Este programa guía al usuario durante la instalación, pidiendo primero que se cierren el resto de programas que se estén usando para proceder a la misma. En el caso del Análisis de Frecuencias,

ANEXO A: INSTALACIÓN Y ADMINISTRACIÓN DEL PROGRAMA PAFE

Pafe1, la instalación crea la carpeta *trabajo* en el disco duro. Mientras que con Estudio de Entornos, Pafe2, la instalación crea la carpeta *trabajo2* en el disco duro. Tras la instalación aparece en el disco duro otra carpeta adicional llamada *PAFE-BD usuarios* que contiene a su vez otra carpeta llamada *bd*. La aplicación PAFE-AF se encuentra dentro de la carpeta *trabajo* con un icono llamado *PROJECT1* y la aplicación PAFE-EE se encuentra dentro de la carpeta *trabajo2* con un icono llamado *PROJ_DIDAC*. Conviene crear accesos directos para ambas aplicaciones que se pueden usar de forma independiente o llamando al Estudio de Entornos desde el Análisis de Frecuencias (*Figura 106*).



Figura 106.- Iconos de las aplicaciones PAFE-AF y PAFE-EE.

Dentro de la carpeta PAFE-BD usuarios se encuentran dos carpetas:

- *C4*, que es el usuario administrador del programa.
- *tablasvacias*, que contiene una copia de las tablas vacías que usan las bases de datos del programa PAFE. Ningún usuario deberá usar esta carpeta como carpeta de trabajo. Cada vez que se cree un usuario nuevo en el programa se copiarán todos los archivos de esta carpeta dentro de la carpeta del nuevo usuario.

Para que el programa funcione es necesario copiar los archivos *plur.DBF* y *plur.MDX*, que se encuentran en *PAFE-BD usuarios/tablasvacias/*, en la carpeta *PAFE-BD usuarios* (*Figura 107*).

ANEXO A: INSTALACIÓN Y ADMINISTRACIÓN DEL PROGRAMA PAFE



Figura 107.- Estructura final de la carpeta PAFE-BD usuarios.

Para terminar se introduce el CD con la versión de septiembre del 2006, y se copia el contenido de las carpetas *trabajo* y *trabajo2* dentro de las carpetas del mismo nombre que han sido instaladas previamente.

Cada vez que se crea un nuevo usuario su carpeta aparece dentro de *PAFE-BD usuarios* (Figura 108).

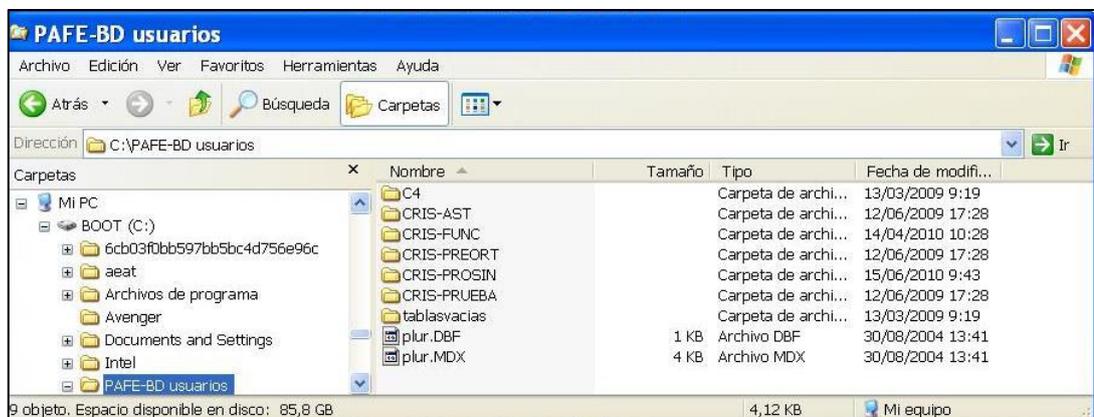


Figura 108.- Usuarios nuevos del PAFE.

Para crear un nuevo usuario se entra en el PAFE mediante el icono llamado *PROJECT1* (es decir, en el Análisis de Frecuencias), con la contraseña del administrador (C4). A través del menú de la Base de Datos se accede a la Inserción, Baja o Listado de usuarios (Figura 109).

ANEXO A: INSTALACIÓN Y ADMINISTRACIÓN DEL PROGRAMA PAFE

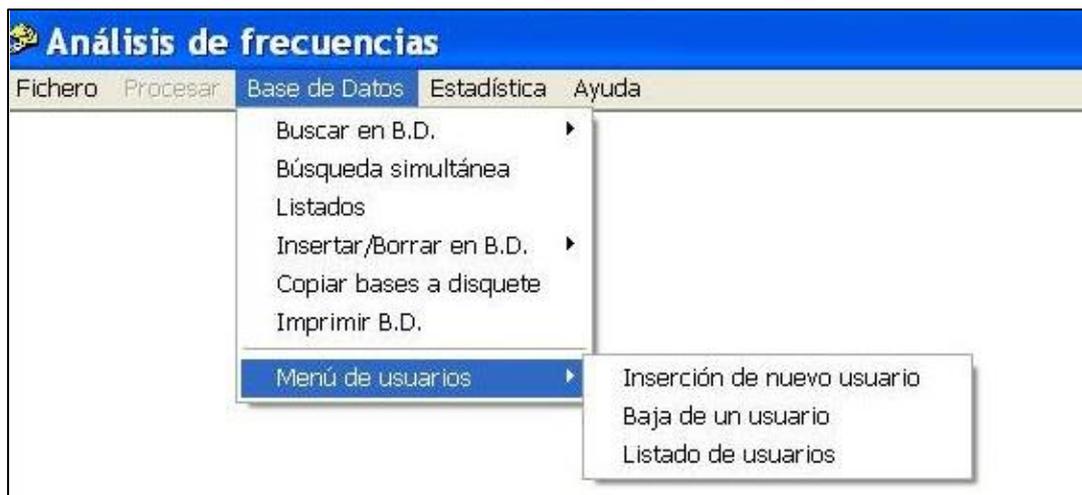


Figura 109.- Submenú de Bases de Datos

La Inserción de Datos conduce a una ventana donde se deben rellenar datos de contacto del investigador (Figura 110).

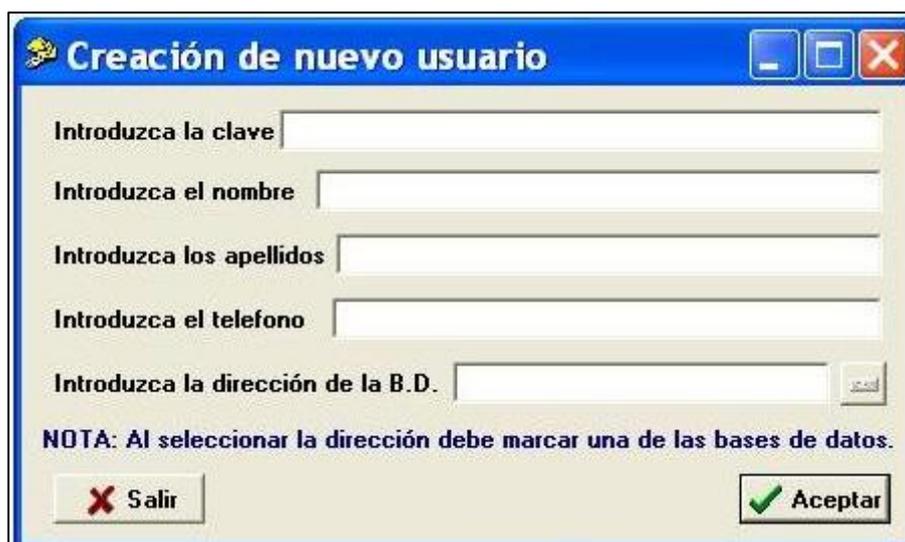
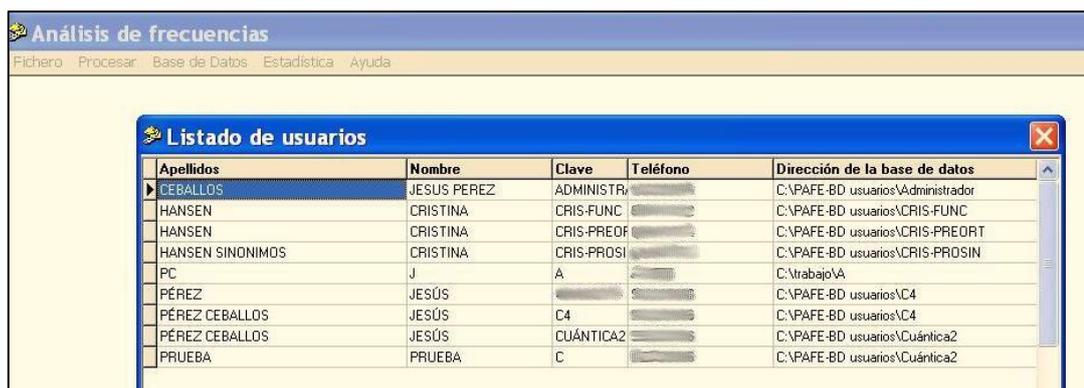


Figura 110.- Ventana de relleno de datos del investigador.

El Listado de usuarios ofrece una lista con datos de contacto de los investigadores que están usando el PAFE (Figura 111). Una vez insertado el usuario aparecerá una nueva carpeta dentro de PAFE-BD usuarios con el nombre dado en la *clave*.

ANEXO A: INSTALACIÓN Y ADMINISTRACIÓN DEL PROGRAMA PAFE



The screenshot shows a software window titled 'Análisis de frecuencias' with a menu bar containing 'Fichero', 'Procesar', 'Base de Datos', 'Estadística', and 'Ayuda'. A sub-window titled 'Listado de usuarios' is open, displaying a table with the following columns: 'Apellidos', 'Nombre', 'Clave', 'Teléfono', and 'Dirección de la base de datos'. The table contains 10 rows of user information.

Apellidos	Nombre	Clave	Teléfono	Dirección de la base de datos
CEBALLOS	JESUS PEREZ	ADMINISTR	[REDACTED]	C:\PAFE-BD usuarios\Administrador
HANSEN	CRISTINA	CRIS-FUNC	[REDACTED]	C:\PAFE-BD usuarios\CRIS-FUNC
HANSEN	CRISTINA	CRIS-PREDI	[REDACTED]	C:\PAFE-BD usuarios\CRIS-PREDI
HANSEN SINONIMOS	CRISTINA	CRIS-PROSI	[REDACTED]	C:\PAFE-BD usuarios\CRIS-PROSIN
PC	J	A	[REDACTED]	C:\trabajo\A
PÉREZ	JESÚS	[REDACTED]	[REDACTED]	C:\PAFE-BD usuarios\C4
PÉREZ CEBALLOS	JESÚS	C4	[REDACTED]	C:\PAFE-BD usuarios\C4
PÉREZ CEBALLOS	JESÚS	CUÁNTICA2	[REDACTED]	C:\PAFE-BD usuarios\Cuántica2
PRUEBA	PRUEBA	C	[REDACTED]	C:\PAFE-BD usuarios\Cuántica2

Figura 111.- Listado de usuarios.

ANEXO B: “Medida de la Distancia y la Edad de un Cúmulo Globular de Estrellas”



Introducción

Las estrellas

Una estrella es una bola gigantesca de gas que tiene luz propia con propiedades físicas tales como la masa, la temperatura o el radio. También es de interés para los astrónomos la distancia de la estrella a la Tierra. La estrella más cercana — y por ende, más estudiada — es, por supuesto, nuestro propio Sol.

La combustión del hidrógeno

La luz emitida por la mayoría de las estrellas es un subproducto del proceso de fusión termonuclear que se produce en el núcleo interno de las estrellas. Una estrella normal como el Sol se compone de aproximadamente un 74% de hidrógeno y un 25% de helio, siendo el restante 1% una mezcla de elementos más pesados. El proceso de fusión más común en una estrella como el Sol es la “combustión del hidrógeno”, en la cual cuatro núcleos de hidrógeno se fusionan para formar un núcleo de helio. El proceso sucede en varias etapas, que se ilustran en la Fig. 2. En el primer paso del proceso, dos protones se fusionan para formar deuterio, una forma pesada del hidrógeno. Este es un suceso muy raro, incluso para el núcleo denso de la estrella, en donde la temperatura es de unos po-

cos millones de grados. Es por esto por lo que las estrellas como el Sol no explotan en una reacción violenta cuando comienza el proceso de fusión, sino que permanecen en esta fase estable de la vida de la estrella durante varios miles de millones de años. Mientras la estrella es estable, la temperatura de su superficie, su radio y su luminosidad son aproximadamente constantes. Las reacciones nucleares del núcleo de la estrella generan la energía justa para mantener un equilibrio entre la presión térmica ejercida hacia fuera y las fuerzas gravitatorias ejercidas hacia dentro.

La masa de un átomo de helio es sólo el 99.3% de la masa de los cuatro núcleos originales de hidrógeno. El proceso de fusión convierte el 0.7% residual de la masa en energía — mayoritariamente en forma de luz. La cantidad de energía se puede calcular usando la famosa ecuación de Einstein $E = Mc^2$. Ya que c^2 es un número muy grande, esta ecuación implica que incluso una pequeña cantidad de materia puede convertirse en una formidable cantidad de energía. El 0.7% residual de la masa de los cuatro núcleos de hidrógeno involucrados en una sola reacción puede parecer pequeño, pero cuando se considera el número total de reacciones de todo el proceso de fusión, hay implicada una masa total (y por tanto energía) considerable.

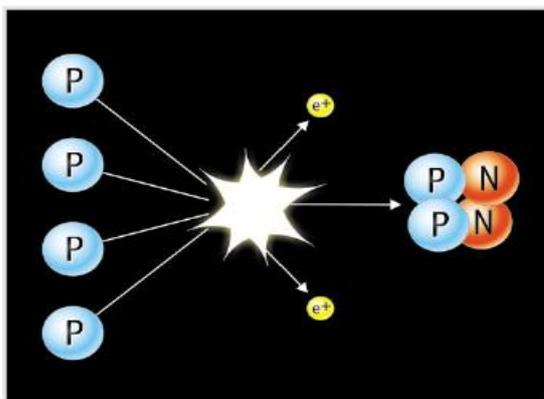


Figura 2: Combustión del Hidrógeno
La forma más simple de “producción” de energía en las estrellas tiene lugar por la fusión de cuatro núcleos de hidrógeno en un núcleo de helio. El proceso tiene varios pasos, pero aquí se muestra el resultado global.

Los cúmulos de estrellas

El término “cúmulo de estrellas” se usa para dos tipos diferentes de agrupaciones de estrellas: cúmulos abiertos de estrellas y cúmulos globulares de estrellas.

Los cúmulos abiertos de estrellas son colecciones no compactas de estrellas relativamente jóvenes que van desde un centenar hasta unas pocas miles de estrellas. Tienen una edad de unos pocos cientos de millones de años, una fracción pequeña del tiempo total de la vida de la estrella (unos pocos miles de millones de años). Estos cúmulos se encuentran en el disco de nuestra galaxia, la Vía Láctea, y a menudo contienen nubes de gas y polvo donde se forman nuevas estrellas. El diámetro típico de un cúmulo abierto de estrellas es de aproximadamente 30 años luz (10 parsecs).

Introducción

Figura 112.- Primera página del texto escolar “Medida de la distancia y la edad de un cúmulo globular de estrellas”.



Introducción



Figura 3: Las Pleiades (Messier 45) en la constelación de Tauro
Este es uno de los más famosos cúmulos de estrellas del cielo. Las Pleiades pueden verse a simple vista incluso desde la mayoría de las ciudades pese a la contaminación lumínica. Es uno de los más brillantes y más próximos cúmulos abiertos. El cúmulo de las Pleiades contiene más de 3000 estrellas, está a una distancia de unos 400 años luz y tiene un tamaño de sólo 13 años luz (cortesía de Bruno Stampfer y Rainer Eisendle).

Introducción

Los cúmulos globulares — las estructuras más viejas de la Vía Láctea

En el halo y en el disco de nuestra Vía Láctea existen unos pocos cientos de cúmulos esféricos y compactos, llamados cúmulos globulares,

y que están ligados gravitatoriamente a nuestra galaxia.

Cada cúmulo globular se compone de un grupo esférico de hasta un millón de estrellas y tiene un diámetro típico de unos 100 años luz. La mayoría de los cúmulos globulares son muy vie-

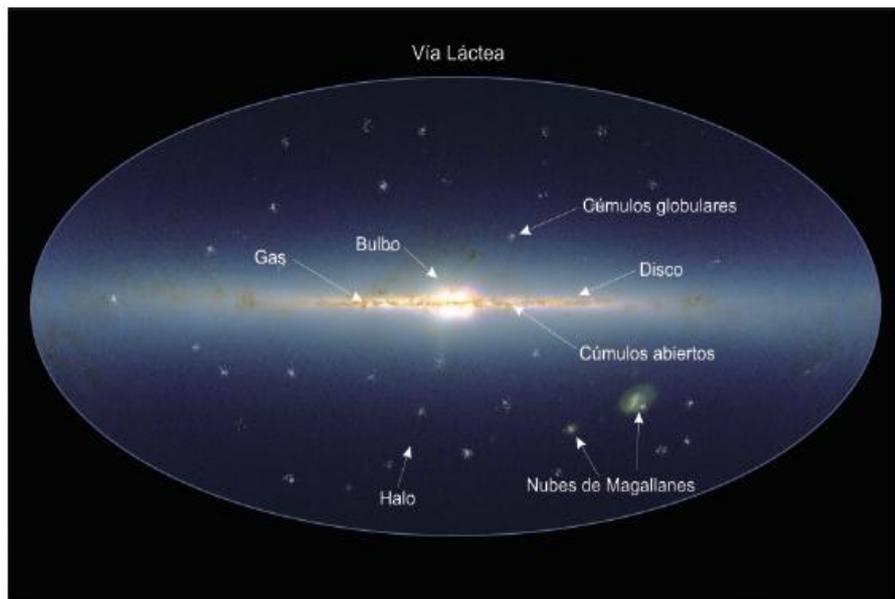


Figura 4: La Vía Láctea
Esta ilustración da una perspectiva general de la galaxia Vía Láctea. Se indican los diferentes componentes de este complicado sistema de estrellas, gas y polvo. El plano del disco se encuentra a lo largo de la línea central horizontal. Los cúmulos globulares se distribuyen en un halo esférico alrededor del centro de la galaxia. Se cree que esta distribución tiene que ver con el hecho de que estos cúmulos de estrellas se formaron muy pronto en la historia de la galaxia.

4

Figura 113.- Segunda página del texto escolar “Medida de la distancia y la edad de un cúmulo globular de estrellas”.

ANEXO B: “MEDIDA DE LA DISTANCIA Y LA EDAD DE UN CÚMULO GLOBULAR DE ESTRELLAS”



Introducción

jos y lo más probable es que sean anteriores a la formación de la propia galaxia que tuvo lugar hace aproximadamente 12 mil millones de años cuando la mayoría del material protogaláctico se depositó en el disco.

Muchos cúmulos globulares probablemente se han destruido con el paso de los miles de millones de años por las repetidas colisiones e interacciones entre sus estrellas y con la Vía Láctea. Los cúmulos globulares supervivientes son más viejos que cualquier otra estructura de nuestra Vía Láctea.

El estudio astrofísico de los cúmulos globulares es una parte importante del programa de investigación de la de la comunidad astronómica internacional. Estos cúmulos de estrellas son importantes, no sólo como un valioso banco de pruebas de valor para las teorías de estructura y evolución estelar, sino también porque se encuentran entre los pocos objetos de la galaxia para los cuales se puede determinar su edad de una forma relativamente precisa. Por ser tan longevos proporcionan un límite inferior muy



Figura 5: La región exterior del cúmulo globular M12

Esta imagen de dos colores se generó a partir de observaciones hechas a través del filtro azul (B) y del filtro verde (V) por el Telescopio VLT del ESO. La imagen B se muestra en azul y la imagen V en rojo en esta composición de imágenes. Algunas de las estrellas son claramente más brillante en la imagen B (vistas como estrellas azuladas), mientras que otras son más brillantes en la imagen V (vistas como estrellas amarillentas).

Introducción

5

Figura 114.- Tercera página del texto escolar “Medida de la distancia y la edad de un cúmulo globular de estrellas”.



Introducción

útil a la edad del Universo. La distribución de sus edades y la correlación entre la edad de un cúmulo y la abundancia en él de los distintos elementos químicos hace de estos sistemas una prueba valiosa en los procesos de formación estelar.

Todas las estrellas que componen un cúmulo globular comparten una historia común y difieren entre sí sólo en sus masas. Por lo tanto, los cúmulos globulares son lugares ideales para estudiar la evolución de las estrellas. En los ejercicios siguientes, determinarás algunas de las propiedades de un cúmulo globular particular, el Messier 12.

Introducción

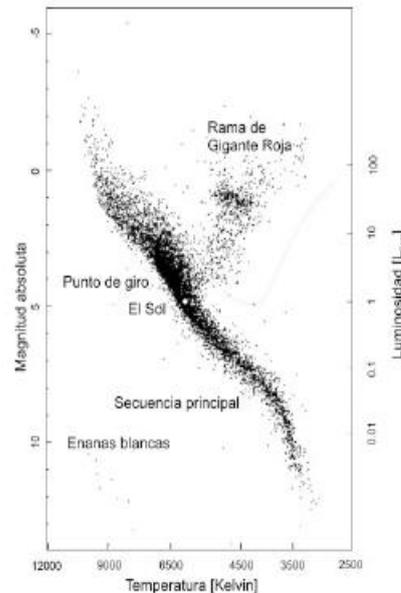


Figura 6: Un diagrama de Hertzsprung-Russell de estrellas próximas

El diagrama H-R muestra la relación entre temperatura superficial y luminosidad de las estrellas. Fíjate en la prominencia de la Secuencia Principal y las regiones donde dominan las gigantes rojas y las enanas blancas. Se indica la posición del Sol así como la “ruta” que una estrella de una masa solar seguirá durante las diferentes fases de su vida. La posición del Sol en el diagrama está determinada por su temperatura superficial de 5800 K y su magnitud absoluta de +4.8.

6

El cúmulo globular Messier 12

El cúmulo globular Messier 12 (o M12), también llamado NGC 6218, fue descubierto en 1764 por Charles Messier de forma que es el objeto decimosegundo de Messier. Como muchos otros cúmulos globulares, Messier lo describió como una “Nebulosa sin estrellas”, una consecuencia del modesto poder de resolución de su telescopio. William Herschel fue el primero que resolvió los cúmulos en estrellas individuales en 1783.

M12 se encuentra en la constelación de Ofiuco y se puede ver con prismáticos desde lugares con muy baja polución luminica. La magnitud visible del conjunto del cúmulo globular es 6.7 (ver más sobre magnitudes en la sección Herramientas Astronómicas, página 2) y la estrella más brillante del cúmulo tiene una magnitud visible de 12.

El NGC (Nuevo Catálogo General) fue publicado en 1888. En él aparece un listado de cúmulos globulares y cúmulos abiertos de estrellas, Nebulosas planetarias y difusas, remanentes de supernova, galaxias de todos los tipos e incluso algunos errores que no correspondían a ningún objeto.

Diagrama de Hertzsprung-Russell

Se llama diagrama de Hertzsprung-Russell (o de forma más concisa, diagrama H-R) a la gráfica que muestra la luminosidad L (o magnitud absoluta M) frente a la temperatura superficial T de las estrellas. La Fig. 6 muestra un ejemplo general que se ha construido a partir de observaciones de estrellas en la vecindad de los cúmulos donde se conocen las distancias (a partir de las medidas de HIPPARCOS). La temperatura superficial T de una estrella se puede derivar de los valores medidos de su color ($m_b - m_v$) (ver la sección Herramientas Astronómicas).

Se ve claramente del diagramas H-R que las medidas (L, T) para diferentes estrellas forman un curioso patrón cuando se colocan sobre el diagrama. Las estrellas se concentran en áreas específicas (indicadas en la figura). El diagrama H-R nos da la clave para comprender como las estrellas evolucionan en el tiempo. Las estrellas, dependiendo de sus masas, se moverán a través del diagrama a lo largo de caminos diferentes.

Figura 115.- Cuarta página del texto escolar “Medida de la distancia y la edad de un cúmulo globular de estrellas”.

ANEXO B: “MEDIDA DE LA DISTANCIA Y LA EDAD DE UN CÚMULO GLOBULAR DE ESTRELLAS”



Introducción

La evolución estelar en el diagrama H-R

Las estrellas pasan la mayor parte de su vida en la Secuencia Principal, quemando el hidrógeno lentamente en un estado de equilibrio estable. Obviamente, esta es la razón por la que la mayoría de las estrellas se localizan en la Secuencia Principal, aproximadamente una línea recta desde el vértice superior izquierdo al vértice inferior derecho en el diagrama. Cuando el suministro de hidrógeno en el núcleo de la estrella se agota, la combustión del hidrógeno ya no es posible. Este es el final de la fase de la secuencia principal de la vida de la estrella y el equilibrio entre la presión del gas y la contracción gravitatoria en el núcleo estelar ya no es estable. La fusión del hidrógeno tiene lugar entonces en la capa circundante mientras el núcleo comienza a contraerse. Al contraerse el núcleo se elevan su presión y su temperatura central, de manera que los núcleos de helio del núcleo de la estrella comienzan a fusionarse y forman elementos más pesados. Este ciclo se puede repetir usando elementos cada vez más pesados a medida que los elementos más ligeros se van acabando en el núcleo. Durante esta fase la estrella aparece como una gigante roja. Tales estrellas se muestran en el diagrama H-R fuera de la línea de la secuencia principal, en la parte superior derecha. El aumento de la temperatura central hace que las capas externas de la estrella se expandan y enfrien de forma que la tempera-

tura superficial decae. El conjunto de la estrella llega a ser muy grande y, debido a la inferior temperatura superficial, emite hacia el espacio sobre todo radiación de longitudes de onda mayores, de forma que la estrella parece roja. A pesar de su baja temperatura superficial T , todas las gigantes rojas tienen una alta luminosidad L debido a su enorme radio R . Este resultado se deduce de la Ley de Radiación de Stefan-Boltzmann para la radiación del cuerpo negro:

$$L = \sigma 4\pi R^2 T^4$$

donde σ es la constante de Stefan-Boltzmann. Valores típicos para las gigantes rojas son $R \sim 10^2 R_{\text{solar}}$, $T \sim (3 \dots 4) 10^3 \text{ K}$, de forma que L es aproximadamente $10^3 L_{\text{solar}}$.

Cuando ya no pueden mantenerse los procesos avanzados de fusión en el núcleo estelar, el núcleo se colapsa de nuevo. Otra vez la temperatura del núcleo se incrementa y ahora se expulsan las capas externas de la estrella. Las llamadas nebulosas planetarias se forman a partir de los restos de las capas externas (ver Ejercicio 3 de esta serie de ejercicios de astronomía de ESA/ESO). El núcleo colapsado está muy caliente (blanco) y la estrella es muy pequeña. A estas estrellas se las denominan muy adecuadamente enanas blancas y es la fase final de la vida de una estrella normal de tipo solar

Para hacer una estimación aproximada de la relación entre la luminosidad L y la temperatura superficial T de todas las estrellas de la secuencia principal, veamos el diagrama H-R (Fig.6). La línea más o menos recta de la Secuencia Principal se extiende aproximadamente un orden de magnitud en temperatura: $(3 \times 10^3 \dots 3 \times 10^4) \text{ K}$. El rango de luminosidades se extiende aproximadamente seis órdenes de magnitud: $(10^{-2} \dots 10^4) L_{\text{solar}}$. Por tanto podemos estimar de forma aproximada: $L \propto T^6$ para las estrellas de la secuencia principal.

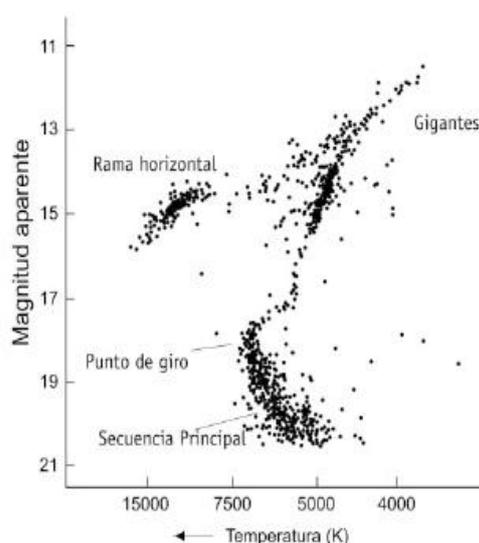


Figura 7: Diagrama típico de Hertzsprung-Russell de un cúmulo globular
Tras miles de millones de años de evolución, el diagrama H-R de un cúmulo globular muestra una Secuencia Principal (MS) corta en la parte inferior derecha. El área llamada rama de las gigantes rojas comienza a partir de la MS y se dirige hacia la parte superior derecha del diagrama. El punto de unión entre la rama MS y la Rama de Gigante Roja se denomina punto de giro.

Introducción

Figura 116.- Quinta página del texto escolar “Medida de la distancia y la edad de un cúmulo globular de estrellas”.



Introducción

Para dar algunos ejemplos:

- Una estrella de gran masa de la secuencia principal con una temperatura superficial de aproximadamente $T_{\text{estrella}} = 1.0 \times 10^4$ K tiene una luminosidad de aproximadamente $L_{\text{estrella}} = (10/5.8)^2 \cdot L_{\text{solar}}$, o aproximadamente 26 veces la luminosidad del Sol. (La luminosidad del Sol tiene un valor estándar de 1 en la escala de luminosidad).
- Una estrella de baja masa con $T_{\text{estrella}} = 3.5 \times 10^3$ K tiene una luminosidad de aproximadamente sólo el 5% de la luminosidad del Sol.

El índice de color B–V: Una pista de la temperatura superficial

Toda la información que podemos extraer de las estrellas proviene de la radiación que recibimos de ellas. Como se explica en la sección Herramientas Astronómicas, diferentes filtros y sistemas de color se pueden usar para medir el brillo de una estrella. En este ejercicio usamos una imagen B y una imagen V. En tu análisis de estas imágenes encontrarás las magnitudes aparentes m_B y m_V de una muestra de estrellas en el cúmulo. Luego podrás calcular los valores m_B and m_V (el índice de color B–V). Finalmente podrás determinar la temperatura superficial de las estrellas (ver la sección Herramientas Astronómicas).

Para un cúmulo, la clave está en el diagrama H-R

Un cúmulo es un grupo de estrellas. La vida de un cúmulo está determinada por la vida de los diferentes tipos de estrellas que lo componen. Las observaciones han mostrado que en los cúmulos globulares queda muy poca cantidad de gas y polvo, de manera que es muy raro el nacimiento de nuevas estrellas en tales cúmulos. Las

estrellas que vemos en un cúmulo globular son todas “adultas” y han evolucionado de formas diferentes en función de sus masas.

La mayoría de las estrellas de baja masa se encuentran en la Secuencia Principal. Esto se debe a que las estrellas de baja masa van consumiendo su energía muy lentamente. Queman sus reservas de hidrógeno poco a poco y continuarán de este modo durante miles de millones de años. Por consiguiente, estarán en la Secuencia Principal durante mucho tiempo.

Por el contrario, las estrellas más pesadas del cúmulo ya han transformado el hidrógeno de sus núcleos y se han convertido en gigantes rojas. Todo esto sucedió hace mucho tiempo, así que hoy en día no hay estrellas calientes masivas, lo que deja un hueco en la mitad superior de la Secuencia Principal (ver Fig. 7). Estas estrellas se encuentran ahora en el área diagonal que comienza en la Secuencia Principal y se desplaza hacia la parte superior derecha del diagrama conocida como rama de las gigantes rojas.

El punto en el cual se encuentran la Secuencia Principal y la Rama de Gigante Roja se llama punto de giro, y nos da una pista importante a la hora de determinar la edad del cúmulo. En el siguiente ejercicio, medirás las coordenadas de este punto sobre tu diagrama y determinarás la edad de M12.

Figura 117.- Sexta página del texto escolar “Medida de la distancia y la edad de un cúmulo globular de estrellas”.

ANEXO C: Texto Preparado FESTESE001V

Una estrella es una bola gigantesca de gas que tiene luz propia con propiedades físicas tales como la masa, la temperatura o el radio. También es de interés para los astrónomos la distancia de la estrella a la Tierra. La estrella más cercana- y por ende, más estudiadaa- es, por supuesto l, nuestro propio Sol.

La luz emitidaa por la mayoría de las estrellas es un subproducto del proceso de fusión termonuclear que se produce en el núcleo interno de las estrellas. Una estrella normal como el Sol se compone de aproximadamente un 74 por ciento de hidrógeno y un 25 por ciento de helio, siendo el restante 1 por ciento una mezcla1 de elementos más pesadosa. El proceso de fusión más común en una estrella como el Sol es la “combustión del hidrógeno”, en la cual cuatro núcleos de hidrógeno se fusionan para formar unaa núcleo de helio. El proceso sucede en varias etapas, que se ilustran en la Fig 2 . En el primer paso1 del proceso, dos protones se fusionan para formar deuterio, una forma1 pesadaa del hidrógeno. Este es un suceso muy raro, incluso para el núcleo denso de la estrella, en donde la temperatura es de unos pocos millones de grados. Es por esto por lo que las estrellas como el Sol no explotan en una reacción violenta cuando comienza el proceso de fusión, sino que permanecen en esta fase estable de la vida de la estrella durante varios miles de millones de años. Mientras la estrella es estable, la temperatura de su superficie, su radio y su luminosidad son aproximadamente constantesa. Las reacciones nucleares del núcleo de la estrella generan la energía justa para mantener un equilibrio entre la

presión térmica ejercidaa hacia fuera y las fuerzas gravitatorias ejercidasa hacia dentro.

La masa de unaa átomo de helio es sólo el 99coma3 por ciento de la masa de los cuatro núcleos originales de hidrógeno. El proceso de fusión convierte el 0coma7 por ciento residual de la masa en energía- mayoritariamente en forma1 de luz. La cantidad de energía se puedex calcular usando la famosa ecuación de Einstein $E = M \times c^2$. Ya que c^2 es un número muy grande, esta ecuación implica que incluso una pequeña cantidad de materia puedex convertirse en una formidable cantidad de energía. El 0coma7 por ciento residual de la masa de los cuatro núcleos de hidrógeno involucradosa en unaa sola reacción puedex parecer pequeño, pero cuando se considera el número total de reacciones de todo el proceso de fusión, hayx implicadaa una masa total (y por tanto energía) considerable.

El término “cúmulo de estrellas” se usa para dos tipos diferentes de agrupaciones de estrellas: cúmulos abiertosa de estrellas y cúmulos globulares de estrellas.

Los cúmulos abiertosa de estrellas son colecciones no compactas de estrellas relativamente jóvenes que van desde un centenar hasta unas pocas miles de estrellas. Tienen una edad de unos pocos cientos de millones de años, una fracción pequeña del tiempo total de la vida de la estrella (unos pocos miles de millones de años). Estos cúmulos se encuentran en el disco de nuestra galaxia, la Vía Láctea, y a menudo contienen nubes de gas y polvo donde se forman nuevas estrellas. El diámetro típico de un cúmulo abiertoa de estrellas es de aproximadamente 30 años luz (10 parsecs).

ANEXO C: TEXTO PREPARADO FESTESE001V

Este es uno de los más famosos cúmulos de estrellas del cielo. Las Pléyades pueden verse a simple vista incluso desde la mayoría de las ciudades pese a la contaminación lumínica. Es uno de los más brillantes y más próximos cúmulos abiertos. El cúmulo de las Pléyades contiene más de 3000 estrellas, está a una distancia de unos 400 años luz y tiene un tamaño de sólo 13 años luz (cortesía de Bruno Stampfer y Rainer Eisendle).

En el halo y en el disco de nuestra Vía Láctea existen unos pocos cientos de cúmulos esféricos y compactos, llamados cúmulos globulares, y que están ligados gravitatoriamente a nuestra galaxia.

Cada cúmulo globular se compone de un grupo esférico de hasta un millón de estrellas y tiene un diámetro típico de unos 100 años luz. La mayoría de los cúmulos globulares son muy viejos y lo más probable es que sean anteriores a la formación de la propia galaxia que tuvo lugar hace aproximadamente 12 mil millones de años cuando la mayoría del material protogaláctico se depositó en el disco.

Muchos cúmulos globulares probablemente se han destruido con el paso de los miles de millones de años por las repetidas colisiones e interacciones entre sus estrellas y con la Vía Láctea. Los cúmulos globulares supervivientes son más viejos que cualquier otra estructura de nuestra Vía Láctea.

Esta ilustración da una perspectiva general de la galaxia Vía Láctea. Se indican los diferentes componentes de este complicado sistema de estrellas, gas y polvo. El plano del disco se encuentra a lo largo de la línea central horizontal. Los cúmulos globulares se distribuyen en un halo esférico alrededor del centro de la

galaxia. Se cree que esta distribución tiene que ver con el hecho de que estos cúmulos de estrellas se formaron muy pronto en la historia de la galaxia.

El estudio astrofísico de los cúmulos globulares es una parte importante del programa de investigación de la comunidad astronómica internacional. Estos cúmulos de estrellas son importantes, no sólo como un valioso banco de pruebas de valor para las teorías de estructura y evolución estelar, sino también porque se encuentran entre los pocos objetos de la galaxia para los cuales se pueden determinar su edad de una forma relativamente precisa. Por ser tan longevos proporcionan un límite inferior muy útil a la edad del Universo. La distribución de sus edades y la correlación entre la edad de un cúmulo y la abundancia en él de los distintos elementos químicos hace de estos sistemas una prueba valiosa en los procesos de formación estelar.

Todas las estrellas que componen un cúmulo globular comparten una historia común y difieren entre sí sólo en sus masas. Por lo tanto, los cúmulos globulares son lugares ideales para estudiar la evolución de las estrellas. En los ejercicios siguientes, determinarás algunas de las propiedades de un cúmulo globular particular, el Messier 12.

El cúmulo globular Messier 12 (o M12), también llamado NGC 6218, fue descubierto en 1764 por Charles Messier de forma que es el objeto duodécimo de Messier. Como muchos otros cúmulos globulares, Messier lo describió como una “Nebulosa sin estrellas”, una consecuencia del modesto poder de resolución de su telescopio. William Herschel fue el primero que resolvió los cúmulos en estrellas individuales en 1783.

ANEXO C: TEXTO PREPARADO FESTESE001V

M12 se encuentra en la constelación de Ofiuco y se puede ver con prismáticos desde lugares con muy baja contaminación lumínica. La magnitud visible del conjunto del cúmulo globular es 6,7 (ver más sobre magnitudes en la sección Herramientas Astronómicas, página 2) y la estrella más brillante del cúmulo tiene una magnitud visible de 12.

El NGC (Nuevo Catálogo General) fue publicado en 1888. En él aparece una lista de cúmulos globulares y cúmulos abiertos de estrellas, nebulosas planetarias y difusas, remanentes de supernova, galaxias de todos los tipos e incluso algunos errores que no correspondían a ningún objeto.

Se llama diagrama de Hertzsprung-Russell (o de forma más concisa, diagrama H-R) a la gráfica que muestra la luminosidad L (o magnitud absoluta M) frente a la temperatura superficial T de las estrellas. La Fig 6 muestra un ejemplo general que se ha construido a partir de observaciones de estrellas en la vecindad de los cúmulos donde se conocen las distancias (a partir de las medidas de HIPPARCOS). La temperatura superficial T de una estrella se puede derivar de los valores medidos de su color (m_B menos m_V) (ver la sección Herramientas Astronómicas).

Se ve claramente del diagrama H-R que las medidas (L , T) para diferentes estrellas forman un curioso patrón cuando se colocan sobre el diagrama. Las estrellas se concentran en áreas específicas (indicadas en la figura). El diagrama H-R nos da la clave para comprender cómo las estrellas evolucionan en el tiempo. Las estrellas, dependiendo de sus masas, se moverán a través del diagrama a lo largo de caminos diferentes.

ANEXO C: TEXTO PREPARADO FESTESE001V

Las estrellas pasan la mayor parte de su vida en la Secuencia Principal, quemando el hidrógeno lentamente en un estado de equilibrio estable. Obviamente, esta es la razón por la que la mayoría de las estrellas se localizan en la Secuencia Principal, aproximadamente una línea recta desde el vértice superior izquierdo al vértice inferior derecho en el diagrama. Cuando el suministro de hidrógeno en el núcleo de la estrella se agota, la combustión del hidrógeno ya no es posible. Este es el final de la fase de la secuencia principal de la vida de la estrella y el equilibrio entre la presión del gas y la contracción gravitatoria en el núcleo estelar ya no es estable. La fusión del hidrógeno tiene lugar entonces en la capa circundante mientras el núcleo comienza a contraerse. Al contraerse el núcleo se elevan su presión y su temperatura central, de manera que los núcleos de helio del núcleo de la estrella comienzan a fusionarse y forman elementos más pesados. Este ciclo se puede repetir usando elementos cada vez más pesados a medida que los elementos más ligeros se van acabando en el núcleo. Durante esta fase la estrella aparece como una gigante roja. Tales estrellas se muestran en el diagrama H-R fuera de la línea de la secuencia principal, en la parte superior derecha. El aumento de la temperatura central hace que las capas externas de la estrella se expandan y enfríen de forma que la temperatura superficial decae. El conjunto de la estrella llega a ser muy grande y, debido a la inferior temperatura superficial, emite hacia el espacio sobre todo radiación de longitudes de onda mayores, de forma que la estrella parece roja. A pesar de su baja temperatura superficial T , todas las gigantes rojas tienen una alta luminosidad L debido a su enorme radio R . Este resultado se deduce de la Ley de Radiación de Stefan-Boltzmann para la radiación del cuerpo negro:

ANEXO C: TEXTO PREPARADO FESTESE001V

$$L = \sigma \times 4 \times \pi \times R^2 \times T^4$$

donde σ es la constante de Stefan-Boltzmann. Valores típicos para las gigantes rojas son $R \sim 10^2 \times R_{\text{solar}}$, $T \sim (3 \text{ hasta } 4) \times 10^3 \text{ K}$, de forma que L es aproximadamente $10^3 \times L_{\text{solar}}$.

Cuando ya no pueden mantenerse los procesos avanzados de fusión en el núcleo estelar, el núcleo se colapsa de nuevo. Otra vez la temperatura del núcleo se incrementa y ahora se expulsan las capas externas de la estrella. Las llamadas nebulosas planetarias se forman a partir de los restos de las capas externas (ver Ejercicio 3 de esta serie de ejercicios de astronomía de ESA1/ESO1). El núcleo colapsado está muy caliente (blanco) y la estrella es muy pequeña. A estas estrellas se las denominan muy adecuadamente enanas blancas y es la fase final de la vida de una estrella normal de tipo solar.

Para hacer una estimación aproximada de la relación entre la luminosidad L y la temperatura superficial T de todas las estrellas de la secuencia principal, veamos el diagrama H-R (Fig 6). La línea más o menos recta de la Secuencia Principal se extiende aproximadamente un orden de magnitud en temperatura: (3×10^3 hasta 3×10^4) K. El rango de luminosidades se extiende aproximadamente seis órdenes de magnitud: (10^{-2} hasta 10^4) $\times L_{\text{solar}}$. Por tanto podemos estimar de forma aproximada: L proporcional T^6 para las estrellas de la secuencia principal.

Para dar algunos ejemplos:

Una estrella de gran masa de la secuencia principal con una temperatura superficial de aproximadamente $T_{\text{estrella}} = 10^4 \text{ K}$ tiene una

ANEXO C: TEXTO PREPARADO FESTESE001V

luminosidad de aproximadamente $L_{\text{estrella}} = (10 \text{ dividido por } 5 \text{ coma } 8) \cdot L_{\text{solar}}$, o aproximadamente 26 veces la luminosidad del Sol. (La luminosidad del Sol tiene un valor estándar de 1 en la escala de luminosidad).

Una estrella de baja masa con $T_{\text{estrella}} = 3 \text{ coma } 5 \times 10^3 \text{ K}$ tiene una luminosidad de aproximadamente sólo el 5 por ciento de la luminosidad del Sol.

Toda la información que podemos extraer de las estrellas proviene de la radiación que recibimos de ellas. Como se explica en la sección Herramientas Astronómicas, diferentes filtros y sistemas de color se pueden usar para medir el brillo de una estrella. En este ejercicio usamos una imagen B y una imagen V. En tu análisis de estas imágenes encontrarás las magnitudes aparentes m_B y m_V de una muestra de estrellas en el cúmulo. Luego podrás calcular los valores m_B y m_V (el índice de color B menos V). Finalmente podrás determinar la temperatura superficial de las estrellas (ver la sección Herramientas Astronómicas).

Un cúmulo es un grupo de estrellas. La vida de un cúmulo está determinada por la vida de los diferentes tipos de estrellas que lo componen. Las observaciones han mostrado que en los cúmulos globulares queda muy poca cantidad de gas y polvo, de manera que es muy raro el nacimiento de nuevas estrellas en tales cúmulos. Las estrellas que vemos en un cúmulo globular son todas “adultas” y han evolucionado de formas diferentes en función de sus masas.

La mayoría de las estrellas de baja masa se encuentran en la Secuencia Principal. Esto se debe a que las estrellas de baja masa van consumiendo su energía muy lentamente. Queman sus reservas de hidrógeno poco a poco y continuarán de este modo durante miles de millones de años. Por consiguiente, estarán en la Secuencia Principal durante mucho tiempo.

ANEXO C: TEXTO PREPARADO FESTESE001V

Por el contrario, las estrellas más pesadas del cúmulo ya han transformado el hidrógeno de sus núcleos y se han convertido en gigantes rojas. Todo esto sucedió hace mucho tiempo, así que hoy en día no hay estrellas calientes masivas, lo que deja un hueco en la mitad superior de la Secuencia Principal (ver Fig 7) . Estas estrellas se encuentran ahora en el área diagonal que comienza en la Secuencia Principal y se desplaza hacia la parte superior derecha del diagrama conocida como rama de las gigantes rojas.

El punto en el cual se encuentran la Secuencia Principal y la Rama de Gigante Roja se llama punto de giro¹, y nos da una pista importante a la hora de determinar la edad del cúmulo. En el siguiente ejercicio, medirás las coordenadas de este punto sobre tu diagrama y determinarás la edad de M12 .

ANEXO D: La Aplicación Informática PAFE

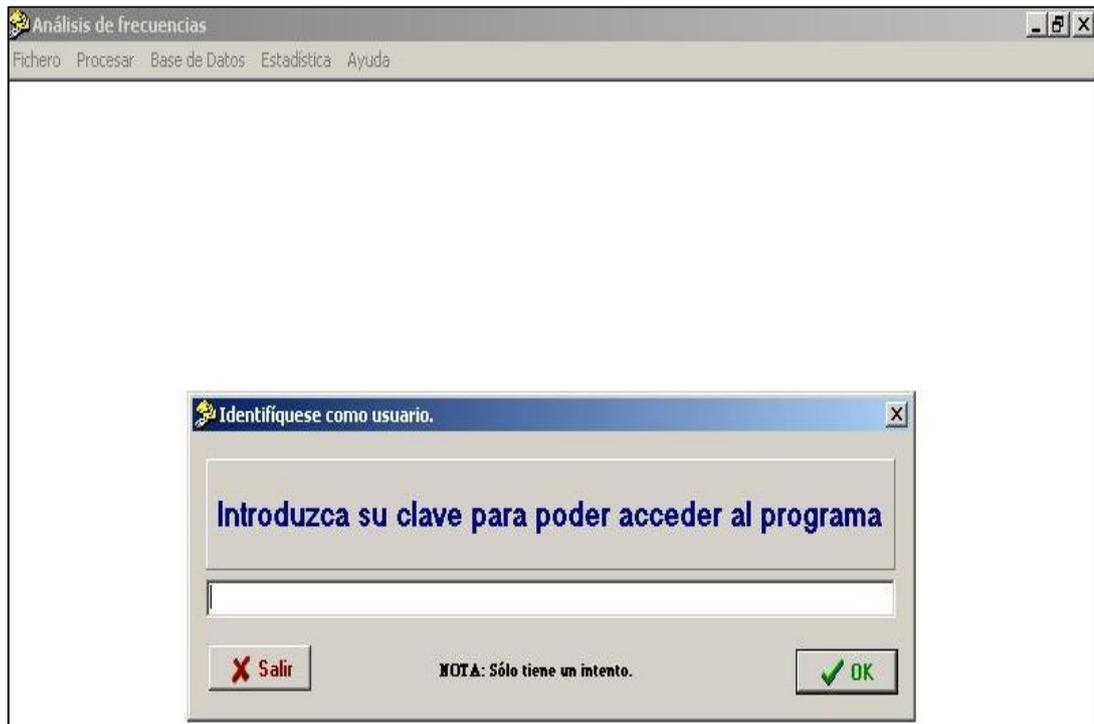


Figura 118.- Entrada en el Análisis de Frecuencias del PAFE: introducción de la clave de usuario.



Figura 119.- Análisis de Frecuencias: selección del texto a procesar.

ANEXO D: LA APLICACIÓN INFORMÁTICA PAFE

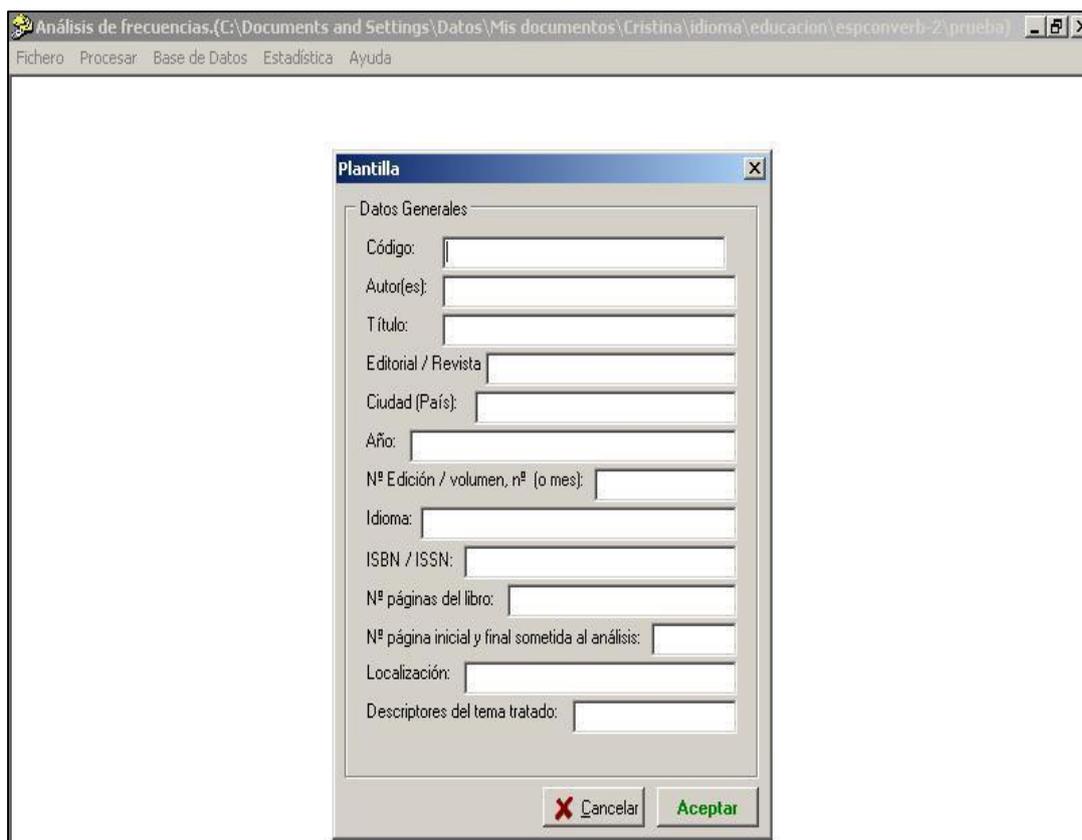


Figura 120.- Análisis de Frecuencias: plantilla de identificación del texto a procesar.

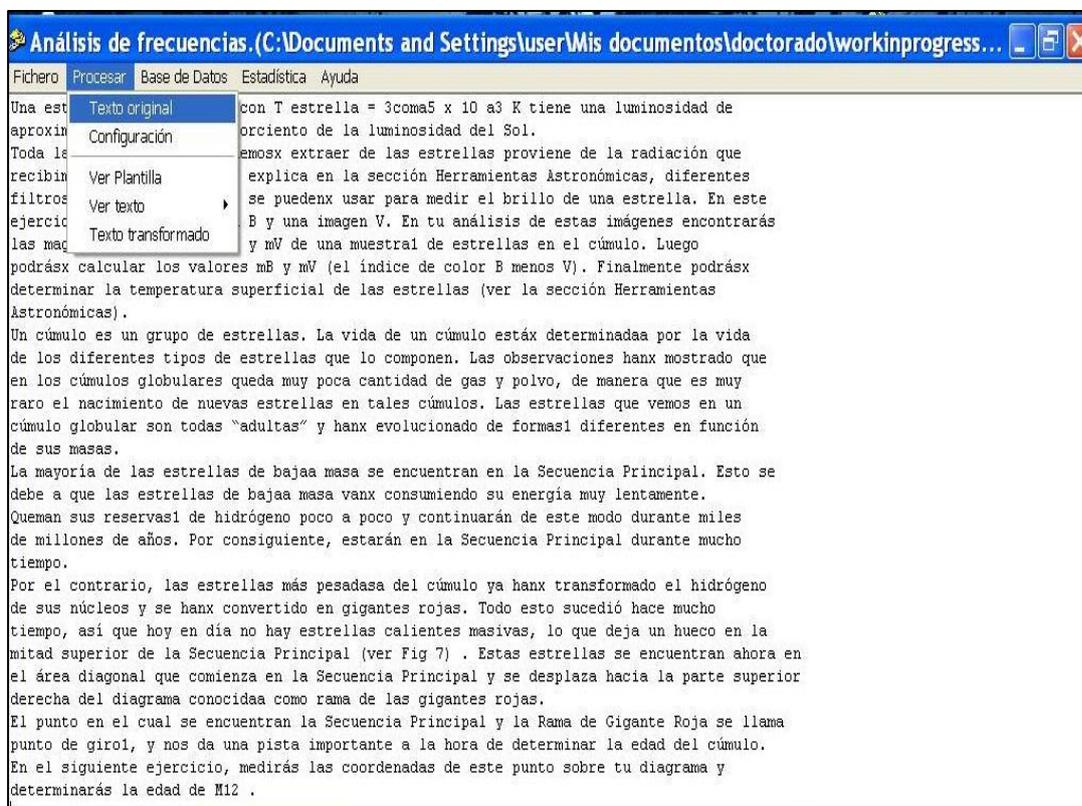


Figura 121.- Análisis de Frecuencias: menú para procesar el texto.

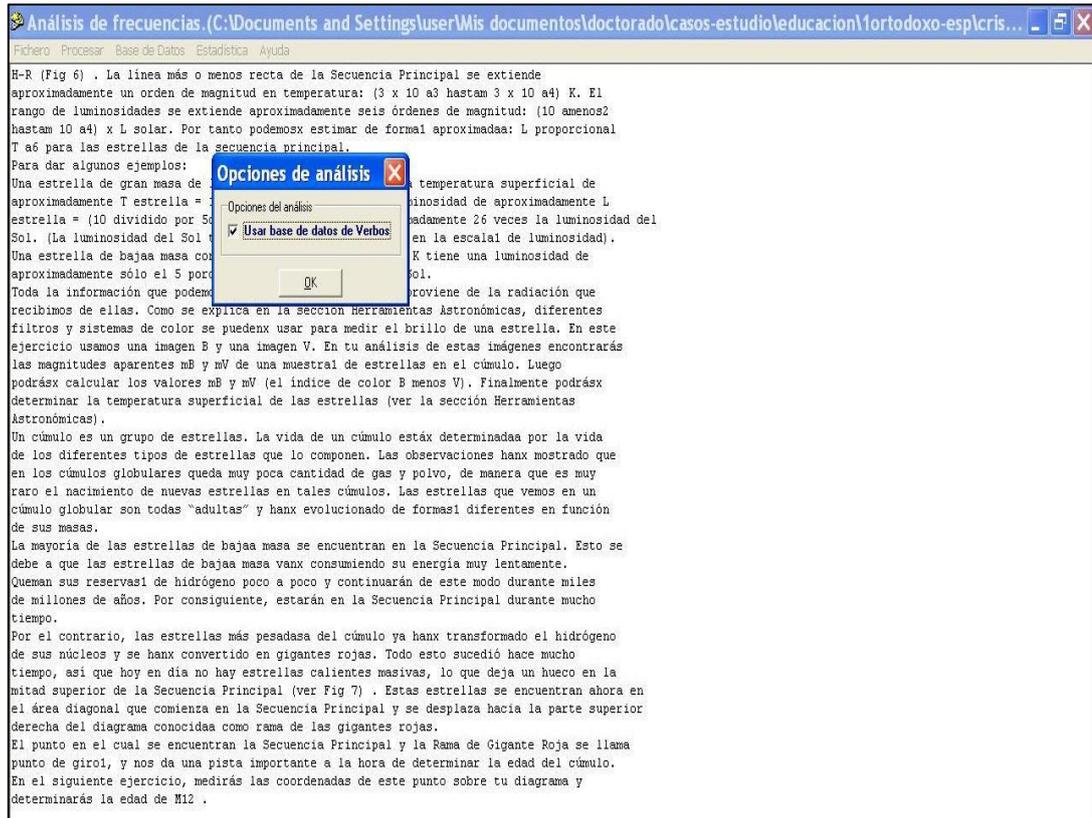


Figura 122.- Análisis de Frecuencias: elección de considerar los verbos dentro del análisis.



Figura 123.- Análisis de Frecuencias: selección de la base *Eliminar* para la palabra que se indica. Esta es también la ventana que se abre para la clasificación previa en cualquiera de las bases de datos. Sólo en el caso de elegir *Cambiar a* o *Verbos* cambia la ventana abierta.

ANEXO D: LA APLICACIÓN INFORMÁTICA PAFE

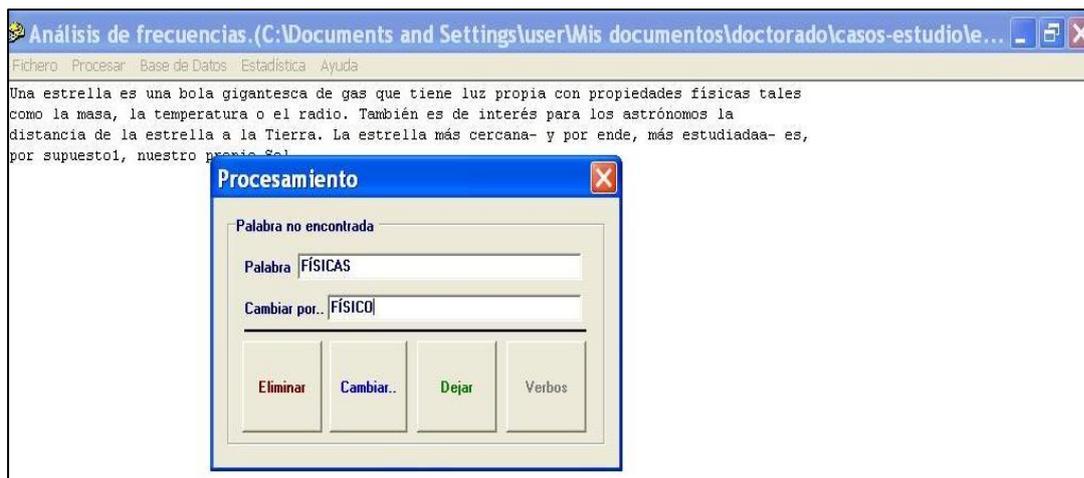


Figura 124.- Análisis de Frecuencias: ventana de lematización de formas no verbales (en *Cambiar a*) y de formas verbales (en *Verbos*).



Figura 125.- Análisis de Frecuencias: menú de las bases de datos.

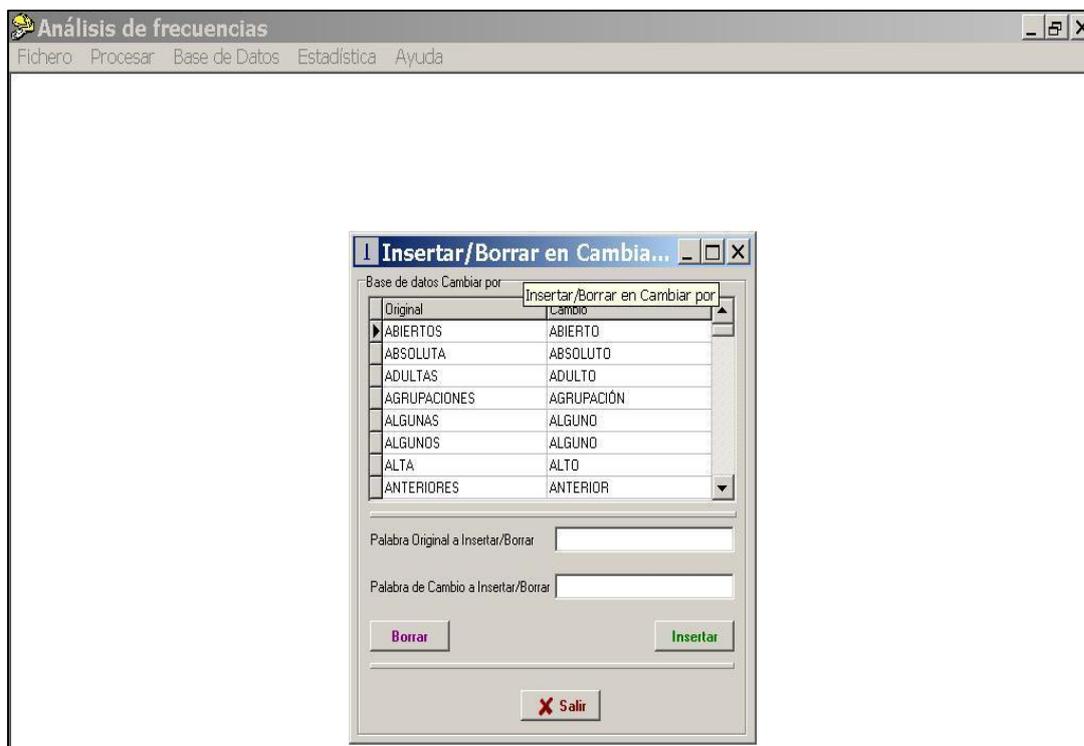


Figura 126.- Análisis de Frecuencias: modificación de las bases de datos.

ANEXO D: LA APLICACIÓN INFORMÁTICA PAFE

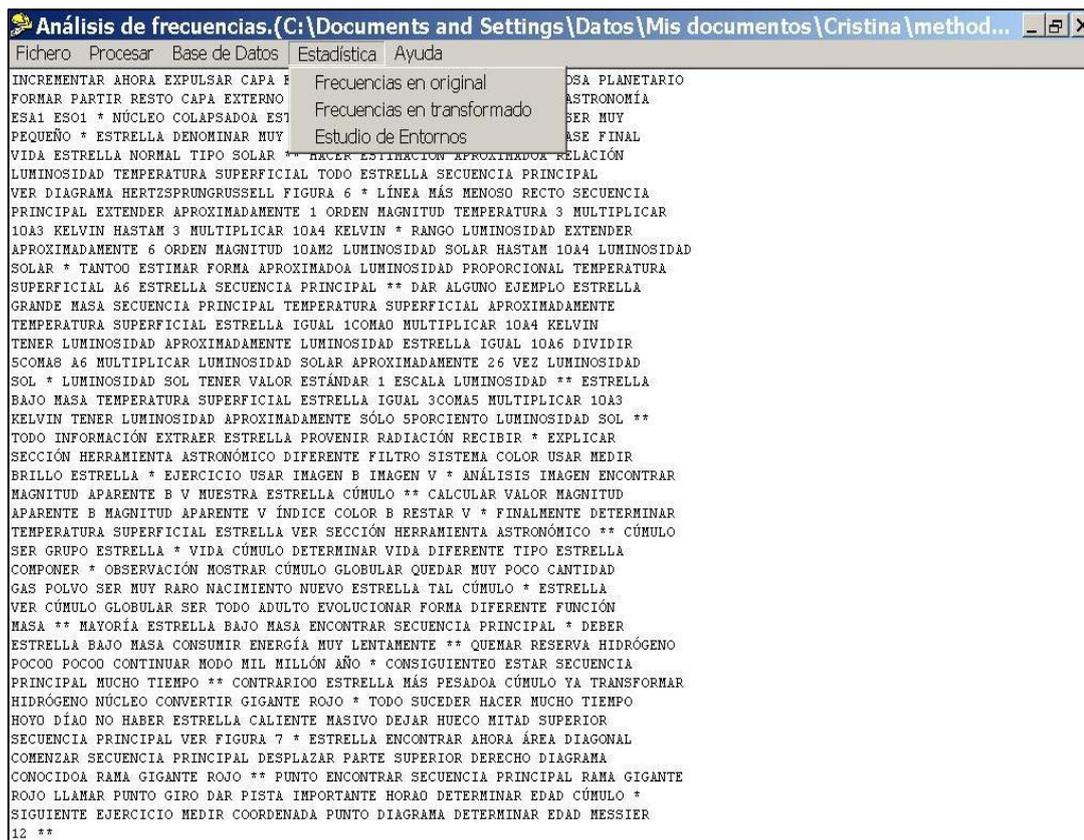


Figura 127.- Análisis de Frecuencias: texto lematizado o transformado, menú de frecuencias y salto al Estudio de Entorno.

The screenshot shows the same application window, but now displaying a list of words and their frequencies. The list includes words like 'TAN', 'TELESCOPIO', 'TEORÍAS', etc., with their absolute frequency (1) and relative frequency (0.419). At the bottom, it shows the total number of words (2388) and the number of different words (670).

Nº	Palabra	Frecuencia Absoluta	Frecuencia Relativa
632	TAN	1	0.419
633	TELESCOPIO	1	0.419
634	TEORÍAS	1	0.419
635	TERMONUCLEAR	1	0.419
636	TIERRA	1	0.419
637	TIPO	1	0.419
638	TODA	1	0.419
639	TODOS	1	0.419
640	TRANSFORMADO	1	0.419
641	TRAVÉS	1	0.419
642	TUVO	1	0.419
643	TÉRMICA	1	0.419
644	TÉRMINO	1	0.419
645	TÍPICOS	1	0.419
646	UNAS	1	0.419
647	UNIVERSO	1	0.419
648	USA	1	0.419
649	USAMOS	1	0.419
650	USAR	1	0.419
651	VALIOSA	1	0.419
652	VALIOSO	1	0.419
653	VANI	1	0.419
654	VARIAS	1	0.419
655	VARIOS	1	0.419
656	VE	1	0.419
657	VEAMOS	1	0.419
658	VECES	1	0.419
659	VECINDAD	1	0.419
660	VEMOS	1	0.419
661	VERSE	1	0.419
662	VIOLENTA	1	0.419
663	VISTA	1	0.419
664	WILLIAM	1	0.419
665	ÁREA	1	0.419
666	ÁREAS	1	0.419
667	ÁTOMO	1	0.419
668	ÍNDICE	1	0.419
669	ÓRDENES	1	0.419
670	ÚTIL	1	0.419

Nº de palabras en total:2388
Nº de palabras diferentes:670

Figura 128.- Análisis de Frecuencias: listado de la frecuencia absoluta y la frecuencia relativa.

ANEXO D: LA APLICACIÓN INFORMÁTICA PAFE



Figura 129.- Estudio de Entorno: acceso a la aplicación desde su icono.



Figura 130.- Estudio de Entorno: selección del texto transformado a estudiar.

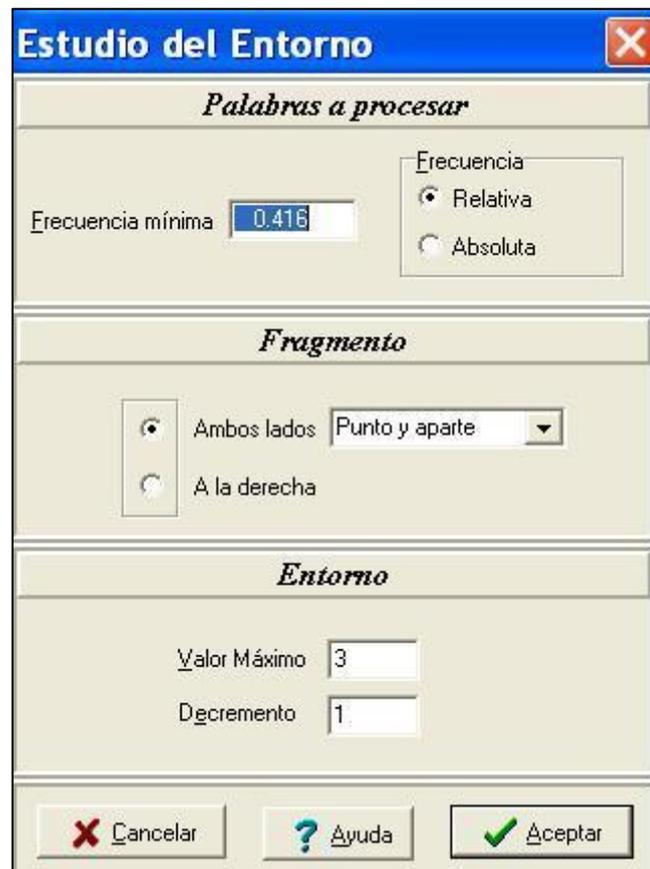


Figura 131.- Estudio de Entorno: determinación de las características del estudio.

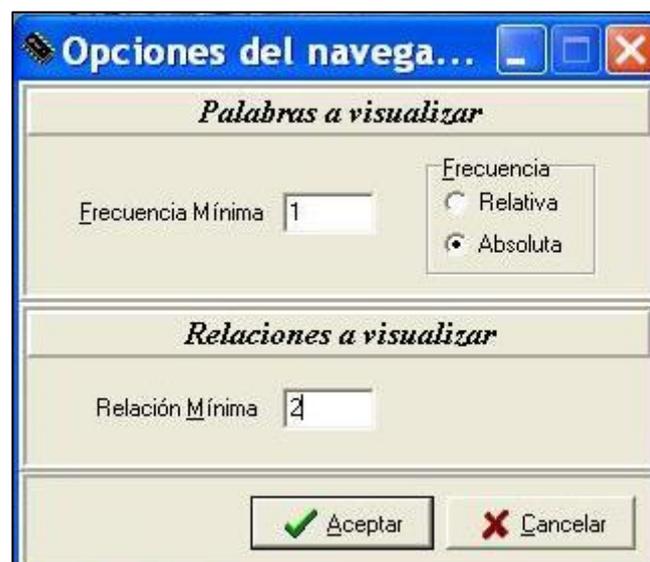


Figura 132.- Estudio de Entorno: selección del valor de relación mínimo a ver.

ANEXO D: LA APLICACIÓN INFORMÁTICA PAFE

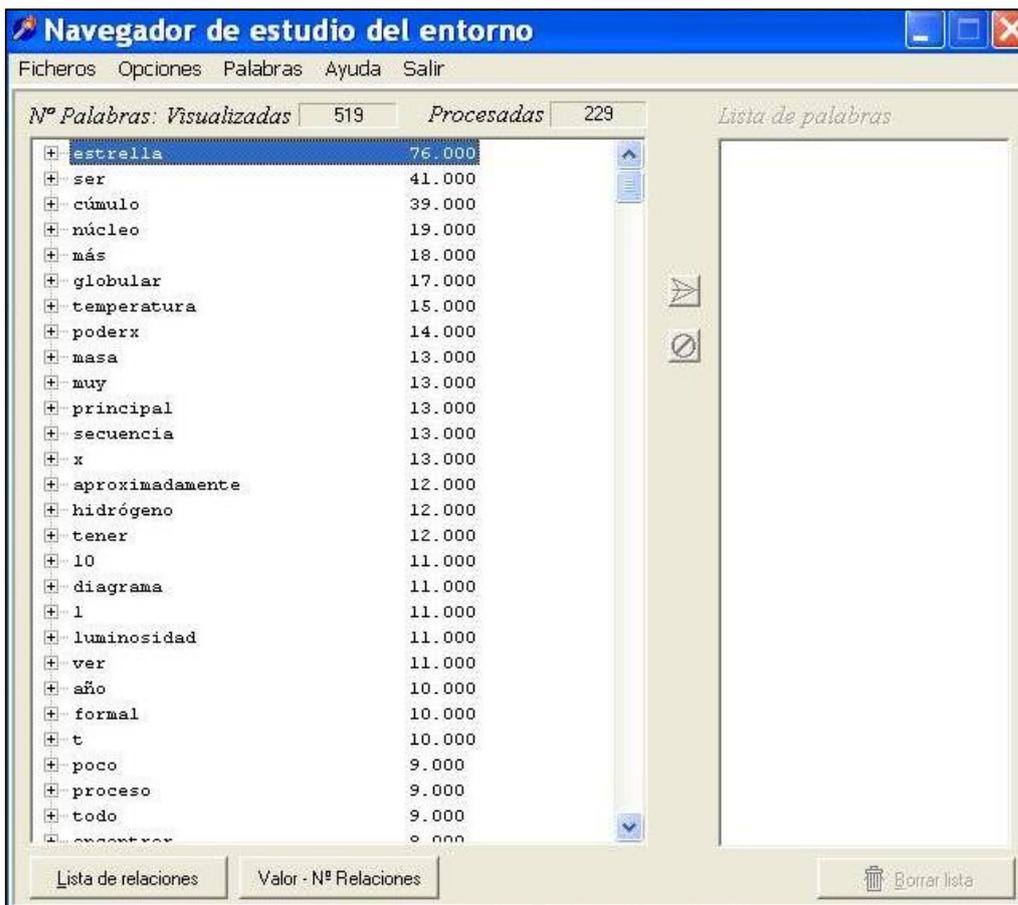


Figura 133.- Estudio de Entorno: listado de ULL del sistema definido.



Figura 134.- Estudio de Entorno: Lista de relaciones, archivo *.qrp con el listado de ULL del sistema definido.

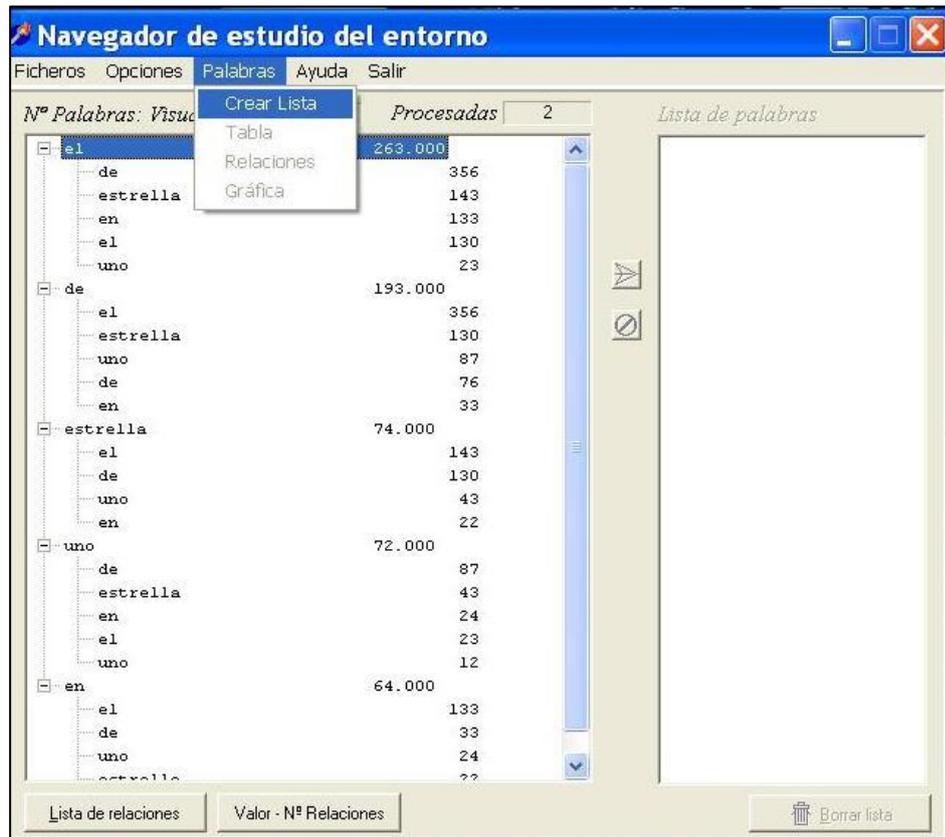


Figura 135.- Estudio de Entorno: crear lista de palabras.

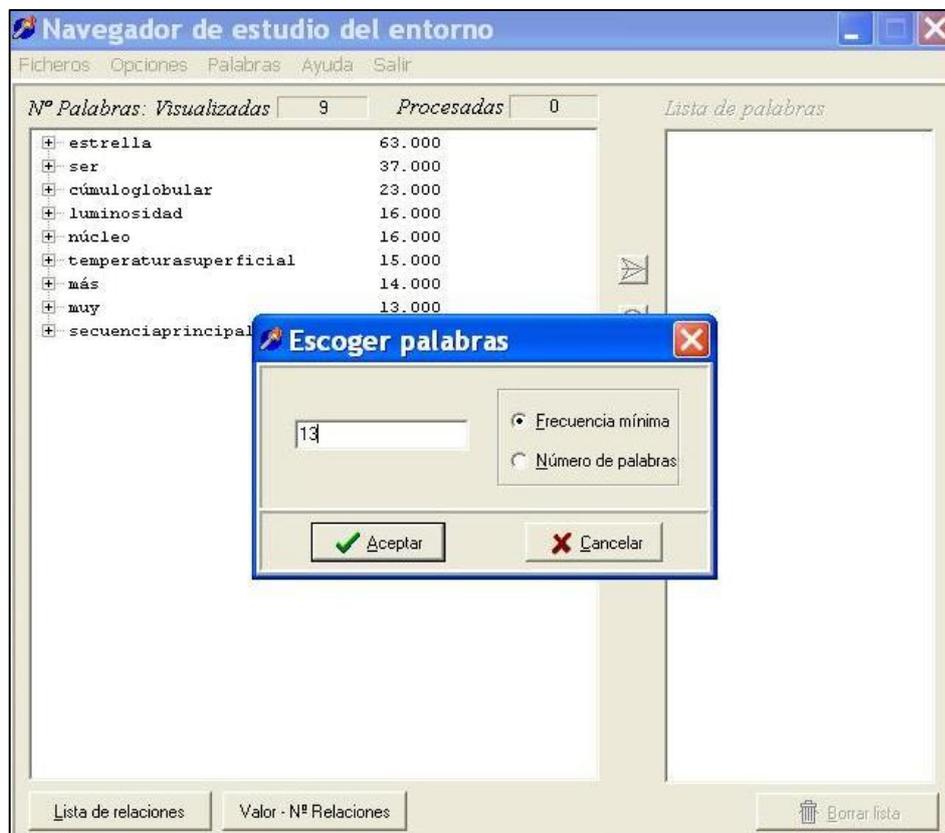


Figura 136.- Estudio de Entorno: crear lista de palabras usando la frecuencia mínima.

ANEXO D: LA APLICACIÓN INFORMÁTICA PAFE

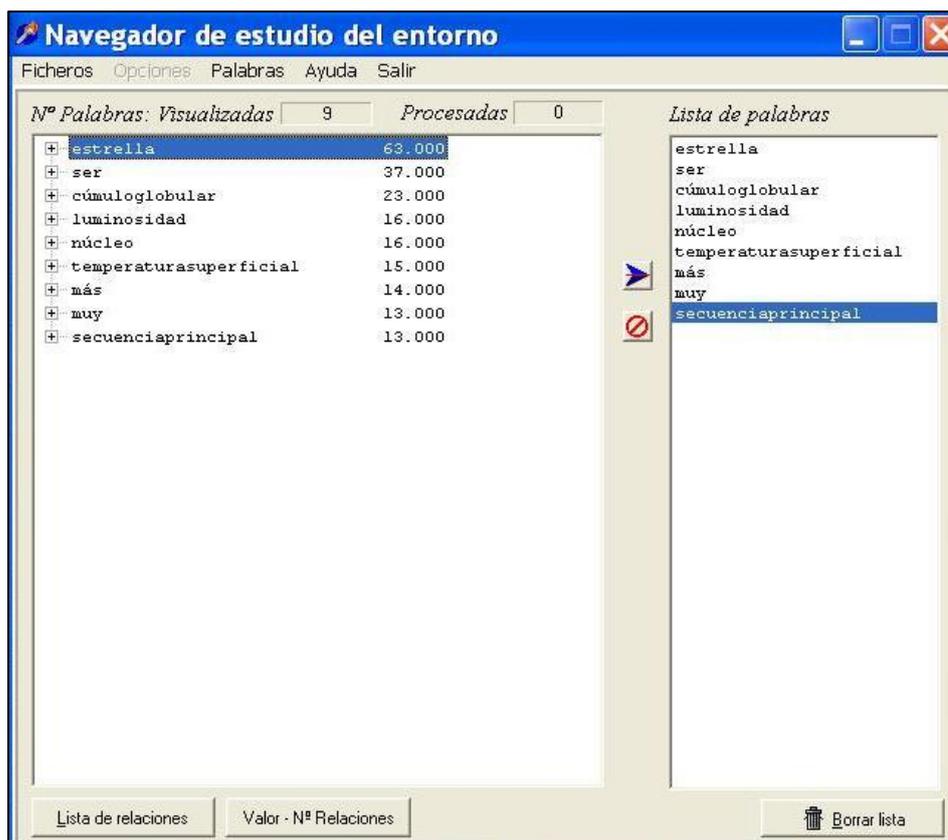


Figura 137.- Estudio de Entorno: lista de palabras en el cuadro situado a la derecha.

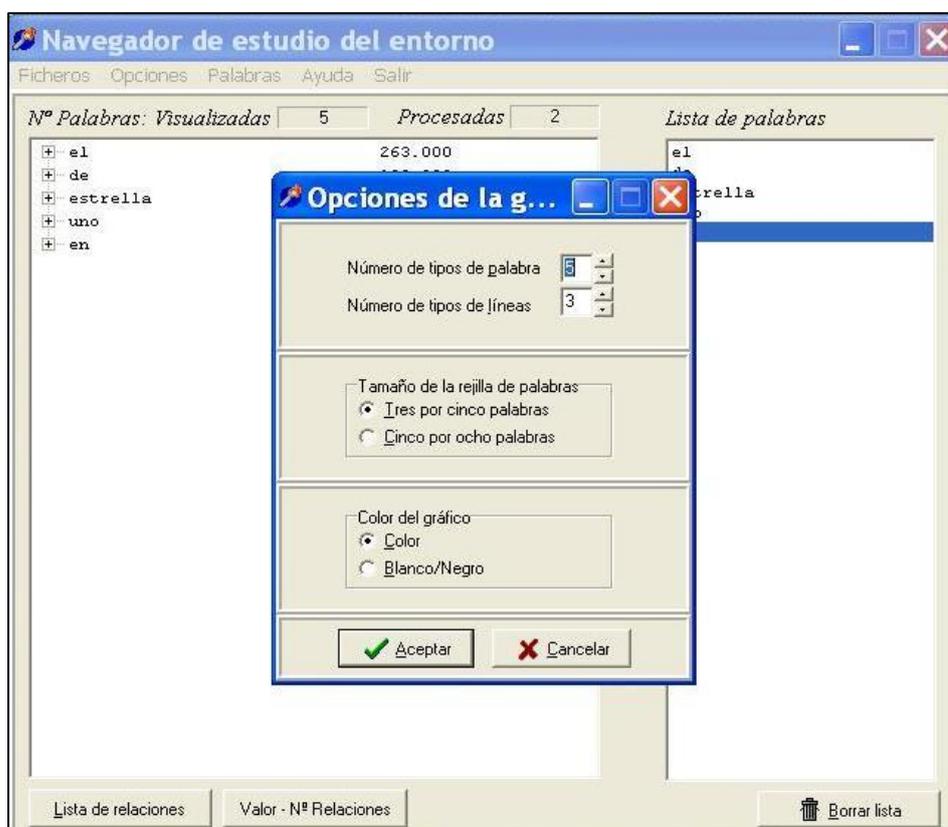


Figura 138.- Estudio de Entorno: características de la gráfica.



Figura 139.- Estudio de Entorno: selección de intervalos de frecuencias y relaciones.

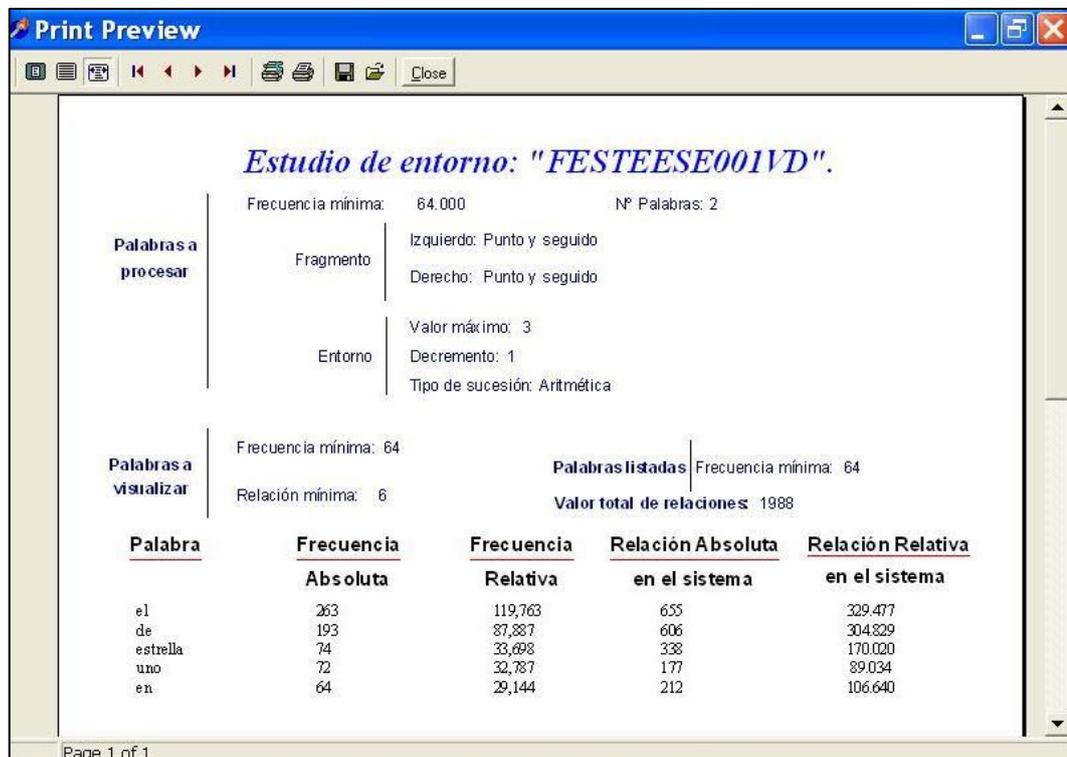


Figura 140.- Estudio de Entorno: datos de las ULL del sistema estudiado.

ANEXO E: Palabras y Locuciones Susceptibles de Ser Eliminadas

En otros estudios no publicados realizados con el PAFE se ha encontrado la posibilidad de eliminar otras palabras además de las funcionales. Se trataría de palabras que no aportan nada al significado: adverbios, alguna forma verbal y locuciones.

Las palabras susceptibles de ser eliminadas son: *a, además, ahí, ahora, al, allá, allí, antes, apenas, aquel, aquello, aquella, aquellos, aquellas, aquél, aquello, aquélla, aquéllos, aquéllas, aquí, así, asimismo, aun, aún, aunque, como, cómo, con, consigo, cual, cuales, cuál, cuáles, cuando, cuándo, cuanto, cuanta, cuantos, cuantas, cuánto, cuánta, cuántos, cuántas, cuyo, cuya, cuyos, cuyas, de del, desde, donde, dónde, e, el, él, ello, ella, ellos, ellas, en, entonces, entre, eso, esa, ese, esos, esas, eses, esto, esta, este, estos, estas, éste, etcétera, finalmente, generalmente, la, las, lo, los, le, les, luego, mas, mi, mis, mientras, nos, nosotros, nosotras, nuestro, nuestros, nuestra, nuestras, o, ó, para, pero, por, porque, porqué, pues, que, qué, quien, quizá, quizás, se, si, sí, sea, siempre, siguiente, siguientes, siquiera, sobre, solamente, su, sus, suyo, suyos, suya, suyas, tal, tales, también, te, todavía, tras, tu, tus, tú, un, unos, una, unas, y, ya, yo*. En este listado se incluyen palabras funcionales, adverbios y alguna forma verbal.

Además se han detectado locuciones susceptibles de ser eliminadas: *acontinuación, ahorabien, alfinalcabo, almenos, amedidaque, amenudo, amodode,*

ANEXO E: PALABRAS Y LOCUCIONES SUSCEPTIBLES DE SER ELIMINADAS

amodoque, apartede, comoseve, comosigue, comoyaesconocido, comoyahemosdicho, comoyahemosindicado, comoyahemosvisto, comoyasabemos, comoyasehavisto, dadoque, dealgunamanera, deesemodo, destaforma, deestamanera, deestemodo, deformaque, dehecho, deigualmanera, demanera, demanerageneral, demaneraque, demodo, demodoque, denuevo, depaso, desde luego, deunaforma u otra, dichodeotramanera, dichodeotromodo, elhechode, elhechodeque, enabsoluto, enbreve, encambio, enconcreto, enconsecuencia, encualquiercaso, endefinitiva, enefecto, enesencia, enespecial, enfin, engeneral, enocasiones, enparticular, enprimerlugar, enprincipio, enrealidad, enresumen, ensegundolugar, entérminos, entodocaso, entreotrascosas, enunapalabra, esdecir, esmás, esto es, hastaahora, hastaentonces, másadelante, másaún, másbien, másconcretamente, másomenos, nisiquiera, noobstante, obien, osea, otraparte, porconsiguiente, porejemplo, porende, porlomenos, porotraparte, porsímismo, porsísolo, porsuparte, porsupuesto, portanto, porúltimo, porunaparte, porunlado, puesbien, puestoque, quieredecir, sibien, sinembargo, sobretodo, talvez, tardeo temprano, unavez más, verapéndice, yadeporsí, yasea. En este listado aparecen palabras que fuera de las locuciones serían susceptibles de entrar en el estudio, como por ejemplo *medida*.

ANEXO F: Listado de Palabras Funcionales

Este, ese, aquel, esta, esa, aquella, estos, esos, aquellos, estas, esas, aquellas. éste, ése, aquél, ésta, ésa, aquélla, éstos, éstos, aquéllos, éstas, éstas, aquéllas, esto, eso, aquello.

Mío, mía, míos, mías, tuyo, tuya, tuyos, tuyas, suyo, suya, suyos, suyas, nuestro, nuestra, nuestros, nuestras, vuestro, vuestra, vuestros, vuestras.

Que, cual, cuales, quien, quienes, cuyo, cuya, cuyos, cuyas, como, donde, cuando, cuanto, cuanta, cuantos, cuantas, cuál, cuáles, qué, cuán, cuándo, cómo, dónde, quién, quiénes, cuánto, cuánta, cuántos, cuántas.

Yo, nosotros, nosotras, tú, vosotros, vosotras, vos, él, ella, ello, ellos, ellas, me, nos, te, os, lo, la, le, los, las, les, se, mí, ti, sí, usted, ustedes, conmigo, contigo, consigo.

Un, uno, una, unos, unas.

El, la los, las, lo.

Mi, tu, su, mis, tus, sus.

A, ante, bajo, con, contra, de, desde, durante, en, entre, hacia, hasta, mediante, para, por, según, sin, sobre, tras, cabe, so.

Y, ni, o, u, e, pero, sino, empero, conque, luego, que, porque, si.

Al, del.

ANEXO G: Texto Lematizado FESTESE001V

En el texto lematizado o transformado, generado por el programa PAFE, una estrella denota una marca de punto y seguido y dos estrellas una marca de punto y aparte. El texto que se presenta a continuación está solamente lematizado, dado que no se ha hecho uso de la base *Eliminar*.

UNO ESTRELLA SER UNO BOLA GIGANTESCO DE GAS QUE
TENER LUZ PROPIO CON PROPIEDAD FÍSICO TAL COMO EL MASA EL
TEMPERATURA O EL RADIO * TAMBIÉN SER DE INTERÉS PARA EL
ASTRÓNOMO EL DISTANCIA DE EL ESTRELLA A EL TIERRA * EL
ESTRELLA MÁS CERCANO Y POR ENDE MÁS ESTUDIADOA SER POR
SUPUESTO1 NUESTRO PROPIO SOL ** EL LUZ EMITIDOA POR EL
MAYORÍA DE EL ESTRELLA SER UNO SUBPRODUCTO DEL PROCESO DE
FUSIÓN TERMONUCLEAR QUE SE PRODUCIR EN EL NÚCLEO INTERNO
DE EL ESTRELLA * UNO ESTRELLA NORMAL COMO EL SOL SE
COMPONER DE APROXIMADAMENTE UNO 74 PORCIENTO DE
HIDRÓGENO Y UNO 25 PORCIENTO DE HELIO SER EL RESTANTE 1
PORCIENTO UNO MEZCLA1 DE ELEMENTO MÁS PESADOA * EL
PROCESO DE FUSIÓN MÁS COMÚN EN UNO ESTRELLA COMO EL SOL
SER EL COMBUSTIÓN DEL HIDRÓGENO EN EL CUAL CUATRO NÚCLEO
DE HIDRÓGENO SE FUSIONAR PARA FORMAR UNOA NÚCLEO DE

ANEXO G: TEXTO LEMATIZADO FESTESE001V

HELIO * EL PROCESO SUCEDER EN VARIO ETAPA QUE SE ILUSTRAR EN EL FIG 2 * EN EL PRIMERO PASO1 DEL PROCESO DOS PROTÓN SE FUSIONAR PARA FORMAR DEUTERIO UNO FORMA1 PESADOA DEL HIDRÓGENO * ESTO SER UNO SUCESO MUY RARO INCLUSO PARA EL NÚCLEO DENSO DE EL ESTRELLA EN DONDE EL TEMPERATURA SER DE UNO POCO MILLÓN DE GRADO * SER POR ESTO POR EL QUE EL ESTRELLA COMO EL SOL NO EXPLOTAR EN UNO REACCIÓN VIOLENTO CUANDO COMENZAR EL PROCESO DE FUSIÓN SINO QUE PERMANECER EN ESTO FASE ESTABLE DE EL VIDA DE EL ESTRELLA DURANTE VARIO MIL DE MILLÓN DE AÑO * MIENTRAS EL ESTRELLA SER ESTABLE EL TEMPERATURA DE SU SUPERFICIE SU RADIO Y SU LUMINOSIDAD SER APROXIMADAMENTE CONSTANTEA * EL REACCIÓN NUCLEAR DEL NÚCLEO DE EL ESTRELLA GENERAR EL ENERGÍA JUSTO PARA MANTENER UNO EQUILIBRIO ENTRE EL PRESIÓN TÉRMICO EJERCIDOA HACIA FUERA Y EL FUERZA GRAVITATORIO EJERCIDOA HACIA DENTRO ** EL MASA DE UNOA ÁTOMO DE HELIO SER SÓLO EL 99COMA3 PORCIENTO DE EL MASA DE EL CUATRO NÚCLEO ORIGINAL DE HIDRÓGENO * EL PROCESO DE FUSIÓN CONVERTIR EL 0COMA7 PORCIENTO RESIDUAL DE EL MASA EN ENERGÍA MAYORITARIAMENTE EN FORMA1 DE LUZ * EL CANTIDAD DE ENERGÍA SE PODERX CALCULAR USAR EL FAMOSO ECUACIÓN DE EINSTEIN $E = M \times C^2$ * YA QUE C^2 SER UNO NÚMERO MUY GRANDE ESTO ECUACIÓN IMPLICAR QUE INCLUSO UNO PEQUEÑO CANTIDAD DE MATERIA PODERX CONVERTIR EN UNO FORMIDABLE CANTIDAD DE ENERGÍA * EL 0COMA7 PORCIENTO

ANEXO G: TEXTO LEMATIZADO FESTESE001V

RESIDUAL DE EL MASA DE EL CUATRO NÚCLEO DE HIDRÓGENO INVOLUCRADO EN UNO A SOLO REACCIÓN PODERX PARECER PEQUEÑO PERO CUANDO SE CONSIDERAR EL NÚMERO TOTAL DE REACCIÓN DE TODO EL PROCESO DE FUSIÓN HABERX IMPLICADO A UNO MASA TOTAL Y POR TANTO ENERGÍA CONSIDERABLE ** EL TÉRMINO CÚMULO DE ESTRELLA SE USAR PARA DOS TIPO DIFERENTE DE AGRUPACIÓN DE ESTRELLA CÚMULO ABIERTO A DE ESTRELLA Y CÚMULO GLOBULAR DE ESTRELLA ** EL CÚMULO ABIERTO A DE ESTRELLA SER COLECCIÓN NO COMPACTO DE ESTRELLA RELATIVAMENTE JOVEN QUE IR DESDE UNO CENTENAR HASTA UNO POCO MIL DE ESTRELLA * TENER UNO EDAD DE UNO POCO CIENTO DE MILLÓN DE AÑO UNO FRACCIÓN PEQUEÑO DEL TIEMPO TOTAL DE EL VIDA DE EL ESTRELLA UNO POCO MIL DE MILLÓN DE AÑO * ESTO CÚMULO SE ENCONTRAR EN EL DISCO DE NUESTRO GALAXIA EL VÍA LÁCTEO Y A MENUDO CONTENER NUBE DE GAS Y POLVO DONDE SE FORMAR NUEVO ESTRELLA * EL DIÁMETRO TÍPICO DE UNO CÚMULO ABIERTO A DE ESTRELLA SER DE APROXIMADAMENTE 30 AÑO LUZ 10 PARSEC ** ESTO SER UNO DE EL MÁS FAMOSO CÚMULO DE ESTRELLA DEL CIELO * EL PLÉYADES PODERX VER A SIMPLE VISTA INCLUSO DESDE EL MAYORÍA DE EL CIUDAD PESAR A EL CONTAMINACIÓN LUMÍNICO * SER UNO DE EL MÁS BRILLANTE Y MÁS PRÓXIMO CÚMULO ABIERTO A * EL CÚMULO DE EL PLÉYADES CONTENER MÁS DE 3000 ESTRELLA ESTAR A UNO DISTANCIA DE UNO 400 AÑO LUZ Y TENER UNO TAMAÑO DE SÓLO 13 AÑO LUZ CORTESÍA DE BRUNO STAMPFER Y RAINER EISENDLE ** EN EL HALO Y EN EL DISCO DE

ANEXO G: TEXTO LEMATIZADO FESTESE001V

NUESTRO VÍA LÁCTEO EXISTIR UNO POCO CIENTO DE CÚMULO ESFÉRICO Y COMPACTO LLAMADO A CÚMULO GLOBULAR Y QUE ESTARX LIGADO A GRAVITATORIAMENTE A NUESTRO GALAXIA ** CADA CÚMULO GLOBULAR SE COMPONER DE UNO GRUPO ESFÉRICO DE HASTA UNO A MILLÓN DE ESTRELLA Y TENER UNO DIÁMETRO TÍPICO DE UNO 100 AÑO LUZ * EL MAYORÍA DE EL CÚMULO GLOBULAR SER MUY VIEJO Y EL MÁS PROBABLE SER QUE SER ANTERIOR A EL FORMACIÓN DE EL PROPIO GALAXIA QUE TENER LUGAR HACER APROXIMADAMENTE 12 MIL MILLÓN DE AÑO CUANDO EL MAYORÍA DEL MATERIAL PROTOGALÁCTICO SE DEPOSITAR EN EL DISCO ** MUCHO CÚMULO GLOBULAR PROBABLEMENTE SE HABERX DESTRUIR CON EL PASO DE EL MIL DE MILLÓN DE AÑO POR EL REPETIDO A COLISIÓN E INTERACCIÓN ENTRE SU ESTRELLA Y CON EL VÍA LÁCTEO * EL CÚMULO GLOBULAR SUPERVIVIENTE SER MÁS VIEJO QUE CUALQUIER OTRO ESTRUCTURA DE NUESTRO VÍA LÁCTEO ** ESTO ILUSTRACIÓN DAR UNO PERSPECTIVA GENERAL DE EL GALAXIA VÍA LÁCTEO * SE INDICAR EL DIFERENTE COMPONENTE DE ESTO COMPLICADO A SISTEMA DE ESTRELLA GAS Y POLVO * EL PLANO DEL DISCO SE ENCONTRAR A EL LARGO DE EL LÍNEA CENTRAL HORIZONTAL * EL CÚMULO GLOBULAR SE DISTRIBUIR EN UNO HALO ESFÉRICO ALREDEDOR DEL CENTRO DE EL GALAXIA * SE CREER QUE ESTO DISTRIBUCIÓN TENER QUE VER CON EL HECHO DE QUE ESTO CÚMULO DE ESTRELLA SE FORMAR MUY PRONTO EN EL HISTORIA DE EL GALAXIA ** EL ESTUDIO ASTROFÍSICO DE EL CÚMULO GLOBULAR SER UNO PARTE IMPORTANTE DEL PROGRAMA

ANEXO G: TEXTO LEMATIZADO FESTESE001V

DE INVESTIGACIÓN DE EL COMUNIDAD ASTRONÓMICO INTERNACIONAL * ESTO CÚMULO DE ESTRELLA SER IMPORTANTE NO SÓLO COMO UNO VALIOSO BANCO DE PRUEBA1 DE VALOR PARA EL TEORÍA DE ESTRUCTURA Y EVOLUCIÓN ESTELAR SINO TAMBIÉN PORQUE SE ENCONTRAR ENTRE EL POCO OBJETO DE EL GALAXIA PARA EL CUAL SE PODERX DETERMINAR SU EDAD DE UNO FORMA1 RELATIVAMENTE PRECISOA * POR SER TANTO LONGEVO PROPORCIONAR UNO LÍMITE INFERIOR MUY ÚTIL A EL EDAD DEL UNIVERSO * EL DISTRIBUCIÓN DE SU EDAD Y EL CORRELACIÓN ENTRE EL EDAD DE UNO CÚMULO Y EL ABUNDANCIA EN ÉL DE EL DISTINTO ELEMENTO QUÍMICO HACER DE ESTO SISTEMA UNO PRUEBA1 VALIOSO EN EL PROCESO DE FORMACIÓN ESTELAR ** TODO EL ESTRELLA QUE COMPONER UNO CÚMULO GLOBULAR COMPARTIR UNO HISTORIA COMÚN Y DIFERIR ENTRE SÍ SÓLO EN SU MASA * POR EL TANTO EL CÚMULO GLOBULAR SER LUGAR IDEAL PARA ESTUDIAR EL EVOLUCIÓN DE EL ESTRELLA * EN EL EJERCICIO SIGUIENTE DETERMINAR ALGUNO DE EL PROPIEDAD DE UNO CÚMULO GLOBULAR PARTICULAR EL MESSIER 12 ** EL CÚMULO GLOBULAR MESSIER 12 O M12 TAMBIÉN LLAMADOA NGC 6218 SERX DESCUBRIR EN 1764 POR CHARLES MESSIER DE FORMA1 QUE SER EL OBJETO DUODÉCIMO DE MESSIER * COMO MUCHO OTRO CÚMULO GLOBULAR MESSIER EL DESCRIBIR COMO UNO NEBULOSA SIN ESTRELLA UNO CONSECUENCIA DEL MODESTO PODER1 DE RESOLUCIÓN DE SU TELESCOPIO * WILLIAM HERSCHEL SER EL PRIMERO QUE RESOLVER EL CÚMULO EN ESTRELLA INDIVIDUAL EN

1783 ** M12 SE ENCONTRAR EN EL CONSTELACIÓN DE OFIUCO Y SE
PODERX VER CON PRISMÁTICO DESDE LUGAR CON MUY BAJOA
POLUCIÓN LUMÍNICO * EL MAGNITUD VISIBLE DEL CONJUNTO DEL
CÚMULO GLOBULAR SER 6COMA7 VER MÁS SOBRE MAGNITUD EN EL
SECCIÓN HERRAMIENTA ASTRONÓMICO PÁGINA 2 Y EL ESTRELLA
MÁS BRILLANTE DEL CÚMULO TENER UNO MAGNITUD VISIBLE DE 12
** EL NGC NUEVO CATÁLOGO GENERAL SERX PUBLICAR EN 1888 * EN
ÉL APARECER UNO LISTA DE CÚMULO GLOBULAR Y CÚMULO
ABIERTO A DE ESTRELLA NEBULOSA PLANETARIO Y DIFUSO
REMANENTE DE SUPERNOVA GALAXIA DE TODO EL TIPO E INCLUSO
ALGUNO ERROR QUE NO CORRESPONDER A NINGUNO OBJETO ** SE
LLAMAR DIAGRAMA DE HERTZSPRUNGRUSSELL O DE FORMA1 MÁS
CONCISO DIAGRAMA HR A EL GRÁFICO QUE MOSTRAR EL
LUMINOSIDAD L O MAGNITUD ABSOLUTO M FRENTE A EL
TEMPERATURA SUPERFICIAL T DE EL ESTRELLA * EL FIG 6 MOSTRAR
UNO EJEMPLO GENERAL QUE SE HABERX CONSTRUIR A PARTIR DE
OBSERVACIÓN DE ESTRELLA EN EL VECINDAD DE EL CÚMULO
DONDE SE CONOCER EL DISTANCIA A PARTIR DE EL MEDIDA1 DE
HIPPARCOS * EL TEMPERATURA SUPERFICIAL T DE UNO ESTRELLA SE
PODERX DERIVAR DE EL VALOR MEDIDOA DE SU COLOR MB MENOS
MV VER EL SECCIÓN HERRAMIENTA ASTRONÓMICO ** SE VER
CLARAMENTE DEL DIAGRAMA HR QUE EL MEDIDA1 L T PARA
DIFERENTE ESTRELLA FORMAR UNO CURIOSO PATRÓN CUANDO SE
COLOCAR SOBRE EL DIAGRAMA * EL ESTRELLA SE CONCENTRAR EN
ÁREA ESPECÍFICO INDICADOA EN EL FIGURA * EL DIAGRAMA HR NOS

ANEXO G: TEXTO LEMATIZADO FESTESE001V

DAR EL CLAVE PARA COMPRENDER COMO EL ESTRELLA
EVOLUCIONAR EN EL TIEMPO * EL ESTRELLA DEPENDER DE SU MASA
SE MOVER A TRAVÉS DEL DIAGRAMA A EL LARGO DE CAMINO
DIFERENTE ** EL ESTRELLA PASAR EL MAYOR PARTE DE SU VIDA EN
EL SECUENCIA PRINCIPAL QUEMAR EL HIDRÓGENO LENTAMENTE EN
UNO ESTADO DE EQUILIBRIO ESTABLE * OBVIAMENTE ESTO SER EL
RAZÓN POR EL QUE EL MAYORÍA DE EL ESTRELLA SE LOCALIZAR EN
EL SECUENCIA PRINCIPAL APROXIMADAMENTE UNO LÍNEA RECTO
DESDE EL VÉRTICE SUPERIOR IZQUIERDO AL VÉRTICE INFERIOR
DERECHO EN EL DIAGRAMA ** CUANDO EL SUMINISTRO1 DE
HIDRÓGENO EN EL NÚCLEO DE EL ESTRELLA SE AGOTAR EL
COMBUSTIÓN DEL HIDRÓGENO YA NO SER POSIBLE * ESTO SER EL
FINAL DE EL FASE DE EL SECUENCIA PRINCIPAL DE EL VIDA DE EL
ESTRELLA Y EL EQUILIBRIO ENTRE EL PRESIÓN DEL GAS Y EL
CONTRACCIÓN GRAVITATORIO EN EL NÚCLEO ESTELAR YA NO SER
ESTABLE * EL FUSIÓN DEL HIDRÓGENO TENER LUGAR ENTONCES EN
EL CAPA CIRCUNDANTE MIENTRAS EL NÚCLEO COMENZARX A
CONTRAER * AL CONTRAER EL NÚCLEO SE ELEVAR SU PRESIÓN Y SU
TEMPERATURA CENTRAL DE MANERA QUE EL NÚCLEO DE HELIO DEL
NÚCLEO DE EL ESTRELLA COMENZARX A FUSIONAR Y FORMAR
ELEMENTO MÁS PESADOA * ESTO CICLO SE PODERX REPETIR USAR
ELEMENTO CADA VEZ MÁS PESADOA A MEDIDOA QUE EL ELEMENTO
MÁS LIGERO SE IRX ACABAR EN EL NÚCLEO * DURANTE ESTO FASE
EL ESTRELLA APARECER COMO UNO GIGANTE ROJO * TAL ESTRELLA
SE MOSTRAR EN EL DIAGRAMA HR FUERA DE EL LÍNEA DE EL

ANEXO G: TEXTO LEMATIZADO FESTESE001V

SECUENCIA PRINCIPAL EN EL PARTE SUPERIOR DERECHO * EL AUMENTO DE LA TEMPERATURA CENTRAL HACER QUE EL CAPA EXTERNO DE LA ESTRELLA SE EXPANDIR Y ENFRIAR DE FORMA QUE LA TEMPERATURA SUPERFICIAL DECAER * EL CONJUNTO DE LA ESTRELLA LLEGAR A SER MUY GRANDE Y DEBER A LA INFERIOR TEMPERATURA SUPERFICIAL EMITIR HACIA EL ESPACIO SOBRE TODO RADIACIÓN DE LONGITUD DE ONDA MAYOR DE FORMA QUE LA ESTRELLA PARECER ROJO * A PESAR DE SU BAJA TEMPERATURA SUPERFICIAL TODO EL GIGANTE ROJO TENER UNO ALTO LUMINOSIDAD L DEBER A SU ENORME RADIO R * ESTO RESULTADO SE DEDUCIR DE LA LEY DE RADIACIÓN DE STEFANBOLTZMANN PARA LA RADIACIÓN DEL CUERPO NEGRO $L = \sigma \times 4 \times \pi \times R^2 \times T^4$ DONDE σ SER EL CONSTANTE DE STEFANBOLTZMANN * VALOR TÍPICO PARA EL GIGANTE ROJO SER $R \sim 10^2 \times R_{SOLAR}$ $T \sim 3$ HASTA 4×10^3 K DE FORMA QUE L SER APROXIMADAMENTE $10^3 \times L_{SOLAR}$ ** CUANDO YA NO PODER MANTENER EL PROCESO AVANZADO DE FUSIÓN EN EL NÚCLEO ESTELAR EL NÚCLEO SE COLAPSAR DE NUEVO * OTRO VEZ LA TEMPERATURA DEL NÚCLEO SE INCREMENTAR Y AHORA SE EXPULSAR EL CAPA EXTERNO DE LA ESTRELLA * EL LLAMADO NEBULOSA PLANETARIO SE FORMAR A PARTIR DE EL RESTO DE LA CAPA EXTERNO VER EJERCICIO 3 DE ESTO SERIE DE EJERCICIO DE ASTRONOMÍA DE ESA ESO * EL NÚCLEO COLAPSADO ESTAR MUY CALIENTE BLANCO Y LA ESTRELLA SER MUY PEQUEÑO * A ESTO ESTRELLA SE LE DENOMINAR MUY ADECUADAMENTE ENANO BLANCO Y SER LA FASE FINAL DE LA

ANEXO G: TEXTO LEMATIZADO FESTESE001V

VIDA DE UNO ESTRELLA NORMAL DE TIPO SOLAR ** PARA HACER UNO ESTIMACIÓN APROXIMADOA DE EL RELACIÓN ENTRE EL LUMINOSIDAD L Y EL TEMPERATURA SUPERFICIAL T DE TODO EL ESTRELLA DE EL SECUENCIA PRINCIPAL VER EL DIAGRAMA HR FIG 6

* EL LÍNEA MÁS O MENOS RECTO DE EL SECUENCIA PRINCIPAL SE EXTENDER APROXIMADAMENTE UNO ORDEN DE MAGNITUD EN TEMPERATURA 3×10^3 HASTAM 3×10^4 K * EL RANGO DE LUMINOSIDAD SE EXTENDER APROXIMADAMENTE SEIS ORDEN DE MAGNITUD 10^{-2} HASTAM 10^4 X L_{SOLAR} * POR TANTO PODERX ESTIMAR DE FORMA1 APROXIMADOA $L \propto T^4$ PARA EL ESTRELLA DE EL SECUENCIA PRINCIPAL ** PARA DAR ALGUNO EJEMPLO UNO ESTRELLA DE GRANDE MASA DE EL SECUENCIA PRINCIPAL CON UNO TEMPERATURA SUPERFICIAL DE APROXIMADAMENTE $T_{ESTRELLA} = 10^4$ K TENER UNO LUMINOSIDAD DE APROXIMADAMENTE $L_{ESTRELLA} = 10^4$ DIVIDIR POR $50 \cdot L_{SOLAR}$ O APROXIMADAMENTE 26 VEZ EL LUMINOSIDAD DEL SOL * EL LUMINOSIDAD DEL SOL TENER UNO VALOR ESTÁNDARA DE 1 EN EL ESCALA1 DE LUMINOSIDAD ** UNO ESTRELLA DE BAJOA MASA CON $T_{ESTRELLA} = 3000$ K TENER UNO LUMINOSIDAD DE APROXIMADAMENTE SÓLO EL 5 PORCIENTO DE EL LUMINOSIDAD DEL SOL ** TODO EL INFORMACIÓN QUE PODERX EXTRAER DE EL ESTRELLA PROVENIR DE EL RADIACIÓN QUE RECIBIR DE ELLO * COMO SE EXPLICAR EN EL SECCIÓN HERRAMIENTA ASTRONÓMICO DIFERENTE FILTRO Y SISTEMA DE COLOR SE PODERX USAR PARA MEDIR EL BRILLO DE UNO ESTRELLA *

ANEXO G: TEXTO LEMATIZADO FESTEESE001V

EN ESTO EJERCICIO USAR UNO IMAGEN B Y UNO IMAGEN V * EN TU ANÁLISIS DE ESTO IMAGEN ENCONTRAR EL MAGNITUD APARENTE MB Y MV DE UNO MUESTRA1 DE ESTRELLA EN EL CÚMULO * LUEGO PODERX CALCULAR EL VALOR MB Y MV EL ÍNDICE DE COLOR B MENOS V * FINALMENTE PODERX DETERMINAR EL TEMPERATURA SUPERFICIAL DE EL ESTRELLA VER EL SECCIÓN HERRAMIENTA ASTRONÓMICO ** UNO CÚMULO SER UNO GRUPO DE ESTRELLA * EL VIDA DE UNO CÚMULO ESTARX DETERMINADOA POR EL VIDA DE EL DIFERENTE TIPO DE ESTRELLA QUE EL COMPONER * EL OBSERVACIÓN HABERX MOSTRAR QUE EN EL CÚMULO GLOBULAR QUEDAR MUY POCO CANTIDAD DE GAS Y POLVO DE MANERA QUE SER MUY RARO EL NACIMIENTO DE NUEVO ESTRELLA EN TAL CÚMULO * EL ESTRELLA QUE VER EN UNO CÚMULO GLOBULAR SER TODO ADULTO Y HABERX EVOLUCIONAR DE FORMA1 DIFERENTE EN FUNCIÓN DE SU MASA ** EL MAYORÍA DE EL ESTRELLA DE BAJOA MASA SE ENCONTRAR EN EL SECUENCIA PRINCIPAL * ESTO SE DEBER A QUE EL ESTRELLA DE BAJOA MASA IRX CONSUMIR SU ENERGÍA MUY LENTAMENTE ** QUEMAR SU RESERVA1 DE HIDRÓGENO POCO A POCO Y CONTINUAR DE ESTO MODO DURANTE MIL DE MILLÓN DE AÑO * POR CONSIGUIENTE ESTAR EN EL SECUENCIA PRINCIPAL DURANTE MUCHO TIEMPO ** POR EL CONTRARIO EL ESTRELLA MÁS PESADOA DEL CÚMULO YA HABERX TRANSFORMAR EL HIDRÓGENO DE SU NÚCLEO Y SE HABERX CONVERTIR EN GIGANTE ROJO * TODO ESTO SUCEDER HACER MUCHO TIEMPO ASÍ QUE HOY EN DÍA NO HABER ESTRELLA CALIENTE MASIVO EL QUE DEJAR UNO HUECO EN

ANEXO G: TEXTO LEMATIZADO FESTESE001V

EL MITAD SUPERIOR DE EL SECUENCIA PRINCIPAL VER FIG 7 * ESTO ESTRELLA SE ENCONTRAR AHORA EN EL ÁREA DIAGONAL QUE COMENZAR EN EL SECUENCIA PRINCIPAL Y SE DESPLAZAR HACIA EL PARTE SUPERIOR DERECHO DEL DIAGRAMA CONOCIDO A COMO RAMA DE EL GIGANTE ROJO ** EL PUNTO EN EL CUAL SE ENCONTRAR EL SECUENCIA PRINCIPAL Y EL RAMA DE GIGANTE ROJO SE LLAMAR PUNTO DE GIRO1 Y NOS DAR UNO PISTA IMPORTANTE A EL HORA DE DETERMINAR EL EDAD DEL CÚMULO ** EN EL SIGUIENTE EJERCICIO MEDIR EL COORDENADA DE ESTO PUNTO SOBRE TU DIAGRAMA Y DETERMINAR EL EDAD DE M12 **

ANEXO H: Frecuencias del Texto Preparado FESTESE001V

Listado de las frecuencias del texto preparado obtenidas con el programa PAFE. En el orden léxicométrico aparecen al principio los números y símbolos y al final las palabras que comienzan por vocales con tilde. En la columna de la categoría gramatical, el lenguaje matemático se denota con “lm” y entre paréntesis se ha añadido, cuando estaba claro, si además correspondía con otro tipo de palabra léxica (“s”= sustantivo, “adj”= adjetivo, “v”=verbo). Los verbos auxiliares se señalan como “f-verbo”. La tabla se encuentra dividida en las tres franjas señaladas en la metodología: alta, media y baja.

Tabla 95

Frecuencias del texto preparado FESTESE001V

OL	Palabra	n_i	f_i (a1000)	N_i	Categoría gramatical
Franja alta					
1	DE	222	92.346	222	funcional
2	LA	128	53.245	350	funcional
3	EL	67	27.87	417	funcional
4	EN	64	26.622	481	funcional
5	SE	54	22.463	535	funcional
6	ESTRELLAS	48	19.967	583	sustantivo
7	Y	48	19.967	631	funcional
8	LAS	41	17.055	672	funcional
9	QUE	39	16.223	711	funcional
10	UNA	38	15.807	749	funcional
11	DEL	30	12.479	779	funcional
12	LOS	29	12.063	808	funcional
13	A	28	11.647	836	funcional
14	ES	28	11.647	864	verbo
15	ESTRELLA	28	11.647	892	sustantivo
16	UN	27	11.231	919	funcional
17	CÚMULOS	23	9.567	942	sustantivo

ANEXO H: FRECUENCIAS DEL TEXTO PREPARADO FESTESE001V

OL	Palabra	n_i	f_i (a1000)	N_i	Categoría gramatical
18	MÁS	18	7.488	960	adverbio
Franja media					
19	PARA	17	7.072	977	funcional
20	CÚMULO	16	6.656	993	sustantivo
21	POR	16	6.656	1009	funcional
22	TEMPERATURA	15	6.24	1024	sustantivo
23	NÚCLEO	14	5.824	1038	sustantivo
24	MUY	13	5.408	1051	adverbio
25	PRINCIPAL	13	5.408	1064	adjetivo
26	SECUENCIA	13	5.408	1077	sustantivo
27	X	13	5.408	1090	lm(v)
28	APROXIMADAMENTE	12	4.992	1102	adverbio
29	HIDRÓGENO	12	4.992	1114	sustantivo
30	SU	12	4.992	1126	funcional
31	10	11	4.576	1137	lm(adj)
32	COMO	11	4.576	1148	funcional
33	DIAGRAMA	11	4.576	1159	sustantivo
34	GLOBULARES	11	4.576	1170	adjetivo
35	L	11	4.576	1181	lm(s)
36	AÑOS	10	4.16	1191	sustantivo
37	LUMINOSIDAD	10	4.16	1201	sustantivo
38	MASA	10	4.16	1211	sustantivo
39	T	10	4.16	1221	lm(s)
40	ESTE	9	3.744	1230	funcional
41	FORMA1	9	3.744	1239	sustantivo
42	TIENE	9	3.744	1248	verbo
43	CON	8	3.328	1256	funcional
44	LO	8	3.328	1264	funcional
45	NO	8	3.328	1272	adverbio
46	SON	8	3.328	1280	verbo
47	SUPERFICIAL	8	3.328	1288	adjetivo
48	DIFERENTES	7	2.912	1295	adjetivo
49	ENTRE	7	2.912	1302	funcional
50	ESTA	7	2.912	1309	funcional
51	FUSIÓN	7	2.912	1316	sustantivo
52	GALAXIA	7	2.912	1323	sustantivo
53	LUZ	7	2.912	1330	sustantivo
54	MILLONES	7	2.912	1337	sustantivo
55	PORCIENTO	7	2.912	1344	lm
56	PROCESO	7	2.912	1351	sustantivo
57	PUEDEX	7	2.912	1358	f-verbo
58	SOL	7	2.912	1365	sustantivo
59	SUS	7	2.912	1372	funcional
60	VER	7	2.912	1379	verbo
61	VIDA	7	2.912	1386	sustantivo
62	CUANDO	6	2.496	1392	funcional
63	EDAD	6	2.496	1398	sustantivo

ANEXO H: FRECUENCIAS DEL TEXTO PREPARADO FESTESE001V

OL	Palabra	n_i	f_i (a1000)	N_i	Categoría gramatical
64	ENERGÍA	6	2.496	1404	sustantivo
65	GLOBULAR	6	2.496	1410	adjetivo
66	MAYORÍA	6	2.496	1416	sustantivo
67	O	6	2.496	1422	funcional
68	UNOS	6	2.496	1428	funcional
69	=	5	2.08	1433	lm(adj)
70	BAJAA	5	2.08	1438	adjetivo
71	ELEMENTOS	5	2.08	1443	sustantivo
72	ENCUENTRAN	5	2.08	1448	verbo
73	GAS	5	2.08	1453	sustantivo
74	HANX	5	2.08	1458	f-verbo
75	HR	5	2.08	1463	sustantivo
76	LÁCTEA	5	2.08	1468	adjetivo
77	MAGNITUD	5	2.08	1473	sustantivo
78	MESSIER	5	2.08	1478	sustantivo
79	MILES	5	2.08	1483	adjetivo
80	NÚCLEOS	5	2.08	1488	sustantivo
81	POCOS	5	2.08	1493	adjetivo
82	SOLAR	5	2.08	1498	adjetivo
83	SÓLO	5	2.08	1503	adverbio
84	VÍA	5	2.08	1508	sustantivo
85	YA	5	2.08	1513	adverbio
86	12	4	1.664	1517	lm(adj)
87	3	4	1.664	1521	lm(adj)
88	A2	4	1.664	1525	lm
89	A3	4	1.664	1529	lm
90	A4	4	1.664	1533	lm
91	ABIERTOSA	4	1.664	1537	adjetivo
92	ASTRONÓMICAS	4	1.664	1541	adjetivo
93	CANTIDAD	4	1.664	1545	sustantivo
94	DESDE	4	1.664	1549	funcional
95	DISCO	4	1.664	1553	sustantivo
96	DONDE	4	1.664	1557	funcional
97	DURANTE	4	1.664	1561	funcional
98	EJERCICIO	4	1.664	1565	sustantivo
99	ESTABLE	4	1.664	1569	adjetivo
100	ESTELAR	4	1.664	1573	adjetivo
101	ESTOS	4	1.664	1577	funcional
102	FASE	4	1.664	1581	sustantivo
103	FIG	4	1.664	1585	sustantivo
104	FORMAN	4	1.664	1589	verbo
105	GIGANTES	4	1.664	1593	adjetivo
106	HACE	4	1.664	1597	verbo
107	HACIA	4	1.664	1601	funcional
108	HELIO	4	1.664	1605	sustantivo
109	HERRAMIENTAS	4	1.664	1609	sustantivo
110	INCLUSO	4	1.664	1613	adverbio
111	K	4	1.664	1617	lm(s)

ANEXO H: FRECUENCIAS DEL TEXTO PREPARADO FESTESE001V

OL	Palabra	n_i	f_i (a1000)	N_i	Categoría gramatical
112	LÍNEA	4	1.664	1621	sustantivo
113	NUESTRA	4	1.664	1625	funcional
114	PARTE	4	1.664	1629	sustantivo
115	R	4	1.664	1633	lm(s)
116	RADIACIÓN	4	1.664	1637	sustantivo
117	ROJAS	4	1.664	1641	adjetivo
118	SECCIÓN	4	1.664	1645	sustantivo
119	SOBRE	4	1.664	1649	funcional
120	SUPERIOR	4	1.664	1653	adjetivo
121	TIEMPO	4	1.664	1657	sustantivo
122	TODAS	4	1.664	1661	adjetivo
123	UNAA	4	1.664	1665	adjetivo
124	CAPAS	3	1.248	1668	sustantivo
125	CENTRAL	3	1.248	1671	adjetivo
126	COLOR	3	1.248	1674	sustantivo
127	CUATRO	3	1.248	1677	adjetivo
128	DA	3	1.248	1680	verbo
129	DETERMINAR	3	1.248	1683	verbo
130	EQUILIBRIO	3	1.248	1686	sustantivo
131	ESTAS	3	1.248	1689	funcional
132	ESTO	3	1.248	1692	funcional
133	EXTERNAS	3	1.248	1695	adjetivo
134	GENERAL	3	1.248	1698	adjetivo
135	HASTAM	3	1.248	1701	lm
136	INFERIOR	3	1.248	1704	adjetivo
137	M12	3	1.248	1707	sustantivo
138	MASAS	3	1.248	1710	sustantivo
139	MB	3	1.248	1713	lm(s)
140	MENOS	3	1.248	1716	adverbio
141	MV	3	1.248	1719	lm(s)
142	PARTIR	3	1.248	1722	verbo
143	PEQUEÑA	3	1.248	1725	adjetivo
144	PESADOSA	3	1.248	1728	adjetivo
145	POLVO	3	1.248	1731	sustantivo
146	PRESIÓN	3	1.248	1734	sustantivo
147	PUEDENX	3	1.248	1737	f-verbo
148	PUNTO	3	1.248	1740	sustantivo
149	RADIO	3	1.248	1743	sustantivo
150	ROJA	3	1.248	1746	adjetivo
151	TALES	3	1.248	1749	adjetivo
152	TAMBIÉN	3	1.248	1752	adverbio
153	TANTO	3	1.248	1755	adjetivo
154	TIPOS	3	1.248	1758	sustantivo
155	TODO	3	1.248	1761	adjetivo
156	TOTAL	3	1.248	1764	adjetivo
157	VALORES	3	1.248	1767	sustantivo
158	OCOMA7	2	0.832	1769	lm(adj)
159	1	2	0.832	1771	lm(adj)

ANEXO H: FRECUENCIAS DEL TEXTO PREPARADO FESTESE001V

OL	Palabra	n_i	f_i (a1000)	N_i	Categoría gramatical
160	2	2	0.832	1773	lm(adj)
161	4	2	0.832	1775	lm(adj)
162	6	2	0.832	1777	lm(adj)
163	A6	2	0.832	1779	lm
164	AHORA	2	0.832	1781	adverbio
165	AL	2	0.832	1783	funcional
166	ALGUNOS	2	0.832	1785	adjetivo
167	APARECE	2	0.832	1787	verbo
168	APROXIMADAA	2	0.832	1789	adjetivo
169	B	2	0.832	1791	lm
170	C	2	0.832	1793	lm(s)
171	CADA	2	0.832	1795	adjetivo
172	CALCULAR	2	0.832	1797	verbo
173	CIENTOS	2	0.832	1799	adjetivo
174	COMBUSTIÓN	2	0.832	1801	sustantivo
175	COMIENZA	2	0.832	1803	verbo
176	COMPONE	2	0.832	1805	verbo
177	COMPONEN	2	0.832	1807	verbo
178	COMÚN	2	0.832	1809	adjetivo
179	CONJUNTO	2	0.832	1811	sustantivo
180	CONTRAERSE	2	0.832	1813	verbo
181	CUAL	2	0.832	1815	funcional
182	DEBIDO	2	0.832	1817	verbo
183	DERECHA	2	0.832	1819	adjetivo
184	DETERMINARÁS	2	0.832	1821	verbo
185	DISTANCIA	2	0.832	1823	sustantivo
186	DISTRIBUCIÓN	2	0.832	1825	sustantivo
187	DIÁMETRO	2	0.832	1827	sustantivo
188	DOS	2	0.832	1829	adjetivo
189	E	2	0.832	1831	funcional
190	ECUACIÓN	2	0.832	1833	sustantivo
191	ENCUENTRA	2	0.832	1835	verbo
192	ESFÉRICO	2	0.832	1837	adjetivo
193	ESTRUCTURA	2	0.832	1839	sustantivo
194	ESTÁ	2	0.832	1841	verbo
195	EVOLUCIÓN	2	0.832	1843	sustantivo
196	EXTIENDE	2	0.832	1845	verbo
197	FORMACIÓN	2	0.832	1847	sustantivo
198	FORMAR	2	0.832	1849	verbo
199	FUERA	2	0.832	1851	adverbio
200	FUEVSERX	2	0.832	1853	f-verbo
201	FUSIONAN	2	0.832	1855	verbo
202	GIGANTE	2	0.832	1857	adjetivo
203	GRANDE	2	0.832	1859	adjetivo
204	GRUPO	2	0.832	1861	sustantivo
205	HALO	2	0.832	1863	sustantivo
206	HASTA	2	0.832	1865	funcional
207	HISTORIA	2	0.832	1867	sustantivo

ANEXO H: FRECUENCIAS DEL TEXTO PREPARADO FESTESE001V

OL	Palabra	n_i	f_i (a1000)	N_i	Categoría gramatical
208	IMAGEN	2	0.832	1869	sustantivo
209	IMPORTANTE	2	0.832	1871	adjetivo
210	LARGO	2	0.832	1873	adjetivo
211	LENTAMENTE	2	0.832	1875	adverbio
212	LLAMA	2	0.832	1877	verbo
213	LUGAR	2	0.832	1879	sustantivo
214	LUGARES	2	0.832	1881	sustantivo
215	LUMÍNICA	2	0.832	1883	adjetivo
216	M	2	0.832	1885	lm(s)
217	MAGNITUDES	2	0.832	1887	sustantivo
218	MANERA	2	0.832	1889	sustantivo
219	MEDIDAS1	2	0.832	1891	sustantivo
220	MIENTRAS	2	0.832	1893	adverbio
221	MUCHO	2	0.832	1895	adjetivo
222	MUCHOS	2	0.832	1897	adjetivo
223	MUESTRAV	2	0.832	1899	verbo
224	NEBULOSAS	2	0.832	1901	sustantivo
225	NGC	2	0.832	1903	sustantivo
226	NORMAL	2	0.832	1905	adjetivo
227	NOS	2	0.832	1907	funcional
228	NUEVAS	2	0.832	1909	adjetivo
229	NUEVO	2	0.832	1911	adjetivo
230	NÚMERO	2	0.832	1913	sustantivo
231	OBJETO	2	0.832	1915	sustantivo
232	OBSERVACIONES	2	0.832	1917	sustantivo
233	OTRA	2	0.832	1919	adjetivo
234	PASO1	2	0.832	1921	sustantivo
235	PLANETARIAS	2	0.832	1923	adjetivo
236	PLÉYADES	2	0.832	1925	sustantivo
237	POCO	2	0.832	1927	adjetivo
238	PODEMOSX	2	0.832	1929	f-verbo
239	PODRÁSX	2	0.832	1931	f-verbo
240	PROCESOS	2	0.832	1933	sustantivo
241	PROPIA	2	0.832	1935	adjetivo
242	PROPIEDADES	2	0.832	1937	sustantivo
243	RAMA	2	0.832	1939	sustantivo
244	RARO	2	0.832	1941	adjetivo
245	REACCIONES	2	0.832	1943	sustantivo
246	REACCIÓN	2	0.832	1945	sustantivo
247	RECTA	2	0.832	1947	adjetivo
248	RELATIVAMENTE	2	0.832	1949	adverbio
249	RESIDUAL	2	0.832	1951	adjetivo
250	SER	2	0.832	1953	verbo
251	SIGMA	2	0.832	1955	sustantivo
252	SINO	2	0.832	1957	funcional
253	SISTEMAS	2	0.832	1959	sustantivo
254	STEFANBOLTZMANN	2	0.832	1961	sustantivo
255	TIENEN	2	0.832	1963	verbo

ANEXO H: FRECUENCIAS DEL TEXTO PREPARADO FESTESE001V

OL	Palabra	n_i	f_i (a1000)	N_i	Categoría gramatical
256	TU	2	0.832	1965	funcional
257	TÍPICO	2	0.832	1967	adjetivo
258	UNO	2	0.832	1969	funcional
259	USANDO	2	0.832	1971	verbo
260	V	2	0.832	1973	lm
261	VALOR	2	0.832	1975	sustantivo
262	VANX	2	0.832	1977	f-verbo
263	VEZ	2	0.832	1979	sustantivo
264	VIEJOS	2	0.832	1981	adjetivo
265	VISIBLE	2	0.832	1983	adjetivo
266	VÉRTICE	2	0.832	1985	sustantivo
267	~	2	0.832	1987	lm
268	ÉL	2	0.832	1989	funcional
Franja baja					
269	100	1	0.416	1990	lm(adj)
270	13	1	0.416	1991	lm(adj)
271	1764	1	0.416	1992	lm(adj)
272	1783	1	0.416	1993	lm(adj)
273	1888	1	0.416	1994	lm(adj)
274	1COMA0	1	0.416	1995	lm(adj)
275	25	1	0.416	1996	lm(adj)
276	26	1	0.416	1997	lm(adj)
277	30	1	0.416	1998	lm(adj)
278	3000	1	0.416	1999	lm(adj)
279	3COMA5	1	0.416	2000	lm(adj)
280	400	1	0.416	2001	lm(adj)
281	5	1	0.416	2002	lm(adj)
282	5COMA8	1	0.416	2003	lm(adj)
283	6218	1	0.416	2004	lm(adj)
284	6COMA7	1	0.416	2005	lm(adj)
285	7	1	0.416	2006	lm(adj)
286	74	1	0.416	2007	lm(adj)
287	99COMA3	1	0.416	2008	lm(adj)
288	ABIERTO A	1	0.416	2009	adjetivo
289	ABSOLUTA	1	0.416	2010	adjetivo
290	ABUNDANCIA	1	0.416	2011	sustantivo
291	ACABANDO	1	0.416	2012	verbo
292	ADECUADAMENTE	1	0.416	2013	adverbio
293	ADULTAS	1	0.416	2014	adjetivo
294	AGOTA	1	0.416	2015	verbo
295	AGRUPACIONES	1	0.416	2016	sustantivo
296	ALGUNAS	1	0.416	2017	adjetivo
297	ALREDEDOR	1	0.416	2018	adverbio
298	ALTA	1	0.416	2019	adjetivo
299	AMENOS2	1	0.416	2020	lm
300	ANTERIORES	1	0.416	2021	adjetivo
301	ANÁLISIS	1	0.416	2022	sustantivo

ANEXO H: FRECUENCIAS DEL TEXTO PREPARADO FESTESE001V

OL	Palabra	n_i	f_i (a1000)	N_i	Categoría gramatical
302	APARENTES	1	0.416	2023	adjetivo
303	ASTROFÍSICO	1	0.416	2024	adjetivo
304	ASTRONOMÍA	1	0.416	2025	sustantivo
305	ASTRONÓMICA	1	0.416	2026	adjetivo
306	ASTRÓNOMOS	1	0.416	2027	sustantivo
307	ASÍ	1	0.416	2028	adverbio
308	AUMENTO1	1	0.416	2029	sustantivo
309	AVANZADOSA	1	0.416	2030	adjetivo
310	BANCO	1	0.416	2031	sustantivo
311	BLANCAS	1	0.416	2032	adjetivo
312	BLANCO	1	0.416	2033	adjetivo
313	BOLA	1	0.416	2034	sustantivo
314	BRILLANTE	1	0.416	2035	adjetivo
315	BRILLANTES	1	0.416	2036	adjetivo
316	BRILLO	1	0.416	2037	sustantivo
317	BRUNO	1	0.416	2038	sustantivo
318	CALIENTE	1	0.416	2039	adjetivo
319	CALIENTES	1	0.416	2040	adjetivo
320	CAMINOS	1	0.416	2041	sustantivo
321	CAPA	1	0.416	2042	sustantivo
322	CATÁLOGO	1	0.416	2043	sustantivo
323	CENTENAR	1	0.416	2044	sustantivo
324	CENTRO	1	0.416	2045	sustantivo
325	CERCANA	1	0.416	2046	adjetivo
326	CHARLES	1	0.416	2047	sustantivo
327	CICLO	1	0.416	2048	sustantivo
328	CIELO	1	0.416	2049	sustantivo
329	CIRCUNDANTE	1	0.416	2050	adjetivo
330	CIUDADES	1	0.416	2051	sustantivo
331	CLARAMENTE	1	0.416	2052	adverbio
332	CLAVE	1	0.416	2053	sustantivo
333	COLAPSA	1	0.416	2054	verbo
334	COLAPSADOA	1	0.416	2055	adjetivo
335	COLECCIONES	1	0.416	2056	sustantivo
336	COLISIONES	1	0.416	2057	sustantivo
337	COLOCAN	1	0.416	2058	verbo
338	COMIENZANX	1	0.416	2059	f-verbo
339	COMIENZAX	1	0.416	2060	f-verbo
340	COMPACTAS	1	0.416	2061	adjetivo
341	COMPACTOS	1	0.416	2062	adjetivo
342	COMPARTEN	1	0.416	2063	verbo
343	COMPLICADOA	1	0.416	2064	adjetivo
344	COMPONENTES	1	0.416	2065	sustantivo
345	COMPRENDER	1	0.416	2066	verbo
346	COMUNIDAD	1	0.416	2067	sustantivo
347	CONCENTRAN	1	0.416	2068	verbo
348	CONCISA	1	0.416	2069	adjetivo
349	CONOCEN	1	0.416	2070	verbo

ANEXO H: FRECUENCIAS DEL TEXTO PREPARADO FESTESE001V

OL	Palabra	n_i	f_i (a1000)	N_i	Categoría gramatical
350	CONOCIDAA	1	0.416	2071	adjetivo
351	CONSECUENCIA	1	0.416	2072	sustantivo
352	CONSIDERA	1	0.416	2073	verbo
353	CONSIDERABLE	1	0.416	2074	adjetivo
354	CONSIGUIENTE	1	0.416	2075	adjetivo
355	CONSTANTE1	1	0.416	2076	sustantivo
356	CONSTANTESA	1	0.416	2077	adjetivo
357	CONSTELACIÓN	1	0.416	2078	sustantivo
358	CONSTRUIDO	1	0.416	2079	verbo
359	CONSUMIENDO	1	0.416	2080	verbo
360	CONTAMINACIÓN	1	0.416	2081	sustantivo
361	CONTIENE	1	0.416	2082	verbo
362	CONTIENEN	1	0.416	2083	verbo
363	CONTINUARÁN	1	0.416	2084	verbo
364	CONTRACCIÓN	1	0.416	2085	sustantivo
365	CONTRARIO	1	0.416	2086	adjetivo
366	CONVERTIDO	1	0.416	2087	verbo
367	CONVERTIRSE	1	0.416	2088	verbo
368	CONVIERTE	1	0.416	2089	verbo
369	COORDENADAS	1	0.416	2090	sustantivo
370	CORRELACIÓN	1	0.416	2091	sustantivo
371	CORRESPONDÍAN	1	0.416	2092	verbo
372	CORTESÍA	1	0.416	2093	sustantivo
373	CREE	1	0.416	2094	verbo
374	CUALES	1	0.416	2095	funcional
375	CUALQUIER	1	0.416	2096	adjetivo
376	CUERPO	1	0.416	2097	sustantivo
377	CURIOSO	1	0.416	2098	adjetivo
378	DAR	1	0.416	2099	verbo
379	DEBE	1	0.416	2100	verbo
380	DECAE	1	0.416	2101	verbo
381	DEDUCE	1	0.416	2102	verbo
382	DEJA	1	0.416	2103	verbo
383	DENOMINAN	1	0.416	2104	verbo
384	DENSO	1	0.416	2105	adjetivo
385	DENTRO	1	0.416	2106	adverbio
386	DEPENDIENDO	1	0.416	2107	verbo
387	DEPOSITÓ	1	0.416	2108	verbo
388	DERECHO	1	0.416	2109	adjetivo
389	DERIVAR	1	0.416	2110	verbo
390	DESCRIBIÓ	1	0.416	2111	verbo
391	DESCUBIERTO	1	0.416	2112	verbo
392	DESPLAZA	1	0.416	2113	verbo
393	DESTRUIDO	1	0.416	2114	verbo
394	DETERMINADAA	1	0.416	2115	adjetivo
395	DEUTERIO	1	0.416	2116	sustantivo
396	DIAGONAL	1	0.416	2117	adjetivo
397	DIFIEREN	1	0.416	2118	verbo

ANEXO H: FRECUENCIAS DEL TEXTO PREPARADO FESTESE001V

OL	Palabra	n_i	f_i (a1000)	N_i	Categoría gramatical
398	DIFUSAS	1	0.416	2119	adjetivo
399	DISTANCIAS	1	0.416	2120	sustantivo
400	DISTINTOS	1	0.416	2121	adjetivo
401	DISTRIBUYEN	1	0.416	2122	verbo
402	DIVIDIDO	1	0.416	2123	verbo
403	DUODÉCIMO	1	0.416	2124	adjetivo
404	DÍA	1	0.416	2125	sustantivo
405	E1	1	0.416	2126	sustantivo
406	EDADES	1	0.416	2127	sustantivo
407	EINSTEIN	1	0.416	2128	sustantivo
408	EISENDLE	1	0.416	2129	sustantivo
409	EJEMPLO	1	0.416	2130	sustantivo
410	EJEMPLOS	1	0.416	2131	sustantivo
411	EJERCICIOS	1	0.416	2132	sustantivo
412	EJERCIDAA	1	0.416	2133	adjetivo
413	EJERCIDASA	1	0.416	2134	adjetivo
414	ELEVAN	1	0.416	2135	verbo
415	ELLAS	1	0.416	2136	funcional
416	EMITE	1	0.416	2137	verbo
417	EMITIDAA	1	0.416	2138	adjetivo
418	ENANAS	1	0.416	2139	adjetivo
419	ENCONTRARÁS	1	0.416	2140	verbo
420	ENDE	1	0.416	2141	adverbio
421	ENFRÍEN	1	0.416	2142	verbo
422	ENORME	1	0.416	2143	adjetivo
423	ENTONCES	1	0.416	2144	adverbio
424	ERRORES	1	0.416	2145	sustantivo
425	ESA1	1	0.416	2146	sustantivo
426	ESCALA1	1	0.416	2147	sustantivo
427	ESFÉRICOS	1	0.416	2148	adjetivo
428	ESO1	1	0.416	2149	sustantivo
429	ESPACIO	1	0.416	2150	sustantivo
430	ESPECÍFICAS	1	0.416	2151	adjetivo
431	ESTADO	1	0.416	2152	sustantivo
432	ESTARÁN	1	0.416	2153	verbo
433	ESTIMACIÓN	1	0.416	2154	sustantivo
434	ESTIMAR	1	0.416	2155	verbo
435	ESTUDIADAA	1	0.416	2156	adjetivo
436	ESTUDIAR	1	0.416	2157	verbo
437	ESTUDIO1	1	0.416	2158	sustantivo
438	ESTÁNDARA	1	0.416	2159	adjetivo
439	ESTÁNX	1	0.416	2160	f-verbo
440	ESTÁX	1	0.416	2161	f-verbo
441	ETAPAS	1	0.416	2162	sustantivo
442	EVOLUCIONADO	1	0.416	2163	verbo
443	EVOLUCIONAN	1	0.416	2164	verbo
444	EXISTEN	1	0.416	2165	verbo
445	EXPANDAN	1	0.416	2166	verbo

ANEXO H: FRECUENCIAS DEL TEXTO PREPARADO FESTESE001V

OL	Palabra	n_i	f_i (a1000)	N_i	Categoría gramatical
446	EXPLICA	1	0.416	2167	verbo
447	EXPLOTAN	1	0.416	2168	verbo
448	EXPULSAN	1	0.416	2169	verbo
449	EXTRAER	1	0.416	2170	verbo
450	FAMOSA	1	0.416	2171	adjetivo
451	FAMOSOS	1	0.416	2172	adjetivo
452	FIGURA	1	0.416	2173	sustantivo
453	FILTROS	1	0.416	2174	sustantivo
454	FINAL	1	0.416	2175	sustantivo
455	FINALA	1	0.416	2176	adjetivo
456	FINALMENTE	1	0.416	2177	adverbio
457	FORMARON	1	0.416	2178	verbo
458	FORMAS1	1	0.416	2179	sustantivo
459	FORMIDABLE	1	0.416	2180	adjetivo
460	FRACCIÓN	1	0.416	2181	sustantivo
461	FRENTE	1	0.416	2182	adverbio
462	FUERZAS	1	0.416	2183	sustantivo
463	FUEVSER	1	0.416	2184	verbo
464	FUNCIÓN	1	0.416	2185	sustantivo
465	FUSIONARSE	1	0.416	2186	verbo
466	FÍSICAS	1	0.416	2187	adjetivo
467	GALAXIAS	1	0.416	2188	sustantivo
468	GENERAN	1	0.416	2189	verbo
469	GIGANTESCA	1	0.416	2190	adjetivo
470	GIRO1	1	0.416	2191	sustantivo
471	GRADOS	1	0.416	2192	sustantivo
472	GRAN	1	0.416	2193	adjetivo
473	GRAVITATORIA	1	0.416	2194	adjetivo
474	GRAVITATORIAMENTE	1	0.416	2195	adverbio
475	GRAVITATORIAS	1	0.416	2196	adjetivo
476	GRÁFICA	1	0.416	2197	sustantivo
477	HACER	1	0.416	2198	verbo
478	HAX	1	0.416	2199	f-verbo
479	HAY	1	0.416	2200	verbo
480	HAYX	1	0.416	2201	f-verbo
481	HECHO1	1	0.416	2202	sustantivo
482	HERSCHEL	1	0.416	2203	sustantivo
483	HERTZSPRUNGRUSSELL	1	0.416	2204	sustantivo
484	HIPPARCOS	1	0.416	2205	sustantivo
485	HORA	1	0.416	2206	sustantivo
486	HORIZONTAL	1	0.416	2207	adjetivo
487	HOY	1	0.416	2208	adverbio
488	HUECO	1	0.416	2209	sustantivo
489	IDEALES	1	0.416	2210	adjetivo
490	ILUSTRACIÓN	1	0.416	2211	sustantivo
491	ILUSTRAN	1	0.416	2212	verbo
492	IMPLICA	1	0.416	2213	verbo
493	IMPLICADAA	1	0.416	2214	adjetivo

ANEXO H: FRECUENCIAS DEL TEXTO PREPARADO FESTESE001V

OL	Palabra	n_i	f_i (a1000)	N_i	Categoría gramatical
494	IMPORTANTES	1	0.416	2215	adjetivo
495	IMÁGENES	1	0.416	2216	sustantivo
496	INCREMENTA	1	0.416	2217	verbo
497	INDICADASA	1	0.416	2218	adjetivo
498	INDICAN	1	0.416	2219	verbo
499	INDIVIDUALES	1	0.416	2220	adjetivo
500	INFORMACIÓN	1	0.416	2221	sustantivo
501	INTERACCIONES	1	0.416	2222	sustantivo
502	INTERNACIONAL	1	0.416	2223	adjetivo
503	INTERNO	1	0.416	2224	adjetivo
504	INTERÉS	1	0.416	2225	sustantivo
505	INVESTIGACIÓN	1	0.416	2226	sustantivo
506	INVOLUCRADOSA	1	0.416	2227	adjetivo
507	IZQUIERDO	1	0.416	2228	adjetivo
508	JUSTA	1	0.416	2229	adjetivo
509	JÓVENES	1	0.416	2230	adjetivo
510	LEY	1	0.416	2231	sustantivo
511	LIGADOSA	1	0.416	2232	adjetivo
512	LIGEROS	1	0.416	2233	adjetivo
513	LISTA	1	0.416	2234	sustantivo
514	LLAMADASA	1	0.416	2235	adjetivo
515	LLAMADOA	1	0.416	2236	adjetivo
516	LLAMADOSA	1	0.416	2237	adjetivo
517	LLEGAX	1	0.416	2238	f-verbo
518	LOCALIZAN	1	0.416	2239	verbo
519	LONGEVOS	1	0.416	2240	adjetivo
520	LONGITUDES	1	0.416	2241	sustantivo
521	LUEGO	1	0.416	2242	funcional
522	LUMINOSIDADES	1	0.416	2243	sustantivo
523	LÍMITE	1	0.416	2244	sustantivo
524	MANTENER	1	0.416	2245	verbo
525	MANTENERSE	1	0.416	2246	verbo
526	MASIVAS	1	0.416	2247	adjetivo
527	MATERIA	1	0.416	2248	sustantivo
528	MATERIAL	1	0.416	2249	sustantivo
529	MAYOR	1	0.416	2250	adjetivo
530	MAYORES	1	0.416	2251	adjetivo
531	MAYORITARIAMENTE	1	0.416	2252	adverbio
532	MEDIDAA	1	0.416	2253	adjetivo
533	MEDIDOSA	1	0.416	2254	adjetivo
534	MEDIR	1	0.416	2255	verbo
535	MEDIRÁS	1	0.416	2256	verbo
536	MENUDO	1	0.416	2257	adjetivo
537	MEZCLA1	1	0.416	2258	sustantivo
538	MIL	1	0.416	2259	adjetivo
539	MILLÓN	1	0.416	2260	sustantivo
540	MITAD	1	0.416	2261	sustantivo
541	MODESTO	1	0.416	2262	adjetivo

ANEXO H: FRECUENCIAS DEL TEXTO PREPARADO FESTESE001V

OL	Palabra	n_i	f_i (a1000)	N_i	Categoría gramatical
542	MODO	1	0.416	2263	sustantivo
543	MOSTRADO	1	0.416	2264	verbo
544	MOVERÁN	1	0.416	2265	verbo
545	MUESTRA1	1	0.416	2266	sustantivo
546	MUESTRAN	1	0.416	2267	verbo
547	NACIMIENTO	1	0.416	2268	sustantivo
548	NEBULOSA	1	0.416	2269	sustantivo
549	NEGRO	1	0.416	2270	adjetivo
550	NINGÚN	1	0.416	2271	adjetivo
551	NUBES	1	0.416	2272	sustantivo
552	NUCLEARES	1	0.416	2273	adjetivo
553	NUESTRO	1	0.416	2274	funcional
554	OBJETOS	1	0.416	2275	sustantivo
555	OBVIAMENTE	1	0.416	2276	adverbio
556	OFIUCO	1	0.416	2277	sustantivo
557	ONDA	1	0.416	2278	sustantivo
558	ORDEN	1	0.416	2279	sustantivo
559	ORIGINALES	1	0.416	2280	adjetivo
560	OTROS	1	0.416	2281	adjetivo
561	PARECE	1	0.416	2282	verbo
562	PARECER	1	0.416	2283	verbo
563	PARSECS	1	0.416	2284	sustantivo
564	PARTICULAR	1	0.416	2285	adjetivo
565	PASAN	1	0.416	2286	verbo
566	PATRÓN	1	0.416	2287	sustantivo
567	PEQUEÑO	1	0.416	2288	adjetivo
568	PERMANECEN	1	0.416	2289	verbo
569	PERO	1	0.416	2290	funcional
570	PERSPECTIVA	1	0.416	2291	sustantivo
571	PESADAA	1	0.416	2292	adjetivo
572	PESADASA	1	0.416	2293	adjetivo
573	PESAR	1	0.416	2294	verbo
574	PESE	1	0.416	2295	verbo
575	PI	1	0.416	2296	sustantivo
576	PISTA	1	0.416	2297	sustantivo
577	PLANO	1	0.416	2298	sustantivo
578	POCA	1	0.416	2299	adjetivo
579	POCAS	1	0.416	2300	adjetivo
580	PODER1	1	0.416	2301	sustantivo
581	POLUCIÓN	1	0.416	2302	sustantivo
582	PORQUE	1	0.416	2303	funcional
583	POSIBLE	1	0.416	2304	adjetivo
584	PRECISAA	1	0.416	2305	adjetivo
585	PRIMER	1	0.416	2306	adjetivo
586	PRIMERO	1	0.416	2307	adjetivo
587	PRISMÁTICOS	1	0.416	2308	adjetivo
588	PROBABLE	1	0.416	2309	adjetivo
589	PROBABLEMENTE	1	0.416	2310	adverbio

ANEXO H: FRECUENCIAS DEL TEXTO PREPARADO FESTESE001V

OL	Palabra	n_i	f_i (a1000)	N_i	Categoría gramatical
590	PRODUCE	1	0.416	2311	verbo
591	PROGRAMA	1	0.416	2312	sustantivo
592	PRONTO	1	0.416	2313	adverbio
593	PROPIO	1	0.416	2314	adjetivo
594	PROPORCIONAL	1	0.416	2315	adjetivo
595	PROPORCIONAN	1	0.416	2316	verbo
596	PROTOGALÁCTICO	1	0.416	2317	adjetivo
597	PROTONES	1	0.416	2318	sustantivo
598	PROVIENE	1	0.416	2319	verbo
599	PRUEBA1	1	0.416	2320	sustantivo
600	PRUEBAS1	1	0.416	2321	sustantivo
601	PRÓXIMOS	1	0.416	2322	adjetivo
602	PUBLICADO	1	0.416	2323	verbo
603	PÁGINA	1	0.416	2324	sustantivo
604	QUEDA	1	0.416	2325	verbo
605	QUEMAN	1	0.416	2326	verbo
606	QUEMANDO	1	0.416	2327	verbo
607	QUÍMICOS	1	0.416	2328	adjetivo
608	RAINER	1	0.416	2329	sustantivo
609	RANGO	1	0.416	2330	sustantivo
610	RAZÓN	1	0.416	2331	sustantivo
611	RECIBIMOS	1	0.416	2332	verbo
612	RELACIÓN	1	0.416	2333	sustantivo
613	REMANENTES	1	0.416	2334	sustantivo
614	REPETIDASA	1	0.416	2335	adjetivo
615	REPETIR	1	0.416	2336	verbo
616	RESERVAS1	1	0.416	2337	sustantivo
617	RESOLUCIÓN	1	0.416	2338	sustantivo
618	RESOLVIÓ	1	0.416	2339	verbo
619	RESTANTE	1	0.416	2340	adjetivo
620	RESTOS	1	0.416	2341	sustantivo
621	RESULTADO1	1	0.416	2342	sustantivo
622	SEAN	1	0.416	2343	verbo
623	SEIS	1	0.416	2344	adjetivo
624	SERIE	1	0.416	2345	sustantivo
625	SIENDO	1	0.416	2346	verbo
626	SIGUIENTE	1	0.416	2347	adjetivo
627	SIGUIENTES	1	0.416	2348	adjetivo
628	SIMPLE	1	0.416	2349	adjetivo
629	SIN	1	0.416	2350	funcional
630	SISTEMA	1	0.416	2351	sustantivo
631	SOLA	1	0.416	2352	adjetivo
632	STAMPFER	1	0.416	2353	sustantivo
633	SUBPRODUCTO	1	0.416	2354	sustantivo
634	SUCEDE	1	0.416	2355	verbo
635	SUCEDIÓ	1	0.416	2356	verbo
636	SUCESO	1	0.416	2357	sustantivo
637	SUMINISTRO1	1	0.416	2358	sustantivo

ANEXO H: FRECUENCIAS DEL TEXTO PREPARADO FESTESE001V

OL	Palabra	n_i	f_i (a1000)	N_i	Categoría gramatical
638	SUPERFICIE	1	0.416	2359	sustantivo
639	SUPERNOVA	1	0.416	2360	sustantivo
640	SUPERVIVIENTES	1	0.416	2361	adjetivo
641	SUPUESTO1	1	0.416	2362	sustantivo
642	SÍ	1	0.416	2363	funcional
643	TAMAÑO	1	0.416	2364	sustantivo
644	TAN	1	0.416	2365	adjetivo
645	TELESCOPIO	1	0.416	2366	sustantivo
646	TEORÍAS	1	0.416	2367	sustantivo
647	TERMONUCLEAR	1	0.416	2368	adjetivo
648	TIERRA	1	0.416	2369	sustantivo
649	TIPO	1	0.416	2370	sustantivo
650	TODA	1	0.416	2371	adjetivo
651	TODOS	1	0.416	2372	adjetivo
652	TRANSFORMADO	1	0.416	2373	verbo
653	TRAVÉS	1	0.416	2374	sustantivo
654	TUVO	1	0.416	2375	verbo
655	TÉRMICA	1	0.416	2376	adjetivo
656	TÉRMINO	1	0.416	2377	sustantivo
657	TÍPICOS	1	0.416	2378	adjetivo
658	UNAS	1	0.416	2379	funcional
659	UNIVERSO	1	0.416	2380	sustantivo
660	USA	1	0.416	2381	verbo
661	USAMOS	1	0.416	2382	verbo
662	USAR	1	0.416	2383	verbo
663	VALIOSA	1	0.416	2384	adjetivo
664	VALIOSO	1	0.416	2385	adjetivo
665	VAN	1	0.416	2386	verbo
666	VARIAS	1	0.416	2387	adjetivo
667	VARIOS	1	0.416	2388	adjetivo
668	VE	1	0.416	2389	verbo
669	VEAMOS	1	0.416	2390	verbo
670	VECES	1	0.416	2391	sustantivo
671	VECINDAD	1	0.416	2392	sustantivo
672	VEMOS	1	0.416	2393	verbo
673	VERSE	1	0.416	2394	verbo
674	VIOLENTA	1	0.416	2395	adjetivo
675	VISTA	1	0.416	2396	sustantivo
676	WILLIAM	1	0.416	2397	sustantivo
677	.	1	0.416	2398	lm
678	ÁREA	1	0.416	2399	sustantivo
679	ÁREAS	1	0.416	2400	sustantivo
680	ÁTOMO	1	0.416	2401	sustantivo
681	ÍNDICE	1	0.416	2402	sustantivo
682	ÓRDENES	1	0.416	2403	sustantivo
683	ÚTIL	1	0.416	2404	adjetivo

Nota. OL= orden lexicométrico; n= frecuencia absoluta; f= frecuencia relativa; N= frecuencia acumulada.

ANEXO I: Frecuencias del Texto Lematizado FESTESE001V

Listado de las frecuencias del texto lematizado obtenidas con el programa PAFE. En el orden léxicométrico aparecen al principio los números y símbolos y al final las palabras que comienzan por vocales con tilde. En la columna de la categoría gramatical, el lenguaje matemático se denota con “lm” y entre paréntesis se ha añadido, cuando estaba claro, si además correspondía con otro tipo de palabra léxica (“s”= sustantivo, “adj”= adjetivo, “v”=verbo). Los verbos auxiliares se señalan como “f-verbo”. La tabla se encuentra dividida en las tres franjas señaladas en la metodología: alta, media y baja.

Tabla 96

Frecuencias del texto lematizado FESTESE001V

OL	Lema	n_i	f_i (a1000)	N_i	Categoría gramatical
Franja alta					
1	EL	273	113.561	273	funcional
Franja media					
2	DE	222	92.346	495	funcional
3	ESTRELLA	76	31.614	571	sustantivo
4	UNO	74	30.782	645	funcional
5	EN	64	26.622	709	funcional
6	SE	54	22.463	763	funcional
7	Y	48	19.967	811	funcional
8	SER	41	17.055	852	verbo
9	CÚMULO	39	16.223	891	sustantivo
10	QUE	39	16.223	930	funcional
11	DEL	30	12.479	960	funcional
12	A	28	11.647	988	funcional
13	ESTO	26	10.815	1014	funcional

ANEXO I: FRECUENCIAS DEL TEXTO LEMATIZADO FESTESE001V

OL	Lema	n _i	f _i (a1000)	N _i	Categoría gramatical
14	NÚCLEO	19	7.903	1033	sustantivo
15	SU	19	7.903	1052	funcional
16	MÁS	18	7.488	1070	adverbio
17	GLOBULAR	17	7.072	1087	adjetivo
18	PARA	17	7.072	1104	funcional
19	POR	16	6.656	1120	funcional
20	TEMPERATURA	15	6.24	1135	sustantivo
21	PODERX	14	5.824	1149	f-verbo
22	MASA	13	5.408	1162	sustantivo
23	MUY	13	5.408	1175	adverbio
24	PRINCIPAL	13	5.408	1188	adjetivo
25	SECUENCIA	13	5.408	1201	sustantivo
26	X	13	5.408	1214	lm(v)
27	APROXIMADAMENTE	12	4.992	1226	adverbio
28	HIDRÓGENO	12	4.992	1238	sustantivo
29	TENER	12	4.992	1250	verbo
30	10	11	4.576	1261	lm(adj)
31	COMO	11	4.576	1272	funcional
32	DIAGRAMA	11	4.576	1283	sustantivo
33	L	11	4.576	1294	lm(s)
34	LUMINOSIDAD	11	4.576	1305	sustantivo
35	VER	11	4.576	1316	verbo
36	AÑO	10	4.16	1326	sustantivo
37	FORMA1	10	4.16	1336	sustantivo
38	T	10	4.16	1346	lm(s)
39	POCO	9	3.744	1355	adjetivo
40	PROCESO	9	3.744	1364	sustantivo
41	TODO	9	3.744	1373	adjetivo
42	CON	8	3.328	1381	funcional
43	ENCONTRAR	8	3.328	1389	verbo
44	GALAXIA	8	3.328	1397	sustantivo
45	MILLÓN	8	3.328	1405	sustantivo
46	NO	8	3.328	1413	adverbio
47	SUPERFICIAL	8	3.328	1421	adjetivo
48	DIFERENTE	7	2.912	1428	adjetivo
49	EDAD	7	2.912	1435	sustantivo
50	ENTRE	7	2.912	1442	funcional
51	FORMAR	7	2.912	1449	verbo
52	FUSIÓN	7	2.912	1456	sustantivo
53	HABERX	7	2.912	1463	f-verbo
54	LUZ	7	2.912	1470	sustantivo
55	MAGNITUD	7	2.912	1477	sustantivo
56	PORCIENTO	7	2.912	1484	lm
57	ROJO	7	2.912	1491	adjetivo
58	SOL	7	2.912	1498	sustantivo
59	VIDA	7	2.912	1505	sustantivo

ANEXO I: FRECUENCIAS DEL TEXTO LEMATIZADO FESTESE001V

OL	Lema	n_i	f_i (a1000)	N_i	Categoría gramatical
		Franja baja			
60	CUANDO	6	2.496	1511	funcional
61	ENERGÍA	6	2.496	1517	sustantivo
62	GIGANTE	6	2.496	1523	adjetivo
63	MAYORÍA	6	2.496	1529	sustantivo
64	MIL	6	2.496	1535	adjetivo
65	O	6	2.496	1541	funcional
66	=	5	2.08	1546	lm(adj)
67	ABIERTOA	5	2.08	1551	adjetivo
68	ASTRONÓMICO	5	2.08	1556	adjetivo
69	BAJOA	5	2.08	1561	adjetivo
70	DETERMINAR	5	2.08	1566	verbo
71	EJERCICIO	5	2.08	1571	sustantivo
72	ELEMENTO	5	2.08	1576	sustantivo
73	GAS	5	2.08	1581	sustantivo
74	HACER	5	2.08	1586	verbo
75	HR	5	2.08	1591	sustantivo
76	LÁCTEO	5	2.08	1596	adjetivo
77	MESSIER	5	2.08	1601	sustantivo
78	NUESTRO	5	2.08	1606	funcional
79	PESADOA	5	2.08	1611	adjetivo
80	SOLAR	5	2.08	1616	adjetivo
81	SÓLO	5	2.08	1621	adverbio
82	USAR	5	2.08	1626	verbo
83	VALOR	5	2.08	1631	sustantivo
84	VÍA	5	2.08	1636	sustantivo
85	YA	5	2.08	1641	adverbio
86	12	4	1.664	1645	lm(adj)
87	3	4	1.664	1649	lm(adj)
88	A2	4	1.664	1653	lm
89	A3	4	1.664	1657	lm
90	A4	4	1.664	1661	lm
91	CANTIDAD	4	1.664	1665	sustantivo
92	CAPA	4	1.664	1669	sustantivo
93	COMPONER	4	1.664	1673	verbo
94	DAR	4	1.664	1677	verbo
95	DESDE	4	1.664	1681	funcional
96	DISCO	4	1.664	1685	sustantivo
97	DONDE	4	1.664	1689	funcional
98	DURANTE	4	1.664	1693	funcional
99	ESTABLE	4	1.664	1697	adjetivo
100	ESTELAR	4	1.664	1701	adjetivo
101	FASE	4	1.664	1705	sustantivo
102	FIG	4	1.664	1709	sustantivo
103	HACIA	4	1.664	1713	funcional
104	HELIO	4	1.664	1717	sustantivo
105	HERRAMIENTA	4	1.664	1721	sustantivo
106	INCLUSO	4	1.664	1725	adverbio

ANEXO I: FRECUENCIAS DEL TEXTO LEMATIZADO FESTESE001V

OL	Lema	n_i	f_i (a1000)	N_i	Categoría gramatical
107	K	4	1.664	1729	lm(s)
108	LUGAR	4	1.664	1733	sustantivo
109	LÍNEA	4	1.664	1737	sustantivo
110	MOSTRAR	4	1.664	1741	verbo
111	MUCHO	4	1.664	1745	adjetivo
112	NUEVO	4	1.664	1749	adjetivo
113	PARTE	4	1.664	1753	sustantivo
114	PEQUEÑO	4	1.664	1757	adjetivo
115	R	4	1.664	1761	lm(s)
116	RADIACIÓN	4	1.664	1765	sustantivo
117	REACCIÓN	4	1.664	1769	sustantivo
118	SECCIÓN	4	1.664	1773	sustantivo
119	SOBRE	4	1.664	1777	funcional
120	SUPERIOR	4	1.664	1781	adjetivo
121	TANTO	4	1.664	1785	adjetivo
122	TIEMPO	4	1.664	1789	sustantivo
123	TIPO	4	1.664	1793	sustantivo
124	UNOA	4	1.664	1797	adjetivo
125	ALGUNO	3	1.248	1800	adjetivo
126	CENTRAL	3	1.248	1803	adjetivo
127	COLOR	3	1.248	1806	sustantivo
128	CONVERTIR	3	1.248	1809	verbo
129	CUAL	3	1.248	1812	funcional
130	CUATRO	3	1.248	1815	adjetivo
131	DEBER	3	1.248	1818	verbo
132	DERECHO	3	1.248	1821	adjetivo
133	DISTANCIA	3	1.248	1824	sustantivo
134	EQUILIBRIO	3	1.248	1827	sustantivo
135	ESFÉRICO	3	1.248	1830	adjetivo
136	ESTAR	3	1.248	1833	verbo
137	EXTERNO	3	1.248	1836	adjetivo
138	FUSIONAR	3	1.248	1839	verbo
139	GENERAL	3	1.248	1842	adjetivo
140	GRANDE	3	1.248	1845	adjetivo
141	HASTAM	3	1.248	1848	lm
142	IMAGEN	3	1.248	1851	sustantivo
143	IMPORTANTE	3	1.248	1854	adjetivo
144	INFERIOR	3	1.248	1857	adjetivo
145	LLAMADOA	3	1.248	1860	adjetivo
146	M12	3	1.248	1863	sustantivo
147	MB	3	1.248	1866	lm(s)
148	MENOS	3	1.248	1869	adverbio
149	MV	3	1.248	1872	lm(s)
150	NEBULOSA	3	1.248	1875	sustantivo
151	OBJETO	3	1.248	1878	sustantivo
152	OTRO	3	1.248	1881	adjetivo
153	PARTIR	3	1.248	1884	verbo
154	POLVO	3	1.248	1887	sustantivo

ANEXO I: FRECUENCIAS DEL TEXTO LEMATIZADO FESTESE001V

OL	Lema	n _i	f _i (a1000)	N _i	Categoría gramatical
155	PRESIÓN	3	1.248	1890	sustantivo
156	PROPIO	3	1.248	1893	adjetivo
157	PUNTO	3	1.248	1896	sustantivo
158	RADIO	3	1.248	1899	sustantivo
159	SISTEMA	3	1.248	1902	sustantivo
160	TAL	3	1.248	1905	adjetivo
161	TAMBIÉN	3	1.248	1908	adverbio
162	TOTAL	3	1.248	1911	adjetivo
163	TÍPICO	3	1.248	1914	adjetivo
164	VEZ	3	1.248	1917	sustantivo
165	0COMA7	2	0.832	1919	lm
166	1	2	0.832	1921	lm(adj)
167	2	2	0.832	1923	lm(adj)
168	4	2	0.832	1925	lm(adj)
169	6	2	0.832	1927	lm(adj)
170	A6	2	0.832	1929	lm
171	AHORA	2	0.832	1931	adverbio
172	AL	2	0.832	1933	funcional
173	APARECER	2	0.832	1935	verbo
174	APROXIMADOA	2	0.832	1937	adjetivo
175	B	2	0.832	1939	lm
176	BLANCO	2	0.832	1941	adjetivo
177	BRILLANTE	2	0.832	1943	adjetivo
178	C	2	0.832	1945	lm(s)
179	CADA	2	0.832	1947	adjetivo
180	CALCULAR	2	0.832	1949	verbo
181	CALIENTE	2	0.832	1951	adjetivo
182	CIENTO	2	0.832	1953	adjetivo
183	COMBUSTIÓN	2	0.832	1955	sustantivo
184	COMENZAR	2	0.832	1957	verbo
185	COMENZARX	2	0.832	1959	f-verbo
186	COMPACTO	2	0.832	1961	adjetivo
187	COMÚN	2	0.832	1963	adjetivo
188	CONJUNTO	2	0.832	1965	sustantivo
189	CONTENER	2	0.832	1967	verbo
190	CONTRAER	2	0.832	1969	verbo
191	DISTRIBUCIÓN	2	0.832	1971	sustantivo
192	DIÁMETRO	2	0.832	1973	sustantivo
193	DOS	2	0.832	1975	adjetivo
194	E	2	0.832	1977	funcional
195	ECUACIÓN	2	0.832	1979	sustantivo
196	EJEMPLO	2	0.832	1981	sustantivo
197	EJERCIDOA	2	0.832	1983	adjetivo
198	ESTARX	2	0.832	1985	f-verbo
199	ESTRUCTURA	2	0.832	1987	sustantivo
200	EVOLUCIONAR	2	0.832	1989	verbo
201	EVOLUCIÓN	2	0.832	1991	sustantivo
202	EXTENDER	2	0.832	1993	verbo

ANEXO I: FRECUENCIAS DEL TEXTO LEMATIZADO FESTESE001V

OL	Lema	n_i	f_i (a1000)	N_i	Categoría gramatical
203	FAMOSO	2	0.832	1995	adjetivo
204	FORMACIÓN	2	0.832	1997	sustantivo
205	FUERA	2	0.832	1999	adverbio
206	GRAVITATORIO	2	0.832	2001	adjetivo
207	GRUPO	2	0.832	2003	sustantivo
208	HALO	2	0.832	2005	sustantivo
209	HASTA	2	0.832	2007	funcional
210	HISTORIA	2	0.832	2009	sustantivo
211	IRX	2	0.832	2011	f-verbo
212	LARGO	2	0.832	2013	adjetivo
213	LENTAMENTE	2	0.832	2015	adverbio
214	LLAMAR	2	0.832	2017	verbo
215	LUMÍNICO	2	0.832	2019	adjetivo
216	M	2	0.832	2021	lm(s)
217	MANERA	2	0.832	2023	sustantivo
218	MANTENER	2	0.832	2025	verbo
219	MAYOR	2	0.832	2027	adjetivo
220	MEDIDA1	2	0.832	2029	sustantivo
221	MEDIDO A	2	0.832	2031	adjetivo
222	MEDIR	2	0.832	2033	verbo
223	MIENTRAS	2	0.832	2035	adverbio
224	NGC	2	0.832	2037	sustantivo
225	NORMAL	2	0.832	2039	adjetivo
226	NOS	2	0.832	2041	funcional
227	NÚMERO	2	0.832	2043	sustantivo
228	OBSERVACIÓN	2	0.832	2045	sustantivo
229	ORDEN	2	0.832	2047	sustantivo
230	PARECER	2	0.832	2049	verbo
231	PASO1	2	0.832	2051	sustantivo
232	PESAR	2	0.832	2053	verbo
233	PLANETARIO	2	0.832	2055	adjetivo
234	PLÉYADES	2	0.832	2057	sustantivo
235	PRIMERO	2	0.832	2059	adjetivo
236	PROPIEDAD	2	0.832	2061	sustantivo
237	PRUEBA1	2	0.832	2063	sustantivo
238	QUEMAR	2	0.832	2065	verbo
239	RAMA	2	0.832	2067	sustantivo
240	RARO	2	0.832	2069	adjetivo
241	RECTO	2	0.832	2071	adjetivo
242	RELATIVAMENTE	2	0.832	2073	adverbio
243	RESIDUAL	2	0.832	2075	adjetivo
244	SERX	2	0.832	2077	f-verbo
245	SIGMA	2	0.832	2079	sustantivo
246	SIGUIENTE	2	0.832	2081	adjetivo
247	SINO	2	0.832	2083	funcional
248	STEFANBOLTZMANN	2	0.832	2085	sustantivo
249	SUCEDER	2	0.832	2087	verbo
250	TU	2	0.832	2089	funcional

ANEXO I: FRECUENCIAS DEL TEXTO LEMATIZADO FESTESE001V

OL	Lema	n _i	f _i (a1000)	N _i	Categoría gramatical
251	V	2	0.832	2091	lm(s)
252	VALIOSO	2	0.832	2093	adjetivo
253	VARIO	2	0.832	2095	adjetivo
254	VIEJO	2	0.832	2097	adjetivo
255	VISIBLE	2	0.832	2099	adjetivo
256	VÉRTICE	2	0.832	2101	sustantivo
257	~	2	0.832	2103	lm
258	ÁREA	2	0.832	2105	sustantivo
259	ÉL	2	0.832	2107	funcional
260	100	1	0.416	2108	lm(adj)
261	13	1	0.416	2109	lm(adj)
262	1764	1	0.416	2110	lm(adj)
263	1783	1	0.416	2111	lm(adj)
264	1888	1	0.416	2112	lm(adj)
265	1COMA0	1	0.416	2113	lm(adj)
266	25	1	0.416	2114	lm(adj)
267	26	1	0.416	2115	lm(adj)
268	30	1	0.416	2116	lm(adj)
269	3000	1	0.416	2117	lm(adj)
270	3COMA5	1	0.416	2118	lm(adj)
271	400	1	0.416	2119	lm(adj)
272	5	1	0.416	2120	lm(adj)
273	5COMA8	1	0.416	2121	lm(adj)
274	6218	1	0.416	2122	lm(adj)
275	6COMA7	1	0.416	2123	lm(adj)
276	7	1	0.416	2124	lm(adj)
277	74	1	0.416	2125	lm(adj)
278	99COMA3	1	0.416	2126	lm(adj)
279	ABSOLUTO	1	0.416	2127	adjetivo
280	ABUNDANCIA	1	0.416	2128	sustantivo
281	ACABAR	1	0.416	2129	verbo
282	ADECUADAMENTE	1	0.416	2130	adverbio
283	ADULTO	1	0.416	2131	adjetivo
284	AGOTAR	1	0.416	2132	verbo
285	AGRUPACIÓN	1	0.416	2133	sustantivo
286	ALREDEDOR	1	0.416	2134	adverbio
287	ALTO	1	0.416	2135	adjetivo
288	AMENOS2	1	0.416	2136	lm
289	ANTERIOR	1	0.416	2137	adjetivo
290	ANÁLISIS	1	0.416	2138	sustantivo
291	APARENTE	1	0.416	2139	adjetivo
292	ASTROFÍSICO	1	0.416	2140	adjetivo
293	ASTRONOMÍA	1	0.416	2141	sustantivo
294	ASTRÓNOMO	1	0.416	2142	sustantivo
295	ASÍ	1	0.416	2143	adverbio
296	AUMENTO1	1	0.416	2144	sustantivo
297	AVANZADOA	1	0.416	2145	adjetivo
298	BANCO	1	0.416	2146	sustantivo

ANEXO I: FRECUENCIAS DEL TEXTO LEMATIZADO FESTESE001V

OL	Lema	n_i	f_i (a1000)	N_i	Categoría gramatical
299	BOLA	1	0.416	2147	sustantivo
300	BRILLO	1	0.416	2148	sustantivo
301	BRUNO	1	0.416	2149	sustantivo
302	CAMINO	1	0.416	2150	sustantivo
303	CATÁLOGO	1	0.416	2151	sustantivo
304	CENTENAR	1	0.416	2152	sustantivo
305	CENTRO	1	0.416	2153	sustantivo
306	CERCANO	1	0.416	2154	adjetivo
307	CHARLES	1	0.416	2155	sustantivo
308	CICLO	1	0.416	2156	sustantivo
309	CIELO	1	0.416	2157	sustantivo
310	CIRCUNDANTE	1	0.416	2158	adjetivo
311	CIUDAD	1	0.416	2159	sustantivo
312	CLARAMENTE	1	0.416	2160	adverbio
313	CLAVE	1	0.416	2161	sustantivo
314	COLAPSADOA	1	0.416	2162	adjetivo
315	COLAPSAR	1	0.416	2163	verbo
316	COLECCIÓN	1	0.416	2164	sustantivo
317	COLISIÓN	1	0.416	2165	sustantivo
318	COLOCAR	1	0.416	2166	verbo
319	COMPARTIR	1	0.416	2167	verbo
320	COMPLICADOA	1	0.416	2168	adjetivo
321	COMPONENTE	1	0.416	2169	sustantivo
322	COMPRENDER	1	0.416	2170	verbo
323	COMUNIDAD	1	0.416	2171	sustantivo
324	CONCENTRAR	1	0.416	2172	verbo
325	CONCISO	1	0.416	2173	adjetivo
326	CONOCER	1	0.416	2174	verbo
327	CONOCIDOA	1	0.416	2175	adjetivo
328	CONSECUENCIA	1	0.416	2176	sustantivo
329	CONSIDERABLE	1	0.416	2177	adjetivo
330	CONSIDERAR	1	0.416	2178	verbo
331	CONSIGUIENTE	1	0.416	2179	adjetivo
332	CONSTANTE1	1	0.416	2180	sustantivo
333	CONSTANTEA	1	0.416	2181	adjetivo
334	CONSTELACIÓN	1	0.416	2182	sustantivo
335	CONSTRUIR	1	0.416	2183	verbo
336	CONSUMIR	1	0.416	2184	verbo
337	CONTAMINACIÓN	1	0.416	2185	sustantivo
338	CONTINUAR	1	0.416	2186	verbo
339	CONTRACCIÓN	1	0.416	2187	sustantivo
340	CONTRARIO	1	0.416	2188	adjetivo
341	COORDENADA	1	0.416	2189	sustantivo
342	CORRELACIÓN	1	0.416	2190	sustantivo
343	CORRESPONDER	1	0.416	2191	verbo
344	CORTESÍA	1	0.416	2192	sustantivo
345	CREER	1	0.416	2193	verbo
346	CUALQUIER	1	0.416	2194	adjetivo

ANEXO I: FRECUENCIAS DEL TEXTO LEMATIZADO FESTESE001V

OL	Lema	n_i	f_i (a1000)	N_i	Categoría gramatical
347	CUERPO	1	0.416	2195	sustantivo
348	CURIOSO	1	0.416	2196	adjetivo
349	DECAER	1	0.416	2197	verbo
350	DEDUCIR	1	0.416	2198	verbo
351	DEJAR	1	0.416	2199	verbo
352	DENOMINAR	1	0.416	2200	verbo
353	DENSO	1	0.416	2201	adjetivo
354	DENTRO	1	0.416	2202	adverbio
355	DEPENDER	1	0.416	2203	verbo
356	DEPOSITAR	1	0.416	2204	verbo
357	DERIVAR	1	0.416	2205	verbo
358	DESCRIBIR	1	0.416	2206	verbo
359	DESCUBRIR	1	0.416	2207	verbo
360	DESPLAZAR	1	0.416	2208	verbo
361	DESTRUIR	1	0.416	2209	verbo
362	DETERMINADOA	1	0.416	2210	adjetivo
363	DEUTERIO	1	0.416	2211	sustantivo
364	DIAGONAL	1	0.416	2212	adjetivo
365	DIFERIR	1	0.416	2213	verbo
366	DIFUSO	1	0.416	2214	adjetivo
367	DISTINTO	1	0.416	2215	adjetivo
368	DISTRIBUIR	1	0.416	2216	verbo
369	DIVIDIR	1	0.416	2217	verbo
370	DUODÉCIMO	1	0.416	2218	adjetivo
371	DÍA	1	0.416	2219	sustantivo
372	E1	1	0.416	2220	sustantivo
373	EINSTEIN	1	0.416	2221	sustantivo
374	EISENDLE	1	0.416	2222	sustantivo
375	ELEVAR	1	0.416	2223	verbo
376	ELLO	1	0.416	2224	funcional
377	EMITIDOA	1	0.416	2225	adjetivo
378	EMITIR	1	0.416	2226	verbo
379	ENANO	1	0.416	2227	adjetivo
380	ENDE	1	0.416	2228	adverbio
381	ENFRIAR	1	0.416	2229	verbo
382	ENORME	1	0.416	2230	adjetivo
383	ENTONCES	1	0.416	2231	adverbio
384	ERROR	1	0.416	2232	sustantivo
385	ESA1	1	0.416	2233	sustantivo
386	ESCALA1	1	0.416	2234	sustantivo
387	ESO1	1	0.416	2235	sustantivo
388	ESPACIO	1	0.416	2236	sustantivo
389	ESPECÍFICO	1	0.416	2237	adjetivo
390	ESTADO	1	0.416	2238	sustantivo
391	ESTIMACIÓN	1	0.416	2239	sustantivo
392	ESTIMAR	1	0.416	2240	verbo
393	ESTUDIADOA	1	0.416	2241	adjetivo
394	ESTUDIAR	1	0.416	2242	verbo

ANEXO I: FRECUENCIAS DEL TEXTO LEMATIZADO FESTESE001V

OL	Lema	n_i	f_i (a1000)	N_i	Categoría gramatical
395	ESTUDIO1	1	0.416	2243	sustantivo
396	ESTÁNDARA	1	0.416	2244	adjetivo
397	ETAPA	1	0.416	2245	sustantivo
398	EXISTIR	1	0.416	2246	verbo
399	EXPANDIR	1	0.416	2247	verbo
400	EXPLICAR	1	0.416	2248	verbo
401	EXPLOTAR	1	0.416	2249	verbo
402	EXPULSAR	1	0.416	2250	verbo
403	EXTRAER	1	0.416	2251	verbo
404	FIGURA	1	0.416	2252	sustantivo
405	FILTRO	1	0.416	2253	sustantivo
406	FINAL	1	0.416	2254	sustantivo
407	FINALA	1	0.416	2255	adjetivo
408	FINALMENTE	1	0.416	2256	adverbio
409	FORMIDABLE	1	0.416	2257	adjetivo
410	FRACCIÓN	1	0.416	2258	sustantivo
411	FRENTE	1	0.416	2259	adverbio
412	FUERZA	1	0.416	2260	sustantivo
413	FUNCIÓN	1	0.416	2261	sustantivo
414	FÍSICO	1	0.416	2262	adjetivo
415	GENERAR	1	0.416	2263	verbo
416	GIGANTESCO	1	0.416	2264	adjetivo
417	GIRO1	1	0.416	2265	sustantivo
418	GRADO	1	0.416	2266	sustantivo
419	GRAVITATORIAMENTE	1	0.416	2267	adverbio
420	GRÁFICO	1	0.416	2268	sustantivo
421	HABER	1	0.416	2269	verbo
422	HECHO1	1	0.416	2270	sustantivo
423	HERSCHEL	1	0.416	2271	sustantivo
424	HERTZSPRUNGRUSSELL	1	0.416	2272	sustantivo
425	HIPPARCOS	1	0.416	2273	sustantivo
426	HORA	1	0.416	2274	sustantivo
427	HORIZONTAL	1	0.416	2275	adjetivo
428	HOY	1	0.416	2276	adverbio
429	HUECO	1	0.416	2277	sustantivo
430	IDEAL	1	0.416	2278	adjetivo
431	ILUSTRACIÓN	1	0.416	2279	sustantivo
432	ILUSTRAR	1	0.416	2280	verbo
433	IMPLICADOA	1	0.416	2281	adjetivo
434	IMPLICAR	1	0.416	2282	verbo
435	INCREMENTAR	1	0.416	2283	verbo
436	INDICADOA	1	0.416	2284	adjetivo
437	INDICAR	1	0.416	2285	verbo
438	INDIVIDUAL	1	0.416	2286	adjetivo
439	INFORMACIÓN	1	0.416	2287	sustantivo
440	INTERACCIÓN	1	0.416	2288	sustantivo
441	INTERNACIONAL	1	0.416	2289	adjetivo
442	INTERNO	1	0.416	2290	adjetivo

ANEXO I: FRECUENCIAS DEL TEXTO LEMATIZADO FESTESE001V

OL	Lema	n_i	f_i (a1000)	N_i	Categoría gramatical
443	INTERÉS	1	0.416	2291	sustantivo
444	INVESTIGACIÓN	1	0.416	2292	sustantivo
445	INVOLUCRADOA	1	0.416	2293	adjetivo
446	IR	1	0.416	2294	verbo
447	IZQUIERDO	1	0.416	2295	adjetivo
448	JOVEN	1	0.416	2296	adjetivo
449	JUSTO	1	0.416	2297	adjetivo
450	LEY	1	0.416	2298	sustantivo
451	LIGADOA	1	0.416	2299	adjetivo
452	LIGERO	1	0.416	2300	adjetivo
453	LISTA	1	0.416	2301	sustantivo
454	LLEGARX	1	0.416	2302	f-verbo
455	LOCALIZAR	1	0.416	2303	verbo
456	LONGEVO	1	0.416	2304	adjetivo
457	LONGITUD	1	0.416	2305	sustantivo
458	LUEGO	1	0.416	2306	funcional
459	LÍMITE	1	0.416	2307	sustantivo
460	MASIVO	1	0.416	2308	adjetivo
461	MATERIA	1	0.416	2309	sustantivo
462	MATERIAL	1	0.416	2310	sustantivo
463	MAYORITARIAMENTE	1	0.416	2311	adverbio
464	MENUDO	1	0.416	2312	adjetivo
465	MEZCLA1	1	0.416	2313	sustantivo
466	MITAD	1	0.416	2314	sustantivo
467	MODESTO	1	0.416	2315	adjetivo
468	MODO	1	0.416	2316	sustantivo
469	MOVER	1	0.416	2317	verbo
470	MUESTRA1	1	0.416	2318	sustantivo
471	NACIMIENTO	1	0.416	2319	sustantivo
472	NEGRO	1	0.416	2320	adjetivo
473	NINGUNO	1	0.416	2321	adjetivo
474	NUBE	1	0.416	2322	sustantivo
475	NUCLEAR	1	0.416	2323	adjetivo
476	OBVIAMENTE	1	0.416	2324	adverbio
477	OFIUCO	1	0.416	2325	sustantivo
478	ONDA	1	0.416	2326	sustantivo
479	ORIGINAL	1	0.416	2327	adjetivo
480	PARSEC	1	0.416	2328	sustantivo
481	PARTICULAR	1	0.416	2329	adjetivo
482	PASAR	1	0.416	2330	verbo
483	PATRÓN	1	0.416	2331	sustantivo
484	PERMANECER	1	0.416	2332	verbo
485	PERO	1	0.416	2333	funcional
486	PERSPECTIVA	1	0.416	2334	sustantivo
487	PI	1	0.416	2335	sustantivo
488	PISTA	1	0.416	2336	sustantivo
489	PLANO	1	0.416	2337	sustantivo
490	PODER1	1	0.416	2338	sustantivo

ANEXO I: FRECUENCIAS DEL TEXTO LEMATIZADO FESTESE001V

OL	Lema	n _i	f _i (a1000)	N _i	Categoría gramatical
491	POLUCIÓN	1	0.416	2339	sustantivo
492	PORQUE	1	0.416	2340	funcional
493	POSIBLE	1	0.416	2341	adjetivo
494	PRECISOA	1	0.416	2342	adjetivo
495	PRISMÁTICO	1	0.416	2343	adjetivo
496	PROBABLE	1	0.416	2344	adjetivo
497	PROBABLEMENTE	1	0.416	2345	adverbio
498	PRODUCIR	1	0.416	2346	verbo
499	PROGRAMA	1	0.416	2347	sustantivo
500	PRONTO	1	0.416	2348	adverbio
501	PROPORCIONAL	1	0.416	2349	adjetivo
502	PROPORCIONAR	1	0.416	2350	verbo
503	PROTOGALÁCTICO	1	0.416	2351	adjetivo
504	PROTÓN	1	0.416	2352	sustantivo
505	PROVENIR	1	0.416	2353	verbo
506	PRÓXIMO	1	0.416	2354	adjetivo
507	PUBLICAR	1	0.416	2355	verbo
508	PÁGINA	1	0.416	2356	sustantivo
509	QUEDAR	1	0.416	2357	verbo
510	QUÍMICO	1	0.416	2358	adjetivo
511	RAINER	1	0.416	2359	sustantivo
512	RANGO	1	0.416	2360	sustantivo
513	RAZÓN	1	0.416	2361	sustantivo
514	RECIBIR	1	0.416	2362	verbo
515	RELACIÓN	1	0.416	2363	sustantivo
516	REMANENTE	1	0.416	2364	sustantivo
517	REPETIDOA	1	0.416	2365	adjetivo
518	REPETIR	1	0.416	2366	verbo
519	RESERVA1	1	0.416	2367	sustantivo
520	RESOLUCIÓN	1	0.416	2368	sustantivo
521	RESOLVER	1	0.416	2369	verbo
522	RESTANTE	1	0.416	2370	adjetivo
523	RESTO	1	0.416	2371	sustantivo
524	RESULTADO1	1	0.416	2372	sustantivo
525	SEIS	1	0.416	2373	adjetivo
526	SERIE	1	0.416	2374	sustantivo
527	SIMPLE	1	0.416	2375	adjetivo
528	SIN	1	0.416	2376	funcional
529	SOLO	1	0.416	2377	adjetivo
530	STAMPFER	1	0.416	2378	sustantivo
531	SUBPRODUCTO	1	0.416	2379	sustantivo
532	SUCESO	1	0.416	2380	sustantivo
533	SUMINISTRO1	1	0.416	2381	sustantivo
534	SUPERFICIE	1	0.416	2382	sustantivo
535	SUPERNOVA	1	0.416	2383	sustantivo
536	SUPERVIVIENTE	1	0.416	2384	adjetivo
537	SUPUESTO1	1	0.416	2385	sustantivo
538	SÍ	1	0.416	2386	funcional

ANEXO I: FRECUENCIAS DEL TEXTO LEMATIZADO FESTESE001V

OL	Lema	n _i	f _i (a1000)	N _i	Categoría gramatical
539	TAMAÑO	1	0.416	2387	sustantivo
540	TELESCOPIO	1	0.416	2388	sustantivo
541	TEORÍA	1	0.416	2389	sustantivo
542	TERMONUCLEAR	1	0.416	2390	adjetivo
543	TIERRA	1	0.416	2391	sustantivo
544	TRANSFORMAR	1	0.416	2392	verbo
545	TRAVÉS	1	0.416	2393	sustantivo
546	TÉRMICO	1	0.416	2394	adjetivo
547	TÉRMINO	1	0.416	2395	sustantivo
548	UNIVERSO	1	0.416	2396	sustantivo
549	VECINDAD	1	0.416	2397	sustantivo
550	VIOLENTO	1	0.416	2398	adjetivo
551	VISTA	1	0.416	2399	sustantivo
552	WILLIAM	1	0.416	2400	sustantivo
553	.	1	0.416	2401	lm
554	ÁTOMO	1	0.416	2402	sustantivo
555	ÍNDICE	1	0.416	2403	sustantivo
556	ÚTIL	1	0.416	2404	adjetivo

Nota. OL= orden lexicométrico; n= frecuencia absoluta; f= frecuencia relativa; N= frecuencia acumulada.

ANEXO J: Hallando Colocaciones en FESTESE001V

A continuación se muestra el listado de los 122 lemas, de los 556 lemas distintos que presentan un mínimo de 2 relaciones con otro. Se señalan en **negrita** las uniones a considerar. La tabla se ordena en primer lugar lexicométricamente de mayor a menor frecuencia absoluta del lema origen, y en cada uno de éstos de mayor a menor relación absoluta del lema que sigue.

Tabla 97

Listado de lemas con un mínimo de dos relaciones con el siguiente

Lema origen	Lema	ra_i	rr_i	n_i	ra_i/n_i	Colocación
el	estrella	36	33.364	273	0.13	
el	secuencia	13	12.048	273	0.05	
el	cúmulo	12	11.121	273	0.04	
el	núcleo	11	10.195	273	0.04	
el	temperatura	10	9.268	273	0.04	
el	proceso	7	6.487	273	0.03	
el	mayoría	6	5.561	273	0.02	
el	vida	6	5.561	273	0.02	
el	diagrama	5	4.634	273	0.02	
el	luminosidad	5	4.634	273	0.02	
el	masa	5	4.634	273	0.02	
el	capa	4	3.707	273	0.01	
el	edad	4	3.707	273	0.01	
el	galaxia	4	3.707	273	0.01	
el	sección	4	3.707	273	0.01	
el	cual	3	2.780	273	0.01	
el	disco	3	2.780	273	0.01	
el	gigante	3	2.780	273	0.01	
el	línea	3	2.780	273	0.01	
el	más	3	2.780	273	0.01	
el	que	3	2.780	273	0.01	
el	sol	3	2.780	273	0.01	
el	0coma7	2	1.854	273	0.01	

ANEXO J: HALLANDO COLOCACIONES EN FESTESE001V

Lema origen	Lema	ra _i	rr _i	n _i	ra _i /n _i	Colocación
el	combustión	2	1.854	273	0.01	
el	cuatro	2	1.854	273	0.01	
el	diferente	2	1.854	273	0.01	
el	distancia	2	1.854	273	0.01	
el	fase	2	1.854	273	0.01	
el	fig	2	1.854	273	0.01	
el	hidrógeno	2	1.854	273	0.01	
el	largo	2	1.854	273	0.01	
el	magnitud	2	1.854	273	0.01	
el	medida1	2	1.854	273	0.01	
el	parte	2	1.854	273	0.01	
el	pléyades	2	1.854	273	0.01	
el	presión	2	1.854	273	0.01	
el	primero	2	1.854	273	0.01	
el	radiación	2	1.854	273	0.01	
el	valor	2	1.854	273	0.01	
el	vía	2	1.854	273	0.01	
de	el	65	60.241	222	0.29	
de	estrella	18	16.682	222	0.08	
de	uno	14	12.975	222	0.06	
de	su	9	8.341	222	0.04	
de	formal	7	6.487	222	0.03	
de	año	6	5.561	222	0.03	
de	esto	6	5.561	222	0.03	
de	fusión	6	5.561	222	0.03	
de	hidrógeno	6	5.561	222	0.03	
de	aproximadamente	5	4.634	222	0.02	
de	millón	5	4.634	222	0.02	
de	helio	4	3.707	222	0.02	
de	bajoa	3	2.780	222	0.01	
de	gas	3	2.780	222	0.01	
de	nuestro	3	2.780	222	0.01	
de	todo	3	2.780	222	0.01	
de	color	2	1.854	222	0.01	
de	cúmulo	2	1.854	222	0.01	
de	energía	2	1.854	222	0.01	
de	luminosidad	2	1.854	222	0.01	
de	magnitud	2	1.854	222	0.01	
de	manera	2	1.854	222	0.01	
de	nuevo	2	1.854	222	0.01	
de	stefanboltzmann	2	1.854	222	0.01	
estrella	se	10	9.268	76	0.13	
estrella	ser	7	6.487	76	0.09	
estrella	de	6	5.561	76	0.08	
estrella	en	4	3.707	76	0.05	
estrella	y	4	3.707	76	0.05	
estrella	=	3	2.780	76	0.04	
estrella	más	3	2.780	76	0.04	
estrella	que	3	2.780	76	0.04	

ANEXO J: HALLANDO COLOCACIONES EN FESTESE001V

Lema origen	Lema	ra _i	rr _i	n _i	ra _i /n _i	Colocación
estrella	como	2	1.854	76	0.03	
estrella	normal	2	1.854	76	0.03	
estrella	uno	2	1.854	76	0.03	
uno	estrella	8	7.414	74	0.11	
uno	cúmulo	7	6.487	74	0.09	
uno	poco	5	4.634	74	0.07	
uno	de	2	1.854	74	0.03	
uno	formal	2	1.854	74	0.03	
uno	grupo	2	1.854	74	0.03	
uno	imagen	2	1.854	74	0.03	
uno	luminosidad	2	1.854	74	0.03	
en	el	37	34.291	64	0.58	colocación
en	uno	6	5.561	64	0.09	
en	él	2	1.854	64	0.03	
en	esto	2	1.854	64	0.03	
se	encontrar	7	6.487	54	0.13	
se	poderx	6	5.561	54	0.11	
se	formar	3	2.780	54	0.06	
se	haberx	3	2.780	54	0.06	
se	componer	2	1.854	54	0.04	
se	extender	2	1.854	54	0.04	
se	fusionar	2	1.854	54	0.04	
se	llamar	2	1.854	54	0.04	
y	el	10	9.268	48	0.21	
y	polvo	3	2.780	48	0.06	
y	se	3	2.780	48	0.06	
y	cúmulo	2	1.854	48	0.04	
y	mv	2	1.854	48	0.04	
y	por	2	1.854	48	0.04	
y	su	2	1.854	48	0.04	
y	tener	2	1.854	48	0.04	
y	uno	2	1.854	48	0.04	
ser	el	8	7.414	41	0.20	
ser	uno	8	7.414	41	0.20	
ser	muy	4	3.707	41	0.10	
ser	de	3	2.780	41	0.07	
ser	aproximadamente	2	1.854	41	0.05	
ser	estable	2	1.854	41	0.05	
ser	por	2	1.854	41	0.05	
cúmulo	globular	17	15.755	39	0.44	
cúmulo	abiertoa	5	4.634	39	0.13	
cúmulo	de	5	4.634	39	0.13	
que	el	10	9.268	39	0.26	
que	se	3	2.780	39	0.08	
que	ser	3	2.780	39	0.08	
que	esto	2	1.854	39	0.05	
que	tener	2	1.854	39	0.05	
que	ver	2	1.854	39	0.05	
del	cúmulo	4	3.707	30	0.13	

ANEXO J: HALLANDO COLOCACIONES EN FESTESE001V

Lema origen	Lema	ra _i	rr _i	n _i	ra _i /n _i	Colocación
del	hidrógeno	4	3.707	30	0.13	
del	diagrama	3	2.780	30	0.10	
del	núcleo	3	2.780	30	0.10	
del	sol	3	2.780	30	0.10	
del	proceso	2	1.854	30	0.07	
a	el	10	9.268	28	0.36	
a	partir	3	2.780	28	0.11	
esto	ser	4	3.707	26	0.15	
esto	cúmulo	3	2.780	26	0.12	
esto	estrella	2	1.854	26	0.08	
esto	fase	2	1.854	26	0.08	
núcleo	de	7	6.487	19	0.37	
núcleo	se	3	2.780	19	0.16	
núcleo	estelar	2	1.854	19	0.11	
su	masa	3	2.780	19	0.16	
su	edad	2	1.854	19	0.11	
más	pesadoa	4	3.707	18	0.22	
más	brillante	2	1.854	18	0.11	
globular	ser	5	4.634	17	0.29	
globular	messier	2	1.854	17	0.12	
globular	se	2	1.854	17	0.12	
globular	y	2	1.854	17	0.12	
para	el	7	6.487	17	0.41	
para	formar	2	1.854	17	0.12	
por	el	7	6.487	16	0.44	
por	tanto	2	1.854	16	0.13	
temperatura	superficial	8	7.414	15	0.53	colocación
temperatura	central	2	1.854	15	0.13	
poderx	calcular	2	1.854	14	0.14	
poderx	determinar	2	1.854	14	0.14	
poderx	ver	2	1.854	14	0.14	
masa	de	4	3.707	13	0.31	
masa	se	2	1.854	13	0.15	
muy	grande	2	1.854	13	0.15	
muy	raro	2	1.854	13	0.15	
principal	ver	2	1.854	13	0.15	
principal	y	2	1.854	13	0.15	
secuencia	principal	13	12.048	13	1.00	colocación
x	10	5	4.634	13	0.38	
x	l	2	1.854	13	0.15	
x	r	2	1.854	13	0.15	
aproximadamente	uno	3	2.780	12	0.25	
hidrógeno	en	2	1.854	12	0.17	
tener	uno	8	7.414	12	0.67	colocación
tener	lugar	2	1.854	12	0.17	
10	a3	4	3.707	11	0.36	
10	a4	3	2.780	11	0.27	
como	el	5	4.634	11	0.45	
como	uno	3	2.780	11	0.27	

ANEXO J: HALLANDO COLOCACIONES EN FESTESE001V

Lema origen	Lema	ra _i	rr _i	n _i	ra _i /n _i	Colocación
diagrama	hr	5	4.634	11	0.45	
l	solar	3	2.780	11	0.27	
luminosidad	del	3	2.780	11	0.27	
luminosidad	l	3	2.780	11	0.27	
luminosidad	de	2	1.854	11	0.18	
ver	el	3	2.780	11	0.27	
ver	con	2	1.854	11	0.18	
año	luz	4	3.707	10	0.40	
formal	que	4	3.707	10	0.40	
t	de	3	2.780	10	0.30	
t	estrella	2	1.854	10	0.20	
poco	ciento	2	1.854	9	0.22	
poco	mil	2	1.854	9	0.22	
proceso	de	6	5.561	9	0.67	colocación
todo	el	6	5.561	9	0.67	colocación
con	el	3	2.780	8	0.38	
encontrar	en	3	2.780	8	0.38	
encontrar	el	2	1.854	8	0.25	
millón	de	8	7.414	8	1.00	colocación
no	ser	2	1.854	8	0.25	
superficial	t	4	3.707	8	0.50	colocación
superficial	de	2	1.854	8	0.25	
edad	de	4	3.707	7	0.57	colocación
edad	del	2	1.854	7	0.29	
entre	el	5	4.634	7	0.71	colocación
magnitud	en	2	1.854	7	0.29	
magnitud	visible	2	1.854	7	0.29	
porciento	de	4	3.707	7	0.57	colocación
porciento	residual	2	1.854	7	0.29	
vida	de	6	5.561	7	0.86	colocación
cuando	el	2	1.854	6	0.33	
cuando	se	2	1.854	6	0.33	
gigante	rojo	6	5.561	6	1.00	colocación
mayoría	de	5	4.634	6	0.83	colocación
mil	de	5	4.634	6	0.83	colocación
abiertoa	de	4	3.707	5	0.80	colocación
bajoa	masa	3	2.780	5	0.60	colocación
determinar	el	3	2.780	5	0.60	colocación
elemento	más	3	2.780	5	0.60	colocación
gas	y	4	3.707	5	0.80	colocación
messier	12	2	1.854	5	0.40	
nuestro	galaxia	2	1.854	5	0.40	
nuestro	vía	2	1.854	5	0.40	
pesadoa	del	2	1.854	5	0.40	
sólo	el	2	1.854	5	0.40	
usar	para	2	1.854	5	0.40	
vía	lácteo	5	4.634	5	1.00	colocación
ya	no	3	2.780	5	0.60	colocación
3	x	2	1.854	4	0.50	colocación

ANEXO J: HALLANDO COLOCACIONES EN FESTESE001V

Lema origen	Lema	ra _i	rr _i	n _i	ra _i /n _i	Colocación
a2	x	2	1.854	4	0.50	colocación
a3	k	2	1.854	4	0.50	colocación
a4	k	2	1.854	4	0.50	colocación
cantidad	de	4	3.707	4	1.00	colocación
capa	externo	3	2.780	4	0.75	colocación
componer	de	2	1.854	4	0.50	colocación
dar	uno	2	1.854	4	0.50	colocación
desde	el	2	1.854	4	0.50	colocación
disco	de	2	1.854	4	0.50	colocación
donde	se	2	1.854	4	0.50	colocación
fig	6	2	1.854	4	0.50	colocación
hacia	el	2	1.854	4	0.50	colocación
helio	ser	2	1.854	4	0.50	colocación
herramienta	astronómico	4	3.707	4	1.00	colocación
k	tener	2	1.854	4	0.50	colocación
mucho	tiempo	2	1.854	4	0.50	colocación
nuevo	estrella	2	1.854	4	0.50	colocación
parte	superior	2	1.854	4	0.50	colocación
radiación	de	2	1.854	4	0.50	colocación
sección	herramienta	4	3.707	4	1.00	colocación
superior	derecho	2	1.854	4	0.50	colocación
convertir	en	2	1.854	3	0.67	colocación
cual	se	2	1.854	3	0.67	colocación
cuatro	núcleo	3	2.780	3	1.00	colocación
deber	a	3	2.780	3	1.00	colocación
distancia	de	2	1.854	3	0.67	colocación
equilibrio	entre	2	1.854	3	0.67	colocación
externo	de	2	1.854	3	0.67	colocación
fusionar	para	2	1.854	3	0.67	colocación
mb	y	2	1.854	3	0.67	colocación
nebulosa	planetario	2	1.854	3	0.67	colocación
partir	de	3	2.780	3	1.00	colocación
sistema	de	2	1.854	3	0.67	colocación
típico	de	2	1.854	3	0.67	colocación
total	de	2	1.854	3	0.67	colocación
vez	el	2	1.854	3	0.67	colocación
0coma7	porciento	2	1.854	2	1.00	colocación
4	x	2	1.854	2	1.00	colocación
blanco	y	2	1.854	2	1.00	colocación
c	a2	2	1.854	2	1.00	colocación
ciento	de	2	1.854	2	1.00	colocación
combustión	del	2	1.854	2	1.00	colocación
comenzarx	a	2	1.854	2	1.00	colocación
diámetro	típico	2	1.854	2	1.00	colocación
ejercidoa	hacia	2	1.854	2	1.00	colocación
extender	aproximadamente	2	1.854	2	1.00	colocación
largo	de	2	1.854	2	1.00	colocación
manera	que	2	1.854	2	1.00	colocación
medir	el	2	1.854	2	1.00	colocación

ANEXO J: HALLANDO COLOCACIONES EN FESTESE001V

Lema origen	Lema	ra_i	rr_i	n_i	ra_i/n_i	Colocación
mientras	el	2	1.854	2	1.00	colocación
nos	dar	2	1.854	2	1.00	colocación
orden	de	2	1.854	2	1.00	colocación
rama	de	2	1.854	2	1.00	colocación
residual	de	2	1.854	2	1.00	colocación

Nota. ra_i = relación absoluta; rr_i = relación relativa; n_i = frecuencia absoluta.

ANEXO K: Nuevo Texto Transformado FESTESE001VD

UNO ESTRELLA SER UNO BOLA GIGANDESCO DE GAS QUE TENER LUZ PROPIO CON PROPIEDAD FÍSICO TAL COMO EL MASA EL TEMPERATURA O EL RADIO. TAMBIÉN SER DE INTERÉS PARA EL ASTRÓNOMO EL DISTANCIA DE EL ESTRELLA A EL TIERRA. EL ESTRELLA MÁS CERCANO Y POR ENDE MÁS ESTUDIADOA SER POR SUPUESTO1 NUESTRO PROPIO SOL.

EL LUZ EMITIDOA POR EL MAYORÍA DE EL ESTRELLA SER UNO SUBPRODUCTO DEL PROCESO DE FUSIÓN TERMONUCLEAR QUE SE PRODUCIR EN EL NÚCLEO INTERNO DE EL ESTRELLA. UNO ESTRELLA NORMAL COMO EL SOL SE COMPONER DE APROXIMADAMENTE UNO 74 PORCIENTODE HIDRÓGENO Y UNO 25 PORCIENTODE HELIOSER EL RESTANTE 1 PORCIENTO UNO MEZCLA1 DE ELEMENTOMÁSPESADOA. EL PROCESO DE FUSIÓN MÁS COMÚN EN UNO ESTRELLA COMO EL SOL SER EL COMBUSTIÓNDELHIDRÓGENO EN EL CUAL CUATRONÚCLEO DE HIDRÓGENO SE FUSIONAR PARA FORMAR UNOA NÚCLEO DE HELIO. EL PROCESO SUCEDER EN VARIO ETAPA QUE SE ILUSTRAR EN EL FIG 2 . EN EL PRIMERO PASO1 DEL PROCESO DOS PROTÓN SE FUSIONAR PARA FORMAR DEUTERIO UNO FORMA1 PESADOA DEL HIDRÓGENO. ESTO SER UNO SUCESO MUY RARO INCLUSO PARA EL NÚCLEO DENSO DE EL ESTRELLA EN DONDE EL TEMPERATURA SER DE UNO POCO MILLÓNDE GRADO. SER POR ESTO POR EL QUE EL ESTRELLA COMO EL SOL NO EXPLOTAR EN UNO REACCIÓN VIOLENTO CUANDO COMENZAR EL PROCESO DE FUSIÓN SINO QUE PERMANECER EN ESTO FASE ESTABLE DE EL VIDA DE EL ESTRELLA DURANTE VARIO MIL DE MILLÓNDEAÑO. MIENTRASEL ESTRELLA SER ESTABLE EL TEMPERATURA DE SU SUPERFICIE SU RADIO Y SU LUMINOSIDAD SER APROXIMADAMENTE CONSTANTEA. EL REACCIÓN NUCLEAR DEL NÚCLEO DE EL

ANEXO K: NUEVO TEXTO TRANSFORMADO FESTESE001V

ESTRELLA GENERAR EL ENERGÍA JUSTO PARA MANTENER UNO EQUILIBRIO ENTRE EL PRESIÓN TÉRMICO EJERCIDO AHACIA FUERA Y EL FUERZA GRAVITATORIO EJERCIDO AHACIA DENTRO.

EL MASA DE UNO A ÁTOMO DE HELIO SER SÓLO EL 99 COMA 3 POR CIENTO DE EL MASA DE EL CUATRO NÚCLEO ORIGINAL DE HIDRÓGENO. EL PROCESO DE FUSIÓN CONVERTIR EL 0 COMA 7 POR CIENTO RESIDUAL DE LA MASA EN ENERGÍA MAYORITARIAMENTE EN FORMA DE LUZ. EL CANTIDAD DE ENERGÍA SE PODERÁ CALCULAR USAR EL FAMOSO ECUACIÓN DE EINSTEIN $E = M \times CA^2$. YA QUE CA^2 SER UNO NÚMERO MUY GRANDE ESTO ECUACIÓN IMPLICAR QUE INCLUSO UNO PEQUEÑO CANTIDAD DE MATERIA PODERÁ CONVERTIR EN UNO FORMIDABLE CANTIDAD DE ENERGÍA. EL 0 COMA 7 POR CIENTO RESIDUAL DE LA MASA DE EL CUATRO NÚCLEO DE HIDRÓGENO INVOLUCRADO EN UNO SOLO REACCIÓN PODERÁ PARECER PEQUEÑO PERO CUANDO SE CONSIDERAR EL NÚMERO TOTAL DE REACCIÓN DE TODO EL PROCESO DE FUSIÓN HABERÁ IMPLICADO UNO MASA TOTAL Y POR TANTO ENERGÍA CONSIDERABLE.

EL TÉRMINO CÚMULO DE ESTRELLA SE USAR PARA DOS TIPO DIFERENTE DE AGRUPACIÓN DE ESTRELLA CÚMULO ABIERTO DE ESTRELLA Y CÚMULO GLOBULAR DE ESTRELLA.

EL CÚMULO ABIERTO DE ESTRELLA SER COLECCIÓN NO COMPACTO DE ESTRELLA RELATIVAMENTE JOVEN QUE IR DESDE UNO CENTENAR HASTA UNO POCO MIL DE ESTRELLA. TENER UNO EDAD DE UNO POCO CIENTO DE MILLÓN DE AÑO UNO FRACCIÓN PEQUEÑO DEL TIEMPO TOTAL DE EL VIDA DE EL ESTRELLA UNO POCO MIL DE MILLÓN DE AÑO. ESTO CÚMULO SE ENCONTRAR EN EL DISCO DE NUESTRO GALAXIA EL VÍALÁCTEO Y A MENUDO CONTENER NUBE DE GAS Y POLVO DONDE SE FORMAR NUEVO ESTRELLA. EL DIÁMETRO TÍPICO DE UNO CÚMULO ABIERTO DE ESTRELLA SER DE APROXIMADAMENTE 30 AÑO LUZ 10 PARSEC.

ESTO SER UNO DE EL MÁS FAMOSO CÚMULO DE ESTRELLA DEL CIELO. EL PLÉYADES PODERÁ VER A SIMPLE VISTA INCLUSO DESDE EL MAYORÍA DE EL

ANEXO K: NUEVO TEXTO TRANSFORMADO FESTESE001V

CIUDAD PESAR A EL CONTAMINACIÓN LUMÍNICO. SER UNO DE EL MÁS BRILLANTE Y MÁS PRÓXIMO CÚMULO ABIERTO. EL CÚMULO DE EL PLÉYADES CONTENER MÁS DE 3000 ESTRELLA ESTAR A UNO DISTANCIA DE UNO 400 AÑOS LUZ Y TENER UNO TAMAÑO DE SÓLO 13 AÑOS LUZ. CORTESÍA DE BRUNO STAMPFER Y RAINER EISENDLE.

EN EL HALO Y EN EL DISCO DE NUESTRO VÍALÁCTEO EXISTIR UNO POCO CIENTOS DE CÚMULO ESFÉRICO Y COMPACTO LLAMADO CÚMULO GLOBULAR Y QUE ESTAR LIGADO GRAVITATORIAMENTE A NUESTRA GALAXIA.

CADA CÚMULO GLOBULAR SE COMPONER DE UNO GRUPO ESFÉRICO DE HASTA UNO MILLÓN DE ESTRELLA Y TENER UNO DIÁMETRO TÍPICO DE UNO 100 AÑOS LUZ. EL MAYORÍA DE EL CÚMULO GLOBULAR SER MUY VIEJO Y EL MÁS PROBABLE SER QUE SER ANTERIOR A EL FORMACIÓN DE EL PROPIA GALAXIA QUE TENER LUGAR HACER APROXIMADAMENTE 12 MIL MILLÓN DE AÑOS CUANDO EL MAYORÍA DEL MATERIAL PROTO GALÁCTICO SE DEPOSITAR EN EL DISCO.

MUCHO CÚMULO GLOBULAR PROBABLEMENTE SE HABER DESTRUIR CON EL PASO DE EL MIL DE MILLÓN DE AÑOS POR EL REPETIDA COLISIÓN E INTERACCIÓN ENTRE SU ESTRELLA Y CON EL VÍALÁCTEO. EL CÚMULO GLOBULAR SUPERVIVIENTE SER MÁS VIEJO QUE CUALQUIER OTRO ESTRUCTURA DE NUESTRO VÍALÁCTEO.

ESTO ILUSTRACIÓN DAR UNO PERSPECTIVA GENERAL DE EL GALAXIA VÍALÁCTEO. SE INDICAR EL DIFERENTE COMPONENTE DE ESTO COMPLICADO SISTEMA DE ESTRELLA GAS Y POLVO. EL PLANO DEL DISCO SE ENCONTRAR A EL LARGO DE EL LÍNEA CENTRAL HORIZONTAL. EL CÚMULO GLOBULAR SE DISTRIBUIR EN UNO HALO ESFÉRICO ALREDEDOR DEL CENTRO DE EL GALAXIA. SE CREER QUE ESTO DISTRIBUCIÓN TENER QUE VER CON EL HECHO DE QUE ESTO CÚMULO DE ESTRELLA SE FORMAR MUY PRONTO EN EL HISTORIA DE EL GALAXIA.

EL ESTUDIO ASTROFÍSICO DE EL CÚMULO GLOBULAR SER UNO PARTE IMPORTANTE DEL PROGRAMA DE INVESTIGACIÓN DE EL COMUNIDAD ASTRONÓMICO INTERNACIONAL. ESTO CÚMULO DE ESTRELLA SER IMPORTANTE

ANEXO K: NUEVO TEXTO TRANSFORMADO FESTESE001V

NO SÓLO COMO UNO VALIOSO BANCO DE PRUEBA¹ DE VALOR PARA EL TEORÍA DE ESTRUCTURA Y EVOLUCIÓN ESTELAR SINO TAMBIÉN PORQUE SE ENCONTRAR ENTRE EL POCO OBJETO DE EL GALAXIA PARA EL CUAL SE PODERX DETERMINAR SU EDAD DE UNO FORMA¹ RELATIVAMENTE PRECISOA. POR SER TANTO LONGEVO PROPORCIONAR UNO LÍMITE INFERIOR MUY ÚTIL A EL EDAD DEL UNIVERSO. EL DISTRIBUCIÓN DE SU EDAD Y EL CORRELACIÓN ENTRE EL EDAD DE UNO CÚMULO Y EL ABUNDANCIA EN ÉL DE EL DISTINTO ELEMENTO QUÍMICO HACER DE ESTO SISTEMA UNO PRUEBA¹ VALIOSO EN EL PROCESO DE FORMACIÓN ESTELAR.

TODO EL ESTRELLA QUE COMPONER UNO CÚMULOGLOBULAR COMPARTIR UNO HISTORIA COMÚN Y DIFERIR ENTRE SÍ SÓLO EN SU MASA. POR EL TANTO EL CÚMULOGLOBULAR SER LUGAR IDEAL PARA ESTUDIAR EL EVOLUCIÓN DE EL ESTRELLA. EN EL EJERCICIO SIGUIENTE DETERMINAR ALGUNO DE EL PROPIEDAD DE UNO CÚMULOGLOBULAR PARTICULAR EL MESSIER¹² .

EL CÚMULOGLOBULAR MESSIER¹² O M¹² TAMBIÉN LLAMADOA NGC 6218 SERX DESCUBRIR EN 1764 POR CHARLES MESSIER DE FORMA¹ QUE SER EL OBJETO DUODÉCIMO DE MESSIER. COMO MUCHO OTRO CÚMULOGLOBULAR MESSIER EL DESCRIBIR COMO UNO NEBULOSA SIN ESTRELLA UNO CONSECUENCIA DEL MODESTO PODER¹ DE RESOLUCIÓN DE SU TELESCOPIO. WILLIAM HERSCHEL SER EL PRIMERO QUE RESOLVER EL CÚMULO EN ESTRELLA INDIVIDUAL EN 1783 .

M¹² SE ENCONTRAR EN EL CONSTELACIÓN DE OFIUCO Y SE PODERX VER CON PRISMÁTICO DESDE LUGAR CON MUY BAJOA POLUCIÓN LUMÍNICO. EL MAGNITUDVISIBLE DEL CONJUNTO DEL CÚMULOGLOBULAR SER 6COMA⁷ VER MÁS SOBRE MAGNITUD EN EL SECHERRASTRONÓMICO PÁGINA 2 Y EL ESTRELLA MÁS BRILLANTE DEL CÚMULO TENER UNO MAGNITUDVISIBLE DE 12 .

EL NGC NUEVO CATÁLOGO GENERAL SERX PUBLICAR EN 1888 . EN ÉL APARECER UNO LISTA DE CÚMULOGLOBULAR Y CÚMULOABIERTO A DE ESTRELLA NEBULOSAPLANETARIO Y DIFUSO REMANENTE DE SUPERNOVA GALAXIA DE TODO EL TIPO E INCLUSO ALGUNO ERROR QUE NO CORRESPONDER A NINGUNO OBJETO.

ANEXO K: NUEVO TEXTO TRANSFORMADO FESTESE001V

SE LLAMAR DIAGRAMA DE HERTZSPRUNGRUSSELL O DE FORMA MÁS CONCISO DIAGRAMA HR A EL GRÁFICO QUE MUESTRA LA LUMINOSIDAD L O MAGNITUD ABSOLUTA M FRENTE A LA TEMPERATURA SUPERFICIAL T DE LA ESTRELLA. EL FIG 6 MUESTRA UNO EJEMPLO GENERAL QUE SE HABER CONSTRUIR A PARTIR DE OBSERVACIÓN DE ESTRELLA EN EL VICINIDAD DE EL CÚMULO DONDE SE CONOCER LA DISTANCIA A PARTIR DE LA MEDIDA DE HIPPARCOS. LA TEMPERATURA SUPERFICIAL DE UNA ESTRELLA SE PUEDE DERIVAR DE EL VALOR MEDIDO DE SU COLOR M_B MENOS M_V VER EL SECHERRASTRONÓMICO.

SE VE CLARAMENTE DEL DIAGRAMA HR QUE LA MEDIDA L T PARA DIFERENTE ESTRELLA FORMAR UNO CURIOSO PATRÓN CUANDO SE COLOCAR SOBRE EL DIAGRAMA. LA ESTRELLA SE CONCENTRAR EN ÁREA ESPECÍFICO INDICADO EN EL FIGURA. EL DIAGRAMA HR NOS DAR LA CLAVE PARA COMPRENDER COMO LA ESTRELLA EVOLUCIONAR EN EL TIEMPO. LA ESTRELLA DEPENDER DE SU MASA SE MOVER A TRAVÉS DEL DIAGRAMA A LA LARGO DEL CAMINO DIFERENTE.

LA ESTRELLA PASAR LA MAYOR PARTE DE SU VIDA EN LA SECUENCIA PRINCIPAL QUE BURN EL HIDRÓGENO LENTAMENTE EN UNO ESTADO DE EQUILIBRIO ESTABLE. OBVIAMENTE ESTO SER LA RAZÓN POR LA QUE LA MAYORÍA DE LA ESTRELLA SE LOCALIZAR EN LA SECUENCIA PRINCIPAL APROXIMADAMENTE UNO LÍNEA RECTO DESDE EL VÉRTICE SUPERIOR IZQUIERDO AL VÉRTICE INFERIOR DERECHO EN EL DIAGRAMA.

CUANDO EL SUMINISTRO DE HIDRÓGENO EN EL NÚCLEO DE LA ESTRELLA SE AGOTAR EL COMBUSTIÓN DEL HIDRÓGENO Y ANOSER POSIBLE. ESTO SER EL FINAL DE LA FASE DE LA SECUENCIA PRINCIPAL DE LA VIDA DE LA ESTRELLA Y EL EQUILIBRIO ENTRE LA PRESIÓN DEL GAS Y LA CONTRACCIÓN GRAVITATORIO EN EL NÚCLEO ESTELAR Y ANOSER ESTABLE. LA FUSIÓN DEL HIDRÓGENO TENER LUGAR ENTONCES EN LA CAPA CIRCUNDANTE MIENTRAS EL NÚCLEO COMENZAR A CONTRAER. AL CONTRAER EL NÚCLEO SE ELEVAR SU PRESIÓN Y SU TEMPERATURA CENTRAL DE MANERA QUE EL NÚCLEO DE HELIO DEL NÚCLEO DE LA ESTRELLA COMENZAR A FUSIONAR Y FORMAR ELEMENTO MÁS PESADO. ESTO

ANEXO K: NUEVO TEXTO TRANSFORMADO FESTESE001V

CICLO SE PODERX REPETIR USAR ELEMENTO CADA VEZ MÁS PESADO A MEDIDA QUE EL ELEMENTO MÁS LIGERO SE IRX ACABAR EN EL NÚCLEO. DURANTE ESTO FASE EL ESTRELLA APARECER COMO UNO GIGANTEROJO. TAL ESTRELLA SE MOSTRAR EN EL DIAGRAMA HR FUERA DE EL LÍNEA DE EL SECUENCIA PRINCIPAL EN EL PARTE SUPERIOR DERECHO. EL AUMENTO DE EL TEMPERATURA CENTRAL HACER QUE EL CAPA EXTERNO DE EL ESTRELLA SE EXPANDIR Y ENFRIAR DE FORMA QUE EL TEMPERATURA SUPERFICIAL DECAER. EL CONJUNTO DE EL ESTRELLA LLEGAR A SER MUY GRANDE Y DEBERA EL INFERIOR TEMPERATURA SUPERFICIAL EMITIR HACIA EL ESPACIO SOBRE TODO RADIACIÓN DE LONGITUD DE ONDA MAYOR DE FORMA QUE EL ESTRELLA PARECER ROJO. A PESAR DE SU BAJA TEMPERATURA SUPERFICIAL TODO EL GIGANTEROJO TENER UNO ALTO LUMINOSIDAD DEBERA SU ENORME RADIO R. ESTO RESULTADO SE DEDUCIR DE EL LEY DE RADIACIÓN DE STEFAN-BOLTZMANN PARA EL RADIACIÓN DEL CUERPO NEGRO $L = \sigma \times 4\pi \times R^2 \times T^4$ DONDE σ SER EL CONSTANTE DE STEFAN-BOLTZMANN. VALOR TÍPICO PARA EL GIGANTEROJO SER $R \sim 10^2 \times R_{SOLAR}$ $T \sim 3$ HASTA 4×10^3 K DE FORMA QUE L SER APROXIMADAMENTE $10^3 \times L_{SOLAR}$.

CUANDO YA NO PODERX MANTENER EL PROCESO AVANZADO DE FUSIÓN EN EL NÚCLEO ESTELAR EL NÚCLEO SE COLAPSAR DE NUEVO. OTRO VEZ EL TEMPERATURA DEL NÚCLEO SE INCREMENTAR Y AHORA SE EXPULSAR EL CAPA EXTERNO DE EL ESTRELLA. EL LLAMADO NEBULOSA PLANETARIO SE FORMAR A PARTIR DE EL RESTO DE EL CAPA EXTERNO VER EJERCICIO 3 DE ESTO SERIE DE EJERCICIO DE ASTRONOMÍA DE ESA1 ESO1. EL NÚCLEO COLAPSADO ESTAR MUY CALIENTE BLANCO Y EL ESTRELLA SER MUY PEQUEÑO. A ESTO ESTRELLA SE EL DENOMINAR MUY ADECUADAMENTE ENANO BLANCO Y SER EL FASE FINAL DE EL VIDA DE UNO ESTRELLA NORMAL DE TIPO SOLAR.

PARA HACER UNO ESTIMACIÓN APROXIMADO DE EL RELACIÓN ENTRE EL LUMINOSIDAD L Y EL TEMPERATURA SUPERFICIAL T DE TODO EL ESTRELLA DE EL SECUENCIA PRINCIPAL VER EL DIAGRAMA HR FIG 6. EL LÍNEA MÁS O MENOS RECTO DE EL SECUENCIA PRINCIPAL SE EXTENDER APROXIMADAMENTE UNO

ANEXO K: NUEVO TEXTO TRANSFORMADO FESTESE001V

ORDEN DE MAGNITUD EN TEMPERATURA 3×10^3 HASTA 3×10^4 K. EL RANGO DE LUMINOSIDAD SE EXTENDE APROXIMADAMENTE SEIS ORDENES DE MAGNITUD 10 MENOS HASTA 10^4 L_{SOLAR} . POR TANTO PODEMOS ESTIMAR DE FORMA APROXIMADA LA PROPORCIÓN DE LA TEMPERATURA PARA LA ESTRELLA DE LA SECUENCIA PRINCIPAL.

PARA DAR ALGUNO EJEMPLO DE UNA ESTRELLA DE GRANDE MASA DE LA SECUENCIA PRINCIPAL CON UNA TEMPERATURA SUPERFICIAL DE APROXIMADAMENTE $T_{\text{ESTRELLA}} = 10^4$ K TENER UNA LUMINOSIDAD DE APROXIMADAMENTE $L_{\text{ESTRELLA}} = 10$ DIVIDIR POR $10^8 \cdot L_{\text{SOLAR}}$ O APROXIMADAMENTE 26 VECES LA LUMINOSIDAD DEL SOL. EL VALOR ESTÁNDAR DE LA ESCALA DE LUMINOSIDAD.

UNA ESTRELLA DE BAJA MASA CON $T_{\text{ESTRELLA}} = 3 \times 10^3$ K TENER UNA LUMINOSIDAD DE APROXIMADAMENTE SÓLO EL 5 POR CIENTO DE LA LUMINOSIDAD DEL SOL.

TODO EL INFORMACIÓN QUE PODEMOS EXTRAER DE LA ESTRELLA PROVIENE DE LA RADIACIÓN QUE RECIBIMOS DE ELLO. COMO SE EXPLICARÁ EN EL SECCION ASTRONÓMICO DIFERENTE FILTRO Y SISTEMA DE COLOR SE PODEMOS USAR PARA MEDIR EL BRILLO DE UNA ESTRELLA. EN ESTO EJERCICIO USAR UNA IMAGEN B Y UNA IMAGEN V. EN TU ANÁLISIS DE ESTAS IMAGENES ENCONTRAR EL MAGNITUD APARENTE M_B Y M_V DE UNA MUESTRA DE ESTRELLA EN EL CÚMULO. LUEGO PODEMOS CALCULAR EL VALOR M_B Y M_V EL ÍNDICE DE COLOR $B - V$. FINALMENTE PODEMOS DETERMINAR LA TEMPERATURA SUPERFICIAL DE LA ESTRELLA VER EL SECCION ASTRONÓMICO.

UN CÚMULO ES UN GRUPO DE ESTRELLAS. LA VIDA DE UN CÚMULO ESTÁ DETERMINADA POR LA VIDA DE LOS DIFERENTES TIPOS DE ESTRELLAS QUE LO COMPUEN. LA OBSERVACIÓN HAY QUE MOSTRAR QUE EN EL CÚMULO GLOBULAR QUEDA MUY POCO CANTIDAD DE GAS Y POLVO DE MANERA QUE SEA MUY RARO EL NACIMIENTO DE NUEVAS ESTRELLAS EN TAL CÚMULO. LAS ESTRELLAS QUE VEMOS EN

ANEXO K: NUEVO TEXTO TRANSFORMADO FESTESE001V

UNO CÚMULO GLOBULAR SER TODO ADULTO Y HABERX EVOLUCIONAR DE FORMA DIFERENTE EN FUNCIÓN DE SU MASA.

EL MAYORÍA DE EL ESTRELLA DE BAJOAMASA SE ENCONTRAR EN EL SECUENCIA PRINCIPAL. ESTO SE DEBERA QUE EL ESTRELLA DE BAJOAMASA IRX CONSUMIR SU ENERGÍA MUY LENTAMENTE.

QUEMAR SU RESERVA DE HIDRÓGENO POCO A POCO Y CONTINUAR DE ESTO MODO DURANTE MIL DE MILLÓN DE AÑO. POR CONSIGUIENTE ESTAR EN EL SECUENCIA PRINCIPAL DURANTE MUCHO TIEMPO.

POR EL CONTRARIO EL ESTRELLA MÁS PESADA DEL CÚMULO YA HABERX TRANSFORMAR EL HIDRÓGENO DE SU NÚCLEO Y SE HABERX CONVERTIR EN GIGANTEROJO. TODO ESTO SUCEDER HACER MUCHO TIEMPO ASÍ QUE HOY EN DÍA NO HABER ESTRELLA CALIENTE MASIVO EL QUE DEJAR UNO HUECO EN EL MITAD SUPERIOR DE EL SECUENCIA PRINCIPAL VER FIG 7. ESTO ESTRELLA SE ENCONTRAR AHORA EN EL ÁREA DIAGONAL QUE COMENZAR EN EL SECUENCIA PRINCIPAL Y SE DESPLAZAR HACIA EL PARTE SUPERIOR DERECHO DEL DIAGRAMA CONOCIDO COMO RAMA DE EL GIGANTEROJO.

EL PUNTO EN EL CUAL SE ENCONTRAR EL SECUENCIA PRINCIPAL Y EL RAMA DE GIGANTEROJO SE LLAMAR PUNTO DE GIRO Y NOS DAR UNO PISTA IMPORTANTE A EL HORA DE DETERMINAR EL EDAD DEL CÚMULO.

EN EL SIGUIENTE EJERCICIO MEDIR EL COORDENADA DE ESTO PUNTO SOBRE TU DIAGRAMA Y DETERMINAR EL EDAD DE M12.

ANEXO L: Texto con Unificación de Grafías: FESTESE001V-S

Una estrella es una bola gigantesca de gas que tiene luz propia con propiedades físicas tales como la masa, la temperatura o el radio. También es de interés para los astrónomos la distancia de la estrella a la Tierra. La estrella más cercana- y por ende, más estudiada- es, por supuesto, nuestro propio Sol.

La luz emitida por la mayoría de las estrellas es un subproducto del proceso de fusión termonuclear que se produce en el núcleo interno de las estrellas. Una estrella normal como el Sol se compone de aproximadamente un 74 por ciento de hidrógeno y un 25 por ciento de helio, siendo el restante 1 por ciento una mezcla de elementos más pesados. El proceso de fusión más común en una estrella como el Sol es la “combustión del hidrógeno”, en la cual 4 núcleos de hidrógeno se fusionan para formar 1 núcleo de helio. El proceso sucede en varias etapas, que se ilustran en la Figura 2. En el primer paso del proceso, 2 protones se fusionan para formar deuterio, una forma pesada del hidrógeno. Este es un suceso muy raro, incluso para el núcleo denso de la estrella, en donde la temperatura es de unos pocos millones de grados. Es por esto por lo que las estrellas como el Sol no explotan en una reacción violenta cuando comienza el proceso de fusión, sino que permanecen en esta fase estable de la vida de la estrella durante varios miles de millones de años. Mientras la estrella es estable, la temperatura de su superficie, su radio y su luminosidad son aproximadamente constantes. Las reacciones nucleares del núcleo

ANEXO L: TEXTO CON UNIFICACIÓN DE GRAFÍAS FESTEESE001V-S

de la estrella generan la energía justa para mantener un equilibrio entre la presión térmica ejercidaa hacia fuera y las fuerzas gravitatorias ejercidasa hacia dentro.

La masa de 1 átomo de helio es sólo el 99coma3 por ciento de la masa de los 4 núcleos originales de hidrógeno. El proceso de fusión convierte el 0coma7 por ciento residual de la masa en energía- mayoritariamente en forma1 de luz. La cantidad de energía se puedex calcular usando la famosa ecuación de Einstein $ENERGÍA = MASA \times c^2$. Ya que c^2 es un número muy grande, esta ecuación implica que incluso una pequeña cantidad de materia puedex convertirse en una formidable cantidad de energía. El 0coma7 por ciento residual de la masa de los 4 núcleos de hidrógeno involucradosa en 1 sola reacción puedex parecer pequeño, pero cuando se considera el número total de reacciones de todo el proceso de fusión, hayx implicadaa una masa total (y por tanto energía) considerable.

El término “cúmulo de estrellas” se usa para 2 tipos diferentes de agrupaciones de estrellas: cúmulos abiertosa de estrellas y cúmulos globulares de estrellas.

Los cúmulos abiertosa de estrellas son colecciones no compactas de estrellas relativamente jóvenes que van desde un centenar hasta unas pocas miles de estrellas. Los cúmulos abiertosa de estrellas tienen una edad de unos pocos cientos de millones de años, una fracción pequeña del tiempo total de la vida de la estrella (unos pocos miles de millones de años). Los cúmulos abiertosa de estrellas se encuentran en el disco de la Vía Láctea y a menudo contienen nubes de gas y polvo donde se forman nuevas estrellas. El diámetro típico de un cúmulo abierta de estrellas es de aproximadamente 30 años luz (10 parsecs).

ANEXO L: TEXTO CON UNIFICACIÓN DE GRAFÍAS FESTESE001V-S

Este es uno de los más famosos cúmulos de estrellas del cielo. Las Pléyades pueden verse a simple vista incluso desde la mayoría de las ciudades pese a la contaminación lumínica. Las Pléyades es uno de los más brillantes y más próximos cúmulos abiertos. El cúmulo de las Pléyades contiene más de 3000 estrellas, está a una distancia de unos 400 años luz y tiene un tamaño de sólo 13 años luz (cortesía de Bruno Stampfer y Rainer Eisendle).

En el halo y en el disco de nuestra Vía Láctea existen unos pocos cientos de cúmulos esféricos y compactos, llamados cúmulos globulares, y que están ligados gravitatoriamente a la Vía Láctea.

Cada cúmulo globular se compone de un grupo esférico de hasta 1 millón de estrellas y tiene un diámetro típico de unos 100 años luz. La mayoría de los cúmulos globulares son muy viejos y lo más probable es que sean anteriores a la formación de la propia galaxia que tuvo lugar hace aproximadamente 12 mil millones de años cuando la mayoría del material protogaláctico se depositó en el disco.

Muchos cúmulos globulares probablemente se han destruido con el paso de los miles de millones de años por las repetidas colisiones e interacciones entre sus estrellas y con la Vía Láctea. Los cúmulos globulares supervivientes son más viejos que cualquier otra estructura de nuestra Vía Láctea.

Esta ilustración da una perspectiva general de la galaxia Vía Láctea. Se indican los diferentes componentes de este complicado sistema de estrellas, gas y polvo. El plano del disco se encuentra a lo largo de la línea central horizontal. Los cúmulos globulares se distribuyen en un halo esférico alrededor del centro de la

ANEXO L: TEXTO CON UNIFICACIÓN DE GRAFÍAS FESTEESE001V-S

galaxia. Se cree que esta distribución tiene que ver con el hecho de que los cúmulos globulares de estrellas se formaron muy pronto en la historia de la galaxia.

El estudio astrofísico de los cúmulos globulares es una parte importante del programa de investigación de la comunidad astronómica internacional. Los cúmulos globulares de estrellas son importantes, no sólo como un valioso banco de pruebas de valor para las teorías de estructura y evolución estelar, sino también porque se encuentran entre los pocos objetos de la galaxia para los cuales se pueden determinar su edad de una forma relativamente precisa. Por ser tan longevos los cúmulos globulares proporcionan un límite inferior muy útil a la edad del Universo. La distribución de sus edades y la correlación entre la edad de un cúmulo y la abundancia en él de los distintos elementos químicos hace de estos sistemas una prueba valiosa en los procesos de formación estelar.

Todas las estrellas que componen un cúmulo globular comparten una historia común y difieren entre sí sólo en sus masas. Por lo tanto, los cúmulos globulares son lugares ideales para estudiar la evolución de las estrellas. En los ejercicios siguientes, determinarás algunas de las propiedades de un cúmulo globular particular, el Messier 12 .

El cúmulo globular Messier 12, también llamado NGC 6218, fue descubierto en 1764 por Charles Messier de forma que es el objeto duodécimo de Messier. Como muchos otros cúmulos globulares, Messier lo describió como una “Nebulosa sin estrellas”, una consecuencia del modesto poder de resolución de su telescopio. William Herschel fue el primero que resolvió los cúmulos en estrellas individuales en 1783 .

ANEXO L: TEXTO CON UNIFICACIÓN DE GRAFÍAS FESTESE001V-S

MESSIER 12 se encuentra en la constelación de Ofiuco y se puede ver con prismáticos desde lugares con muy baja contaminación lumínica. La magnitud visible del conjunto del cúmulo globular es 6,7 (ver más sobre magnitudes en la sección Herramientas Astronómicas, página 2) y la estrella más brillante del cúmulo globular tiene una magnitud visible de 12.

El NGC fue publicado en 1888. En él aparece una lista de cúmulos globulares y cúmulos abiertos de estrellas, nebulosas planetarias y difusas, remanentes de supernova, galaxias de todos los tipos e incluso algunos errores que no correspondían a ningún objeto.

Se llama diagrama H-R a la gráfica que muestra la luminosidad (o magnitud absoluta) frente a la temperatura superficial de las estrellas. La Figura 6 muestra un ejemplo general que se ha construido a partir de observaciones de estrellas en la vecindad de los cúmulos donde se conocen las distancias (a partir de las medidas de HIPPARCOS). La temperatura superficial de una estrella se puede derivar de los valores medidos de su color (magnitud aparente del índice de color B menos magnitud aparente del índice de color V) (ver la sección Herramientas Astronómicas).

Se ve claramente del diagrama H-R que las medidas (luminosidad, temperatura superficial) para diferentes estrellas forman un curioso patrón cuando se colocan sobre el diagrama H-R. Las estrellas se concentran en áreas específicas (indicadas en la figura). El diagrama H-R nos da la clave para comprender cómo las estrellas evolucionan en el tiempo. Las estrellas, dependiendo de sus masas, se moverán a través del diagrama H-R a lo largo de caminos diferentes.

ANEXO L: TEXTO CON UNIFICACIÓN DE GRAFÍAS FESTESE001V-S

Las estrellas pasan la mayor parte de su vida en la Secuencia Principal, quemando el hidrógeno lentamente en un estado de equilibrio estable. Obviamente, esta es la razón por la que la mayoría de las estrellas se localizan en la Secuencia Principal, aproximadamente una línea recta desde el vértice superior izquierdo al vértice inferior derecho en el diagrama H-R. Cuando el suministro de hidrógeno en el núcleo de la estrella se agota, la combustión del hidrógeno ya no es posible. Este es el final de la fase de la secuencia principal de la vida de la estrella y el equilibrio entre la presión del gas y la contracción gravitatoria en el núcleo estelar ya no es estable. La fusión del hidrógeno tiene lugar entonces en la capa circundante mientras el núcleo comienza a contraerse. Al contraerse el núcleo se elevan su presión y su temperatura central, de manera que los núcleos de helio del núcleo de la estrella comienzan a fusionarse y forman elementos más pesados. Este ciclo se puede repetir usando elementos cada vez más pesados a medida que los elementos más ligeros se van acabando en el núcleo. Durante esta fase la estrella aparece como una gigante roja. Tales estrellas se muestran en el diagrama H-R fuera de la línea de la secuencia principal, en la parte superior derecha. El aumento de la temperatura central hace que las capas externas de la estrella se expandan y enfríen de forma que la temperatura superficial decae. El conjunto de la estrella llega a ser muy grande y, debido a la baja temperatura superficial, emite hacia el espacio sobre todo radiación de longitudes de onda mayores, de forma que la estrella parece roja. A pesar de su baja temperatura superficial, todas las gigantes rojas tienen una alta luminosidad debido a su enorme radio. Este resultado se deduce de la Ley de Radiación de Stefan-Boltzmann para la radiación del cuerpo negro:

ANEXO L: TEXTO CON UNIFICACIÓN DE GRAFÍAS FESTESE001V-S

$\text{luminosidad} = \sigma \times 4 \times \pi \times \text{radio}^2 \times \text{temperatura superficial}^4$

donde σ es la constante de Stefan-Boltzmann. Valores típicos para las gigantes rojas son $\text{radio} \sim 10^2 \times \text{radio solar}$, $\text{temperatura superficial} \sim (3 \text{ hasta } 4) \times 10^3 \text{ K}$, de forma que la luminosidad es aproximadamente $10^3 \times \text{luminosidad solar}$.

Cuando ya no pueden mantenerse los procesos avanzados de fusión en el núcleo estelar, el núcleo se colapsa de nuevo. Otra vez la temperatura del núcleo se incrementa y ahora se expulsan las capas externas de la estrella. Las llamadas nebulosas planetarias se forman a partir de los restos de las capas externas (ver Ejercicio 3 de esta serie de ejercicios de astronomía de ESA1/ESO1). El núcleo colapsado está muy caliente (blanco) y la estrella es muy pequeña. A estas estrellas se las denominan muy adecuadamente enanas blancas y es la fase final de la vida de una estrella normal de tipo solar.

Para hacer una estimación aproximada de la relación entre la luminosidad y la temperatura superficial de todas las estrellas de la secuencia principal, veamos el diagrama H-R (Figura 6). La línea más o menos recta de la Secuencia Principal se extiende aproximadamente un orden de magnitud en temperatura superficial: $(3 \times 10^3 \text{ hasta } 3 \times 10^4) \text{ K}$. El rango de luminosidades se extiende aproximadamente seis órdenes de magnitud: $(10^{-2} \text{ hasta } 10^4) \times \text{luminosidad solar}$. Por tanto podemos estimar de forma aproximada: $\text{luminosidad} \propto \text{temperatura superficial}^6$ para las estrellas de la secuencia principal.

Para dar algunos ejemplos:

ANEXO L: TEXTO CON UNIFICACIÓN DE GRAFÍAS FESTESE001V-S

Una estrella de gran masa de la secuencia principal con una temperatura superficial de aproximadamente $T_{\text{superficial estrella}} = 10^4 \text{ K}$ tiene una luminosidad de aproximadamente $L_{\text{estrella}} = (10 \text{ dividido por } 5) \cdot L_{\text{luminosidad solar}}$, o aproximadamente 26 veces la luminosidad del Sol. (La luminosidad del Sol tiene un valor estándar de 1 en la escala de luminosidad).

Una estrella de baja masa con temperatura superficial $T_{\text{estrella}} = 3 \times 10^3 \text{ K}$ tiene una luminosidad de aproximadamente sólo el 5 por ciento de la luminosidad del Sol.

Toda la información que podemos extraer de las estrellas proviene de la radiación que recibimos de ellas. Como se explica en la sección Herramientas Astronómicas, diferentes filtros y sistemas de color se pueden usar para medir el brillo de una estrella. En este ejercicio usamos una imagen B y una imagen V. En tu análisis de estas imágenes encontrarás la magnitud aparente del índice de color B y magnitud aparente del índice de color V de una muestra de estrellas en el cúmulo. Luego podrás calcular los valores magnitud aparente del índice de color B y magnitud aparente del índice de color V (el índice de color B menos V). Finalmente podrás determinar la temperatura superficial de las estrellas (ver la sección Herramientas Astronómicas).

Un cúmulo es un grupo de estrellas. La vida de un cúmulo está determinada por la vida de los diferentes tipos de estrellas que lo componen. Las observaciones han mostrado que en los cúmulos globulares queda muy poca cantidad de gas y polvo, de manera que es muy raro el nacimiento de nuevas estrellas en tales cúmulos. Las estrellas que vemos en un cúmulo globular son todas “adultas” y han evolucionado de formas diferentes en función de sus masas.

ANEXO L: TEXTO CON UNIFICACIÓN DE GRAFÍAS FESTESE001V-S

La mayoría de las estrellas de baja masa se encuentran en la Secuencia Principal. Esto se debe a que las estrellas de baja masa van consumiendo su energía muy lentamente. Las estrellas queman sus reservas de hidrógeno poco a poco y continuarán de este modo durante miles de millones de años. Por consiguiente, las estrellas estarán en la Secuencia Principal durante mucho tiempo.

Por el contrario, las estrellas más pesadas del cúmulo globular ya han transformado el hidrógeno de sus núcleos y se han convertido en gigantes rojas. Todo esto sucedió hace mucho tiempo, así que hoy en día no hay estrellas calientes masivas, lo que deja un hueco en la mitad superior de la Secuencia Principal (ver Figura 7). Estas estrellas se encuentran ahora en el área diagonal que comienza en la Secuencia Principal y se desplaza hacia la parte superior derecha del diagrama H-R conocida como rama de las gigantes rojas.

El punto en el cual se encuentran la Secuencia Principal y la Rama de Gigante Roja se llama punto de giro, y nos da una pista importante a la hora de determinar la edad del cúmulo globular. En el siguiente ejercicio, medirás las coordenadas de este punto sobre tu diagrama H-R y determinarás la edad de MESSIER 12.

ANEXO M: Texto Transformado FESTESE001V-F

Este es el texto transformado del caso de estudio en el que se han clasificado en la base *Eliminar* las palabras funcionales y los verbos auxiliares. En el texto transformado, generado por el programa PAFE, una estrella denota una marca de punto y seguido y dos estrellas de punto y aparte.

ESTRELLA SER BOLA GIGANDESCO GAS TENER LUZ PROPIO
PROPIEDAD FÍSICO TAL MASA TEMPERATURA RADIO * TAMBIÉN SER
INTERÉS ASTRÓNOMO DISTANCIA ESTRELLA TIERRA * ESTRELLA
MÁS CERCANO ENDE MÁS ESTUDIADOA SER SUPUESTO1 PROPIO SOL
** LUZ EMITIDOA MAYORÍA ESTRELLA SER SUBPRODUCTO PROCESO
FUSIÓN TERMONUCLEAR PRODUCIR NÚCLEO INTERNO ESTRELLA *
ESTRELLA NORMAL SOL COMPONER APROXIMADAMENTE 74
PORCIENTO HIDRÓGENO 25 PORCIENTO HELIO SER RESTANTE 1
PORCIENTO MEZCLA1 ELEMENTO MÁS PESADOA * PROCESO FUSIÓN
MÁS COMÚN ESTRELLA SOL SER COMBUSTIÓN HIDRÓGENO CUATRO
NÚCLEO HIDRÓGENO FUSIONAR FORMAR UNOA NÚCLEO HELIO *
PROCESO SUCEDER VARIO ETAPA ILUSTRAR FIG 2 * PRIMERO PASO1
PROCESO DOS PROTÓN FUSIONAR FORMAR DEUTERIO FORMA1
PESADOA HIDRÓGENO * SER SUCESO MUY RARO INCLUSO NÚCLEO

ANEXO M: TEXTO TRANSFORMADO FESTESE001V-F

DENSO ESTRELLA TEMPERATURA SER POCO MILLÓN GRADO * SER
ESTRELLA SOL NO EXPLOTAR REACCIÓN VIOLENTO COMENZAR
PROCESO FUSIÓN PERMANECER FASE ESTABLE VIDA ESTRELLA
VARIO MIL MILLÓN AÑO * MIENTRAS ESTRELLA SER ESTABLE
TEMPERATURA SUPERFICIE RADIO LUMINOSIDAD SER
APROXIMADAMENTE CONSTANTEA * REACCIÓN NUCLEAR NÚCLEO
ESTRELLA GENERAR ENERGÍA JUSTO MANTENER EQUILIBRIO
PRESIÓN TÉRMICO EJERCIDOA FUERA FUERZA GRAVITATORIO
EJERCIDOA DENTRO ** MASA UNOA ÁTOMO HELIO SER SÓLO
99COMA3 PORCIENTO MASA CUATRO NÚCLEO ORIGINAL HIDRÓGENO
* PROCESO FUSIÓN CONVERTIR 0COMA7 PORCIENTO RESIDUAL MASA
ENERGÍA MAYORITARIAMENTE FORMA1 LUZ * CANTIDAD ENERGÍA
CALCULAR USAR FAMOSO ECUACIÓN EINSTEIN $E = M \times C^2$ * YA C
A2 SER NÚMERO MUY GRANDE ECUACIÓN IMPLICAR INCLUSO
PEQUEÑO CANTIDAD MATERIA CONVERTIR FORMIDABLE CANTIDAD
ENERGÍA * 0COMA7 PORCIENTO RESIDUAL MASA CUATRO NÚCLEO
HIDRÓGENO INVOLUCRADOA UNOA SOLO REACCIÓN PARECER
PEQUEÑO CONSIDERAR NÚMERO TOTAL REACCIÓN TODO PROCESO
FUSIÓN IMPLICADOA MASA TOTAL TANTO ENERGÍA CONSIDERABLE
** TÉRMINO CÚMULO ESTRELLA USAR DOS TIPO DIFERENTE
AGRUPACIÓN ESTRELLA CÚMULO ABIERTO A ESTRELLA CÚMULO
GLOBULAR ESTRELLA ** CÚMULO ABIERTO A ESTRELLA SER
COLECCIÓN NO COMPACTO ESTRELLA RELATIVAMENTE JOVEN IR
CENTENAR POCO MIL ESTRELLA * TENER EDAD POCO CIENTO
MILLÓN AÑO FRACCIÓN PEQUEÑO TIEMPO TOTAL VIDA ESTRELLA

ANEXO M: TEXTO TRANSFORMADO FESTESE001V-F

POCO MIL MILLÓN AÑO * CÚMULO ENCONTRAR DISCO GALAXIA VÍA
LÁCTEO MENUDO CONTENER NUBE GAS POLVO FORMAR NUEVO
ESTRELLA * DIÁMETRO TÍPICO CÚMULO ABIERTO A ESTRELLA SER
APROXIMADAMENTE 30 AÑO LUZ 10 PARSEC ** SER MÁS FAMOSO
CÚMULO ESTRELLA CIELO * PLÉYADES VER SIMPLE VISTA INCLUSO
MAYORÍA CIUDAD PESAR CONTAMINACIÓN LUMÍNICO * SER MÁS
BRILLANTE MÁS PRÓXIMO CÚMULO ABIERTO A * CÚMULO PLÉYADES
CONTENER MÁS 3000 ESTRELLA ESTAR DISTANCIA 400 AÑO LUZ
TENER TAMAÑO SÓLO 13 AÑO LUZ CORTESÍA BRUNO STAMPFER
RAINER EISENDLE ** HALO DISCO VÍA LÁCTEO EXISTIR POCO CIENTO
CÚMULO ESFÉRICO COMPACTO LLAMADO A CÚMULO GLOBULAR
LIGADO A GRAVITATORIAMENTE GALAXIA ** CADA CÚMULO
GLOBULAR COMPONER GRUPO ESFÉRICO UNO A MILLÓN ESTRELLA
TENER DIÁMETRO TÍPICO 100 AÑO LUZ * MAYORÍA CÚMULO
GLOBULAR SER MUY VIEJO MÁS PROBABLE SER ANTERIOR
FORMACIÓN PROPIO GALAXIA TENER LUGAR HACER
APROXIMADAMENTE 12 MIL MILLÓN AÑO MAYORÍA MATERIAL
PROTOGALÁCTICO DEPOSITAR DISCO ** MUCHO CÚMULO GLOBULAR
PROBABLEMENTE DESTRUIR PASO 1 MIL MILLÓN AÑO REPETIDO A
COLISIÓN INTERACCIÓN ESTRELLA VÍA LÁCTEO * CÚMULO
GLOBULAR SUPERVIVIENTE SER MÁS VIEJO CUALQUIER OTRO
ESTRUCTURA VÍA LÁCTEO ** ILUSTRACIÓN DAR PERSPECTIVA
GENERAL GALAXIA VÍA LÁCTEO * INDICAR DIFERENTE COMPONENTE
COMPLICADO A SISTEMA ESTRELLA GAS POLVO * PLANO DISCO
ENCONTRAR LARGO LÍNEA CENTRAL HORIZONTAL * CÚMULO

ANEXO M: TEXTO TRANSFORMADO FESTESE001V-F

GLOBULAR DISTRIBUIR HALO ESFÉRICO ALREDEDOR CENTRO
GALAXIA * CREER DISTRIBUCIÓN TENER VER HECHO1 CÚMULO
ESTRELLA FORMAR MUY PRONTO HISTORIA GALAXIA ** ESTUDIO1
ASTROFÍSICO CÚMULO GLOBULAR SER PARTE IMPORTANTE
PROGRAMA INVESTIGACIÓN COMUNIDAD ASTRONÓMICO
INTERNACIONAL * CÚMULO ESTRELLA SER IMPORTANTE NO SÓLO
VALIOSO BANCO PRUEBA1 VALOR TEORÍA ESTRUCTURA EVOLUCIÓN
ESTELAR TAMBIÉN ENCONTRAR POCO OBJETO GALAXIA
DETERMINAR EDAD FORMA1 RELATIVAMENTE PRECISOA * SER
TANTO LONGEVO PROPORCIONAR LÍMITE INFERIOR MUY ÚTIL EDAD
UNIVERSO * DISTRIBUCIÓN EDAD CORRELACIÓN EDAD CÚMULO
ABUNDANCIA DISTINTO ELEMENTO QUÍMICO HACER SISTEMA
PRUEBA1 VALIOSO PROCESO FORMACIÓN ESTELAR ** TODO
ESTRELLA COMPONER CÚMULO GLOBULAR COMPARTIR HISTORIA
COMÚN DIFERIR SÓLO MASA * TANTO CÚMULO GLOBULAR SER
LUGAR IDEAL ESTUDIAR EVOLUCIÓN ESTRELLA * EJERCICIO
SIGUIENTE DETERMINAR ALGUNO PROPIEDAD CÚMULO GLOBULAR
PARTICULAR MESSIER 12 ** CÚMULO GLOBULAR MESSIER 12 M12
TAMBIÉN LLAMADOA NGC 6218 DESCUBRIR 1764 CHARLES MESSIER
FORMA1 SER OBJETO DUODÉCIMO MESSIER * MUCHO OTRO CÚMULO
GLOBULAR MESSIER DESCRIBIR NEBULOSA ESTRELLA
CONSECUENCIA MODESTO PODER1 RESOLUCIÓN TELESCOPIO *
WILLIAM HERSCHEL SER PRIMERO RESOLVER CÚMULO ESTRELLA
INDIVIDUAL 1783 ** M12 ENCONTRAR CONSTELACIÓN OFIUCO VER
PRISMÁTICO LUGAR MUY BAJOA POLUCIÓN LUMÍNICO * MAGNITUD

ANEXO M: TEXTO TRANSFORMADO FESTESE001V-F

VISIBLE CONJUNTO CÚMULO GLOBULAR SER 6COMA7 VER MÁS
MAGNITUD SECCIÓN HERRAMIENTA ASTRONÓMICO PÁGINA 2
ESTRELLA MÁS BRILLANTE CÚMULO TENER MAGNITUD VISIBLE 12 **
NGC NUEVO CATÁLOGO GENERAL PUBLICAR 1888 * APARECER LISTA
CÚMULO GLOBULAR CÚMULO ABIERTO A ESTRELLA NEBULOSA
PLANETARIO DIFUSO REMANENTE SUPERNOVA GALAXIA TODO TIPO
INCLUSO ALGUNO ERROR NO CORRESPONDER NINGUNO OBJETO **
LLAMAR DIAGRAMA HERTZSPRUNGRUSSELL FORMA1 MÁS CONCISO
DIAGRAMA HR GRÁFICO MOSTRAR LUMINOSIDAD L MAGNITUD
ABSOLUTO M FRENTE TEMPERATURA SUPERFICIAL T ESTRELLA * FIG
6 MOSTRAR EJEMPLO GENERAL CONSTRUIR PARTIR OBSERVACIÓN
ESTRELLA VECINDAD CÚMULO CONOCER DISTANCIA PARTIR
MEDIDA1 HIPPARCOS * TEMPERATURA SUPERFICIAL T ESTRELLA
DERIVAR VALOR MEDIDO A COLOR MB MENOS MV VER SECCIÓN
HERRAMIENTA ASTRONÓMICO ** VER CLARAMENTE DIAGRAMA HR
MEDIDA1 L T DIFERENTE ESTRELLA FORMAR CURIOSO PATRÓN
COLOCAR DIAGRAMA * ESTRELLA CONCENTRAR ÁREA ESPECÍFICO
INDICADO A FIGURA * DIAGRAMA HR DAR CLAVE COMPRENDER
ESTRELLA EVOLUCIONAR TIEMPO * ESTRELLA DEPENDER MASA
MOVER TRAVÉS DIAGRAMA LARGO CAMINO DIFERENTE ** ESTRELLA
PASAR MAYOR PARTE VIDA SECUENCIA PRINCIPAL QUEMAR
HIDRÓGENO LENTAMENTE ESTADO EQUILIBRIO ESTABLE *
OBVIAMENTE SER RAZÓN MAYORÍA ESTRELLA LOCALIZAR
SECUENCIA PRINCIPAL APROXIMADAMENTE LÍNEA RECTO VÉRTICE
SUPERIOR IZQUIERDO VÉRTICE INFERIOR DERECHO DIAGRAMA **

ANEXO M: TEXTO TRANSFORMADO FESTESE001V-F

SUMINISTRO1 HIDRÓGENO NÚCLEO ESTRELLA AGOTAR COMBUSTIÓN
HIDRÓGENO YA NO SER POSIBLE * SER FINAL FASE SECUENCIA
PRINCIPAL VIDA ESTRELLA EQUILIBRIO PRESIÓN GAS CONTRACCIÓN
GRAVITATORIO NÚCLEO ESTELAR YA NO SER ESTABLE * FUSIÓN
HIDRÓGENO TENER LUGAR ENTONCES CAPA CIRCUNDANTE
MIENTRAS NÚCLEO CONTRAER * CONTRAER NÚCLEO ELEVAR
PRESIÓN TEMPERATURA CENTRAL MANERA NÚCLEO HELIO NÚCLEO
ESTRELLA FUSIONAR FORMAR ELEMENTO MÁS PESADOA * CICLO
REPETIR USAR ELEMENTO CADA VEZ MÁS PESADOA MEDIDOA
ELEMENTO MÁS LIGERO ACABAR NÚCLEO * FASE ESTRELLA
APARECER GIGANTE ROJO * TAL ESTRELLA MOSTRAR DIAGRAMA HR
FUERA LÍNEA SECUENCIA PRINCIPAL PARTE SUPERIOR DERECHO *
AUMENTO1 TEMPERATURA CENTRAL HACER CAPA EXTERNO
ESTRELLA EXPANDIR ENFRIAR FORMA1 TEMPERATURA SUPERFICIAL
DECAER * CONJUNTO ESTRELLA SER MUY GRANDE DEBER INFERIOR
TEMPERATURA SUPERFICIAL EMITIR ESPACIO TODO RADIACIÓN
LONGITUD ONDA MAYOR FORMA1 ESTRELLA PARECER ROJO * PESAR
BAJOA TEMPERATURA SUPERFICIAL T TODO GIGANTE ROJO TENER
ALTO LUMINOSIDAD L DEBER ENORME RADIO R * RESULTADO1
DEDUCIR LEY RADIACIÓN STEFANBOLTZMANN RADIACIÓN CUERPO
NEGRO $L = \sigma \times 4 \times \pi \times R^2 \times T^4$ SIGMA SER CONSTANTE1
STEFANBOLTZMANN * VALOR TÍPICO GIGANTE ROJO SER $R \sim 10^2 \times R_{\text{SOLAR}}$
 $T \sim 3 \text{ HASTA } 4 \times 10^4 \text{ K}$ FORMA1 L SER
APROXIMADAMENTE $10^3 \times L_{\text{SOLAR}}$ ** YA NO MANTENER PROCESO
AVANZADOA FUSIÓN NÚCLEO ESTELAR NÚCLEO COLAPSAR NUEVO *

ANEXO M: TEXTO TRANSFORMADO FESTESE001V-F

OTRO VEZ TEMPERATURA NÚCLEO INCREMENTAR AHORA EXPULSAR
CAPA EXTERNO ESTRELLA * LLAMADOA NEBULOSA PLANETARIO
FORMAR PARTIR RESTO CAPA EXTERNO VER EJERCICIO 3 SERIE
EJERCICIO ASTRONOMÍA ESA1 ESO1 * NÚCLEO COLAPSADOA ESTAR
MUY CALIENTE BLANCO ESTRELLA SER MUY PEQUEÑO * ESTRELLA
DENOMINAR MUY ADECUADAMENTE ENANO BLANCO SER FASE
FINALA VIDA ESTRELLA NORMAL TIPO SOLAR ** HACER ESTIMACIÓN
APROXIMADOA RELACIÓN LUMINOSIDAD L TEMPERATURA
SUPERFICIAL T TODO ESTRELLA SECUENCIA PRINCIPAL VER
DIAGRAMA HR FIG 6 * LÍNEA MÁS MENOS RECTO SECUENCIA
PRINCIPAL EXTENDER APROXIMADAMENTE ORDEN MAGNITUD
TEMPERATURA 3×10^3 A 3×10^4 K * RANGO LUMINOSIDAD
EXTENDER APROXIMADAMENTE SEIS ORDEN MAGNITUD 10 AMENOS2
HASTAM 10^4 X L SOLAR * TANTO ESTIMAR FORMA1 APROXIMADOA
L PROPORCIONAL T A6 ESTRELLA SECUENCIA PRINCIPAL ** DAR
ALGUNO EJEMPLO ESTRELLA GRANDE MASA SECUENCIA PRINCIPAL
TEMPERATURA SUPERFICIAL APROXIMADAMENTE T ESTRELLA =
 10^4 X 10^4 K TENER LUMINOSIDAD APROXIMADAMENTE L
ESTRELLA = 10 DIVIDIR 5×10^8 A6 · L SOLAR APROXIMADAMENTE 26
VEZ LUMINOSIDAD SOL * LUMINOSIDAD SOL TENER VALOR
ESTÁNDARA 1 ESCALA1 LUMINOSIDAD ** ESTRELLA BAJOA MASA T
ESTRELLA = 3×10^5 X 10^3 K TENER LUMINOSIDAD
APROXIMADAMENTE SÓLO 5 PORCIENTO LUMINOSIDAD SOL ** TODO
INFORMACIÓN EXTRAER ESTRELLA PROVENIR RADIACIÓN RECIBIR *
EXPLICAR SECCIÓN HERRAMIENTA ASTRONÓMICO DIFERENTE

ANEXO M: TEXTO TRANSFORMADO FESTESE001V-F

FILTRO SISTEMA COLOR USAR MEDIR BRILLO ESTRELLA * EJERCICIO
USAR IMAGEN B IMAGEN V * ANÁLISIS IMAGEN ENCONTRAR
MAGNITUD APARENTE MB MV MUESTRA1 ESTRELLA CÚMULO *
CALCULAR VALOR MB MV ÍNDICE COLOR B MENOS V * FINALMENTE
DETERMINAR TEMPERATURA SUPERFICIAL ESTRELLA VER SECCIÓN
HERRAMIENTA ASTRONÓMICO ** CÚMULO SER GRUPO ESTRELLA *
VIDA CÚMULO DETERMINADO A VIDA DIFERENTE TIPO ESTRELLA
COMPONER * OBSERVACIÓN MOSTRAR CÚMULO GLOBULAR QUEDAR
MUY POCO CANTIDAD GAS POLVO MANERA SER MUY RARO
NACIMIENTO NUEVO ESTRELLA TAL CÚMULO * ESTRELLA VER
CÚMULO GLOBULAR SER TODO ADULTO EVOLUCIONAR FORMA1
DIFERENTE FUNCIÓN MASA ** MAYORÍA ESTRELLA BAJOA MASA
ENCONTRAR SECUENCIA PRINCIPAL * DEBER ESTRELLA BAJOA MASA
CONSUMIR ENERGÍA MUY LENTAMENTE ** QUEMAR RESERVA1
HIDRÓGENO POCO POCO CONTINUAR MODO MIL MILLÓN AÑO *
CONSIGUIENTE ESTAR SECUENCIA PRINCIPAL MUCHO TIEMPO **
CONTRARIO ESTRELLA MÁS PESADO A CÚMULO YA TRANSFORMAR
HIDRÓGENO NÚCLEO CONVERTIR GIGANTE ROJO * TODO SUCEDER
HACER MUCHO TIEMPO ASÍ HOY DÍA NO HABER ESTRELLA CALIENTE
MASIVO DEJAR HUECO MITAD SUPERIOR SECUENCIA PRINCIPAL VER
FIG 7 * ESTRELLA ENCONTRAR AHORA ÁREA DIAGONAL COMENZAR
SECUENCIA PRINCIPAL DESPLAZAR PARTE SUPERIOR DERECHO
DIAGRAMA CONOCIDO A RAMA GIGANTE ROJO ** PUNTO ENCONTRAR
SECUENCIA PRINCIPAL RAMA GIGANTE ROJO LLAMAR PUNTO GIRO1
DAR PISTA IMPORTANTE HORA DETERMINAR EDAD CÚMULO **

ANEXO M: TEXTO TRANSFORMADO FESTESE001V-F

SIGUIENTE EJERCICIO MEDIR COORDENADA PUNTO DIAGRAMA

DETERMINAR EDAD M12 **

ANEXO N: Texto Transformado FESTESE001V-S

ESTRELLA SER BOLA GIGANDESCO GAS TENER LUZ PROPIO
PROPIEDAD FÍSICO TAL MASA TEMPERATURA RADIO * TAMBIÉN SER
INTERÉS ASTRÓNOMO DISTANCIA ESTRELLA TIERRA * ESTRELLA
MÁS CERCANO ENDE MÁS ESTUDIADOA SER SUPUESTO1 PROPIO SOL
** LUZ EMITIDOA MAYORÍA ESTRELLA SER SUBPRODUCTO PROCESO
FUSIÓN TERMONUCLEAR PRODUCIR NÚCLEO INTERNO ESTRELLA *
ESTRELLA NORMAL SOL COMPONER APROXIMADAMENTE 74
PORCIENTO HIDRÓGENO 25 PORCIENTO HELIO SER RESTANTE 1
PORCIENTO MEZCLA1 ELEMENTO MÁS PESADOA * PROCESO FUSIÓN
MÁS COMÚN ESTRELLA SOL SER COMBUSTIÓN HIDRÓGENO 4 NÚCLEO
HIDRÓGENO FUSIONAR FORMAR 1 NÚCLEO HELIO * PROCESO
SUCEDER VARIO ETAPA ILUSTRAR FIGURA 2 * PRIMERO PASO1
PROCESO 2 PROTÓN FUSIONAR FORMAR DEUTERIO FORMA1 PESADOA
HIDRÓGENO * SER SUCESO MUY RARO INCLUSO NÚCLEO DENSO
ESTRELLA TEMPERATURA SER POCO MILLÓN GRADO * SER ESTRELLA
SOL NO EXPLOTAR REACCIÓN VIOLENTO COMENZAR PROCESO
FUSIÓN PERMANECER FASE ESTABLE VIDA ESTRELLA VARIO MIL
MILLÓN AÑO * MIENTRAS ESTRELLA SER ESTABLE TEMPERATURA
SUPERFICIE RADIO LUMINOSIDAD SER APROXIMADAMENTE

ANEXO N: TEXTO TRANSFORMADO FESTESE001V-S

CONSTANTEA * REACCIÓN NUCLEAR NÚCLEO ESTRELLA GENERAR
ENERGÍA JUSTO MANTENER EQUILIBRIO PRESIÓN TÉRMICO
EJERCIDOA FUERA FUERZA GRAVITATORIO EJERCIDOA DENTRO **
MASA 1 ÁTOMO HELIO SER SÓLO 99COMA3 PORCIENTO MASA 4
NÚCLEO ORIGINAL HIDRÓGENO * PROCESO FUSIÓN CONVERTIR
0COMA7 PORCIENTO RESIDUAL MASA ENERGÍA MAYORITARIAMENTE
FORMA1 LUZ * CANTIDAD ENERGÍA CALCULAR USAR FAMOSO
ECUACIÓN EINSTEIN ENERGÍA = MASA X C A2 * YA C A2 SER NÚMERO
MUY GRANDE ECUACIÓN IMPLICAR INCLUSO PEQUEÑO CANTIDAD
MATERIA CONVERTIR FORMIDABLE CANTIDAD ENERGÍA * 0COMA7
PORCIENTO RESIDUAL MASA 4 NÚCLEO HIDRÓGENO INVOLUCRADOA
1 SOLO REACCIÓN PARECER PEQUEÑO CONSIDERAR NÚMERO TOTAL
REACCIÓN TODO PROCESO FUSIÓN IMPLICADOA MASA TOTAL TANTO
ENERGÍA CONSIDERABLE ** TÉRMINO CÚMULO ESTRELLA USAR 2
TIPO DIFERENTE AGRUPACIÓN ESTRELLA CÚMULO ABIERTOA
ESTRELLA CÚMULO GLOBULAR ESTRELLA ** CÚMULO ABIERTOA
ESTRELLA SER COLECCIÓN NO COMPACTO ESTRELLA
RELATIVAMENTE JOVEN IR CENTENAR POCO MIL ESTRELLA *
CÚMULO ABIERTOA ESTRELLA TENER EDAD POCO CIENTO MILLÓN
AÑO FRACCIÓN PEQUEÑO TIEMPO TOTAL VIDA ESTRELLA POCO MIL
MILLÓN AÑO * CÚMULO ABIERTOA ESTRELLA ENCONTRAR DISCO
VÍA LÁCTEO MENUDO CONTENER NUBE GAS POLVO FORMAR NUEVO
ESTRELLA * DIÁMETRO TÍPICO CÚMULO ABIERTOA ESTRELLA SER
APROXIMADAMENTE 30 AÑO LUZ 10 PARSEC ** SER MÁS FAMOSO
CÚMULO ESTRELLA CIELO * PLÉYADES VER SIMPLE VISTA INCLUSO

ANEXO N: TEXTO TRANSFORMADO FESTESE001V-S

MAYORÍA CIUDAD PESAR CONTAMINACIÓN LUMÍNICO ** PLÉYADES
SER MÁS BRILLANTE MÁS PRÓXIMO CÚMULO ABIERTO * CÚMULO
PLÉYADES CONTENER MÁS 3000 ESTRELLA ESTAR DISTANCIA 400
AÑO LUZ TENER TAMAÑO SÓLO 13 AÑO LUZ CORTESÍA BRUNO
STAMPFER RAINER EISENDLE ** HALO DISCO VÍA LÁCTEO EXISTIR
POCO CIENTO CÚMULO ESFÉRICO COMPACTO LLAMADOA CÚMULO
GLOBULAR LIGADOA GRAVITATORIAMENTE VÍA LÁCTEO ** CADA
CÚMULO GLOBULAR COMPONER GRUPO ESFÉRICO 1 MILLÓN
ESTRELLA TENER DIÁMETRO TÍPICO 100 AÑO LUZ * MAYORÍA
CÚMULO GLOBULAR SER MUY VIEJO MÁS PROBABLE SER SER
ANTERIOR FORMACIÓN PROPIO GALAXIA TENER LUGAR HACER
APROXIMADAMENTE 12 MIL MILLÓN AÑO MAYORÍA MATERIAL
PROTOGALÁCTICO DEPOSITAR DISCO ** MUCHO CÚMULO GLOBULAR
PROBABLEMENTE DESTRUIR PASO1 MIL MILLÓN AÑO REPETIDOA
COLISIÓN INTERACCIÓN ESTRELLA VÍA LÁCTEO * CÚMULO
GLOBULAR SUPERVIVIENTE SER MÁS VIEJO CUALQUIER OTRO
ESTRUCTURA VÍA LÁCTEO ** ILUSTRACIÓN DAR PERSPECTIVA
GENERAL GALAXIA VÍA LÁCTEO * INDICAR DIFERENTE COMPONENTE
COMPLICADOA SISTEMA ESTRELLA GAS POLVO * PLANO DISCO
ENCONTRAR LARGO LÍNEA CENTRAL HORIZONTAL * CÚMULO
GLOBULAR DISTRIBUIR HALO ESFÉRICO ALREDEDOR CENTRO
GALAXIA * CREER DISTRIBUCIÓN TENER VER HECHO1 CÚMULO
GLOBULAR ESTRELLA FORMAR MUY PRONTO HISTORIA GALAXIA **
ESTUDIO1 ASTROFÍSICO CÚMULO GLOBULAR SER PARTE
IMPORTANTE PROGRAMA INVESTIGACIÓN COMUNIDAD

ANEXO N: TEXTO TRANSFORMADO FESTESE001V-S

ASTRONÓMICO INTERNACIONAL * CÚMULO GLOBULAR ESTRELLA
SER IMPORTANTE NO SÓLO VALIOSO BANCO PRUEBA1 VALOR
TEORÍA ESTRUCTURA EVOLUCIÓN ESTELAR TAMBIÉN ENCONTRAR
POCO OBJETO GALAXIA DETERMINAR EDAD FORMA1
RELATIVAMENTE PRECISOA * SER TANTO LONGEVO CÚMULO
GLOBULAR PROPORCIONAR LÍMITE INFERIOR MUY ÚTIL EDAD
UNIVERSO * DISTRIBUCIÓN EDAD CORRELACIÓN EDAD CÚMULO
ABUNDANCIA DISTINTO ELEMENTO QUÍMICO HACER SISTEMA
PRUEBA1 VALIOSO PROCESO FORMACIÓN ESTELAR ** TODO
ESTRELLA COMPONER CÚMULO GLOBULAR COMPARTIR HISTORIA
COMÚN DIFERIR SÓLO MASA * TANTO CÚMULO GLOBULAR SER
LUGAR IDEAL ESTUDIAR EVOLUCIÓN ESTRELLA * EJERCICIO
SIGUIENTE DETERMINAR ALGUNO PROPIEDAD CÚMULO GLOBULAR
PARTICULAR MESSIER 12 ** CÚMULO GLOBULAR MESSIER 12
TAMBIÉN LLAMADOA NGC 6218 DESCUBRIR 1764 CHARLES MESSIER
FORMA1 SER OBJETO DUODÉCIMO MESSIER * MUCHO OTRO CÚMULO
GLOBULAR MESSIER DESCRIBIR NEBULOSA ESTRELLA
CONSECUENCIA MODESTO PODER1 RESOLUCIÓN TELESCOPIO *
WILLIAM HERSCHEL SER PRIMERO RESOLVER CÚMULO ESTRELLA
INDIVIDUAL 1783 ** MESSIER 12 ENCONTRAR CONSTELACIÓN OFIUCO
VER PRISMÁTICO LUGAR MUY BAJOA POLUCIÓN LUMÍNICO *
MAGNITUD VISIBLE CONJUNTO CÚMULO GLOBULAR SER 6COMA7
VER MÁS MAGNITUD SECCIÓN HERRAMIENTA ASTRONÓMICO
PÁGINA 2 ESTRELLA MÁS BRILLANTE CÚMULO GLOBULAR TENER
MAGNITUD VISIBLE 12 ** NGC PUBLICAR 1888 * APARECER LISTA

ANEXO N: TEXTO TRANSFORMADO FESTESE001V-S

CÚMULO GLOBULAR CÚMULO ABIERTO A ESTRELLA NEBULOSA
PLANETARIO DIFUSO REMANENTE SUPERNOVA GALAXIA TODO TIPO
INCLUSO ALGUNO ERROR NO CORRESPONDER NINGUNO OBJETO **
LLAMAR DIAGRAMA HR GRÁFICO MOSTRAR LUMINOSIDAD
MAGNITUD ABSOLUTO FRENTE TEMPERATURA SUPERFICIAL
ESTRELLA * FIGURA 6 MOSTRAR EJEMPLO GENERAL CONSTRUIR
PARTIR OBSERVACIÓN ESTRELLA VICINDAD CÚMULO CONOCER
DISTANCIA PARTIR MEDIDA HIPPARCOS * TEMPERATURA
SUPERFICIAL ESTRELLA DERIVAR VALOR MEDIDO A COLOR
MAGNITUD APARENTE ÍNDICE COLOR B MENOS MAGNITUD
APARENTE ÍNDICE COLOR V VER SECCIÓN HERRAMIENTA
ASTRONÓMICO ** VER CLARAMENTE DIAGRAMA HR MEDIDA
LUMINOSIDAD TEMPERATURA SUPERFICIAL DIFERENTE ESTRELLA
FORMAR CURIOSO PATRÓN COLOCAR DIAGRAMA HR ** ESTRELLA
CONCENTRAR ÁREA ESPECÍFICO INDICADO A FIGURA * DIAGRAMA HR
DAR CLAVE COMPRENDER ESTRELLA EVOLUCIONAR TIEMPO *
ESTRELLA DEPENDER MASA MOVER TRAVÉS DIAGRAMA HR LARGO
CAMINO DIFERENTE ** ESTRELLA PASAR MAYOR PARTE VIDA
SECUENCIA PRINCIPAL QUEMAR HIDRÓGENO LENTAMENTE ESTADO
EQUILIBRIO ESTABLE * OBVIAMENTE SER RAZÓN MAYORÍA
ESTRELLA LOCALIZAR SECUENCIA PRINCIPAL APROXIMADAMENTE
LÍNEA RECTO VÉRTICE SUPERIOR IZQUIERDO VÉRTICE INFERIOR
DERECHO DIAGRAMA HR ** SUMINISTRO HIDRÓGENO NÚCLEO
ESTRELLA AGOTAR COMBUSTIÓN HIDRÓGENO YA NO SER POSIBLE *
SER FINAL FASE SECUENCIA PRINCIPAL VIDA ESTRELLA EQUILIBRIO

ANEXO N: TEXTO TRANSFORMADO FESTESE001V-S

PRESIÓN GAS CONTRACCIÓN GRAVITATORIO NÚCLEO ESTELAR YA
NO SER ESTABLE * FUSIÓN HIDRÓGENO TENER LUGAR ENTONCES
CAPA CIRCUNDANTE MIENTRAS NÚCLEO CONTRAER * CONTRAER
NÚCLEO ELEVAR PRESIÓN TEMPERATURA CENTRAL MANERA
NÚCLEO HELIO NÚCLEO ESTRELLA FUSIONAR FORMAR ELEMENTO
MÁS PESADOA * CICLO REPETIR USAR ELEMENTO CADA VEZ MÁS
PESADOA MEDIDO A ELEMENTO MÁS LIGERO ACABAR NÚCLEO * FASE
ESTRELLA APARECER GIGANTE ROJO * TAL ESTRELLA MOSTRAR
DIAGRAMA HR FUERA LÍNEA SECUENCIA PRINCIPAL PARTE SUPERIOR
DERECHO * AUMENTO TEMPERATURA CENTRAL HACER CAPA
EXTERNO ESTRELLA EXPANDIR ENFRIAR FORMA TEMPERATURA
SUPERFICIAL DECAER * CONJUNTO ESTRELLA SER MUY GRANDE
DEBER INFERIOR TEMPERATURA SUPERFICIAL EMITIR ESPACIO TODO
RADIACIÓN LONGITUD ONDA MAYOR FORMA ESTRELLA PARECER
ROJO * PESAR BAJA TEMPERATURA SUPERFICIAL TODO GIGANTE
ROJO TENER ALTO LUMINOSIDAD DEBER ENORME RADIO *
RESULTADO DEDUCIR LEY RADIACIÓN STEFANBOLTZMANN
RADIACIÓN CUERPO NEGRO LUMINOSIDAD = $\sigma \times 4 \times \pi \times \text{RADIO}$
 $A^2 \times \text{TEMPERATURA SUPERFICIAL}^4$ SIGMA SER CONSTANTE
STEFANBOLTZMANN * VALOR TÍPICO GIGANTE ROJO SER RADIO ~ 10
 $A^2 \times \text{RADIO SOLAR}$ TEMPERATURA SUPERFICIAL $\sim 3 \text{ HASTA } 4 \times 10$
 $A^3 \text{ K}$ FORMA LUMINOSIDAD SER APROXIMADAMENTE $10^3 \times$
LUMINOSIDAD SOLAR ** YA NO MANTENER PROCESO AVANZADO
FUSIÓN NÚCLEO ESTELAR NÚCLEO COLAPSAR NUEVO * OTRO VEZ
TEMPERATURA NÚCLEO INCREMENTAR AHORA EXPULSAR CAPA

ANEXO N: TEXTO TRANSFORMADO FESTESE001V-S

EXTERNO ESTRELLA * LLAMADOA NEBULOSA PLANETARIO FORMAR
PARTIR RESTO CAPA EXTERNO VER EJERCICIO 3 SERIE EJERCICIO
ASTRONOMÍA ESA1 ESO1 * NÚCLEO COLAPSADOA ESTAR MUY
CALIENTE BLANCO ESTRELLA SER MUY PEQUEÑO * ESTRELLA
DENOMINAR MUY ADECUADAMENTE ENANO BLANCO SER FASE
FINALA VIDA ESTRELLA NORMAL TIPO SOLAR ** HACER ESTIMACIÓN
APROXIMADOA RELACIÓN LUMINOSIDAD TEMPERATURA
SUPERFICIAL TODO ESTRELLA SECUENCIA PRINCIPAL VER
DIAGRAMA HR FIGURA 6 * LÍNEA MÁS MENOS RECTO SECUENCIA
PRINCIPAL EXTENDER APROXIMADAMENTE ORDEN MAGNITUD
TEMPERATURA SUPERFICIAL 3×10^3 HASTAM 3×10^4 K * RANGO
LUMINOSIDAD EXTENDER APROXIMADAMENTE SEIS ORDEN
MAGNITUD 10 AMENOS 2 HASTAM 10 A4 X LUMINOSIDAD SOLAR *
TANTO ESTIMAR FORMA1 APROXIMADOA LUMINOSIDAD
PROPORCIONAL TEMPERATURA SUPERFICIAL A6 ESTRELLA
SECUENCIA PRINCIPAL ** DAR ALGUNO EJEMPLO ESTRELLA GRANDE
MASA SECUENCIA PRINCIPAL TEMPERATURA SUPERFICIAL
APROXIMADAMENTE TEMPERATURA SUPERFICIAL ESTRELLA =
 10^4 X 10^4 K TENER LUMINOSIDAD APROXIMADAMENTE
LUMINOSIDAD ESTRELLA = 10 DIVIDIR 5 COMA 8 A6 · LUMINOSIDAD
SOLAR APROXIMADAMENTE 26 VEZ LUMINOSIDAD SOL *
LUMINOSIDAD SOL TENER VALOR ESTÁNDARA 1 ESCALA1
LUMINOSIDAD ** ESTRELLA BAJOA MASAS TEMPERATURA
SUPERFICIAL ESTRELLA = 3×10^3 X 10^3 K TENER LUMINOSIDAD
APROXIMADAMENTE SÓLO 5 PORCIENTO LUMINOSIDAD SOL ** TODO

ANEXO N: TEXTO TRANSFORMADO FESTESE001V-S

INFORMACIÓN EXTRAER ESTRELLA PROVENIR RADIACIÓN RECIBIR *
EXPLICAR SECCIÓN HERRAMIENTA ASTRONÓMICO DIFERENTE
FILTRO SISTEMA COLOR USAR MEDIR BRILLO ESTRELLA * EJERCICIO
USAR IMAGEN B IMAGEN V * ANÁLISIS IMAGEN ENCONTRAR
MAGNITUD APARENTE ÍNDICE COLOR B MAGNITUD APARENTE
ÍNDICE COLOR V MUESTRA1 ESTRELLA CÚMULO * CALCULAR VALOR
MAGNITUD APARENTE ÍNDICE COLOR B MAGNITUD APARENTE
ÍNDICE COLOR V ÍNDICE COLOR B MENOS V * FINALMENTE
DETERMINAR TEMPERATURA SUPERFICIAL ESTRELLA VER SECCIÓN
HERRAMIENTA ASTRONÓMICO ** CÚMULO SER GRUPO ESTRELLA *
VIDA CÚMULO DETERMINADO A VIDA DIFERENTE TIPO ESTRELLA
COMPONER * OBSERVACIÓN MOSTRAR CÚMULO GLOBULAR QUEDAR
MUY POCO CANTIDAD GAS POLVO MANERA SER MUY RARO
NACIMIENTO NUEVO ESTRELLA TAL CÚMULO * ESTRELLA VER
CÚMULO GLOBULAR SER TODO ADULTO EVOLUCIONAR FORMA1
DIFERENTE FUNCIÓN MASA ** MAYORÍA ESTRELLA BAJOA MASA
ENCONTRAR SECUENCIA PRINCIPAL * DEBER ESTRELLA BAJOA MASA
CONSUMIR ENERGÍA MUY LENTAMENTE * ESTRELLA QUEMAR
RESERVA1 HIDRÓGENO POCO POCO CONTINUAR MODO MIL MILLÓN
AÑO * CONSIGUIENTE ESTRELLA ESTAR SECUENCIA PRINCIPAL
MUCHO TIEMPO ** CONTRARIO ESTRELLA MÁS PESADO A CÚMULO
GLOBULAR YA TRANSFORMAR HIDRÓGENO NÚCLEO CONVERTIR
GIGANTE ROJO * TODO SUCEDER HACER MUCHO TIEMPO ASÍ HOY DÍA
NO HABER ESTRELLA CALIENTE MASIVO DEJAR HUECO MITAD
SUPERIOR SECUENCIA PRINCIPAL VER FIGURA 7 * ESTRELLA

ANEXO N: TEXTO TRANSFORMADO FESTESE001V-S

ENCONTRAR AHORA ÁREA DIAGONAL COMENZAR SECUENCIA
PRINCIPAL DESPLAZAR PARTE SUPERIOR DERECHO DIAGRAMA HR
CONOCIDO A RAMA GIGANTE ROJO ** PUNTO ENCONTRAR
SECUENCIA PRINCIPAL RAMA GIGANTE ROJO LLAMAR PUNTO GIRO I
DAR PISTA IMPORTANTE HORA DETERMINAR EDAD CÚMULO
GLOBULAR * SIGUIENTE EJERCICIO MEDIR COORDENADA PUNTO
DIAGRAMA HR DETERMINAR EDAD MESSIER 12 **

ANEXO O: Frecuencias del Texto Preparado FESTESE001V-S

Listado de las frecuencias del texto transformado de FESTESE001V-S obtenidas con el programa PAFE. En el orden léxicométrico aparecen al principio los números y símbolos y al final las palabras que comienzan por vocales con tilde. En la columna de la categoría gramatical, el lenguaje matemático se denota con “lm” y entre paréntesis se ha añadido, cuando estaba claro, si además correspondía con otro tipo de palabra léxica (“s”= sustantivo, “adj”= adjetivo, “v”=verbo). Los verbos auxiliares se señalan como “f-verbo”. La tabla se encuentra dividida en las tres franjas señaladas en la metodología: alta, media y baja.

Tabla 98

Frecuencias del texto preparado FESTESE001V-S

OL	Palabra	n_i	f_i (a1000)	N_i	Categoría gramatical
Franja alta					
1	DE	228	93.023	228	funcional
2	LA	130	53.04	358	funcional
3	EL	67	27.336	425	funcional
4	EN	64	26.112	489	funcional
5	SE	54	22.032	543	funcional
6	ESTRELLAS	52	21.216	595	sustantivo
7	Y	48	19.584	643	funcional
8	LAS	43	17.544	686	funcional
9	QUE	39	15.912	725	funcional
10	UNA	38	15.504	763	funcional
11	DEL	36	14.688	799	funcional
12	LOS	34	13.872	833	funcional
13	A	28	11.424	861	funcional
14	ES	28	11.424	889	verbo
15	ESTRELLA	28	11.424	917	sustantivo
16	UN	27	11.016	944	funcional

ANEXO O: FRECUENCIAS DEL TEXTO PREPARADO FESTESE001V-

OL	Palabra	n_i	f_i (a1000)	N_i	Categoría gramatical
17	CÚMULOS	25	10.2	969	sustantivo
Franja media					
18	TEMPERATURA	21	8.568	990	sustantivo
19	LUMINOSIDAD	18	7.344	1008	sustantivo
20	MÁS	17	6.936	1025	adverbio
21	PARA	17	6.936	1042	funcional
22	CÚMULO	16	6.528	1058	sustantivo
23	POR	16	6.528	1074	funcional
24	SUPERFICIAL	15	6.12	1089	adjetivo
25	GLOBULARES	14	5.712	1103	adjetivo
26	NÚCLEO	14	5.712	1117	sustantivo
27	MUY	13	5.304	1130	adverbio
28	PRINCIPAL	13	5.304	1143	adjetivo
29	SECUENCIA	13	5.304	1156	sustantivo
30	X	13	5.304	1169	lm(v)
31	APROXIMADAMENTE	12	4.896	1181	adverbio
32	HIDRÓGENO	12	4.896	1193	sustantivo
33	SU	12	4.896	1205	funcional
34	10	11	4.488	1216	lm(adj)
35	COMO	11	4.488	1227	funcional
36	MAGNITUD	11	4.488	1238	sustantivo
37	MASA	11	4.488	1249	sustantivo
38	AÑOS	10	4.08	1259	sustantivo
39	DIAGRAMA	10	4.08	1269	sustantivo
40	HR	10	4.08	1279	sustantivo
41	COLOR	9	3.672	1288	sustantivo
42	ESTE	9	3.672	1297	funcional
43	GLOBULAR	9	3.672	1306	adjetivo
44	TIENE	9	3.672	1315	verbo
45	CON	8	3.264	1323	funcional
46	FORMA1	8	3.264	1331	sustantivo
47	LO	8	3.264	1339	funcional
48	NO	8	3.264	1347	adverbio
49	SON	8	3.264	1355	verbo
50	DIFERENTES	7	2.856	1362	adjetivo
51	ENERGÍA	7	2.856	1369	sustantivo
52	ENTRE	7	2.856	1376	funcional
53	ESTA	7	2.856	1383	funcional
54	FUSIÓN	7	2.856	1390	sustantivo
55	LUZ	7	2.856	1397	sustantivo
56	MESSIER	7	2.856	1404	sustantivo
57	MILLONES	7	2.856	1411	sustantivo
58	PORCIENTO	7	2.856	1418	lm
59	PROCESO	7	2.856	1425	sustantivo
60	PUEDEX	7	2.856	1432	f-verbo
61	SOL	7	2.856	1439	sustantivo
62	SUS	7	2.856	1446	funcional

ANEXO O: FRECUENCIAS DEL TEXTO PREPARADO FESTESE001V-S

OL	Palabra	n_i	f_i (a1000)	N_i	Categoría gramatical
63	VER	7	2.856	1453	verbo
64	VIDA	7	2.856	1460	sustantivo
65	ÍNDICE	7	2.856	1467	sustantivo
66	1	6	2.448	1473	lm(adj)
67	12	6	2.448	1479	lm(adj)
68	ABIERTOSA	6	2.448	1485	adjetivo
69	APARENTE	6	2.448	1491	adjetivo
70	CUANDO	6	2.448	1497	funcional
71	EDAD	6	2.448	1503	sustantivo
72	LÁCTEA	6	2.448	1509	adjetivo
73	MAYORÍA	6	2.448	1515	sustantivo
74	RADIO	6	2.448	1521	sustantivo
75	UNOS	6	2.448	1527	funcional
76	VÍA	6	2.448	1533	sustantivo
77	4	5	2.04	1538	lm(adj)
78	=	5	2.04	1543	lm(adj)
79	B	5	2.04	1548	lm
80	BAJAA	5	2.04	1553	adjetivo
81	ELEMENTOS	5	2.04	1558	sustantivo
82	ENCUENTRAN	5	2.04	1563	verbo
83	FIGURA	5	2.04	1568	sustantivo
84	GALAXIA	5	2.04	1573	sustantivo
85	GAS	5	2.04	1578	sustantivo
86	HANX	5	2.04	1583	f-verbo
87	MILES	5	2.04	1588	adjetivo
88	NÚCLEOS	5	2.04	1593	sustantivo
89	POCOS	5	2.04	1598	adjetivo
90	SOLAR	5	2.04	1603	adjetivo
91	SÓLO	5	2.04	1608	adverbio
92	V	5	2.04	1613	lm
93	YA	5	2.04	1618	adverbio
94	2	4	1.632	1622	lm(adj)
95	3	4	1.632	1626	lm(adj)
96	A2	4	1.632	1630	lm
97	A3	4	1.632	1634	lm
98	A4	4	1.632	1638	lm
99	ASTRONÓMICAS	4	1.632	1642	adjetivo
100	CANTIDAD	4	1.632	1646	sustantivo
101	DESDE	4	1.632	1650	funcional
102	DISCO	4	1.632	1654	sustantivo
103	DONDE	4	1.632	1658	funcional
104	DURANTE	4	1.632	1662	funcional
105	EJERCICIO	4	1.632	1666	sustantivo
106	ESTABLE	4	1.632	1670	adjetivo
107	ESTELAR	4	1.632	1674	adjetivo
108	FASE	4	1.632	1678	sustantivo
109	FORMAN	4	1.632	1682	verbo
110	GIGANTES	4	1.632	1686	adjetivo

ANEXO O: FRECUENCIAS DEL TEXTO PREPARADO FESTESE001V-

OL	Palabra	n_i	f_i (a1000)	N_i	Categoría gramatical
111	HACE	4	1.632	1690	verbo
112	HACIA	4	1.632	1694	funcional
113	HELIO	4	1.632	1698	sustantivo
114	HERRAMIENTAS	4	1.632	1702	sustantivo
115	INCLUSO	4	1.632	1706	adverbio
116	K	4	1.632	1710	lm(s)
117	LÍNEA	4	1.632	1714	sustantivo
118	O	4	1.632	1718	funcional
119	PARTE	4	1.632	1722	sustantivo
120	RADIACIÓN	4	1.632	1726	sustantivo
121	ROJAS	4	1.632	1730	adjetivo
122	SECCIÓN	4	1.632	1734	sustantivo
123	SOBRE	4	1.632	1738	funcional
124	SUPERIOR	4	1.632	1742	adjetivo
125	TIEMPO	4	1.632	1746	sustantivo
126	TODAS	4	1.632	1750	adjetivo
127	CAPAS	3	1.224	1753	sustantivo
128	CENTRAL	3	1.224	1756	adjetivo
129	DA	3	1.224	1759	verbo
130	DETERMINAR	3	1.224	1762	verbo
131	EQUILIBRIO	3	1.224	1765	sustantivo
132	ESTAS	3	1.224	1768	funcional
133	ESTO	3	1.224	1771	funcional
134	EXTERNAS	3	1.224	1774	adjetivo
135	HASTAM	3	1.224	1777	lm
136	INFERIOR	3	1.224	1780	adjetivo
137	MASAS	3	1.224	1783	sustantivo
138	MENOS	3	1.224	1786	adverbio
139	PARTIR	3	1.224	1789	verbo
140	PEQUEÑA	3	1.224	1792	adjetivo
141	PESADOSA	3	1.224	1795	adjetivo
142	PLÉYADES	3	1.224	1798	sustantivo
143	POLVO	3	1.224	1801	sustantivo
144	PRESIÓN	3	1.224	1804	sustantivo
145	PUEDENX	3	1.224	1807	f-verbo
146	PUNTO	3	1.224	1810	sustantivo
147	ROJA	3	1.224	1813	adjetivo
148	TALES	3	1.224	1816	adjetivo
149	TAMBIÉN	3	1.224	1819	adverbio
150	TANTO	3	1.224	1822	adjetivo
151	TIPOS	3	1.224	1825	sustantivo
152	TODO	3	1.224	1828	adjetivo
153	TOTAL	3	1.224	1831	adjetivo
154	VALORES	3	1.224	1834	sustantivo
155	OCOMA7	2	0.816	1836	lm(adj)
156	6	2	0.816	1838	lm(adj)
157	A6	2	0.816	1840	lm
158	AHORA	2	0.816	1842	adverbio

ANEXO O: FRECUENCIAS DEL TEXTO PREPARADO FESTESE001V-S

OL	Palabra	n_i	f_i (a1000)	N_i	Categoría gramatical
159	AL	2	0.816	1844	funcional
160	ALGUNOS	2	0.816	1846	adjetivo
161	APARECE	2	0.816	1848	verbo
162	APROXIMADAA	2	0.816	1850	adjetivo
163	C	2	0.816	1852	lm(s)
164	CADA	2	0.816	1854	adjetivo
165	CALCULAR	2	0.816	1856	verbo
166	CIENTOS	2	0.816	1858	adjetivo
167	COMBUSTIÓN	2	0.816	1860	sustantivo
168	COMIENZA	2	0.816	1862	verbo
169	COMPONE	2	0.816	1864	verbo
170	COMPONEN	2	0.816	1866	verbo
171	COMÚN	2	0.816	1868	adjetivo
172	CONJUNTO	2	0.816	1870	sustantivo
173	CONTRAERSE	2	0.816	1872	verbo
174	CUAL	2	0.816	1874	funcional
175	DEBIDO	2	0.816	1876	verbo
176	DERECHA	2	0.816	1878	adjetivo
177	DETERMINARÁS	2	0.816	1880	verbo
178	DISTANCIA	2	0.816	1882	sustantivo
179	DISTRIBUCIÓN	2	0.816	1884	sustantivo
180	DIÁMETRO	2	0.816	1886	sustantivo
181	E	2	0.816	1888	funcional
182	ECUACIÓN	2	0.816	1890	sustantivo
183	ENCUENTRA	2	0.816	1892	verbo
184	ESFÉRICO	2	0.816	1894	adjetivo
185	ESTRUCTURA	2	0.816	1896	sustantivo
186	ESTÁ	2	0.816	1898	verbo
187	EVOLUCIÓN	2	0.816	1900	sustantivo
188	EXTIENDE	2	0.816	1902	verbo
189	FORMACIÓN	2	0.816	1904	sustantivo
190	FORMAR	2	0.816	1906	verbo
191	FUERA	2	0.816	1908	adverbio
192	FUEVSERX	2	0.816	1910	f-verbo
193	FUSIONAN	2	0.816	1912	verbo
194	GENERAL	2	0.816	1914	adjetivo
195	GIGANTE	2	0.816	1916	adjetivo
196	GRANDE	2	0.816	1918	adjetivo
197	GRUPO	2	0.816	1920	sustantivo
198	HALO	2	0.816	1922	sustantivo
199	HASTA	2	0.816	1924	funcional
200	HISTORIA	2	0.816	1926	sustantivo
201	IMAGEN	2	0.816	1928	sustantivo
202	IMPORTANTE	2	0.816	1930	adjetivo
203	LARGO	2	0.816	1932	adjetivo
204	LENTAMENTE	2	0.816	1934	adverbio
205	LLAMA	2	0.816	1936	verbo
206	LUGAR	2	0.816	1938	sustantivo

ANEXO O: FRECUENCIAS DEL TEXTO PREPARADO FESTESE001V-

OL	Palabra	n_i	f_i (a1000)	N_i	Categoría gramatical
207	LUGARES	2	0.816	1940	sustantivo
208	LUMÍNICA	2	0.816	1942	adjetivo
209	MANERA	2	0.816	1944	sustantivo
210	MEDIDAS1	2	0.816	1946	sustantivo
211	MIENTRAS	2	0.816	1948	adverbio
212	MUCHO	2	0.816	1950	adjetivo
213	MUCHOS	2	0.816	1952	adjetivo
214	MUESTRAV	2	0.816	1954	verbo
215	NEBULOSAS	2	0.816	1956	sustantivo
216	NGC	2	0.816	1958	sustantivo
217	NORMAL	2	0.816	1960	adjetivo
218	NOS	2	0.816	1962	funcional
219	NUESTRA	2	0.816	1964	funcional
220	NUEVAS	2	0.816	1966	adjetivo
221	NÚMERO	2	0.816	1968	sustantivo
222	OBJETO	2	0.816	1970	sustantivo
223	OBSERVACIONES	2	0.816	1972	sustantivo
224	OTRA	2	0.816	1974	adjetivo
225	PASO1	2	0.816	1976	sustantivo
226	PLANETARIAS	2	0.816	1978	adjetivo
227	POCO	2	0.816	1980	adjetivo
228	PODEMOSX	2	0.816	1982	f-verbo
229	PODRÁSX	2	0.816	1984	f-verbo
230	PROCESOS	2	0.816	1986	sustantivo
231	PROPIA	2	0.816	1988	adjetivo
232	PROPIEDADES	2	0.816	1990	sustantivo
233	RAMA	2	0.816	1992	sustantivo
234	RARO	2	0.816	1994	adjetivo
235	REACCIONES	2	0.816	1996	sustantivo
236	REACCIÓN	2	0.816	1998	sustantivo
237	RECTA	2	0.816	2000	adjetivo
238	RELATIVAMENTE	2	0.816	2002	adverbio
239	RESIDUAL	2	0.816	2004	adjetivo
240	SER	2	0.816	2006	verbo
241	SIGMA	2	0.816	2008	sustantivo
242	SINO	2	0.816	2010	funcional
243	SISTEMAS	2	0.816	2012	sustantivo
244	STEFANBOLTZMANN	2	0.816	2014	sustantivo
245	TIENEN	2	0.816	2016	verbo
246	TU	2	0.816	2018	funcional
247	TÍPICO	2	0.816	2020	adjetivo
248	UNO	2	0.816	2022	funcional
249	USANDO	2	0.816	2024	verbo
250	VALOR	2	0.816	2026	sustantivo
251	VANX	2	0.816	2028	f-verbo
252	VEZ	2	0.816	2030	sustantivo
253	VIEJOS	2	0.816	2032	adjetivo
254	VISIBLE	2	0.816	2034	adjetivo

ANEXO O: FRECUENCIAS DEL TEXTO PREPARADO FESTESE001V-S

OL	Palabra	n _i	f _i (a1000)	N _i	Categoría gramatical
255	VÉRTICE	2	0.816	2036	sustantivo
256	~	2	0.816	2038	lm
257	ÉL	2	0.816	2040	funcional
Franja baja					
258	100	1	0.408	2041	lm(adj)
259	13	1	0.408	2042	lm(adj)
260	1764	1	0.408	2043	lm(adj)
261	1783	1	0.408	2044	lm(adj)
262	1888	1	0.408	2045	lm(adj)
263	1COMA0	1	0.408	2046	lm(adj)
264	25	1	0.408	2047	lm(adj)
265	26	1	0.408	2048	lm(adj)
266	30	1	0.408	2049	lm(adj)
267	3000	1	0.408	2050	lm(adj)
268	3COMA5	1	0.408	2051	lm(adj)
269	400	1	0.408	2052	lm(adj)
270	5	1	0.408	2053	lm(adj)
271	5COMA8	1	0.408	2054	lm(adj)
272	6218	1	0.408	2055	lm(adj)
273	6COMA7	1	0.408	2056	lm(adj)
274	7	1	0.408	2057	lm(adj)
275	74	1	0.408	2058	lm(adj)
276	99COMA3	1	0.408	2059	lm(adj)
277	ABIERTO A	1	0.408	2060	adjetivo
278	ABSOLUTA	1	0.408	2061	adjetivo
279	ABUNDANCIA	1	0.408	2062	sustantivo
280	ACABANDO	1	0.408	2063	verbo
281	ADECUADAMENTE	1	0.408	2064	adverbio
282	ADULTAS	1	0.408	2065	adjetivo
283	AGOTA	1	0.408	2066	verbo
284	AGRUPACIONES	1	0.408	2067	sustantivo
285	ALGUNAS	1	0.408	2068	adjetivo
286	ALREDEDOR	1	0.408	2069	adverbio
287	ALTA	1	0.408	2070	adjetivo
288	AMENOS2	1	0.408	2071	lm
289	ANTERIORES	1	0.408	2072	adjetivo
290	ANÁLISIS	1	0.408	2073	sustantivo
291	ASTROFÍSICO	1	0.408	2074	adjetivo
292	ASTRONOMÍA	1	0.408	2075	sustantivo
293	ASTRONÓMICA	1	0.408	2076	adjetivo
294	ASTRÓNOMOS	1	0.408	2077	sustantivo
295	ASÍ	1	0.408	2078	adverbio
296	AUMENTO1	1	0.408	2079	sustantivo
297	AVANZADOS A	1	0.408	2080	adjetivo
298	BANCO	1	0.408	2081	sustantivo
299	BLANCAS	1	0.408	2082	adjetivo
300	BLANCO	1	0.408	2083	adjetivo

ANEXO O: FRECUENCIAS DEL TEXTO PREPARADO FESTESE001V-

OL	Palabra	n_i	f_i (a1000)	N_i	Categoría gramatical
301	BOLA	1	0.408	2084	sustantivo
302	BRILLANTE	1	0.408	2085	adjetivo
303	BRILLANTES	1	0.408	2086	adjetivo
304	BRILLO	1	0.408	2087	sustantivo
305	BRUNO	1	0.408	2088	sustantivo
306	CALIENTE	1	0.408	2089	adjetivo
307	CALIENTES	1	0.408	2090	adjetivo
308	CAMINOS	1	0.408	2091	sustantivo
309	CAPA	1	0.408	2092	sustantivo
310	CENTENAR	1	0.408	2093	sustantivo
311	CENTRO	1	0.408	2094	sustantivo
312	CERCANA	1	0.408	2095	adjetivo
313	CHARLES	1	0.408	2096	sustantivo
314	CICLO	1	0.408	2097	sustantivo
315	CIELO	1	0.408	2098	sustantivo
316	CIRCUNDANTE	1	0.408	2099	adjetivo
317	CIUDADES	1	0.408	2100	sustantivo
318	CLARAMENTE	1	0.408	2101	adverbio
319	CLAVE	1	0.408	2102	sustantivo
320	COLAPSA	1	0.408	2103	verbo
321	COLAPSADOA	1	0.408	2104	adjetivo
322	COLECCIONES	1	0.408	2105	sustantivo
323	COLISIONES	1	0.408	2106	sustantivo
324	COLOCAN	1	0.408	2107	verbo
325	COMIENZANX	1	0.408	2108	f-verbo
326	COMIENZAX	1	0.408	2109	f-verbo
327	COMPACTAS	1	0.408	2110	adjetivo
328	COMPACTOS	1	0.408	2111	adjetivo
329	COMPARTEN	1	0.408	2112	verbo
330	COMPLICADOA	1	0.408	2113	adjetivo
331	COMPONENTES	1	0.408	2114	sustantivo
332	COMPRENDER	1	0.408	2115	verbo
333	COMUNIDAD	1	0.408	2116	sustantivo
334	CONCENTRAN	1	0.408	2117	verbo
335	CONOCEN	1	0.408	2118	verbo
336	CONOCIDAA	1	0.408	2119	adjetivo
337	CONSECUENCIA	1	0.408	2120	sustantivo
338	CONSIDERA	1	0.408	2121	verbo
339	CONSIDERABLE	1	0.408	2122	adjetivo
340	CONSIGUIENTE	1	0.408	2123	adjetivo
341	CONSTANTEI	1	0.408	2124	sustantivo
342	CONSTANTESA	1	0.408	2125	adjetivo
343	CONSTELACIÓN	1	0.408	2126	sustantivo
344	CONSTRUIDO	1	0.408	2127	verbo
345	CONSUMIENDO	1	0.408	2128	verbo
346	CONTAMINACIÓN	1	0.408	2129	sustantivo
347	CONTIENE	1	0.408	2130	verbo
348	CONTIENEN	1	0.408	2131	verbo

ANEXO O: FRECUENCIAS DEL TEXTO PREPARADO FESTESE001V-S

OL	Palabra	n_i	f_i (a1000)	N_i	Categoría gramatical
349	CONTINUARÁN	1	0.408	2132	verbo
350	CONTRACCIÓN	1	0.408	2133	sustantivo
351	CONTRARIO	1	0.408	2134	adjetivo
352	CONVERTIDO	1	0.408	2135	verbo
353	CONVERTIRSE	1	0.408	2136	verbo
354	CONVIERTE	1	0.408	2137	verbo
355	COORDENADAS	1	0.408	2138	sustantivo
356	CORRELACIÓN	1	0.408	2139	sustantivo
357	CORRESPONDÍAN	1	0.408	2140	verbo
358	CORTESÍA	1	0.408	2141	sustantivo
359	CREE	1	0.408	2142	verbo
360	CUALES	1	0.408	2143	funcional
361	CUALQUIER	1	0.408	2144	adjetivo
362	CUERPO	1	0.408	2145	sustantivo
363	CURIOSO	1	0.408	2146	adjetivo
364	DAR	1	0.408	2147	verbo
365	DEBE	1	0.408	2148	verbo
366	DECAE	1	0.408	2149	verbo
367	DEDUCE	1	0.408	2150	verbo
368	DEJA	1	0.408	2151	verbo
369	DENOMINAN	1	0.408	2152	verbo
370	DENSO	1	0.408	2153	adjetivo
371	DENTRO	1	0.408	2154	adverbio
372	DEPENDIENDO	1	0.408	2155	verbo
373	DEPOSITÓ	1	0.408	2156	verbo
374	DERECHO	1	0.408	2157	adjetivo
375	DERIVAR	1	0.408	2158	verbo
376	DESCRIBIÓ	1	0.408	2159	verbo
377	DESCUBIERTO	1	0.408	2160	verbo
378	DESPLAZA	1	0.408	2161	verbo
379	DESTRUIDO	1	0.408	2162	verbo
380	DETERMINADAA	1	0.408	2163	adjetivo
381	DEUTERIO	1	0.408	2164	sustantivo
382	DIAGONAL	1	0.408	2165	adjetivo
383	DIFIEREN	1	0.408	2166	verbo
384	DIFUSAS	1	0.408	2167	adjetivo
385	DISTANCIAS	1	0.408	2168	sustantivo
386	DISTINTOS	1	0.408	2169	adjetivo
387	DISTRIBUYEN	1	0.408	2170	verbo
388	DIVIDIDO	1	0.408	2171	verbo
389	DUODÉCIMO	1	0.408	2172	adjetivo
390	DÍA	1	0.408	2173	sustantivo
391	EDADES	1	0.408	2174	sustantivo
392	EINSTEIN	1	0.408	2175	sustantivo
393	EISENDLE	1	0.408	2176	sustantivo
394	EJEMPLO	1	0.408	2177	sustantivo
395	EJEMPLOS	1	0.408	2178	sustantivo
396	EJERCICIOS	1	0.408	2179	sustantivo

ANEXO O: FRECUENCIAS DEL TEXTO PREPARADO FESTESE001V-

OL	Palabra	n_i	f_i (a1000)	N_i	Categoría gramatical
397	EJERCIDAA	1	0.408	2180	adjetivo
398	EJERCIDASA	1	0.408	2181	adjetivo
399	ELEVAN	1	0.408	2182	verbo
400	ELLAS	1	0.408	2183	funcional
401	EMITE	1	0.408	2184	verbo
402	EMITIDAA	1	0.408	2185	adjetivo
403	ENANAS	1	0.408	2186	adjetivo
404	ENCONTRARÁS	1	0.408	2187	verbo
405	ENDE	1	0.408	2188	adverbio
406	ENFRÍEN	1	0.408	2189	verbo
407	ENORME	1	0.408	2190	adjetivo
408	ENTONCES	1	0.408	2191	adverbio
409	ERRORES	1	0.408	2192	sustantivo
410	ESA1	1	0.408	2193	sustantivo
411	ESCALA1	1	0.408	2194	sustantivo
412	ESFÉRICOS	1	0.408	2195	adjetivo
413	ESO1	1	0.408	2196	sustantivo
414	ESPACIO	1	0.408	2197	sustantivo
415	ESPECÍFICAS	1	0.408	2198	adjetivo
416	ESTADO	1	0.408	2199	sustantivo
417	ESTARÁN	1	0.408	2200	verbo
418	ESTIMACIÓN	1	0.408	2201	sustantivo
419	ESTIMAR	1	0.408	2202	verbo
420	ESTOS	1	0.408	2203	funcional
421	ESTUDIADAA	1	0.408	2204	adjetivo
422	ESTUDIAR	1	0.408	2205	verbo
423	ESTUDIO1	1	0.408	2206	sustantivo
424	ESTÁNDARA	1	0.408	2207	adjetivo
425	ESTÁNX	1	0.408	2208	f-verbo
426	ESTÁX	1	0.408	2209	f-verbo
427	ETAPAS	1	0.408	2210	sustantivo
428	EVOLUCIONADO	1	0.408	2211	verbo
429	EVOLUCIONAN	1	0.408	2212	verbo
430	EXISTEN	1	0.408	2213	verbo
431	EXPANDAN	1	0.408	2214	verbo
432	EXPLICA	1	0.408	2215	verbo
433	EXPLOTAN	1	0.408	2216	verbo
434	EXPULSAN	1	0.408	2217	verbo
435	EXTRAER	1	0.408	2218	verbo
436	FAMOSA	1	0.408	2219	adjetivo
437	FAMOSOS	1	0.408	2220	adjetivo
438	FILTROS	1	0.408	2221	sustantivo
439	FINAL	1	0.408	2222	sustantivo
440	FINALA	1	0.408	2223	adjetivo
441	FINALMENTE	1	0.408	2224	adverbio
442	FORMARON	1	0.408	2225	verbo
443	FORMAS1	1	0.408	2226	sustantivo
444	FORMIDABLE	1	0.408	2227	adjetivo

ANEXO O: FRECUENCIAS DEL TEXTO PREPARADO FESTESE001V-S

OL	Palabra	n_i	f_i (a1000)	N_i	Categoría gramatical
445	FRACCIÓN	1	0.408	2228	sustantivo
446	FRENTE	1	0.408	2229	adverbio
447	FUERZAS	1	0.408	2230	sustantivo
448	FUEVSER	1	0.408	2231	verbo
449	FUNCIÓN	1	0.408	2232	sustantivo
450	FUSIONARSE	1	0.408	2233	verbo
451	FÍSICAS	1	0.408	2234	adjetivo
452	GALAXIAS	1	0.408	2235	sustantivo
453	GENERAN	1	0.408	2236	verbo
454	GIGANTESCA	1	0.408	2237	adjetivo
455	GIRO1	1	0.408	2238	sustantivo
456	GRADOS	1	0.408	2239	sustantivo
457	GRAN	1	0.408	2240	adjetivo
458	GRAVITATORIA	1	0.408	2241	adjetivo
459	GRAVITATORIAMENTE	1	0.408	2242	adverbio
460	GRAVITATORIAS	1	0.408	2243	adjetivo
461	GRÁFICA	1	0.408	2244	sustantivo
462	HACER	1	0.408	2245	verbo
463	HAX	1	0.408	2246	f-verbo
464	HAY	1	0.408	2247	verbo
465	HAYX	1	0.408	2248	f-verbo
466	HECHO1	1	0.408	2249	sustantivo
467	HERSCHEL	1	0.408	2250	sustantivo
468	HIPPARCOS	1	0.408	2251	sustantivo
469	HORA	1	0.408	2252	sustantivo
470	HORIZONTAL	1	0.408	2253	adjetivo
471	HOY	1	0.408	2254	adverbio
472	HUECO	1	0.408	2255	sustantivo
473	IDEALES	1	0.408	2256	adjetivo
474	ILUSTRACIÓN	1	0.408	2257	sustantivo
475	ILUSTRAN	1	0.408	2258	verbo
476	IMPLICA	1	0.408	2259	verbo
477	IMPLICADAA	1	0.408	2260	adjetivo
478	IMPORTANTES	1	0.408	2261	adjetivo
479	IMÁGENES	1	0.408	2262	sustantivo
480	INCREMENTA	1	0.408	2263	verbo
481	INDICADASA	1	0.408	2264	adjetivo
482	INDICAN	1	0.408	2265	verbo
483	INDIVIDUALES	1	0.408	2266	adjetivo
484	INFORMACIÓN	1	0.408	2267	sustantivo
485	INTERACCIONES	1	0.408	2268	sustantivo
486	INTERNACIONAL	1	0.408	2269	adjetivo
487	INTERNO	1	0.408	2270	adjetivo
488	INTERÉS	1	0.408	2271	sustantivo
489	INVESTIGACIÓN	1	0.408	2272	sustantivo
490	INVOLUCRADOSA	1	0.408	2273	adjetivo
491	IZQUIERDO	1	0.408	2274	adjetivo
492	JUSTA	1	0.408	2275	adjetivo

ANEXO O: FRECUENCIAS DEL TEXTO PREPARADO FESTESE001V-

OL	Palabra	n_i	f_i (a1000)	N_i	Categoría gramatical
493	JÓVENES	1	0.408	2276	adjetivo
494	LEY	1	0.408	2277	sustantivo
495	LIGADOSA	1	0.408	2278	adjetivo
496	LIGEROS	1	0.408	2279	adjetivo
497	LISTA	1	0.408	2280	sustantivo
498	LLAMADASA	1	0.408	2281	adjetivo
499	LLAMADOA	1	0.408	2282	adjetivo
500	LLAMADOSA	1	0.408	2283	adjetivo
501	LLEGAX	1	0.408	2284	f-verbo
502	LOCALIZAN	1	0.408	2285	verbo
503	LONGEVOS	1	0.408	2286	adjetivo
504	LONGITUDES	1	0.408	2287	sustantivo
505	LUEGO	1	0.408	2288	funcional
506	LUMINOSIDADES	1	0.408	2289	sustantivo
507	LÍMITE	1	0.408	2290	sustantivo
508	MAGNITUDES	1	0.408	2291	sustantivo
509	MANTENER	1	0.408	2292	verbo
510	MANTENERSE	1	0.408	2293	verbo
511	MASIVAS	1	0.408	2294	adjetivo
512	MATERIA	1	0.408	2295	sustantivo
513	MATERIAL	1	0.408	2296	sustantivo
514	MAYOR	1	0.408	2297	adjetivo
515	MAYORES	1	0.408	2298	adjetivo
516	MAYORITARIAMENTE	1	0.408	2299	adverbio
517	MEDIDAA	1	0.408	2300	adjetivo
518	MEDIDOSA	1	0.408	2301	adjetivo
519	MEDIR	1	0.408	2302	verbo
520	MEDIRÁS	1	0.408	2303	verbo
521	MENUDO	1	0.408	2304	adjetivo
522	MEZCLA1	1	0.408	2305	sustantivo
523	MIL	1	0.408	2306	adjetivo
524	MILLÓN	1	0.408	2307	sustantivo
525	MITAD	1	0.408	2308	sustantivo
526	MODESTO	1	0.408	2309	adjetivo
527	MODO	1	0.408	2310	sustantivo
528	MOSTRADO	1	0.408	2311	verbo
529	MOVERÁN	1	0.408	2312	verbo
530	MUESTRA1	1	0.408	2313	sustantivo
531	MUESTRAN	1	0.408	2314	verbo
532	NACIMIENTO	1	0.408	2315	sustantivo
533	NEBULOSA	1	0.408	2316	sustantivo
534	NEGRO	1	0.408	2317	adjetivo
535	NINGÚN	1	0.408	2318	adjetivo
536	NUBES	1	0.408	2319	sustantivo
537	NUCLEARES	1	0.408	2320	adjetivo
538	NUESTRO	1	0.408	2321	funcional
539	NUEVO	1	0.408	2322	adjetivo
540	OBJETOS	1	0.408	2323	sustantivo

ANEXO O: FRECUENCIAS DEL TEXTO PREPARADO FESTESE001V-S

OL	Palabra	n_i	f_i (a1000)	N_i	Categoría gramatical
541	OBVIAMENTE	1	0.408	2324	adverbio
542	OFIUCO	1	0.408	2325	sustantivo
543	ONDA	1	0.408	2326	sustantivo
544	ORDEN	1	0.408	2327	sustantivo
545	ORIGINALES	1	0.408	2328	adjetivo
546	OTROS	1	0.408	2329	adjetivo
547	PARECE	1	0.408	2330	verbo
548	PARECER	1	0.408	2331	verbo
549	PARSECS	1	0.408	2332	sustantivo
550	PARTICULAR	1	0.408	2333	adjetivo
551	PASAN	1	0.408	2334	verbo
552	PATRÓN	1	0.408	2335	sustantivo
553	PEQUEÑO	1	0.408	2336	adjetivo
554	PERMANECEN	1	0.408	2337	verbo
555	PERO	1	0.408	2338	funcional
556	PERSPECTIVA	1	0.408	2339	sustantivo
557	PESADAA	1	0.408	2340	adjetivo
558	PESADASA	1	0.408	2341	adjetivo
559	PESAR	1	0.408	2342	verbo
560	PESE	1	0.408	2343	verbo
561	PI	1	0.408	2344	sustantivo
562	PISTA	1	0.408	2345	sustantivo
563	PLANO	1	0.408	2346	sustantivo
564	POCA	1	0.408	2347	adjetivo
565	POCAS	1	0.408	2348	adjetivo
566	PODER1	1	0.408	2349	sustantivo
567	POLUCIÓN	1	0.408	2350	sustantivo
568	PORQUE	1	0.408	2351	funcional
569	POSIBLE	1	0.408	2352	adjetivo
570	PRECISAA	1	0.408	2353	adjetivo
571	PRIMER	1	0.408	2354	adjetivo
572	PRIMERO	1	0.408	2355	adjetivo
573	PRISMÁTICOS	1	0.408	2356	adjetivo
574	PROBABLE	1	0.408	2357	adjetivo
575	PROBABLEMENTE	1	0.408	2358	adverbio
576	PRODUCE	1	0.408	2359	verbo
577	PROGRAMA	1	0.408	2360	sustantivo
578	PRONTO	1	0.408	2361	adverbio
579	PROPIO	1	0.408	2362	adjetivo
580	PROPORCIONAL	1	0.408	2363	adjetivo
581	PROPORCIONAN	1	0.408	2364	verbo
582	PROTOGALÁCTICO	1	0.408	2365	adjetivo
583	PROTONES	1	0.408	2366	sustantivo
584	PROVIENE	1	0.408	2367	verbo
585	PRUEBA1	1	0.408	2368	sustantivo
586	PRUEBAS1	1	0.408	2369	sustantivo
587	PRÓXIMOS	1	0.408	2370	adjetivo
588	PUBLICADO	1	0.408	2371	verbo

ANEXO O: FRECUENCIAS DEL TEXTO PREPARADO FESTESE001V-

OL	Palabra	n_i	f_i (a1000)	N_i	Categoría gramatical
589	PÁGINA	1	0.408	2372	sustantivo
590	QUEDA	1	0.408	2373	verbo
591	QUEMAN	1	0.408	2374	verbo
592	QUEMANDO	1	0.408	2375	verbo
593	QUÍMICOS	1	0.408	2376	adjetivo
594	RAINER	1	0.408	2377	sustantivo
595	RANGO	1	0.408	2378	sustantivo
596	RAZÓN	1	0.408	2379	sustantivo
597	RECIBIMOS	1	0.408	2380	verbo
598	RELACIÓN	1	0.408	2381	sustantivo
599	REMANENTES	1	0.408	2382	sustantivo
600	REPETIDASA	1	0.408	2383	adjetivo
601	REPETIR	1	0.408	2384	verbo
602	RESERVAS1	1	0.408	2385	sustantivo
603	RESOLUCIÓN	1	0.408	2386	sustantivo
604	RESOLVIÓ	1	0.408	2387	verbo
605	RESTANTE	1	0.408	2388	adjetivo
606	RESTOS	1	0.408	2389	sustantivo
607	RESULTADO1	1	0.408	2390	sustantivo
608	SEAN	1	0.408	2391	verbo
609	SEIS	1	0.408	2392	adjetivo
610	SERIE	1	0.408	2393	sustantivo
611	SIENDO	1	0.408	2394	verbo
612	SIGUIENTE	1	0.408	2395	adjetivo
613	SIGUIENTES	1	0.408	2396	adjetivo
614	SIMPLE	1	0.408	2397	adjetivo
615	SIN	1	0.408	2398	funcional
616	SISTEMA	1	0.408	2399	sustantivo
617	SOLA	1	0.408	2400	adjetivo
618	STAMPFER	1	0.408	2401	sustantivo
619	SUBPRODUCTO	1	0.408	2402	sustantivo
620	SUCEDE	1	0.408	2403	verbo
621	SUCEDIÓ	1	0.408	2404	verbo
622	SUCESO	1	0.408	2405	sustantivo
623	SUMINISTRO1	1	0.408	2406	sustantivo
624	SUPERFICIE	1	0.408	2407	sustantivo
625	SUPERNOVA	1	0.408	2408	sustantivo
626	SUPERVIVIENTES	1	0.408	2409	adjetivo
627	SUPUESTO1	1	0.408	2410	sustantivo
628	SÍ	1	0.408	2411	funcional
629	TAMAÑO	1	0.408	2412	sustantivo
630	TAN	1	0.408	2413	adjetivo
631	TELESCOPIO	1	0.408	2414	sustantivo
632	TEORÍAS	1	0.408	2415	sustantivo
633	TERMONUCLEAR	1	0.408	2416	adjetivo
634	TIERRA	1	0.408	2417	sustantivo
635	TIPO	1	0.408	2418	sustantivo
636	TODA	1	0.408	2419	adjetivo

ANEXO O: FRECUENCIAS DEL TEXTO PREPARADO FESTESE001V-S

OL	Palabra	n_i	f_i (a1000)	N_i	Categoría gramatical
637	TODOS	1	0.408	2420	adjetivo
638	TRANSFORMADO	1	0.408	2421	verbo
639	TRAVÉS	1	0.408	2422	sustantivo
640	TUVO	1	0.408	2423	verbo
641	TÉRMICA	1	0.408	2424	adjetivo
642	TÉRMINO	1	0.408	2425	sustantivo
643	TÍPICOS	1	0.408	2426	adjetivo
644	UNAS	1	0.408	2427	funcional
645	UNIVERSO	1	0.408	2428	sustantivo
646	USA	1	0.408	2429	verbo
647	USAMOS	1	0.408	2430	verbo
648	USAR	1	0.408	2431	verbo
649	VALIOSA	1	0.408	2432	adjetivo
650	VALIOSO	1	0.408	2433	adjetivo
651	VAN	1	0.408	2434	verbo
652	VARIAS	1	0.408	2435	adjetivo
653	VARIOS	1	0.408	2436	adjetivo
654	VE	1	0.408	2437	verbo
655	VEAMOS	1	0.408	2438	verbo
656	VECES	1	0.408	2439	sustantivo
657	VECINDAD	1	0.408	2440	sustantivo
658	VEMOS	1	0.408	2441	verbo
659	VERSE	1	0.408	2442	verbo
660	VIOLENTA	1	0.408	2443	adjetivo
661	VISTA	1	0.408	2444	sustantivo
662	WILLIAM	1	0.408	2445	sustantivo
663	.	1	0.408	2446	lm
664	ÁREA	1	0.408	2447	sustantivo
665	ÁREAS	1	0.408	2448	sustantivo
666	ÁTOMO	1	0.408	2449	sustantivo
667	ÓRDENES	1	0.408	2450	sustantivo
668	ÚTIL	1	0.408	2451	adjetivo

Nota. OL= orden lexicométrico; n= frecuencia absoluta; f= frecuencia relativa; N= frecuencia acumulada.

Anexo P: Frecuencias del Texto Transformado FESTESE001V-F

Listado de las frecuencias del texto transformado de FESTESE001V-F obtenidas con el programa PAFE. En el orden léxicométrico aparecen al principio los números y símbolos y al final las palabras que comienzan por vocales con tilde. En la columna de la categoría gramatical, el lenguaje matemático se denota con “lm” y entre paréntesis se ha añadido, cuando estaba claro, si además correspondía con otro tipo de palabra léxica (“s”= sustantivo, “adj”= adjetivo, “v”=verbo). Los verbos auxiliares se señalan como “f-verbo”. La tabla se encuentra dividida en las tres franjas señaladas en la metodología: alta, media y baja.

Tabla 99

Frecuencias del texto transformado FESTESE001V-F

OL	Lema	n_i	f_i (a1000)	N_i	Categoría gramatical
Franja alta					
1	ESTRELLA	76	55.152	76	sustantivo
2	SER	41	29.753	117	verbo
3	CÚMULO	39	28.302	156	sustantivo
4	NÚCLEO	19	13.788	175	sustantivo
5	MÁS	18	13.062	193	adverbio
6	GLOBULAR	17	12.337	210	adjetivo
7	TEMPERATURA	15	10.885	225	sustantivo
8	MASA	13	9.434	238	sustantivo
9	MUY	13	9.434	251	adverbio
10	PRINCIPAL	13	9.434	264	adjetivo
11	SECUENCIA	13	9.434	277	sustantivo
12	X	13	9.434	290	lm(v)
Franja media					
13	APROXIMADAMENTE	12	8.708	302	adverbio

ANEXO P: FRECUENCIAS DEL TEXTO TRANSFORMADO FESTESE001V-F

OL	Lema	n_i	f_i (a1000)	N_i	Categoría gramatical
14	HIDRÓGENO	12	8.708	314	sustantivo
15	TENER	12	8.708	326	verbo
16	10	11	7.983	337	lm(adj)
17	DIAGRAMA	11	7.983	348	sustantivo
18	L	11	7.983	359	lm(s)
19	LUMINOSIDAD	11	7.983	370	sustantivo
20	VER	11	7.983	381	verbo
21	AÑO	10	7.257	391	sustantivo
22	FORMA1	10	7.257	401	sustantivo
23	T	10	7.257	411	lm(s)
24	POCO	9	6.531	420	adjetivo
25	PROCESO	9	6.531	429	sustantivo
26	TODO	9	6.531	438	adjetivo
27	ENCONTRAR	8	5.806	446	verbo
28	GALAXIA	8	5.806	454	sustantivo
29	MILLÓN	8	5.806	462	sustantivo
30	NO	8	5.806	470	adverbio
31	SUPERFICIAL	8	5.806	478	adjetivo
32	DIFERENTE	7	5.08	485	adjetivo
33	EDAD	7	5.08	492	sustantivo
34	FORMAR	7	5.08	499	verbo
35	FUSIÓN	7	5.08	506	sustantivo
36	LUZ	7	5.08	513	sustantivo
37	MAGNITUD	7	5.08	520	sustantivo
38	PORCIENTO	7	5.08	527	lm
39	ROJO	7	5.08	534	adjetivo
40	SOL	7	5.08	541	sustantivo
41	VIDA	7	5.08	548	sustantivo
42	ENERGÍA	6	4.354	554	sustantivo
43	GIGANTE	6	4.354	560	adjetivo
44	MAYORÍA	6	4.354	566	sustantivo
45	MIL	6	4.354	572	adjetivo
46	=	5	3.628	577	lm(adj)
47	ABIERTO A	5	3.628	582	adjetivo
48	ASTRONÓMICO	5	3.628	587	adjetivo
49	BAJO A	5	3.628	592	adjetivo
50	DETERMINAR	5	3.628	597	verbo
51	EJERCICIO	5	3.628	602	sustantivo
52	ELEMENTO	5	3.628	607	sustantivo
53	GAS	5	3.628	612	sustantivo
54	HACER	5	3.628	617	verbo
55	HR	5	3.628	622	sustantivo
56	LÁCTEO	5	3.628	627	adjetivo
57	MESSIER	5	3.628	632	sustantivo
58	PESADO A	5	3.628	637	adjetivo
59	SOLAR	5	3.628	642	adjetivo
60	SÓLO	5	3.628	647	adverbio
61	USAR	5	3.628	652	verbo

ANEXO P: FRECUENCIAS DEL TEXTO TRANSFORMADO FESTESE001V-F

OL	Lema	n_i	f_i (a1000)	N_i	Categoría gramatical
62	VALOR	5	3.628	657	sustantivo
63	VÍA	5	3.628	662	sustantivo
64	YA	5	3.628	667	adverbio
65	12	4	2.903	671	lm(adj)
66	3	4	2.903	675	lm(adj)
67	A2	4	2.903	679	lm
68	A3	4	2.903	683	lm
69	A4	4	2.903	687	lm
70	CANTIDAD	4	2.903	691	sustantivo
71	CAPA	4	2.903	695	sustantivo
72	COMPONER	4	2.903	699	verbo
73	DAR	4	2.903	703	verbo
74	DISCO	4	2.903	707	sustantivo
75	ESTABLE	4	2.903	711	adjetivo
76	ESTELAR	4	2.903	715	adjetivo
77	FASE	4	2.903	719	sustantivo
78	FIG	4	2.903	723	sustantivo
79	HELIO	4	2.903	727	sustantivo
80	HERRAMIENTA	4	2.903	731	sustantivo
81	INCLUSO	4	2.903	735	adverbio
82	K	4	2.903	739	lm(s)
83	LUGAR	4	2.903	743	sustantivo
84	LÍNEA	4	2.903	747	sustantivo
85	MOSTRAR	4	2.903	751	verbo
86	MUCHO	4	2.903	755	adjetivo
87	NUEVO	4	2.903	759	adjetivo
88	PARTE	4	2.903	763	sustantivo
89	PEQUEÑO	4	2.903	767	adjetivo
90	R	4	2.903	771	lm(s)
91	RADIACIÓN	4	2.903	775	sustantivo
92	REACCIÓN	4	2.903	779	sustantivo
93	SECCIÓN	4	2.903	783	sustantivo
94	SUPERIOR	4	2.903	787	adjetivo
95	TANTO	4	2.903	791	adjetivo
96	TIEMPO	4	2.903	795	sustantivo
97	TIPO	4	2.903	799	sustantivo
98	UNOA	4	2.903	803	adjetivo
99	ALGUNO	3	2.177	806	adjetivo
100	CENTRAL	3	2.177	809	adjetivo
101	COLOR	3	2.177	812	sustantivo
102	CONVERTIR	3	2.177	815	verbo
103	CUATRO	3	2.177	818	adjetivo
104	DEBER	3	2.177	821	verbo
105	DERECHO	3	2.177	824	adjetivo
106	DISTANCIA	3	2.177	827	sustantivo
107	EQUILIBRIO	3	2.177	830	sustantivo
108	ESFÉRICO	3	2.177	833	adjetivo
109	ESTAR	3	2.177	836	verbo

ANEXO P: FRECUENCIAS DEL TEXTO TRANSFORMADO FESTESE001V-F

OL	Lema	n_i	f_i (a1000)	N_i	Categoría gramatical
110	EXTERNO	3	2.177	839	adjetivo
111	FUSIONAR	3	2.177	842	verbo
112	GENERAL	3	2.177	845	adjetivo
113	GRANDE	3	2.177	848	adjetivo
114	HASTAM	3	2.177	851	lm
115	IMAGEN	3	2.177	854	sustantivo
116	IMPORTANTE	3	2.177	857	adjetivo
117	INFERIOR	3	2.177	860	adjetivo
118	LLAMADOA	3	2.177	863	adjetivo
119	M12	3	2.177	866	sustantivo
120	MB	3	2.177	869	lm(s)
121	MENOS	3	2.177	872	adverbio
122	MV	3	2.177	875	lm(s)
123	NEBULOSA	3	2.177	878	sustantivo
124	OBJETO	3	2.177	881	sustantivo
125	OTRO	3	2.177	884	adjetivo
126	PARTIR	3	2.177	887	verbo
127	POLVO	3	2.177	890	sustantivo
128	PRESIÓN	3	2.177	893	sustantivo
129	PROPIO	3	2.177	896	adjetivo
130	PUNTO	3	2.177	899	sustantivo
131	RADIO	3	2.177	902	sustantivo
132	SISTEMA	3	2.177	905	sustantivo
133	TAL	3	2.177	908	adjetivo
134	TAMBIÉN	3	2.177	911	adverbio
135	TOTAL	3	2.177	914	adjetivo
136	TÍPICO	3	2.177	917	adjetivo
137	VEZ	3	2.177	920	sustantivo
Franja baja					
138	OCOMA7	2	1.451	922	lm
139	1	2	1.451	924	lm(adj)
140	2	2	1.451	926	lm(adj)
141	4	2	1.451	928	lm(adj)
142	6	2	1.451	930	lm(adj)
143	A6	2	1.451	932	lm
144	AHORA	2	1.451	934	adverbio
145	APARECER	2	1.451	936	verbo
146	APROXIMADOA	2	1.451	938	adjetivo
147	B	2	1.451	940	lm
148	BLANCO	2	1.451	942	adjetivo
149	BRILLANTE	2	1.451	944	adjetivo
150	C	2	1.451	946	lm(s)
151	CADA	2	1.451	948	adjetivo
152	CALCULAR	2	1.451	950	verbo
153	CALIENTE	2	1.451	952	adjetivo
154	CIENTO	2	1.451	954	adjetivo
155	COMBUSTIÓN	2	1.451	956	sustantivo

ANEXO P: FRECUENCIAS DEL TEXTO TRANSFORMADO FESTESE001V-F

OL	Lema	n_i	f_i (a1000)	N_i	Categoría gramatical
156	COMENZAR	2	1.451	958	verbo
157	COMPACTO	2	1.451	960	adjetivo
158	COMÚN	2	1.451	962	adjetivo
159	CONJUNTO	2	1.451	964	sustantivo
160	CONTENER	2	1.451	966	verbo
161	CONTRAER	2	1.451	968	verbo
162	DISTRIBUCIÓN	2	1.451	970	sustantivo
163	DIÁMETRO	2	1.451	972	sustantivo
164	DOS	2	1.451	974	adjetivo
165	ECUACIÓN	2	1.451	976	sustantivo
166	EJEMPLO	2	1.451	978	sustantivo
167	EJERCIDOA	2	1.451	980	adjetivo
168	ESTRUCTURA	2	1.451	982	sustantivo
169	EVOLUCIONAR	2	1.451	984	verbo
170	EVOLUCIÓN	2	1.451	986	sustantivo
171	EXTENDER	2	1.451	988	verbo
172	FAMOSO	2	1.451	990	adjetivo
173	FORMACIÓN	2	1.451	992	sustantivo
174	FUERA	2	1.451	994	adverbio
175	GRAVITATORIO	2	1.451	996	adjetivo
176	GRUPO	2	1.451	998	sustantivo
177	HALO	2	1.451	1000	sustantivo
178	HISTORIA	2	1.451	1002	sustantivo
179	LARGO	2	1.451	1004	adjetivo
180	LENTAMENTE	2	1.451	1006	adverbio
181	LLAMAR	2	1.451	1008	verbo
182	LUMÍNICO	2	1.451	1010	adjetivo
183	M	2	1.451	1012	lm(s)
184	MANERA	2	1.451	1014	sustantivo
185	MANTENER	2	1.451	1016	verbo
186	MAYOR	2	1.451	1018	adjetivo
187	MEDIDA1	2	1.451	1020	sustantivo
188	MEDIDOA	2	1.451	1022	adjetivo
189	MEDIR	2	1.451	1024	verbo
190	MIENTRAS	2	1.451	1026	adverbio
191	NGC	2	1.451	1028	sustantivo
192	NORMAL	2	1.451	1030	adjetivo
193	NÚMERO	2	1.451	1032	sustantivo
194	OBSERVACIÓN	2	1.451	1034	sustantivo
195	ORDEN	2	1.451	1036	sustantivo
196	PARECER	2	1.451	1038	verbo
197	PASO1	2	1.451	1040	sustantivo
198	PESAR	2	1.451	1042	verbo
199	PLANETARIO	2	1.451	1044	adjetivo
200	PLÉYADES	2	1.451	1046	sustantivo
201	PRIMERO	2	1.451	1048	adjetivo
202	PROPIEDAD	2	1.451	1050	sustantivo
203	PRUEBA1	2	1.451	1052	sustantivo

ANEXO P: FRECUENCIAS DEL TEXTO TRANSFORMADO FESTESE001V-F

OL	Lema	n_i	f_i (a1000)	N_i	Categoría gramatical
204	QUEMAR	2	1.451	1054	verbo
205	RAMA	2	1.451	1056	sustantivo
206	RARO	2	1.451	1058	adjetivo
207	RECTO	2	1.451	1060	adjetivo
208	RELATIVAMENTE	2	1.451	1062	adverbio
209	RESIDUAL	2	1.451	1064	adjetivo
210	SIGMA	2	1.451	1066	sustantivo
211	SIGUIENTE	2	1.451	1068	adjetivo
212	STEFANBOLTZMANN	2	1.451	1070	sustantivo
213	SUCEDER	2	1.451	1072	verbo
214	V	2	1.451	1074	lm(s)
215	VALIOSO	2	1.451	1076	adjetivo
216	VARIO	2	1.451	1078	adjetivo
217	VIEJO	2	1.451	1080	adjetivo
218	VISIBLE	2	1.451	1082	adjetivo
219	VÉRTICE	2	1.451	1084	sustantivo
220	~	2	1.451	1086	lm
221	ÁREA	2	1.451	1088	sustantivo
222	100	1	0.726	1089	lm(adj)
223	13	1	0.726	1090	lm(adj)
224	1764	1	0.726	1091	lm(adj)
225	1783	1	0.726	1092	lm(adj)
226	1888	1	0.726	1093	lm(adj)
227	1COMA0	1	0.726	1094	lm(adj)
228	25	1	0.726	1095	lm(adj)
229	26	1	0.726	1096	lm(adj)
230	30	1	0.726	1097	lm(adj)
231	3000	1	0.726	1098	lm(adj)
232	3COMA5	1	0.726	1099	lm(adj)
233	400	1	0.726	1100	lm(adj)
234	5	1	0.726	1101	lm(adj)
235	5COMA8	1	0.726	1102	lm(adj)
236	6218	1	0.726	1103	lm(adj)
237	6COMA7	1	0.726	1104	lm(adj)
238	7	1	0.726	1105	lm(adj)
239	74	1	0.726	1106	lm(adj)
240	99COMA3	1	0.726	1107	lm(adj)
241	ABSOLUTO	1	0.726	1108	adjetivo
242	ABUNDANCIA	1	0.726	1109	sustantivo
243	ACABAR	1	0.726	1110	verbo
244	ADECUADAMENTE	1	0.726	1111	adverbio
245	ADULTO	1	0.726	1112	adjetivo
246	AGOTAR	1	0.726	1113	verbo
247	AGRUPACIÓN	1	0.726	1114	sustantivo
248	ALREDEDOR	1	0.726	1115	adverbio
249	ALTO	1	0.726	1116	adjetivo
250	AMENOS2	1	0.726	1117	lm
251	ANTERIOR	1	0.726	1118	adjetivo

ANEXO P: FRECUENCIAS DEL TEXTO TRANSFORMADO FESTESE001V-F

OL	Lema	n_i	f_i (a1000)	N_i	Categoría gramatical
252	ANÁLISIS	1	0.726	1119	sustantivo
253	APARENTE	1	0.726	1120	adjetivo
254	ASTROFÍSICO	1	0.726	1121	adjetivo
255	ASTRONOMÍA	1	0.726	1122	sustantivo
256	ASTRÓNOMO	1	0.726	1123	sustantivo
257	ASÍ	1	0.726	1124	adverbio
258	AUMENTO1	1	0.726	1125	sustantivo
259	AVANZADOA	1	0.726	1126	adjetivo
260	BANCO	1	0.726	1127	sustantivo
261	BOLA	1	0.726	1128	sustantivo
262	BRILLO	1	0.726	1129	sustantivo
263	BRUNO	1	0.726	1130	sustantivo
264	CAMINO	1	0.726	1131	sustantivo
265	CATÁLOGO	1	0.726	1132	sustantivo
266	CENTENAR	1	0.726	1133	sustantivo
267	CENTRO	1	0.726	1134	sustantivo
268	CERCANO	1	0.726	1135	adjetivo
269	CHARLES	1	0.726	1136	sustantivo
270	CICLO	1	0.726	1137	sustantivo
271	CIELO	1	0.726	1138	sustantivo
272	CIRCUNDANTE	1	0.726	1139	adjetivo
273	CIUDAD	1	0.726	1140	sustantivo
274	CLARAMENTE	1	0.726	1141	adverbio
275	CLAVE	1	0.726	1142	sustantivo
276	COLAPSADOA	1	0.726	1143	adjetivo
277	COLAPSAR	1	0.726	1144	verbo
278	COLECCIÓN	1	0.726	1145	sustantivo
279	COLISIÓN	1	0.726	1146	sustantivo
280	COLOCAR	1	0.726	1147	verbo
281	COMPARTIR	1	0.726	1148	verbo
282	COMPLICADOA	1	0.726	1149	adjetivo
283	COMPONENTE	1	0.726	1150	sustantivo
284	COMPRENDER	1	0.726	1151	verbo
285	COMUNIDAD	1	0.726	1152	sustantivo
286	CONCENTRAR	1	0.726	1153	verbo
287	CONCISO	1	0.726	1154	adjetivo
288	CONOCER	1	0.726	1155	verbo
289	CONOCIDOA	1	0.726	1156	adjetivo
290	CONSECUENCIA	1	0.726	1157	sustantivo
291	CONSIDERABLE	1	0.726	1158	adjetivo
292	CONSIDERAR	1	0.726	1159	verbo
293	CONSIGUIENTE	1	0.726	1160	adjetivo
294	CONSTANTE1	1	0.726	1161	sustantivo
295	CONSTANTEA	1	0.726	1162	adjetivo
296	CONSTELACIÓN	1	0.726	1163	sustantivo
297	CONSTRUIR	1	0.726	1164	verbo
298	CONSUMIR	1	0.726	1165	verbo
299	CONTAMINACIÓN	1	0.726	1166	sustantivo

ANEXO P: FRECUENCIAS DEL TEXTO TRANSFORMADO FESTESE001V-F

OL	Lema	n_i	f_i (a1000)	N_i	Categoría gramatical
300	CONTINUAR	1	0.726	1167	verbo
301	CONTRACCIÓN	1	0.726	1168	sustantivo
302	CONTRARIO	1	0.726	1169	adjetivo
303	COORDENADA	1	0.726	1170	sustantivo
304	CORRELACIÓN	1	0.726	1171	sustantivo
305	CORRESPONDER	1	0.726	1172	verbo
306	CORTESÍA	1	0.726	1173	sustantivo
307	CREER	1	0.726	1174	verbo
308	CUALQUIER	1	0.726	1175	adjetivo
309	CUERPO	1	0.726	1176	sustantivo
310	CURIOSO	1	0.726	1177	adjetivo
311	DECAER	1	0.726	1178	verbo
312	DEDUCIR	1	0.726	1179	verbo
313	DEJAR	1	0.726	1180	verbo
314	DENOMINAR	1	0.726	1181	verbo
315	DENSO	1	0.726	1182	adjetivo
316	DENTRO	1	0.726	1183	adverbio
317	DEPENDER	1	0.726	1184	verbo
318	DEPOSITAR	1	0.726	1185	verbo
319	DERIVAR	1	0.726	1186	verbo
320	DESCRIBIR	1	0.726	1187	verbo
321	DESCUBRIR	1	0.726	1188	verbo
322	DESPLAZAR	1	0.726	1189	verbo
323	DESTRUIR	1	0.726	1190	verbo
324	DETERMINADOA	1	0.726	1191	adjetivo
325	DEUTERIO	1	0.726	1192	sustantivo
326	DIAGONAL	1	0.726	1193	adjetivo
327	DIFERIR	1	0.726	1194	verbo
328	DIFUSO	1	0.726	1195	adjetivo
329	DISTINTO	1	0.726	1196	adjetivo
330	DISTRIBUIR	1	0.726	1197	verbo
331	DIVIDIR	1	0.726	1198	verbo
332	DUODÉCIMO	1	0.726	1199	adjetivo
333	DÍA	1	0.726	1200	sustantivo
334	E1	1	0.726	1201	sustantivo
335	EINSTEIN	1	0.726	1202	sustantivo
336	EISENDLE	1	0.726	1203	sustantivo
337	ELEVAR	1	0.726	1204	verbo
338	EMITIDOA	1	0.726	1205	adjetivo
339	EMITIR	1	0.726	1206	verbo
340	ENANO	1	0.726	1207	adjetivo
341	ENDE	1	0.726	1208	adverbio
342	ENFRIAR	1	0.726	1209	verbo
343	ENORME	1	0.726	1210	adjetivo
344	ENTONCES	1	0.726	1211	adverbio
345	ERROR	1	0.726	1212	sustantivo
346	ESA1	1	0.726	1213	sustantivo
347	ESCALA1	1	0.726	1214	sustantivo

ANEXO P: FRECUENCIAS DEL TEXTO TRANSFORMADO FESTESE001V-F

OL	Lema	n_i	f_i (a1000)	N_i	Categoría gramatical
348	ESO1	1	0.726	1215	sustantivo
349	ESPACIO	1	0.726	1216	sustantivo
350	ESPECÍFICO	1	0.726	1217	adjetivo
351	ESTADO	1	0.726	1218	sustantivo
352	ESTIMACIÓN	1	0.726	1219	sustantivo
353	ESTIMAR	1	0.726	1220	verbo
354	ESTUDIADOA	1	0.726	1221	adjetivo
355	ESTUDIAR	1	0.726	1222	verbo
356	ESTUDIO1	1	0.726	1223	sustantivo
357	ESTÁNDARA	1	0.726	1224	adjetivo
358	ETAPA	1	0.726	1225	sustantivo
359	EXISTIR	1	0.726	1226	verbo
360	EXPANDIR	1	0.726	1227	verbo
361	EXPLICAR	1	0.726	1228	verbo
362	EXPLOTAR	1	0.726	1229	verbo
363	EXPULSAR	1	0.726	1230	verbo
364	EXTRAER	1	0.726	1231	verbo
365	FIGURA	1	0.726	1232	sustantivo
366	FILTRO	1	0.726	1233	sustantivo
367	FINAL	1	0.726	1234	sustantivo
368	FINALA	1	0.726	1235	adjetivo
369	FINALMENTE	1	0.726	1236	adverbio
370	FORMIDABLE	1	0.726	1237	adjetivo
371	FRACCIÓN	1	0.726	1238	sustantivo
372	FRENTE	1	0.726	1239	adverbio
373	FUERZA	1	0.726	1240	sustantivo
374	FUNCIÓN	1	0.726	1241	sustantivo
375	FÍSICO	1	0.726	1242	adjetivo
376	GENERAR	1	0.726	1243	verbo
377	GIGANTESCO	1	0.726	1244	adjetivo
378	GIRO1	1	0.726	1245	sustantivo
379	GRADO	1	0.726	1246	sustantivo
380	GRAVITATORIAMENTE	1	0.726	1247	adverbio
381	GRÁFICO	1	0.726	1248	sustantivo
382	HABER	1	0.726	1249	verbo
383	HECHO1	1	0.726	1250	sustantivo
384	HERSCHEL	1	0.726	1251	sustantivo
385	HERTZSPRUNGRUSSELL	1	0.726	1252	sustantivo
386	HIPPARCOS	1	0.726	1253	sustantivo
387	HORA	1	0.726	1254	sustantivo
388	HORIZONTAL	1	0.726	1255	adjetivo
389	HOY	1	0.726	1256	adverbio
390	HUECO	1	0.726	1257	sustantivo
391	IDEAL	1	0.726	1258	adjetivo
392	ILUSTRACIÓN	1	0.726	1259	sustantivo
393	ILUSTRAR	1	0.726	1260	verbo
394	IMPLICADOA	1	0.726	1261	adjetivo
395	IMPLICAR	1	0.726	1262	verbo

ANEXO P: FRECUENCIAS DEL TEXTO TRANSFORMADO FESTESE001V-F

OL	Lema	n_i	f_i (a1000)	N_i	Categoría gramatical
396	INCREMENTAR	1	0.726	1263	verbo
397	INDICADOA	1	0.726	1264	adjetivo
398	INDICAR	1	0.726	1265	verbo
399	INDIVIDUAL	1	0.726	1266	adjetivo
400	INFORMACIÓN	1	0.726	1267	sustantivo
401	INTERACCIÓN	1	0.726	1268	sustantivo
402	INTERNACIONAL	1	0.726	1269	adjetivo
403	INTERNO	1	0.726	1270	adjetivo
404	INTERÉS	1	0.726	1271	sustantivo
405	INVESTIGACIÓN	1	0.726	1272	sustantivo
406	INVOLUCRADOA	1	0.726	1273	adjetivo
407	IR	1	0.726	1274	verbo
408	IZQUIERDO	1	0.726	1275	adjetivo
409	JOVEN	1	0.726	1276	adjetivo
410	JUSTO	1	0.726	1277	adjetivo
411	LEY	1	0.726	1278	sustantivo
412	LIGADOA	1	0.726	1279	adjetivo
413	LIGERO	1	0.726	1280	adjetivo
414	LISTA	1	0.726	1281	sustantivo
415	LOCALIZAR	1	0.726	1282	verbo
416	LONGEVO	1	0.726	1283	adjetivo
417	LONGITUD	1	0.726	1284	sustantivo
418	LÍMITE	1	0.726	1285	sustantivo
419	MASIVO	1	0.726	1286	adjetivo
420	MATERIA	1	0.726	1287	sustantivo
421	MATERIAL	1	0.726	1288	sustantivo
422	MAYORITARIAMENTE	1	0.726	1289	adverbio
423	MENUDO	1	0.726	1290	adjetivo
424	MEZCLA1	1	0.726	1291	sustantivo
425	MITAD	1	0.726	1292	sustantivo
426	MODESTO	1	0.726	1293	adjetivo
427	MODO	1	0.726	1294	sustantivo
428	MOVER	1	0.726	1295	verbo
429	MUESTRA1	1	0.726	1296	sustantivo
430	NACIMIENTO	1	0.726	1297	sustantivo
431	NEGRO	1	0.726	1298	adjetivo
432	NINGUNO	1	0.726	1299	adjetivo
433	NUBE	1	0.726	1300	sustantivo
434	NUCLEAR	1	0.726	1301	adjetivo
435	OBVIAMENTE	1	0.726	1302	adverbio
436	OFIUCO	1	0.726	1303	sustantivo
437	ONDA	1	0.726	1304	sustantivo
438	ORIGINAL	1	0.726	1305	adjetivo
439	PARSEC	1	0.726	1306	sustantivo
440	PARTICULAR	1	0.726	1307	adjetivo
441	PASAR	1	0.726	1308	verbo
442	PATRÓN	1	0.726	1309	sustantivo
443	PERMANECER	1	0.726	1310	verbo

ANEXO P: FRECUENCIAS DEL TEXTO TRANSFORMADO FESTESE001V-F

OL	Lema	n_i	f_i (a1000)	N_i	Categoría gramatical
444	PERSPECTIVA	1	0.726	1311	sustantivo
445	PI	1	0.726	1312	sustantivo
446	PISTA	1	0.726	1313	sustantivo
447	PLANO	1	0.726	1314	sustantivo
448	PODER1	1	0.726	1315	sustantivo
449	POLUCIÓN	1	0.726	1316	sustantivo
450	POSIBLE	1	0.726	1317	adjetivo
451	PRECISOA	1	0.726	1318	adjetivo
452	PRISMÁTICO	1	0.726	1319	adjetivo
453	PROBABLE	1	0.726	1320	adjetivo
454	PROBABLEMENTE	1	0.726	1321	adverbio
455	PRODUCIR	1	0.726	1322	verbo
456	PROGRAMA	1	0.726	1323	sustantivo
457	PRONTO	1	0.726	1324	adverbio
458	PROPORCIONAL	1	0.726	1325	adjetivo
459	PROPORCIONAR	1	0.726	1326	verbo
460	PROTOGALÁCTICO	1	0.726	1327	adjetivo
461	PROTÓN	1	0.726	1328	sustantivo
462	PROVENIR	1	0.726	1329	verbo
463	PRÓXIMO	1	0.726	1330	adjetivo
464	PUBLICAR	1	0.726	1331	verbo
465	PÁGINA	1	0.726	1332	sustantivo
466	QUEDAR	1	0.726	1333	verbo
467	QUÍMICO	1	0.726	1334	adjetivo
468	RAINER	1	0.726	1335	sustantivo
469	RANGO	1	0.726	1336	sustantivo
470	RAZÓN	1	0.726	1337	sustantivo
471	RECIBIR	1	0.726	1338	verbo
472	RELACIÓN	1	0.726	1339	sustantivo
473	REMANENTE	1	0.726	1340	sustantivo
474	REPETIDOA	1	0.726	1341	adjetivo
475	REPETIR	1	0.726	1342	verbo
476	RESERVA1	1	0.726	1343	sustantivo
477	RESOLUCIÓN	1	0.726	1344	sustantivo
478	RESOLVER	1	0.726	1345	verbo
479	RESTANTE	1	0.726	1346	adjetivo
480	RESTO	1	0.726	1347	sustantivo
481	RESULTADO1	1	0.726	1348	sustantivo
482	SEIS	1	0.726	1349	adjetivo
483	SERIE	1	0.726	1350	sustantivo
484	SIMPLE	1	0.726	1351	adjetivo
485	SOLO	1	0.726	1352	adjetivo
486	STAMPFER	1	0.726	1353	sustantivo
487	SUBPRODUCTO	1	0.726	1354	sustantivo
488	SUCESO	1	0.726	1355	sustantivo
489	SUMINISTRO1	1	0.726	1356	sustantivo
490	SUPERFICIE	1	0.726	1357	sustantivo
491	SUPERNOVA	1	0.726	1358	sustantivo

ANEXO P: FRECUENCIAS DEL TEXTO TRANSFORMADO FESTESE001V-F

OL	Lema	n_i	f_i (a1000)	N_i	Categoría gramatical
492	SUPERVIVIENTE	1	0.726	1359	adjetivo
493	SUPUESTO1	1	0.726	1360	sustantivo
494	TAMAÑO	1	0.726	1361	sustantivo
495	TELESCOPIO	1	0.726	1362	sustantivo
496	TEORÍA	1	0.726	1363	sustantivo
497	TERMONUCLEAR	1	0.726	1364	adjetivo
498	TIERRA	1	0.726	1365	sustantivo
499	TRANSFORMAR	1	0.726	1366	verbo
500	TRAVÉS	1	0.726	1367	sustantivo
501	TÉRMICO	1	0.726	1368	adjetivo
502	TÉRMINO	1	0.726	1369	sustantivo
503	UNIVERSO	1	0.726	1370	sustantivo
504	VECINDAD	1	0.726	1371	sustantivo
505	VIOLENTO	1	0.726	1372	adjetivo
506	VISTA	1	0.726	1373	sustantivo
507	WILLIAM	1	0.726	1374	sustantivo
508	.	1	0.726	1375	lm
509	ÁTOMO	1	0.726	1376	sustantivo
510	ÍNDICE	1	0.726	1377	sustantivo
511	ÚTIL	1	0.726	1378	adjetivo

Nota. OL= orden lexicométrico; n= frecuencia absoluta; f= frecuencia relativa; N= frecuencia acumulada.

ANEXO Q: Frecuencias del Texto Transformado FESTESE001V-S

Listado de las frecuencias del texto transformado de FESTESE001V-S obtenidas con el programa PAFE. En el orden léxicométrico aparecen al principio los números y símbolos y al final las palabras que comienzan por vocales con tilde. En la columna de la categoría gramatical, el lenguaje matemático se denota con “lm” y entre paréntesis se ha añadido, cuando estaba claro, si además correspondía con otro tipo de palabra léxica (“s”= sustantivo, “adj”= adjetivo, “v”=verbo). Los verbos auxiliares se señalan como “f-verbo”. La tabla se encuentra dividida en las tres franjas señaladas en la metodología: alta, media y baja.

Tabla 100

Frecuencias del texto transformado FESTESE001V-S

OL	Lema	n_i	f_i (a1000)	N_i	Categoría gramatical
Franja alta					
1	ESTRELLA	80	56.697	80	sustantivo
2	CÚMULO	41	29.057	121	sustantivo
3	SER	41	29.057	162	verbo
4	GLOBULAR	23	16.3	185	adjetivo
5	TEMPERATURA	21	14.883	206	sustantivo
6	LUMINOSIDAD	19	13.466	225	sustantivo
7	NÚCLEO	19	13.466	244	sustantivo
8	MÁS	17	12.048	261	adverbio
9	SUPERFICIAL	15	10.631	276	adjetivo
Franja media					
10	MASA	14	9.922	290	sustantivo
11	MUY	13	9.213	303	adverbio
12	PRINCIPAL	13	9.213	316	adjetivo
13	SECUENCIA	13	9.213	329	sustantivo

ANEXO Q: FRECUENCIAS DEL TEXTO TRANSFORMADO FESTESE001V-S

OL	Lema	n _i	f _i (a1000)	N _i	Categoría gramatical
14	X	13	9.213	342	lm(v)
15	APROXIMADAMENTE	12	8.505	354	adverbio
16	HIDRÓGENO	12	8.505	366	sustantivo
17	MAGNITUD	12	8.505	378	sustantivo
18	TENER	12	8.505	390	verbo
19	10	11	7.796	401	lm(adj)
20	VER	11	7.796	412	verbo
21	AÑO	10	7.087	422	sustantivo
22	DIAGRAMA	10	7.087	432	sustantivo
23	HR	10	7.087	442	sustantivo
24	COLOR	9	6.378	451	sustantivo
25	FORMA1	9	6.378	460	sustantivo
26	POCO	9	6.378	469	adjetivo
27	PROCESO	9	6.378	478	sustantivo
28	TODO	9	6.378	487	adjetivo
29	ENCONTRAR	8	5.67	495	verbo
30	MILLÓN	8	5.67	503	sustantivo
31	NO	8	5.67	511	adverbio
32	ABIERTOA	7	4.961	518	adjetivo
33	DIFERENTE	7	4.961	525	adjetivo
34	EDAD	7	4.961	532	sustantivo
35	ENERGÍA	7	4.961	539	sustantivo
36	FORMAR	7	4.961	546	verbo
37	FUSIÓN	7	4.961	553	sustantivo
38	LUZ	7	4.961	560	sustantivo
39	MESSIER	7	4.961	567	sustantivo
40	PORCIENTO	7	4.961	574	lm
41	ROJO	7	4.961	581	adjetivo
42	SOL	7	4.961	588	sustantivo
43	VIDA	7	4.961	595	sustantivo
44	ÍNDICE	7	4.961	602	sustantivo
45	1	6	4.252	608	lm(adj)
46	12	6	4.252	614	lm(adj)
47	APARENTE	6	4.252	620	adjetivo
48	GALAXIA	6	4.252	626	sustantivo
49	GIGANTE	6	4.252	632	adjetivo
50	LÁCTEO	6	4.252	638	adjetivo
51	MAYORÍA	6	4.252	644	sustantivo
52	MIL	6	4.252	650	adjetivo
53	RADIO	6	4.252	656	sustantivo
54	VÍA	6	4.252	662	sustantivo
55	4	5	3.544	667	lm(adj)
56	=	5	3.544	672	lm(adj)
57	ASTRONÓMICO	5	3.544	677	adjetivo
58	B	5	3.544	682	lm
59	BAJOA	5	3.544	687	adjetivo
60	DETERMINAR	5	3.544	692	verbo
61	EJERCICIO	5	3.544	697	sustantivo

ANEXO Q: FRECUENCIAS DEL TEXTO TRANSFORMADO FESTESE001V-S

OL	Lema	n_i	f_i (a1000)	N_i	Categoría gramatical
62	ELEMENTO	5	3.544	702	sustantivo
63	FIGURA	5	3.544	707	sustantivo
64	GAS	5	3.544	712	sustantivo
65	HACER	5	3.544	717	verbo
66	PESADOA	5	3.544	722	adjetivo
67	SOLAR	5	3.544	727	adjetivo
68	SÓLO	5	3.544	732	adverbio
69	USAR	5	3.544	737	verbo
70	V	5	3.544	742	lm
71	VALOR	5	3.544	747	sustantivo
72	YA	5	3.544	752	adverbio
73	2	4	2.835	756	lm(adj)
74	3	4	2.835	760	lm(adj)
75	A2	4	2.835	764	lm
76	A3	4	2.835	768	lm
77	A4	4	2.835	772	lm
78	CANTIDAD	4	2.835	776	sustantivo
79	CAPA	4	2.835	780	sustantivo
80	COMPONER	4	2.835	784	verbo
81	DAR	4	2.835	788	verbo
82	DISCO	4	2.835	792	sustantivo
83	ESTABLE	4	2.835	796	adjetivo
84	ESTELAR	4	2.835	800	adjetivo
85	FASE	4	2.835	804	sustantivo
86	HELIO	4	2.835	808	sustantivo
87	HERRAMIENTA	4	2.835	812	sustantivo
88	INCLUSO	4	2.835	816	adverbio
89	K	4	2.835	820	lm(s)
90	LUGAR	4	2.835	824	sustantivo
91	LÍNEA	4	2.835	828	sustantivo
92	MOSTRAR	4	2.835	832	verbo
93	MUCHO	4	2.835	836	adjetivo
94	PARTE	4	2.835	840	sustantivo
95	PEQUEÑO	4	2.835	844	adjetivo
96	RADIACIÓN	4	2.835	848	sustantivo
97	REACCIÓN	4	2.835	852	sustantivo
98	SECCIÓN	4	2.835	856	sustantivo
99	SUPERIOR	4	2.835	860	adjetivo
100	TANTO	4	2.835	864	adjetivo
101	TIEMPO	4	2.835	868	sustantivo
102	TIPO	4	2.835	872	sustantivo
Franja baja					
103	ALGUNO	3	2.126	875	adjetivo
104	CENTRAL	3	2.126	878	adjetivo
105	CONVERTIR	3	2.126	881	verbo
106	DEBER	3	2.126	884	verbo
107	DERECHO	3	2.126	887	adjetivo

ANEXO Q: FRECUENCIAS DEL TEXTO TRANSFORMADO FESTESE001V-S

OL	Lema	n_i	f_i (a1000)	N_i	Categoría gramatical
108	DISTANCIA	3	2.126	890	sustantivo
109	EQUILIBRIO	3	2.126	893	sustantivo
110	ESFÉRICO	3	2.126	896	adjetivo
111	ESTAR	3	2.126	899	verbo
112	EXTERNO	3	2.126	902	adjetivo
113	FUSIONAR	3	2.126	905	verbo
114	GRANDE	3	2.126	908	adjetivo
115	HASTAM	3	2.126	911	lm
116	IMAGEN	3	2.126	914	sustantivo
117	IMPORTANTE	3	2.126	917	adjetivo
118	INFERIOR	3	2.126	920	adjetivo
119	LLAMADOA	3	2.126	923	adjetivo
120	MENOS	3	2.126	926	adverbio
121	NEBULOSA	3	2.126	929	sustantivo
122	NUEVO	3	2.126	932	adjetivo
123	OBJETO	3	2.126	935	sustantivo
124	OTRO	3	2.126	938	adjetivo
125	PARTIR	3	2.126	941	verbo
126	PLÉYADES	3	2.126	944	sustantivo
127	POLVO	3	2.126	947	sustantivo
128	PRESIÓN	3	2.126	950	sustantivo
129	PROPIO	3	2.126	953	adjetivo
130	PUNTO	3	2.126	956	sustantivo
131	SISTEMA	3	2.126	959	sustantivo
132	TAL	3	2.126	962	adjetivo
133	TAMBIÉN	3	2.126	965	adverbio
134	TOTAL	3	2.126	968	adjetivo
135	TÍPICO	3	2.126	971	adjetivo
136	VEZ	3	2.126	974	sustantivo
137	0COMA7	2	1.417	976	lm(adj)
138	6	2	1.417	978	lm(adj)
139	A6	2	1.417	980	lm
140	AHORA	2	1.417	982	adverbio
141	APARECER	2	1.417	984	verbo
142	APROXIMADOA	2	1.417	986	adjetivo
143	BLANCO	2	1.417	988	adjetivo
144	BRILLANTE	2	1.417	990	adjetivo
145	C	2	1.417	992	lm(s)
146	CADA	2	1.417	994	adjetivo
147	CALCULAR	2	1.417	996	verbo
148	CALIENTE	2	1.417	998	adjetivo
149	CIENTO	2	1.417	1000	adjetivo
150	COMBUSTIÓN	2	1.417	1002	sustantivo
151	COMENZAR	2	1.417	1004	verbo
152	COMPACTO	2	1.417	1006	adjetivo
153	COMÚN	2	1.417	1008	adjetivo
154	CONJUNTO	2	1.417	1010	sustantivo
155	CONTENER	2	1.417	1012	verbo

ANEXO Q: FRECUENCIAS DEL TEXTO TRANSFORMADO FESTESE001V-S

OL	Lema	n_i	f_i (a1000)	N_i	Categoría gramatical
156	CONTRAER	2	1.417	1014	verbo
157	DISTRIBUCIÓN	2	1.417	1016	sustantivo
158	DIÁMETRO	2	1.417	1018	sustantivo
159	ECUACIÓN	2	1.417	1020	sustantivo
160	EJEMPLO	2	1.417	1022	sustantivo
161	EJERCIDOA	2	1.417	1024	adjetivo
162	ESTRUCTURA	2	1.417	1026	sustantivo
163	EVOLUCIONAR	2	1.417	1028	verbo
164	EVOLUCIÓN	2	1.417	1030	sustantivo
165	EXTENDER	2	1.417	1032	verbo
166	FAMOSO	2	1.417	1034	adjetivo
167	FORMACIÓN	2	1.417	1036	sustantivo
168	FUERA	2	1.417	1038	adverbio
169	GENERAL	2	1.417	1040	adjetivo
170	GRAVITATORIO	2	1.417	1042	adjetivo
171	GRUPO	2	1.417	1044	sustantivo
172	HALO	2	1.417	1046	sustantivo
173	HISTORIA	2	1.417	1048	sustantivo
174	LARGO	2	1.417	1050	adjetivo
175	LENTAMENTE	2	1.417	1052	adverbio
176	LLAMAR	2	1.417	1054	verbo
177	LUMÍNICO	2	1.417	1056	adjetivo
178	MANERA	2	1.417	1058	sustantivo
179	MANTENER	2	1.417	1060	verbo
180	MAYOR	2	1.417	1062	adjetivo
181	MEDIDA1	2	1.417	1064	sustantivo
182	MEDIDOA	2	1.417	1066	adjetivo
183	MEDIR	2	1.417	1068	verbo
184	MIENTRAS	2	1.417	1070	adverbio
185	NGC	2	1.417	1072	sustantivo
186	NORMAL	2	1.417	1074	adjetivo
187	NÚMERO	2	1.417	1076	sustantivo
188	OBSERVACIÓN	2	1.417	1078	sustantivo
189	ORDEN	2	1.417	1080	sustantivo
190	PARECER	2	1.417	1082	verbo
191	PASO1	2	1.417	1084	sustantivo
192	PESAR	2	1.417	1086	verbo
193	PLANETARIO	2	1.417	1088	adjetivo
194	PRIMERO	2	1.417	1090	adjetivo
195	PROPIEDAD	2	1.417	1092	sustantivo
196	PRUEBA1	2	1.417	1094	sustantivo
197	QUEMAR	2	1.417	1096	verbo
198	RAMA	2	1.417	1098	sustantivo
199	RARO	2	1.417	1100	adjetivo
200	RECTO	2	1.417	1102	adjetivo
201	RELATIVAMENTE	2	1.417	1104	adverbio
202	RESIDUAL	2	1.417	1106	adjetivo
203	SIGMA	2	1.417	1108	sustantivo

ANEXO Q: FRECUENCIAS DEL TEXTO TRANSFORMADO FESTESE001V-S

OL	Lema	n_i	f_i (a1000)	N_i	Categoría gramatical
204	SIGUIENTE	2	1.417	1110	adjetivo
205	STEFANBOLTZMANN	2	1.417	1112	sustantivo
206	SUCEDER	2	1.417	1114	verbo
207	VALIOSO	2	1.417	1116	adjetivo
208	VARIO	2	1.417	1118	adjetivo
209	VIEJO	2	1.417	1120	adjetivo
210	VISIBLE	2	1.417	1122	adjetivo
211	VÉRTICE	2	1.417	1124	sustantivo
212	~	2	1.417	1126	lm
213	ÁREA	2	1.417	1128	sustantivo
214	100	1	0.709	1129	lm(adj)
215	13	1	0.709	1130	lm(adj)
216	1764	1	0.709	1131	lm(adj)
217	1783	1	0.709	1132	lm(adj)
218	1888	1	0.709	1133	lm(adj)
219	1COMA0	1	0.709	1134	lm(adj)
220	25	1	0.709	1135	lm(adj)
221	26	1	0.709	1136	lm(adj)
222	30	1	0.709	1137	lm(adj)
223	3000	1	0.709	1138	lm(adj)
224	3COMA5	1	0.709	1139	lm(adj)
225	400	1	0.709	1140	lm(adj)
226	5	1	0.709	1141	lm(adj)
227	5COMA8	1	0.709	1142	lm(adj)
228	6218	1	0.709	1143	lm(adj)
229	6COMA7	1	0.709	1144	lm(adj)
230	7	1	0.709	1145	lm(adj)
231	74	1	0.709	1146	lm(adj)
232	99COMA3	1	0.709	1147	lm(adj)
233	ABSOLUTO	1	0.709	1148	adjetivo
234	ABUNDANCIA	1	0.709	1149	sustantivo
235	ACABAR	1	0.709	1150	verbo
236	ADECUADAMENTE	1	0.709	1151	adverbio
237	ADULTO	1	0.709	1152	adjetivo
238	AGOTAR	1	0.709	1153	verbo
239	AGRUPACIÓN	1	0.709	1154	sustantivo
240	ALREDEDOR	1	0.709	1155	adverbio
241	ALTO	1	0.709	1156	adjetivo
242	AMENOS2	1	0.709	1157	lm
243	ANTERIOR	1	0.709	1158	adjetivo
244	ANÁLISIS	1	0.709	1159	sustantivo
245	ASTROFÍSICO	1	0.709	1160	adjetivo
246	ASTRONOMÍA	1	0.709	1161	sustantivo
247	ASTRÓNOMO	1	0.709	1162	sustantivo
248	ASÍ	1	0.709	1163	adverbio
249	AUMENTO1	1	0.709	1164	sustantivo
250	AVANZADOA	1	0.709	1165	adjetivo
251	BANCO	1	0.709	1166	sustantivo

ANEXO Q: FRECUENCIAS DEL TEXTO TRANSFORMADO FESTESE001V-S

OL	Lema	n_i	f_i (a1000)	N_i	Categoría gramatical
252	BOLA	1	0.709	1167	sustantivo
253	BRILLO	1	0.709	1168	sustantivo
254	BRUNO	1	0.709	1169	sustantivo
255	CAMINO	1	0.709	1170	sustantivo
256	CENTENAR	1	0.709	1171	sustantivo
257	CENTRO	1	0.709	1172	sustantivo
258	CERCANO	1	0.709	1173	adjetivo
259	CHARLES	1	0.709	1174	sustantivo
260	CICLO	1	0.709	1175	sustantivo
261	CIELO	1	0.709	1176	sustantivo
262	CIRCUNDANTE	1	0.709	1177	adjetivo
263	CIUDAD	1	0.709	1178	sustantivo
264	CLARAMENTE	1	0.709	1179	adverbio
265	CLAVE	1	0.709	1180	sustantivo
266	COLAPSADOA	1	0.709	1181	adjetivo
267	COLAPSAR	1	0.709	1182	verbo
268	COLECCIÓN	1	0.709	1183	sustantivo
269	COLISIÓN	1	0.709	1184	sustantivo
270	COLOCAR	1	0.709	1185	verbo
271	COMPARTIR	1	0.709	1186	verbo
272	COMPLICADOA	1	0.709	1187	adjetivo
273	COMPONENTE	1	0.709	1188	sustantivo
274	COMPRENDER	1	0.709	1189	verbo
275	COMUNIDAD	1	0.709	1190	sustantivo
276	CONCENTRAR	1	0.709	1191	verbo
277	CONOCER	1	0.709	1192	verbo
278	CONOCIDOA	1	0.709	1193	adjetivo
279	CONSECUENCIA	1	0.709	1194	sustantivo
280	CONSIDERABLE	1	0.709	1195	adjetivo
281	CONSIDERAR	1	0.709	1196	verbo
282	CONSIGUIENTE	1	0.709	1197	adjetivo
283	CONSTANTE1	1	0.709	1198	sustantivo
284	CONSTANTEA	1	0.709	1199	adjetivo
285	CONSTELACIÓN	1	0.709	1200	sustantivo
286	CONSTRUIR	1	0.709	1201	verbo
287	CONSUMIR	1	0.709	1202	verbo
288	CONTAMINACIÓN	1	0.709	1203	sustantivo
289	CONTINUAR	1	0.709	1204	verbo
290	CONTRACCIÓN	1	0.709	1205	sustantivo
291	CONTRARIO	1	0.709	1206	adjetivo
292	COORDENADA	1	0.709	1207	sustantivo
293	CORRELACIÓN	1	0.709	1208	sustantivo
294	CORRESPONDER	1	0.709	1209	verbo
295	CORTESÍA	1	0.709	1210	sustantivo
296	CREER	1	0.709	1211	verbo
297	CUALQUIER	1	0.709	1212	adjetivo
298	CUERPO	1	0.709	1213	sustantivo
299	CURIOSO	1	0.709	1214	adjetivo

ANEXO Q: FRECUENCIAS DEL TEXTO TRANSFORMADO FESTESE001V-S

OL	Lema	n_i	f_i (a1000)	N_i	Categoría gramatical
300	DECAER	1	0.709	1215	verbo
301	DEDUCIR	1	0.709	1216	verbo
302	DEJAR	1	0.709	1217	verbo
303	DENOMINAR	1	0.709	1218	verbo
304	DENSO	1	0.709	1219	adjetivo
305	DENTRO	1	0.709	1220	adverbio
306	DEPENDER	1	0.709	1221	verbo
307	DEPOSITAR	1	0.709	1222	verbo
308	DERIVAR	1	0.709	1223	verbo
309	DESCRIBIR	1	0.709	1224	verbo
310	DESCUBRIR	1	0.709	1225	verbo
311	DESPLAZAR	1	0.709	1226	verbo
312	DESTRUIR	1	0.709	1227	verbo
313	DETERMINADOA	1	0.709	1228	adjetivo
314	DEUTERIO	1	0.709	1229	sustantivo
315	DIAGONAL	1	0.709	1230	adjetivo
316	DIFERIR	1	0.709	1231	verbo
317	DIFUSO	1	0.709	1232	adjetivo
318	DISTINTO	1	0.709	1233	adjetivo
319	DISTRIBUIR	1	0.709	1234	verbo
320	DIVIDIR	1	0.709	1235	verbo
321	DUODÉCIMO	1	0.709	1236	adjetivo
322	DÍA	1	0.709	1237	sustantivo
323	EINSTEIN	1	0.709	1238	sustantivo
324	EISENDLE	1	0.709	1239	sustantivo
325	ELEVAR	1	0.709	1240	verbo
326	EMITIDOA	1	0.709	1241	adjetivo
327	EMITIR	1	0.709	1242	verbo
328	ENANO	1	0.709	1243	adjetivo
329	ENDE	1	0.709	1244	adverbio
330	ENFRIAR	1	0.709	1245	verbo
331	ENORME	1	0.709	1246	adjetivo
332	ENTONCES	1	0.709	1247	adverbio
333	ERROR	1	0.709	1248	sustantivo
334	ESA1	1	0.709	1249	sustantivo
335	ESCALA1	1	0.709	1250	sustantivo
336	ESO1	1	0.709	1251	sustantivo
337	ESPACIO	1	0.709	1252	sustantivo
338	ESPECÍFICO	1	0.709	1253	adjetivo
339	ESTADO	1	0.709	1254	sustantivo
340	ESTIMACIÓN	1	0.709	1255	sustantivo
341	ESTIMAR	1	0.709	1256	verbo
342	ESTUDIADOA	1	0.709	1257	adjetivo
343	ESTUDIAR	1	0.709	1258	verbo
344	ESTUDIO1	1	0.709	1259	sustantivo
345	ESTÁNDARA	1	0.709	1260	adjetivo
346	ETAPA	1	0.709	1261	sustantivo
347	EXISTIR	1	0.709	1262	verbo

ANEXO Q: FRECUENCIAS DEL TEXTO TRANSFORMADO FESTESE001V-S

OL	Lema	n_i	f_i (a1000)	N_i	Categoría gramatical
348	EXPANDIR	1	0.709	1263	verbo
349	EXPLICAR	1	0.709	1264	verbo
350	EXPLOTAR	1	0.709	1265	verbo
351	EXPULSAR	1	0.709	1266	verbo
352	EXTRAER	1	0.709	1267	verbo
353	FILTRO	1	0.709	1268	sustantivo
354	FINAL	1	0.709	1269	sustantivo
355	FINALA	1	0.709	1270	adjetivo
356	FINALMENTE	1	0.709	1271	adverbio
357	FORMIDABLE	1	0.709	1272	adjetivo
358	FRACCIÓN	1	0.709	1273	sustantivo
359	FRENTE	1	0.709	1274	adverbio
360	FUERZA	1	0.709	1275	sustantivo
361	FUNCIÓN	1	0.709	1276	sustantivo
362	FÍSICO	1	0.709	1277	adjetivo
363	GENERAR	1	0.709	1278	verbo
364	GIGANTESCO	1	0.709	1279	adjetivo
365	GIRO1	1	0.709	1280	sustantivo
366	GRADO	1	0.709	1281	sustantivo
367	GRAVITATORIAMENTE	1	0.709	1282	adverbio
368	GRÁFICO	1	0.709	1283	sustantivo
369	HABER	1	0.709	1284	verbo
370	HECHO1	1	0.709	1285	sustantivo
371	HERSCHEL	1	0.709	1286	sustantivo
372	HIPPARCOS	1	0.709	1287	sustantivo
373	HORA	1	0.709	1288	sustantivo
374	HORIZONTAL	1	0.709	1289	adjetivo
375	HOY	1	0.709	1290	adverbio
376	HUECO	1	0.709	1291	sustantivo
377	IDEAL	1	0.709	1292	adjetivo
378	ILUSTRACIÓN	1	0.709	1293	sustantivo
379	ILUSTRAR	1	0.709	1294	verbo
380	IMPLICADOA	1	0.709	1295	adjetivo
381	IMPLICAR	1	0.709	1296	verbo
382	INCREMENTAR	1	0.709	1297	verbo
383	INDICADOA	1	0.709	1298	adjetivo
384	INDICAR	1	0.709	1299	verbo
385	INDIVIDUAL	1	0.709	1300	adjetivo
386	INFORMACIÓN	1	0.709	1301	sustantivo
387	INTERACCIÓN	1	0.709	1302	sustantivo
388	INTERNACIONAL	1	0.709	1303	adjetivo
389	INTERNO	1	0.709	1304	adjetivo
390	INTERÉS	1	0.709	1305	sustantivo
391	INVESTIGACIÓN	1	0.709	1306	sustantivo
392	INVOLUCRADOA	1	0.709	1307	adjetivo
393	IR	1	0.709	1308	verbo
394	IZQUIERDO	1	0.709	1309	adjetivo
395	JOVEN	1	0.709	1310	adjetivo

ANEXO Q: FRECUENCIAS DEL TEXTO TRANSFORMADO FESTESE001V-S

OL	Lema	n_i	f_i (a1000)	N_i	Categoría gramatical
396	JUSTO	1	0.709	1311	adjetivo
397	LEY	1	0.709	1312	sustantivo
398	LIGADOA	1	0.709	1313	adjetivo
399	LIGERO	1	0.709	1314	adjetivo
400	LISTA	1	0.709	1315	sustantivo
401	LOCALIZAR	1	0.709	1316	verbo
402	LONGEVO	1	0.709	1317	adjetivo
403	LONGITUD	1	0.709	1318	sustantivo
404	LÍMITE	1	0.709	1319	sustantivo
405	MASIVO	1	0.709	1320	adjetivo
406	MATERIA	1	0.709	1321	sustantivo
407	MATERIAL	1	0.709	1322	sustantivo
408	MAYORITARIAMENTE	1	0.709	1323	adverbio
409	MENUDO	1	0.709	1324	adjetivo
410	MEZCLA1	1	0.709	1325	sustantivo
411	MITAD	1	0.709	1326	sustantivo
412	MODESTO	1	0.709	1327	adjetivo
413	MODO	1	0.709	1328	sustantivo
414	MOVER	1	0.709	1329	verbo
415	MUESTRA1	1	0.709	1330	sustantivo
416	NACIMIENTO	1	0.709	1331	sustantivo
417	NEGRO	1	0.709	1332	adjetivo
418	NINGUNO	1	0.709	1333	adjetivo
419	NUBE	1	0.709	1334	sustantivo
420	NUCLEAR	1	0.709	1335	adjetivo
421	OBVIAMENTE	1	0.709	1336	adverbio
422	OFIUCO	1	0.709	1337	sustantivo
423	ONDA	1	0.709	1338	sustantivo
424	ORIGINAL	1	0.709	1339	adjetivo
425	PARSEC	1	0.709	1340	sustantivo
426	PARTICULAR	1	0.709	1341	adjetivo
427	PASAR	1	0.709	1342	verbo
428	PATRÓN	1	0.709	1343	sustantivo
429	PERMANECER	1	0.709	1344	verbo
430	PERSPECTIVA	1	0.709	1345	sustantivo
431	PI	1	0.709	1346	sustantivo
432	PISTA	1	0.709	1347	sustantivo
433	PLANO	1	0.709	1348	sustantivo
434	PODER1	1	0.709	1349	sustantivo
435	POLUCIÓN	1	0.709	1350	sustantivo
436	POSIBLE	1	0.709	1351	adjetivo
437	PRECISOA	1	0.709	1352	adjetivo
438	PRISMÁTICO	1	0.709	1353	adjetivo
439	PROBABLE	1	0.709	1354	adjetivo
440	PROBABLEMENTE	1	0.709	1355	adverbio
441	PRODUCIR	1	0.709	1356	verbo
442	PROGRAMA	1	0.709	1357	sustantivo
443	PRONTO	1	0.709	1358	adverbio

ANEXO Q: FRECUENCIAS DEL TEXTO TRANSFORMADO FESTESE001V-S

OL	Lema	n_i	f_i (a1000)	N_i	Categoría gramatical
444	PROPORCIONAL	1	0.709	1359	adjetivo
445	PROPORCIONAR	1	0.709	1360	verbo
446	PROTOGALÁCTICO	1	0.709	1361	adjetivo
447	PROTÓN	1	0.709	1362	sustantivo
448	PROVENIR	1	0.709	1363	verbo
449	PRÓXIMO	1	0.709	1364	adjetivo
450	PUBLICAR	1	0.709	1365	verbo
451	PÁGINA	1	0.709	1366	sustantivo
452	QUEDAR	1	0.709	1367	verbo
453	QUÍMICO	1	0.709	1368	adjetivo
454	RAINER	1	0.709	1369	sustantivo
455	RANGO	1	0.709	1370	sustantivo
456	RAZÓN	1	0.709	1371	sustantivo
457	RECIBIR	1	0.709	1372	verbo
458	RELACIÓN	1	0.709	1373	sustantivo
459	REMANENTE	1	0.709	1374	sustantivo
460	REPETIDOA	1	0.709	1375	adjetivo
461	REPETIR	1	0.709	1376	verbo
462	RESERVA1	1	0.709	1377	sustantivo
463	RESOLUCIÓN	1	0.709	1378	sustantivo
464	RESOLVER	1	0.709	1379	verbo
465	RESTANTE	1	0.709	1380	adjetivo
466	RESTO	1	0.709	1381	sustantivo
467	RESULTADO1	1	0.709	1382	sustantivo
468	SEIS	1	0.709	1383	adjetivo
469	SERIE	1	0.709	1384	sustantivo
470	SIMPLE	1	0.709	1385	adjetivo
471	SOLO	1	0.709	1386	adjetivo
472	STAMPFER	1	0.709	1387	sustantivo
473	SUBPRODUCTO	1	0.709	1388	sustantivo
474	SUCESO	1	0.709	1389	sustantivo
475	SUMINISTRO1	1	0.709	1390	sustantivo
476	SUPERFICIE	1	0.709	1391	sustantivo
477	SUPERNOVA	1	0.709	1392	sustantivo
478	SUPERVIVIENTE	1	0.709	1393	adjetivo
479	SUPUESTO1	1	0.709	1394	sustantivo
480	TAMAÑO	1	0.709	1395	sustantivo
481	TELESCOPIO	1	0.709	1396	sustantivo
482	TEORÍA	1	0.709	1397	sustantivo
483	TERMONUCLEAR	1	0.709	1398	adjetivo
484	TIERRA	1	0.709	1399	sustantivo
485	TRANSFORMAR	1	0.709	1400	verbo
486	TRAVÉS	1	0.709	1401	sustantivo
487	TÉRMICO	1	0.709	1402	adjetivo
488	TÉRMINO	1	0.709	1403	sustantivo
489	UNIVERSO	1	0.709	1404	sustantivo
490	VECINDAD	1	0.709	1405	sustantivo
491	VIOLENTO	1	0.709	1406	adjetivo

ANEXO Q: FRECUENCIAS DEL TEXTO TRANSFORMADO FESTESE001V-S

OL	Lema	n_i	f_i (a1000)	N_i	Categoría gramatical
492	VISTA	1	0.709	1407	sustantivo
493	WILLIAM	1	0.709	1408	sustantivo
494	.	1	0.709	1409	lm
495	ÁTOMO	1	0.709	1410	sustantivo
496	ÚTIL	1	0.709	1411	adjetivo

Nota. OL= orden lexicométrico; n= frecuencia absoluta; f= frecuencia relativa; N= frecuencia acumulada.

ANEXO R: Frecuencias del Texto Transformado FESTESE001VD

Listado de las frecuencias del texto transformado obtenidas con el programa PAFE. En el orden léxicométrico aparecen al principio los números y símbolos y al final las palabras que comienzan por vocales con tilde. En la columna de la categoría gramatical, el lenguaje matemático se denota con “lm” y entre paréntesis se ha añadido, cuando estaba claro, si además correspondía con otro tipo de palabra léxica (“s”= sustantivo, “adj”= adjetivo, “v”=verbo). Los verbos auxiliares se señalan como “f-verbo”. La tabla se encuentra dividida en las tres franjas señaladas en la metodología: alta, media y baja.

Tabla 101

Frecuencias del texto transformado FESTESE001VD

OL	ULL	n_i	f_i (a 1000)	N_i	Categoría gramatical
		Fanja media			
1	EL	265	120.949	265	funcional
2	DE	191	87.175	456	funcional
3	ESTRELLA	74	33.775	530	sustantivo
4	UNO	72	32.862	602	funcional
5	EN	64	29.21	666	funcional
6	SE	54	24.646	720	funcional
7	Y	46	20.995	766	funcional
8	QUE	37	16.887	803	funcional
9	SER	37	16.887	840	verbo
10	DEL	28	12.78	868	funcional
11	ESTO	26	11.867	894	funcional
12	A	23	10.497	917	funcional
13	SU	19	8.672	936	funcional
14	CÚMULO	17	7.759	953	sustantivo
15	CÚMULOGLOBULAR	17	7.759	970	sust-adj
16	PARA	17	7.759	987	funcional
17	NÚCLEO	16	7.303	1003	sustantivo

ANEXO R: FRECUENCIAS DEL TEXTO TRANSFORMADO FESTESE001VD

OL	ULL	n_i	f_i (a 1000)	N_i	Categoría gramatical
18	POR	16	7.303	1019	funcional
19	MÁS	15	6.846	1034	adverbio
20	PODERX	14	6.39	1048	f-verbo
21	MUY	13	5.933	1061	adverbio
22	SECUENCIAPRINCIPAL	13	5.933	1074	sust-adj
23	COMO	11	5.021	1085	funcional
24	LUMINOSIDAD	11	5.021	1096	sustantivo
25	VER	11	5.021	1107	verbo
26	APROXIMADAMENTE	10	4.564	1117	adverbio
27	FORMA1	10	4.564	1127	sustantivo
28	HIDRÓGENO	10	4.564	1137	sustantivo
29	TENER	10	4.564	1147	verbo
30	10	9	4.108	1156	lm(adj)
31	POCO	9	4.108	1165	adjetivo
32	PROCESO	9	4.108	1174	sustantivo
33	TODO	9	4.108	1183	adjetivo
34	CON	8	3.651	1191	funcional
35	ENCONTRAR	8	3.651	1199	verbo
36	L	8	3.651	1207	lm(s)
37	MASA	8	3.651	1215	sustantivo
38	DIFERENTE	7	3.195	1222	adjetivo
39	EDAD	7	3.195	1229	sustantivo
40	ENTRE	7	3.195	1236	funcional
41	FORMAR	7	3.195	1243	verbo
42	FUSIÓN	7	3.195	1250	sustantivo
43	HABERX	7	3.195	1257	f-verbo
44	SOL	7	3.195	1264	sustantivo
45	VIDA	7	3.195	1271	sustantivo
46	X	7	3.195	1278	lm(v)
47	CUANDO	6	2.738	1284	funcional
48	DIAGRAMA	6	2.738	1290	sustantivo
49	GALAXIA	6	2.738	1296	sustantivo
50	GIGANTEROJO	6	2.738	1302	adj-adj
51	MAYORÍA	6	2.738	1308	sustantivo
52	MIL	6	2.738	1314	adjetivo
53	MILLÓNDEAÑO	6	2.738	1320	sust-sust
54	O	6	2.738	1326	funcional
55	T	6	2.738	1332	lm(s)
Franja baja					
56	=	5	2.282	1337	lm(adj)
57	CÚMULOABIERTOA	5	2.282	1342	sust-adj
58	DETERMINAR	5	2.282	1347	verbo
59	DIAGRAMAHR	5	2.282	1352	sust-sust
60	EJERCICIO	5	2.282	1357	sustantivo
61	GAS	5	2.282	1362	sustantivo
62	HACER	5	2.282	1367	verbo
63	NO	5	2.282	1372	adverbio
64	SÓLO	5	2.282	1377	adverbio
65	TEMPERATURA	5	2.282	1382	sustantivo
66	USAR	5	2.282	1387	verbo
67	VALOR	5	2.282	1392	sustantivo

ANEXO R: FRECUENCIAS DEL TEXTO TRANSFORMADO FESTESE001VD

OL	ULL	n_i	f_i (a 1000)	N_i	Categoría gramatical
68	VÍALÁCTEO	5	2.282	1397	sust-adj
69	AÑOLUZ	4	1.826	1401	sust-sust
70	COMPONER	4	1.826	1405	verbo
71	DESDE	4	1.826	1409	funcional
72	DISCO	4	1.826	1413	sustantivo
73	DONDE	4	1.826	1417	funcional
74	DURANTE	4	1.826	1421	funcional
75	ENERGÍA	4	1.826	1425	sustantivo
76	ESTABLE	4	1.826	1429	adjetivo
77	ESTELAR	4	1.826	1433	adjetivo
78	FASE	4	1.826	1437	sustantivo
79	INCLUSO	4	1.826	1441	adverbio
80	LUGAR	4	1.826	1445	sustantivo
81	LÍNEA	4	1.826	1449	sustantivo
82	MOSTRAR	4	1.826	1453	verbo
83	PEQUEÑO	4	1.826	1457	adjetivo
84	PORCIENODE	4	1.826	1461	lm
85	R	4	1.826	1465	lm(s)
86	RADIACIÓN	4	1.826	1469	sustantivo
87	REACCIÓN	4	1.826	1473	sustantivo
88	SECHERRASTRONÓMICO	4	1.826	1477	sust-sust-adj
89	SOBRE	4	1.826	1481	funcional
90	TANTO	4	1.826	1485	adj
91	TEMPERATURASUPERFICIAL	4	1.826	1489	sust-adj
92	TEMPERATURASUPERFICIALT	4	1.826	1493	sust-adj- lm(s)
93	TIPO	4	1.826	1497	sustantivo
94	UNOA	4	1.826	1501	adjetivo
95	ALGUNO	3	1.369	1504	adjetivo
96	BAJOAMASA	3	1.369	1507	adj-sust
97	CAPAEXTERNO	3	1.369	1510	sust-adj
98	COLOR	3	1.369	1513	sustantivo
99	CONVERTIR	3	1.369	1516	verbo
100	CUAL	3	1.369	1519	funcional
101	CUATRONÚCLEO	3	1.369	1522	adj-sust
102	DEBERA	3	1.369	1525	verbo
103	DISTANCIA	3	1.369	1528	sustantivo
104	EQUILIBRIO	3	1.369	1531	sustantivo
105	ESFÉRICO	3	1.369	1534	adjetivo
106	ESTAR	3	1.369	1537	verbo
107	FUSIONAR	3	1.369	1540	verbo
108	GENERAL	3	1.369	1543	adjetivo
109	GRANDE	3	1.369	1546	adjetivo
110	HASTAM	3	1.369	1549	lm
111	IMAGEN	3	1.369	1552	sustantivo
112	IMPORTANTE	3	1.369	1555	adjetivo
113	INFERIOR	3	1.369	1558	adjetivo
114	LLAMADOA	3	1.369	1561	adjetivo
115	LSOLAR	3	1.369	1564	lm-adj
116	LUZ	3	1.369	1567	sustantivo
117	M12	3	1.369	1570	sustantivo
118	MAGNITUD	3	1.369	1573	sustantivo
119	MB	3	1.369	1576	lm(s)

ANEXO R: FRECUENCIAS DEL TEXTO TRANSFORMADO FESTESE001VD

OL	ULL	n_i	f_i (a 1000)	N_i	Categoría gramatical
120	MENOS	3	1.369	1579	adverbio
121	MESSIER	3	1.369	1582	sustantivo
122	MV	3	1.369	1585	lm(s)
123	NUESTRO	3	1.369	1588	funcional
124	OBJETO	3	1.369	1591	sustantivo
125	OTRO	3	1.369	1594	adjetivo
126	PARTIRDE	3	1.369	1597	verbo
127	PESADOA	3	1.369	1600	adjetivo
128	POLVO	3	1.369	1603	sustantivo
129	PRESIÓN	3	1.369	1606	sustantivo
130	PROPIO	3	1.369	1609	adjetivo
131	PUNTO	3	1.369	1612	sustantivo
132	RADIO	3	1.369	1615	sustantivo
133	SISTEMA	3	1.369	1618	sustantivo
134	TAL	3	1.369	1621	adjetivo
135	TAMBIÉN	3	1.369	1624	adverbio
136	TOTAL	3	1.369	1627	adjetivo
137	VEZ	3	1.369	1630	sustantivo
138	0COMA7PCIENTORESIDUALDLMASA	2	0.913	1632	lm-lm-adj- sust
139	1	2	0.913	1634	lm(adj)
140	12	2	0.913	1636	lm(adj)
141	2	2	0.913	1638	lm(adj)
142	3	2	0.913	1640	lm(adj)
143	3X10	2	0.913	1642	lm-lm-lm
144	4X	2	0.913	1644	lm-lm
145	A2X	2	0.913	1646	lm-lm
146	A3	2	0.913	1648	lm
147	A4	2	0.913	1650	lm
148	A6	2	0.913	1652	lm
149	AHORA	2	0.913	1654	adverbio
150	AL	2	0.913	1656	funcional
151	APARECER	2	0.913	1658	verbo
152	APROXIMADOA	2	0.913	1660	adjetivo
153	B	2	0.913	1662	lm
154	BAJOA	2	0.913	1664	adjetivo
155	BLANCOY	2	0.913	1666	adjetivo
156	BRILLANTE	2	0.913	1668	adjetivo
157	CA2	2	0.913	1670	lm-lm
158	CADA	2	0.913	1672	adjetivo
159	CALCULAR	2	0.913	1674	verbo
160	CALIENTE	2	0.913	1676	adjetivo
161	CANTIDADDE	2	0.913	1678	sustantivo
162	CANTIDADDEENERGÍA	2	0.913	1680	sust-sust
163	CIENTODE	2	0.913	1682	adjetivo
164	COMBUSTIÓNDELHIDRÓGENO	2	0.913	1684	sust-sust
165	COMENZAR	2	0.913	1686	verbo
166	COMENZARXA	2	0.913	1688	f-verbo
167	COMPACTO	2	0.913	1690	adjetivo
168	COMÚN	2	0.913	1692	adjetivo
169	CONJUNTO	2	0.913	1694	sustantivo
170	CONTENER	2	0.913	1696	verbo
171	CONTRAER	2	0.913	1698	verbo

ANEXO R: FRECUENCIAS DEL TEXTO TRANSFORMADO FESTESE001VD

OL	ULL	n_i	f_i (a 1000)	N_i	Categoría gramatical
172	DAR	2	0.913	1700	verbo
173	DISTRIBUCIÓN	2	0.913	1702	sustantivo
174	DIÁMETROTÍPICODEUNO	2	0.913	1704	sust-adj
175	DOS	2	0.913	1706	adjetivo
176	E	2	0.913	1708	funcional
177	ECUACIÓN	2	0.913	1710	sustantivo
178	EJEMPLO	2	0.913	1712	sustantivo
179	EJERCIDOAHACIA	2	0.913	1714	adjetivo
180	ELEMENTO	2	0.913	1716	sustantivo
181	ELEMENTOMÁSPESADOA	2	0.913	1718	sust-adv-adj
182	ESTARX	2	0.913	1720	f-verbo
183	ESTRUCTURA	2	0.913	1722	sustantivo
184	EVOLUCIONAR	2	0.913	1724	verbo
185	EVOLUCIÓN	2	0.913	1726	sustantivo
186	EXTENDERAPROXIMADAMENTE	2	0.913	1728	v-adv
187	FAMOSO	2	0.913	1730	adjetivo
188	FIG	2	0.913	1732	sustantivo
189	FIG6	2	0.913	1734	sust-lm
190	FORMACIÓN	2	0.913	1736	sustantivo
191	FUERA	2	0.913	1738	adverbio
192	GRAVITATORIO	2	0.913	1740	adjetivo
193	GRUPO	2	0.913	1742	sustantivo
194	HACIAEL	2	0.913	1744	funcional
195	HALO	2	0.913	1746	sustantivo
196	HASTA	2	0.913	1748	funcional
197	HELIO	2	0.913	1750	sustantivo
198	HELIOSER	2	0.913	1752	sust-v
199	HISTORIA	2	0.913	1754	sustantivo
200	IRX	2	0.913	1756	f-verbo
201	LARGODE	2	0.913	1758	adjetivo
202	LENTAMENTE	2	0.913	1760	adverbio
203	LLAMAR	2	0.913	1762	verbo
204	LUMÍNICO	2	0.913	1764	adjetivo
205	M	2	0.913	1766	lm(s)
206	MAGNITUDVISIBLE	2	0.913	1768	sust-adj
207	MANERAQUE	2	0.913	1770	sustantivo
208	MANTENER	2	0.913	1772	verbo
209	MAYOR	2	0.913	1774	adjetivo
210	MEDIDA1	2	0.913	1776	sustantivo
211	MEDIDOA	2	0.913	1778	adjetivo
212	MEDIREL	2	0.913	1780	verbo
213	MESSIER12	2	0.913	1782	sust-lm
214	MIENTRASEL	2	0.913	1784	adverbio
215	MILLÓNDE	2	0.913	1786	sustantivo
216	MUCHO	2	0.913	1788	adjetivo
217	MUCHOTIEMPO	2	0.913	1790	adj-sust
218	NEBULOSAPLANETARIO	2	0.913	1792	sust-adj
219	NGC	2	0.913	1794	sustantivo
220	NORMAL	2	0.913	1796	adjetivo
221	NOSDAR	2	0.913	1798	verbo
222	NUESTROGALAXIA	2	0.913	1800	sustantivo
223	NUEVO	2	0.913	1802	adjetivo
224	NUEVOESTRELLA	2	0.913	1804	adj-sust

ANEXO R: FRECUENCIAS DEL TEXTO TRANSFORMADO FESTESE001VD

OL	ULL	n_i	f_i (a 1000)	N_i	Categoría gramatical
225	NÚMERO	2	0.913	1806	sustantivo
226	OBSERVACIÓN	2	0.913	1808	sustantivo
227	ORDENDEMAGNITUD	2	0.913	1810	sust-sust
228	PARECER	2	0.913	1812	verbo
229	PARTE	2	0.913	1814	sustantivo
230	PARTESUPERIORDERECHO	2	0.913	1816	sust-adj-adj
231	PASO1	2	0.913	1818	sustantivo
232	PESAR	2	0.913	1820	verbo
233	PLÉYADES	2	0.913	1822	sustantivo
234	PRIMERO	2	0.913	1824	adjetivo
235	PROPIEDAD	2	0.913	1826	sustantivo
236	PRUEBA1	2	0.913	1828	sustantivo
237	QUEMAR	2	0.913	1830	verbo
238	RAMADE	2	0.913	1832	sustantivo
239	RARO	2	0.913	1834	adjetivo
240	RECTO	2	0.913	1836	adjetivo
241	RELATIVAMENTE	2	0.913	1838	adverbio
242	SERX	2	0.913	1840	f-verbo
243	SIGMA	2	0.913	1842	sustantivo
244	SIGUIENTE	2	0.913	1844	adjetivo
245	SINO	2	0.913	1846	funcional
246	SOLAR	2	0.913	1848	adjetivo
247	STEFANBOLTZMANN	2	0.913	1850	sustantivo
248	SUCEDER	2	0.913	1852	verbo
249	SUPERIOR	2	0.913	1854	adjetivo
250	TEMPERATURACENTRAL	2	0.913	1856	sust-adj
251	TIEMPO	2	0.913	1858	sustantivo
252	TU	2	0.913	1860	funcional
253	V	2	0.913	1862	lm(s)
254	VALIOSO	2	0.913	1864	adjetivo
255	VARIO	2	0.913	1866	adjetivo
256	VIEJO	2	0.913	1868	adjetivo
257	VÉRTICE	2	0.913	1870	sustantivo
258	YA	2	0.913	1872	adverbio
259	YANOSER	2	0.913	1874	adv-adv-v
260	~	2	0.913	1876	lm
261	ÁREA	2	0.913	1878	sustantivo
262	ÉL	2	0.913	1880	funcional
263	100	1	0.456	1881	lm(adj)
264	13	1	0.456	1882	lm(adj)
265	1764	1	0.456	1883	lm(adj)
266	1783	1	0.456	1884	lm(adj)
267	1888	1	0.456	1885	lm(adj)
268	1COMA0	1	0.456	1886	lm(adj)
269	25	1	0.456	1887	lm(adj)
270	26	1	0.456	1888	lm(adj)
271	30	1	0.456	1889	lm(adj)
272	3000	1	0.456	1890	lm(adj)
273	3COMA5	1	0.456	1891	lm(adj)
274	400	1	0.456	1892	lm(adj)
275	5	1	0.456	1893	lm(adj)
276	5COMA8	1	0.456	1894	lm(adj)
277	6218	1	0.456	1895	lm(adj)

ANEXO R: FRECUENCIAS DEL TEXTO TRANSFORMADO FESTESE001VD

OL	ULL	n_i	f_i (a 1000)	N_i	Categoría gramatical
278	6COMA7	1	0.456	1896	lm(adj)
279	7	1	0.456	1897	lm(adj)
280	74	1	0.456	1898	lm(adj)
281	99COMA3	1	0.456	1899	lm(adj)
282	A3K	1	0.456	1900	lm-lm
283	A3KTENER	1	0.456	1901	lm-lm-v
284	A4K	1	0.456	1902	lm-lm
285	A4KTENER	1	0.456	1903	lm-lm-v
286	ABSOLUTO	1	0.456	1904	adjetivo
287	ABUNDANCIA	1	0.456	1905	sustantivo
288	ACABAR	1	0.456	1906	verbo
289	ADECUADAMENTE	1	0.456	1907	adverbio
290	ADULTO	1	0.456	1908	adjetivo
291	AGOTAR	1	0.456	1909	verbo
292	AGRUPACIÓN	1	0.456	1910	sustantivo
293	ALREDEDOR	1	0.456	1911	adverbio
294	ALTO	1	0.456	1912	adjetivo
295	AMENOS2	1	0.456	1913	lm
296	ANTERIOR	1	0.456	1914	adjetivo
297	ANÁLISIS	1	0.456	1915	sustantivo
298	APARENTE	1	0.456	1916	adjetivo
299	ASTROFÍSICO	1	0.456	1917	adjetivo
300	ASTRONOMÍA	1	0.456	1918	sustantivo
301	ASTRONÓMICO	1	0.456	1919	adjetivo
302	ASTRÓNOMO	1	0.456	1920	sustantivo
303	ASÍ	1	0.456	1921	adverbio
304	AUMENTO1	1	0.456	1922	sustantivo
305	AVANZADOA	1	0.456	1923	adjetivo
306	BANCO	1	0.456	1924	sustantivo
307	BOLA	1	0.456	1925	sustantivo
308	BRILLO	1	0.456	1926	sustantivo
309	BRUNO	1	0.456	1927	sustantivo
310	CAMINO	1	0.456	1928	sustantivo
311	CAPA	1	0.456	1929	sustantivo
312	CATÁLOGO	1	0.456	1930	sustantivo
313	CENTENAR	1	0.456	1931	sustantivo
314	CENTRAL	1	0.456	1932	adjetivo
315	CENTRO	1	0.456	1933	sustantivo
316	CERCANO	1	0.456	1934	adjetivo
317	CHARLES	1	0.456	1935	sustantivo
318	CICLO	1	0.456	1936	sustantivo
319	CIELO	1	0.456	1937	sustantivo
320	CIRCUNDANTE	1	0.456	1938	adjetivo
321	CIUDAD	1	0.456	1939	sustantivo
322	CLARAMENTE	1	0.456	1940	adverbio
323	CLAVE	1	0.456	1941	sustantivo
324	COLAPSADOA	1	0.456	1942	adjetivo
325	COLAPSAR	1	0.456	1943	verbo
326	COLECCIÓN	1	0.456	1944	sustantivo
327	COLISIÓN	1	0.456	1945	sustantivo
328	COLOCAR	1	0.456	1946	verbo
329	COMPARTIR	1	0.456	1947	verbo
330	COMPLICADOA	1	0.456	1948	adjetivo

ANEXO R: FRECUENCIAS DEL TEXTO TRANSFORMADO FESTESE001VD

OL	ULL	n_i	f_i (a 1000)	N_i	Categoría gramatical
331	COMPONENTE	1	0.456	1949	sustantivo
332	COMPRENDER	1	0.456	1950	verbo
333	COMUNIDAD	1	0.456	1951	sustantivo
334	CONCENTRAR	1	0.456	1952	verbo
335	CONCISO	1	0.456	1953	adjetivo
336	CONOCER	1	0.456	1954	verbo
337	CONOCIDO	1	0.456	1955	adjetivo
338	CONSECUENCIA	1	0.456	1956	sustantivo
339	CONSIDERABLE	1	0.456	1957	adjetivo
340	CONSIDERAR	1	0.456	1958	verbo
341	CONSIGUIENTE	1	0.456	1959	adjetivo
342	CONSTANTE	1	0.456	1960	sustantivo
343	CONSTANTE	1	0.456	1961	adjetivo
344	CONSTELACIÓN	1	0.456	1962	sustantivo
345	CONSTRUIR	1	0.456	1963	verbo
346	CONSUMIR	1	0.456	1964	verbo
347	CONTAMINACIÓN	1	0.456	1965	sustantivo
348	CONTINUAR	1	0.456	1966	verbo
349	CONTRACCIÓN	1	0.456	1967	sustantivo
350	CONTRARIO	1	0.456	1968	adjetivo
351	COORDENADA	1	0.456	1969	sustantivo
352	CORRELACIÓN	1	0.456	1970	sustantivo
353	CORRESPONDER	1	0.456	1971	verbo
354	CORTESÍA	1	0.456	1972	sustantivo
355	CREER	1	0.456	1973	verbo
356	CUALQUIER	1	0.456	1974	adjetivo
357	CUERPO	1	0.456	1975	sustantivo
358	CURIOSO	1	0.456	1976	adjetivo
359	DECAER	1	0.456	1977	verbo
360	DEDUCIR	1	0.456	1978	verbo
361	DEJAR	1	0.456	1979	verbo
362	DENOMINAR	1	0.456	1980	verbo
363	DENSO	1	0.456	1981	adjetivo
364	DENTRO	1	0.456	1982	adverbio
365	DEPENDER	1	0.456	1983	verbo
366	DEPOSITAR	1	0.456	1984	verbo
367	DERECHO	1	0.456	1985	adjetivo
368	DERIVAR	1	0.456	1986	verbo
369	DESCRIBIR	1	0.456	1987	verbo
370	DESCUBRIR	1	0.456	1988	verbo
371	DESPLAZAR	1	0.456	1989	verbo
372	DESTRUIR	1	0.456	1990	verbo
373	DETERMINADO	1	0.456	1991	adjetivo
374	DEUTERIO	1	0.456	1992	sustantivo
375	DIAGONAL	1	0.456	1993	adjetivo
376	DIFERIR	1	0.456	1994	verbo
377	DIFUSO	1	0.456	1995	adjetivo
378	DISTINTO	1	0.456	1996	adjetivo
379	DISTRIBUIR	1	0.456	1997	verbo
380	DIVIDIR	1	0.456	1998	verbo
381	DUODÉCIMO	1	0.456	1999	adjetivo
382	DÍA	1	0.456	2000	sustantivo
383	E1	1	0.456	2001	sustantivo

ANEXO R: FRECUENCIAS DEL TEXTO TRANSFORMADO FESTESE001VD

OL	ULL	n_i	f_i (a 1000)	N_i	Categoría gramatical
384	EINSTEIN	1	0.456	2002	sustantivo
385	EISENDLE	1	0.456	2003	sustantivo
386	ELEMENTOMÁS	1	0.456	2004	sust-adv
387	ELEVAR	1	0.456	2005	verbo
388	ELLO	1	0.456	2006	funcional
389	EMITIDOA	1	0.456	2007	adjetivo
390	EMITIR	1	0.456	2008	verbo
391	ENANO	1	0.456	2009	adjetivo
392	ENDE	1	0.456	2010	adverbio
393	ENFRIAR	1	0.456	2011	verbo
394	ENORME	1	0.456	2012	adjetivo
395	ENTONCES	1	0.456	2013	adverbio
396	ERROR	1	0.456	2014	sustantivo
397	ESA1	1	0.456	2015	sustantivo
398	ESCALA1	1	0.456	2016	sustantivo
399	ESO1	1	0.456	2017	sustantivo
400	ESPACIO	1	0.456	2018	sustantivo
401	ESPECÍFICO	1	0.456	2019	adjetivo
402	ESTADO	1	0.456	2020	sustantivo
403	ESTIMACIÓN	1	0.456	2021	sustantivo
404	ESTIMAR	1	0.456	2022	verbo
405	ESTUDIADOA	1	0.456	2023	adjetivo
406	ESTUDIAR	1	0.456	2024	verbo
407	ESTUDIO1	1	0.456	2025	sustantivo
408	ESTÁNDARA	1	0.456	2026	adjetivo
409	ETAPA	1	0.456	2027	sustantivo
410	EXISTIR	1	0.456	2028	verbo
411	EXPANDIR	1	0.456	2029	verbo
412	EXPLICAR	1	0.456	2030	verbo
413	EXPLOTAR	1	0.456	2031	verbo
414	EXPULSAR	1	0.456	2032	verbo
415	EXTRAER	1	0.456	2033	verbo
416	FIGURA	1	0.456	2034	sustantivo
417	FILTRO	1	0.456	2035	sustantivo
418	FINAL	1	0.456	2036	sustantivo
419	FINALA	1	0.456	2037	adjetivo
420	FINALMENTE	1	0.456	2038	adverbio
421	FORMIDABLE	1	0.456	2039	adjetivo
422	FRACCIÓN	1	0.456	2040	sustantivo
423	FRENTE	1	0.456	2041	adverbio
424	FUERZA	1	0.456	2042	sustantivo
425	FUNCIÓN	1	0.456	2043	sustantivo
426	FÍSICO	1	0.456	2044	adjetivo
427	GENERAR	1	0.456	2045	verbo
428	GIGANTESCO	1	0.456	2046	adjetivo
429	GIRO1	1	0.456	2047	sustantivo
430	GRADO	1	0.456	2048	sustantivo
431	GRAVITATORIAMENTE	1	0.456	2049	adverbio
432	GRÁFICO	1	0.456	2050	sustantivo
433	HABER	1	0.456	2051	verbo
434	HECHO1	1	0.456	2052	sustantivo
435	HERSCHEL	1	0.456	2053	sustantivo
436	HERTZSPRUNGRUSSELL	1	0.456	2054	sustantivo

ANEXO R: FRECUENCIAS DEL TEXTO TRANSFORMADO FESTESE001VD

OL	ULL	n_i	f_i (a 1000)	N_i	Categoría gramatical
437	HIPPARCOS	1	0.456	2055	sustantivo
438	HORA	1	0.456	2056	sustantivo
439	HORIZONTAL	1	0.456	2057	adjetivo
440	HOY	1	0.456	2058	adverbio
441	HUECO	1	0.456	2059	sustantivo
442	IDEAL	1	0.456	2060	adjetivo
443	ILUSTRACIÓN	1	0.456	2061	sustantivo
444	ILUSTRAR	1	0.456	2062	verbo
445	IMPLICADOA	1	0.456	2063	adjetivo
446	IMPLICAR	1	0.456	2064	verbo
447	INCREMENTAR	1	0.456	2065	verbo
448	INDICADOA	1	0.456	2066	adjetivo
449	INDICAR	1	0.456	2067	verbo
450	INDIVIDUAL	1	0.456	2068	adjetivo
451	INFORMACIÓN	1	0.456	2069	sustantivo
452	INTERACCIÓN	1	0.456	2070	sustantivo
453	INTERNACIONAL	1	0.456	2071	adjetivo
454	INTERNO	1	0.456	2072	adjetivo
455	INTERÉS	1	0.456	2073	sustantivo
456	INVESTIGACIÓN	1	0.456	2074	sustantivo
457	INVOLUCRADOA	1	0.456	2075	adjetivo
458	IR	1	0.456	2076	verbo
459	IZQUIERDO	1	0.456	2077	adjetivo
460	JOVEN	1	0.456	2078	adjetivo
461	JUSTO	1	0.456	2079	adjetivo
462	LEY	1	0.456	2080	sustantivo
463	LIGADOA	1	0.456	2081	adjetivo
464	LIGERO	1	0.456	2082	adjetivo
465	LISTA	1	0.456	2083	sustantivo
466	LLEGARX	1	0.456	2084	f-verbo
467	LOCALIZAR	1	0.456	2085	verbo
468	LONGEVO	1	0.456	2086	adjetivo
469	LONGITUD	1	0.456	2087	sustantivo
470	LUEGO	1	0.456	2088	funcional
471	LÍMITE	1	0.456	2089	sustantivo
472	MASIVO	1	0.456	2090	adjetivo
473	MATERIA	1	0.456	2091	sustantivo
474	MATERIAL	1	0.456	2092	sustantivo
475	MAYORITARIAMENTE	1	0.456	2093	adverbio
476	MENUDO	1	0.456	2094	adjetivo
477	MEZCLA1	1	0.456	2095	sustantivo
478	MITAD	1	0.456	2096	sustantivo
479	MODESTO	1	0.456	2097	adjetivo
480	MODO	1	0.456	2098	sustantivo
481	MOVER	1	0.456	2099	verbo
482	MUESTRA1	1	0.456	2100	sustantivo
483	NACIMIENTO	1	0.456	2101	sustantivo
484	NEBULOSA	1	0.456	2102	sustantivo
485	NEGRO	1	0.456	2103	adjetivo
486	NINGUNO	1	0.456	2104	adjetivo
487	NUBE	1	0.456	2105	sustantivo
488	NUCLEAR	1	0.456	2106	adjetivo
489	OBVIAMENTE	1	0.456	2107	adverbio

ANEXO R: FRECUENCIAS DEL TEXTO TRANSFORMADO FESTESE001VD

OL	ULL	n_i	f_i (a 1000)	N_i	Categoría gramatical
490	OFIUCO	1	0.456	2108	sustantivo
491	ONDA	1	0.456	2109	sustantivo
492	ORIGINAL	1	0.456	2110	adjetivo
493	PARSEC	1	0.456	2111	sustantivo
494	PARTICULAR	1	0.456	2112	adjetivo
495	PASAR	1	0.456	2113	verbo
496	PATRÓN	1	0.456	2114	sustantivo
497	PERMANECER	1	0.456	2115	verbo
498	PERO	1	0.456	2116	funcional
499	PERSPECTIVA	1	0.456	2117	sustantivo
500	PI	1	0.456	2118	sustantivo
501	PISTA	1	0.456	2119	sustantivo
502	PLANO	1	0.456	2120	sustantivo
503	PODER1	1	0.456	2121	sustantivo
504	POLUCIÓN	1	0.456	2122	sustantivo
505	PORCIENTO	1	0.456	2123	lm
506	PORQUE	1	0.456	2124	funcional
507	POSIBLE	1	0.456	2125	adjetivo
508	PRECISOA	1	0.456	2126	adjetivo
509	PRISMÁTICO	1	0.456	2127	adjetivo
510	PROBABLE	1	0.456	2128	adjetivo
511	PROBABLEMENTE	1	0.456	2129	adverbio
512	PRODUCIR	1	0.456	2130	verbo
513	PROGRAMA	1	0.456	2131	sustantivo
514	PRONTO	1	0.456	2132	adverbio
515	PROPORCIONAL	1	0.456	2133	adjetivo
516	PROPORCIONAR	1	0.456	2134	verbo
517	PROTOGALÁCTICO	1	0.456	2135	adjetivo
518	PROTÓN	1	0.456	2136	sustantivo
519	PROVENIR	1	0.456	2137	verbo
520	PRÓXIMO	1	0.456	2138	adjetivo
521	PUBLICAR	1	0.456	2139	verbo
522	PÁGINA	1	0.456	2140	sustantivo
523	QUEDAR	1	0.456	2141	verbo
524	QUÍMICO	1	0.456	2142	adjetivo
525	RAINER	1	0.456	2143	sustantivo
526	RANGO	1	0.456	2144	sustantivo
527	RAZÓN	1	0.456	2145	sustantivo
528	RECIBIR	1	0.456	2146	verbo
529	RELACIÓN	1	0.456	2147	sustantivo
530	REMANENTE	1	0.456	2148	sustantivo
531	REPETIDOA	1	0.456	2149	adjetivo
532	REPETIR	1	0.456	2150	verbo
533	RESERVA1	1	0.456	2151	sustantivo
534	RESOLUCIÓN	1	0.456	2152	sustantivo
535	RESOLVER	1	0.456	2153	verbo
536	RESTANTE	1	0.456	2154	adjetivo
537	RESTO	1	0.456	2155	sustantivo
538	RESULTADO1	1	0.456	2156	sustantivo
539	ROJO	1	0.456	2157	adjetivo
540	SEIS	1	0.456	2158	adjetivo
541	SERIE	1	0.456	2159	sustantivo
542	SIMPLE	1	0.456	2160	adjetivo

ANEXO R: FRECUENCIAS DEL TEXTO TRANSFORMADO FESTESE001VD

OL	ULL	n_i	f_i (a 1000)	N_i	Categoría gramatical
543	SIN	1	0.456	2161	funcional
544	SOLO	1	0.456	2162	adjetivo
545	STAMPFER	1	0.456	2163	sustantivo
546	SUBPRODUCTO	1	0.456	2164	sustantivo
547	SUCESO	1	0.456	2165	sustantivo
548	SUMINISTRO1	1	0.456	2166	sustantivo
549	SUPERFICIE	1	0.456	2167	sustantivo
550	SUPERNOVA	1	0.456	2168	sustantivo
551	SUPERVIVIENTE	1	0.456	2169	adjetivo
552	SUPUESTO1	1	0.456	2170	sustantivo
553	SÍ	1	0.456	2171	funcional
554	TAMAÑO	1	0.456	2172	sustantivo
555	TELESCOPIO	1	0.456	2173	sustantivo
556	TEORÍA	1	0.456	2174	sustantivo
557	TERMONUCLEAR	1	0.456	2175	adjetivo
558	TIERRA	1	0.456	2176	sustantivo
559	TRANSFORMAR	1	0.456	2177	verbo
560	TRAVÉS	1	0.456	2178	sustantivo
561	TÉRMICO	1	0.456	2179	adjetivo
562	TÉRMINO	1	0.456	2180	sustantivo
563	TÍPICO	1	0.456	2181	adjetivo
564	UNIVERSO	1	0.456	2182	sustantivo
565	VECINDAD	1	0.456	2183	sustantivo
566	VIOLENTO	1	0.456	2184	adjetivo
567	VISTA	1	0.456	2185	sustantivo
568	WILLIAM	1	0.456	2186	sustantivo
569	YANO	1	0.456	2187	adv-adv
570	.	1	0.456	2188	lm
571	ÁTOMO	1	0.456	2189	sustantivo
572	ÍNDICE	1	0.456	2190	sustantivo
573	ÚTIL	1	0.456	2191	adjetivo

Nota. OL= orden lexicométrico; n = frecuencia absoluta; f = frecuencia relativa; N = frecuencia acumulada.

