DETERMINACIÓN DEL VOCABULARIO USUAL DE LOS ALUMNOS GRANADINOS DE PRIMER CURSO DE EE.MM. UTILIDAD PARA LA ENSEÑANZA DEL ESPAÑOL

Tesis doctoral realizada por Jorge Molero Huertas,
dirigida por el doctor D. Juan Martínez Marín
catedrático de Filología Española de la Universidad
de Granada

A mi padre y a Piedad

AGRADECIMIENTOS 9

INTRODUCCIÓN 11

CAPÍTULO I : MARCO TEÓRICO 21

1. MARCO TEÓRICO 23

1.1. LINGÜÍSTICA APLICADA 23

1.2. LINGÜÍSTICA ESTADÍSTICA 27

1.3. ESTADÍSTICA LÉXICA 28

1.3.1. Léxico y vocabulario 29

1.3.2. La norma léxica 33

1.3.3. Extensión del vocabulario: ley de crecimiento del vocabulario y riqueza léxica 34

1.3.4. Distribución de frecuencias 44

1.4. CLASES DE VOCABULARIO 49

1.4.1. Vocabulario usual, común, básico, fundamental 50

1.4.2. Vocabulario disponible 55

1.5. ESTUDIOS SOBRE EL VOCABULARIO DESDE PERSPECTIVAS CUANTITATIVAS. BREVE VISIÓN HISTÓRICA. 60

1.5.1. Vocabularios de carácter general 60

1.5.2. Vocabularios relativos a la población infantil o juvenil 73

1.5.3. Vocabularios en lengua española 79

1.5.4. Vocabularios escolares en España 90

1.6. CONCLUSIÓN 107

CAPÍTULO II: METODOLOGÍA 113

2. METODOLOGÍA 115

2.1. DETERMINACIÓN DEL TEMA DE TRABAJO 117

2.2. OBJETIVOS 120

2.3. POBLACIÓN 121
2.5. DETERMINACIÓN DE LA MUESTRA 123
2.6. MÉTODO DE MUESTREO 128
2.6.1. Tipo de enseñanza: BUP, FP, Reforma (LOGSE) *130*
2.6.2. Tipo de centro educativo *130*
2.6.3. Variante zona: urbana y rural *131*
2.6.4. Nivel educativo, tipo de centro y zona *131*
2.7. CONSIDERACIONES PREVIAS SOBRE LA RIQUEZA LÉXICA Y LA LOCUACIDAD DE LOS ALUMNOS QUE INTEGRAN LOS DISTINTOS ESTRATOS 137
2.7.1. BUP, FP, Reforma (LOGSE) *137*
2.7.2. Zona urbana, zona rural *137*
2.7.3. Centro público y centro privado *138*
2.7.4. Variante de sexo: alumnos, alumnas *139*
2.8. CORPUS O UNIVERSO LÉXICO 143
2.9. RECOGIDA DE DATOS 148
2.10. INTRODUCCIÓN DE DATOS EN EL PROCESO INFORMÁTICO 151
2.11. LA NORMA LÉXICA DE NUESTRA INVESTIGACIÓN 152
2.11.1. Consideraciones previas: Palabras gramaticales y palabras léxicas *158*
2.11.2. Problemas de N y de V *162*
2.11.3. Palabra como unidad de texto: palabra, ocurrencia, token *163*
2.11.4. La palabra como unidad de léxico: Lema, vocablo, type *174*
2.12. PROCESO DE RECUENTO DE PALABRAS Y VOCABLOS 202
2.13. CODIFICACIÓN 204
2.14. CONCLUSIÓN 205

CAPÍTULO III: DATOS CUANTITATIVOS Y CUALITATIVOS 210

3. DATOS CUANTITATIVOS Y CUALITATIVOS DE NUESTRA INVESTIGACIÓN 212

3.1. LOCUACIDAD Y RIQUEZA LÉXICA 213
3.1.1. Conclusión *226*
3.2. LA DISTRIBUCIÓN DE FRECUENCIAS DE NUESTRO RECUENTO. LEY DE ZIPF 230
3.2.1. Conclusión *238*
3.3. OTROS DATOS CUANTITATIVOS Y CUALITATIVOS. CLASES GRAMATICALES 239
3.4. BREVE ANÁLISIS DEL VOCABULARIO DE LOS ALUMNOS 241
3.4.1. Origen *243*

3.4.2. Formación de palabras: Derivación, composición, parasíntesis *244*
3.4.3. El uso *262*
3.4.4. Unidades fraseológicas *274*
3.4.5. Otros datos *284*
3.4.6. Conclusión *290*

CAPÍTULO IV: ENSEÑANZA DE LA LENGUA **293**

4. UTILIDAD PARA LA ENSEÑANZA DE LA LENGUA **295**

4.1. LA ENSEÑANZA DE LA LENGUA MATERNA 299
4.2. PLANES DE ESTUDIO 310
4.2.1. BUP *310*
4.2.2. Formación Profesional *313*
4.2.3. Reforma de las EE.MM. *314*
4.2.4. La LOGSE: Enseñanza Secundaria Obligatoria *316*
4.3. LA ENSEÑANZA DEL VOCABULARIO 321
4.4. PRINCIPIOS DIDÁCTICOS Y OBJETIVOS DE LA ENSEÑANZA DEL VOCABULARIO 327
4.5. SELECCIÓN DEL VOCABULARIO QUE DEBE ENSEÑARSE 330
4.6. LA FINALIDAD PEDAGÓGICA DE LOS TRABAJOS DE FRECUENCIAS 333
4.7. PLANIFICACIÓN Y PROGRAMACIÓN DEL VOCABULARIO, PARTIENDO DE LOS TRABAJOS DE FRECUENCIAS 338
4.8. UTILIDAD DIDÁCTICA DE NUESTRA INVESTIGACIÓN: PLANIFICACIÓN DE LA ENSEÑANZA DEL VOCABULARIO 344
4.9. MÉTODOS DE APRENDIZAJE DEL VOCABULARIO 367
4.9.1. Aprendizaje del vocabulario a través del contexto *369*
4.9.2. Ortografía y vocabulario usual. Familias de palabras. *373*
4.9.3. Asociaciones léxico-semánticas: campos semánticos, "redes semánticas", sistema de "constelaciones", mapas semánticos *374*
4.9.4. Sinonimia y Antonimia. Variedades diastráticas *380*
4.9.5. Derivación y transformaciones. Familias de palabras *382*
4.9.6. Diccionario y vocabulario *384*
4.9.7. Otros métodos *384*
4.10. EVALUACIÓN 385
4.11. CONCLUSIÓN 388

CAPÍTULO V: CONCLUSIONES GENERALES **391**

5. CONCLUSIONES GENERALES 393

CAPÍTULO VI: PRESENTACIÓN DEL VOCABULARIO 405

6. ORIENTACIONES SOBRE LA PRESENTACIÓN DE LAS LISTAS 407

BIBLIOGRAFÍA 415

ANEXO I 465

ANEXO II 1579

Agradecimientos

Para llevar a cabo la recogida de datos de nuestra investigación tuvimos que visitar numerosos centros de la provincia de Granada. En ellos contactamos con muchos profesores y profesoras de Lengua española que atendieron gentilmente nuestras peticiones. Sin su colaboración no habríamos podido iniciar este trabajo. Queremos expresar así nuestro agradecimiento a Andrés Morell, Ana Mª Linares, Antonio Abril, Antonio Enrique, Benicia Reyes, Cándida Fernández, Carmen Gallegos, Cecilia Morales, Fany López, Felipe Tomás, Francisco Caballero, Humberto, Javier Botrán, José Aguilera, José Luis Martín, Lourdes Blanco, Luis de la Rosa, Mª Dolores Molina, Mª Dolores Pélaez, Mª Dolores Raya, Mª Elena Romero, Mª Victoria González, Marisol Rodríguez, Mercedes López, Miguel Pertíñez, Nieves Lasa, Pastora Ochando, Pío Pélaez, Rosalía, Roberto Rodríguez, Santiago Peinado.

En este sentido, queremos agradecer la ayuda desinteresada de mis compañeros y amigos, y también profesores de Lengua, Fina y Diego: ambos me han prestado una colaboración excepcional. Igualmente mis compañeros de matemáticas Antonia y Leopoldo.

No podemos olvidar tampoco el asesoramiento técnico. El mundo de la informática es un enigma para los que, aun interesados en el lenguaje, desconocen el código mecánico de los ordenadores. De ahí nuestro agradecimiento al profesor de la Universidad de Granada, Vicente Sabido, y a

mi compañera Pilar Revelles, quien amigablemente me ha facilitado operaciones importantes de base de datos.

Pero ante todo hemos de expresar un agradecimiento especial a nuestro amigo Emilio Genaro. Pacientemente ha escuchado nuestras indicaciones y ha ido elaborando los programas informáticos necesarios e imprescindibles para el presente estudio.

INTRODUCCIÓN

Introducción

Después de algunos años de docencia, sentimos la inquietud natural de buscar alguna orientación pedagógica para hacer más efectiva nuestra labor. Atrapados, como muchos compañeros, en las redes subyugantes de la gramática, descuidando así las posibilidades del léxico, decidimos acudir a varios cursos sobre metodología de la enseñanza de la lengua y otros aspectos del español. Uno de ellos fue coordinado por Juan Martínez Marín. Los cursos nos llamaron la atención, sobre todo, por la asombrosa claridad y concreción de sus planteamientos: por primera vez veíamos reflejadas unas orientaciones rigurosas y al mismo tiempo viables para planificar eficazmente la enseñanza. Algunas de las ideas aparecen recogidas en el libro *La enseñanza de la lengua materna*, de Humberto López Morales, quien participó como ponente. En los capítulos de este libro dedicados al léxico, se pone de manifiesto la necesidad de conocer el vocabulario que manejan los escolares como referencia imprescindible para el proceso de enseñanza-aprendizaje. En este sentido, el impulso definitivo nos lo proporcionó el propio Martínez Marín, interesado por el estudio del léxico en general y quien tenía a su cargo un grupo de investigadores analizando los distintos aspectos de este nivel de la lengua. De ahí nos vino la idea, pero necesitábamos un instrumento eficaz, una guía consistente para dar los primeros pasos. La solución nos la aportó un estudio, realizado en nuestra provincia, al que debemos mucho: *El vocabulario usual del niño en el Ciclo Inicial y el Ciclo Medio de la E.G.B.*, de Fernando Justicia. Este trabajo se centra en alumnos de educación básica, pero sus postulados son rigurosos y la metodología ofrece criterios muy precisos, por lo que pueden aplicarse en investigaciones sobre enseñanza secundaria.

Así iniciamos nuestra investigación. Con ella pretendemos determinar el vocabulario usual de los alumnos granadinos de Primero de EE.MM. y Tercero de Reforma (anticipación de la Enseñanza Secundaria Obligatoria -E.S.O.-). Elegimos estos cursos porque son los que iniciaban la Enseñanza Media y el 2º Ciclo de la Enseñanza Secundaria. Respondíamos así al deseo de conocer con cierto rigor la competencia léxica activa que presenta el alumno cuando abandona la EGB y se enfrenta a un nuevo nivel de enseñanza. Evidentemente, ahora la realidad educativa ha cambiado con la implantación de la LOGSE, pero creemos que los resultados de nuestro trabajo serán igualmente valiosos, pues los alumnos cuyo léxico estudiamos tienen 14 ó 15 años, y se pueden equiparar perfectamente con los alumnos del actual 3º de E.S.O. Además, el trabajo puede servir para posteriores análisis, ya que dejará constancia de un sistema escolar que desaparece: el BUP y la FP.

En este sentido, y con relación a nuestro país, no hay demasiados trabajos sobre vocabulario que se ocupen de la Enseñanza Media[1], y que no empleen técnicas propias de los vocabularios de reconocimiento[2]. Pero la novedad, quizás, del nuestro, es recoger una muestra de los alumnos que cursan FP, aspecto no contemplado antes de este modo, y otra muestra de los alumnos que cursan Reforma o experimentación de la LOGSE, lo cual también es, obviamente, novedoso, como lo es, para este nivel de enseñanza, recopilar los datos con técnicas que apuntan al léxico activo.

---

[1] El trabajo de Díaz Castañón y colbs. (1977 y 1978) registra muestras de los alumnos de EGB y de BUP, pero las palabras no aparecen integradas en composiciones sino aisladas, debido al tipo de pruebas aplicadas en la recogida de datos: test de asociación libre y controlada.

[2] Como el trabajo de Márquez Villegas (1975), que a nuestro juicio sí emplea técnicas de los llamados vocabularios de reconocimiento (Cfr. 1.5.2. y 1.5.4.1.).

Nuestro estudio trata de contribuir en algunos aspectos a la enseñanza de la lengua materna, finalidad esencial de la que partimos. Para ello seguimos ideas, a nuestro juicio, esenciales y que todo profesor de lengua se plantea más tarde o más temprano, ideas que esbozaron claramente ilustres lingüistas como Humberto López Morales y Antonio Quilis. Ambos defienden que el conocimiento de la lengua de los alumnos es esencial para una programación seria de la asignatura. Esta idea viene a coincidir con las orientaciones metodológicas del nuevo sistema de enseñanza que propugna la LOGSE. Recordamos, pues, las palabras de A. Quilis (1978: 7-8): "En la programación de la enseñanza de la lengua materna se deberían tener en cuenta determinados factores que habitualmente no se consideran o se consideran sólo intuitivamente. Uno de ellos es el conocimiento de la lengua de los alumnos: la investigación y la acción pedagógica en la enseñanza de la lengua serán eficaces sólo cuando se haya acometido el estudio de la lengua de los alumnos en los diferentes años escolares y se haya profundizado en su descripción. Este estudio proporcionaría datos tanto sobre la competencia y actuación lingüísticas de los escolares en cada uno de los tres niveles de análisis lingüístico -fónico, morfosintáctico y lexicosemántico- como sobre su capacidad de asimilación y abstracción y sobre los centros reales de interés del mundo que los rodea. La información así recogida sería la base cierta de una programación objetiva".

Esa es nuestra intención: contribuir a la planificación de la enseñanza del vocabulario partiendo del que poseen los escolares. Estos constituyen una población joven, que se halla en pleno desarrollo no sólo de su naturaleza fisiológica, sino espiritual, cognitiva, intelectual..., que día a día va ensanchando su universo humano y cultural. Y dentro de la ampliación de sus conocimientos, va progresando y profundizando, con mayor o menor celeridad, en el dominio del vocabulario de su lengua materna. Por lo que recoger éste no es tarea fácil ni definitiva. Ya Rodríguez Bou I. (1952:15), tras recopilar un vastísimo corpus de más de siete millones de palabras, realiza la

siguiente afirmación: "El *Recuento de vocabulario español* es una obra inconclusa en el sentido en que lo es todo recuento de vocabulario. El idioma aumenta cada día su caudal de voces y mañana surgirán palabras que no se incluyen, lógicamente, en los recuentos de hoy". Tratar, por tanto, de determinar el vocabulario de una población cualquiera supone contemplar una serie de postulados esenciales para intentar llevar a buen fin la investigación, teniendo en cuenta al mismo tiempo, dada la naturaleza del terreno movedizo en el que nos movemos -el léxico-, que "es necesario tomar para ello precauciones particulares, no dejarse detener por las dificultades ni por las objeciones de prejuicio y teórico (sic) y aceptar determinadas convenciones" (Müller, Ch., 1973a: 229).

Las dificultades por las que atraviesa un estudio como el que nosotros nos proponemos son de distinta naturaleza. Tal vez la primera de ellas sea reconocer que el nivel léxico tiene un comportamiento distinto y mucho más complicado que el resto de los niveles de la lengua, pues el léxico es el nivel más abierto[3], el que más cambia[4], el que consta de mayor número de unidades[5], el menos estructurado, el más voluble, en el sentido de que varía de un hablante a otro, o en el mismo locutor en situaciones comunicativas

---

[3] Bühler, K. (1966: 430) afirma que el léxico de una lengua es un sistema abierto porque continuamente aparecen y desaparecen palabras nuevas.

[4] Alvar Ezquerra, M. (1990: 72-73) alude a la movilidad del léxico y la necesidad de actualización que tienen los diccionarios; Dubois, Jean et Claude (1971: 110-132) estudian el movimiento del vocabulario francés en algo más de una década, para lo cual analizan un diccionario de uso, el *Petit Larousse*, en sus ediciones de los años 1948 y 1960.

[5] Evidentemente, estos asertos son propios de cualquier estudio sobre el léxico. Véanse, por ejemplo, Salvador, G. (1985: 67-72) quien alude a que mientras que la Gramática o la Fonología tienen sus unidades más o menos limitadas, el léxico "es un monstruo de cien mil cabezas..."; Martinet, A. (1974: 149) se refiere a los monemas léxicos como unidades que integran "inventarios ilimitados".

distintas, por lo que es difícil o imposible asegurar cuándo se agota el léxico de un individuo, de un grupo o de una sociedad determinada (Müller, Ch., 1973a: 227-232). Otras son de naturaleza metodológica, relacionadas con aspectos como la representatividad de la muestra, el corpus de palabras base del recuento, la contemplación de léxico activo o pasivo, registro oral o escrito, las técnicas de recogida de datos, las frecuencias, los criterios léxicos, el proceso de cuantificación, las formas de codificación y de presentación... (Justicia, F., 1995: 10-11; López Morales, H., 1987: 57).

Otras dificultades, en fin, son de orden individual: para realizar el trabajo de campo no hemos podido contar con ningún equipo de colaboradores para repartir la tarea de desplazamiento a las distintas zonas de la geografía granadina, sino que hemos tenido que acudir personalmente a los centros y pedir la ayuda de los profesores de Lengua. Esto ha repercutido, por ejemplo, en el tiempo que nos ha llevado la recogida de datos. También influye el hecho de tener que actuar, aunque sea humildemente, como lexicógrafo, labor "tenaz y persistente, aunque nada estruendosa, por lo general carente de cualquier reconocimiento, y en ocasiones costosa"[6]. Y no olvidemos que hoy en día para confeccionar un vocabulario se necesitan aunar una serie de actividades de diversa índole, para lo cual nos permitimos traer unas palabras de Fontanillo Merino (1986: 651) cuando se plantea una serie de preguntas acerca de la verdadera utilidad de la lingüística: "Para responder a una sola de estas cuestiones con cierta exhaustividad es necesario aunar conocimientos (...), es necesario una experiencia clínica o docente, una larga labor de sociólogo, de estadístico avezado y unos métodos que requieren saber de informática... ¡y no somos más que unos pobres lingüistas!".

---

[6] La cita es de Alvar Ezquerra, M. (1990: 51). En esta y otras páginas, Alvar aporta testimonios históricos de lexicógrafos que apuntan la aridez de su tarea y el escaso reconocimiento que conlleva, por lo que el autor tratará de suplir esas cicaterías reconociendo la labor de los distintos autores.

Pero todos estos obstáculos no deben justificar el abandono de trabajos de esta índole, sino que se ha de seguir indagando y experimentando, sobre todo cuando la finalidad es algo tan importante como el proceso enseñanza-aprendizaje, fundamento y pilar de cualquier sistema educativo, y de la labor que día a día se lleva a cabo en el aula.

El objeto central de nuestra investigación es, pues, la determinación del vocabulario usual. Si hablamos de vocabulario usual hemos de acudir en primer lugar al concepto de frecuencia, ya que ésta indica tanto la usualidad de un vocablo en función del número de veces que aparece como la usualidad media de todos los vocablos de acuerdo con el número total de palabras (N) y de vocablos (V) que integran un determinado trabajo. La usualidad media viene así determinada por la frecuencia media u operación que resulta de dividir el total de N con el total de V. Así operan investigadores anteriores como García Hoz, Lorenzo Delgado y Fernando Justicia (Cfr. 1.4.1.). De ahí también que nuestra investigación adopte la forma y metodología propias de los diccionarios de frecuencias.

La mayor parte del tiempo -exigencia impuesta por la propia naturaleza de la investigación- la hemos dedicado a este punto, el cual constituye uno de los apartados más extensos del presente trabajo. Pero debido a que el tema central necesita de una aplicación lingüística esencial como es la enseñanza de la lengua o, en concreto, la enseñanza del vocabulario, hemos querido también proponer una lista específica de palabras, fundamental a nuestro juicio, y unas orientaciones metodológicas para favorecer su enseñanza-aprendizaje.

Debido a estos dos postulados, determinación del vocabulario y aplicación lingüística, hemos sometido el resultado de nuestro trabajo -el vocabulario de los escolares- a un análisis de diversos aspectos formales. Así, hemos destacado los procedimientos de formación de palabras, el origen de las

las mismas, lo cual a su vez se relaciona con determinados principios de la ley de Zipf, una de las piedras angulares de todo recuento. También hemos analizado otros matices relacionados con el uso; es decir, tomando como referencia obras lexicográficas prestigiosas, hemos señalado los vocablos propios de los llamados registros lingüísticos (coloquial, vulgar, técnico...) y al mismo tiempo, y por los mismos motivos, hemos recogido muestras de la fraseología. Ahora bien, somos conscientes de que en tales aspectos iniciamos una línea de investigación que permanece abierta, ya que el análisis profundo y exhaustivo de cualquiera de ellos constituiría el objeto de un solo y único estudio.

De acuerdo con estos planteamientos, hemos estructurado nuestra investigación en seis capítulos y un anexo. El primer capítulo lo dedicamos a desarrollar el marco teórico en el que nos movemos: la lingüística aplicada, la estadística léxica y los recuentos de vocabulario. En el segundo damos cuenta de la metodología seguida en la investigación para determinar el vocabulario de los alumnos. Aludimos así al sistema de muestreo, al corpus o universo léxico, a la norma de despojo, etc. En el tercero ofrecemos algunos datos cuantitativos y cualitativos de nuestro recuento. Por ejemplo hablamos del total de N y V, de la distribución de frecuencias, de la capacidad de expresión y riqueza léxica de los escolares, del número de sustantivos, adjetivos, verbos, pronombres, etc., y analizamos puntos esenciales del léxico. Operando de esta manera ofrecemos una visión amplia de aspectos importantes que rodean el vocabulario de nuestros alumnos, lo cual sirve de introducción, al mismo tiempo, al capítulo cuarto. En él sugerimos la utilidad de nuestra investigación para el proceso de enseñanza-aprendizaje. En el capítulo quinto, y como balance, se exponen las conclusiones generales. El capítulo sexto, breve, precede a los dos anexos y recoge las orientaciones y explicaciones imprescindibles para comprender debidamente la presentación de las listas.

En el anexo I ofrecemos los vocablos de acuerdo con un orden decreciente de frecuencias.

CAPÍTULO I: MARCO TEÓRICO

1. Marco teórico

1.1. Lingüística aplicada

El trabajo que nos ocupa pretende seguir alguno de los postulados de la Lingüística Aplicada (LA), joven disciplina que ha conseguido numerosos logros. Ha sido y es difícil establecer una concepción o definición clara de Lingüística Aplicada, porque son muchos los temas y perspectivas que en ella se engloban, cosa que impide aunar las distintas consideraciones de los lingüistas o de las diferentes escuelas internacionales[7].

En principio el término (LA) se identificaba con la aplicación de la lingüística general a la enseñanza de idiomas[8], sobre todo, y a la traducción y

---

[7] Vez Jeremías, J. M. (1984: 12-18) realiza una breve introducción acerca del concepto de la Lingüística Aplicada a lo largo del tiempo, desde sus posibles orígenes allá por los años 40-50, hasta nuestros días. En dicha introducción se pone de manifiesto que si bien no hay acuerdo absoluto acerca de su campo de investigación, sí hay coincidencia en puntos esenciales. En este sentido, Ebneter, T. (1982: 9) trata de dar una visión de la Lingüística Aplicada en los siguientes términos: "En la concepción que propugnamos aquí, la Lingüística Aplicada engloba todos los dominios de investigación que resultan de la interferencia de la ciencia del lenguaje con otras disciplinas y de la consideración de todos los factores que operan en la lengua y en el habla. Son los dominios de intersección de la psicolingüística, sociolingüística y etnolingüística, de la adquisición de la lengua materna en conexión con la psicología evolutiva, de las teorías de la enseñanza y aprendizaje, del estudio de los medios, de la lingüística contrastiva, del análisis de errores, etc.".

[8] Marcos Marín, F. y Sánchez Lobato, J. (1988:12); Bennett, W. A. (1975: 77). Ebneter, T. (1982: 14)

a la interpretación[9], pero esta visión hoy queda superada ya que es vastísimo el campo que abarca[10] . Baste recordar los temas principales que polarizaron el Primer Coloquio Internacional de Lingüística Aplicada, que según Ebneter (1982: 13) fueron la semántica y traducción mecánica por un lado y "las posibilidades de aplicación de teorías lingüísticas", por otro. La lista de temas se expande en el segundo Congreso Internacional de Cambridge[11], en 1969, y sigue esta tendencia en el siguiente Congreso, celebrado en Copenhague en 1972. En 1977, como acuerdos mayoritarios, se señalaron las siguientes

---

señala como primeros testimonios de la LA propiamente dicha algunos trabajos de los años 60 "que transforman los conocimientos de la moderna ciencia del lenguaje directamente en material de enseñanza". Vez Jeremías, J. M. (1984: 11) indica que desde la aparición del adjetivo "aplicada" se fue creando la sensación de que la LA "era sinónimo del estudio científico de los principios teóricos y aspectos prácticos de la enseñanza y aprendizaje de las lenguas de ámbito internacional".

[9] Marcos Marín, F. y Sánchez Lobato, J. (1988: 12); Vez Jeremías, J. M. (1984: 8) señala tres frentes de atención preferente de la Lingüística Aplicada: "las lenguas adquiridas en un medio natural, las lenguas aprendidas con ayuda de algún recurso didáctico, y las lenguas en contacto en contextos bilingües o plurilingües".

[10] Vez Jeremías, J. M. (1984: 17) afirma: "En cualquier caso, no resulta arriesgado afirmar que el campo de actividades de la LA se encuentra protagonizado, en la actualidad, por un grupo heterogéneo de expertos que coinciden en su preocupación central por el análisis del lenguaje como parte integral del ejercicio de su profesión. Este grupo incluye profesores de educación especial de trastornos del lenguaje, profesores de lenguas modernas de ámbito internacional, traductores e intérpretes, lexicógrafos, analistas literarios, analistas del discurso, terapeutas de trastornos del habla y un largo etcétera".

[11] Ebneter, T. (1982: 13-14) ofrece la lista siguiente: " 1) Lingüística aplicada a textos literarios. 2) Análisis computacional de textos. 3) Psicología de la adquisición de la primera lengua. 4) Psicología del aprendizaje de segundas lenguas. 5) Investigación del habla y sus aplicaciones. 6) Tecnología en la enseñanza de lenguas. 7) Materiales y metodología de la enseñanza de lenguas. 8) Aspectos lingüísticos de perturbaciones en el lenguaje y terapia. 9) Lexicografía. 10) Materiales de tests lingüísticos. 11) Medición y clasificación de errores en la segunda lengua. 12) Teoría de la traducción. 13) Lingüística contrastiva. 14) Sociolingüística."

líneas de actuación: La Lingüística Aplicada es mediadora entre la teoría y la práctica, es interdisciplinar, está orientada a la resolución de problemas, es científica y educativa[12]. Ebneter (1982: 16) ha precisado claramente dos aspectos del término *aplicado*: 1. La aplicación de la Lingüística Teórica a distintas disciplinas como la Psicología, Sociología, Etnología, etc. de donde resulta una Lingüística de intersección, como la Psicolingüística, Sociolingüística, etc. 2. La transposición de los conocimientos, modelos, etc. de esta lingüística de intersección al terreno de la enseñanza de lenguas. Esta concepción de Lingüística Aplicada como lugar de intersección inspiró el célebre Curso de Lingüística Aplicada de Edimburgo.

Además de las ya mencionadas disciplinas, habría que incluir, según algunos autores, la lingüística cuantitativa[13]. En España y en los distintos congresos de la Asociación Española de Lingüística Aplicada han tenido cabida todas las disciplinas que venimos nombrado y, como no, también la lingüística estadística, como se demuestra en los artículos dedicados desde esta perspectiva tanto a los niveles fónico y gramatical como al nivel léxico (diccionarios de frecuencias, vocabulario disponible...). Jenaro Ortega (1983: 5), autor de un trabajo de estas características, reconoce que "los estudios de índole cuantitativa o estadística sobre el lenguaje adquieren hoy en día, parece, cierta importancia. Es evidente, cada vez más, la costumbre de estimar en trabajos lingüísticos conclusiones extraídas de experiencias cuyos resultados se computan más o menos minuciosamente. Además, algunos

---

[12] Marcos Marín, F. y Sánchez Lobato, J. (1988: 19) recogen los aspectos conceptuales de la LA esbozados por Kaplan. Tales acuerdos fueron fruto del consenso experimentado en la TESOL Convention de Miami, abril de 1977, según cita de Vez Jeremías, J. M. (1984: 15).

[13] Ebneter, T. (1982: 17) señala que Back amplía la consideración de Lingüística Aplicada y "asienta los diferentes campos de intersección, como lingüística cuantitativa, sociolingüística, psicolingüística, entre otros, tanto en la lingüística aplicada, relativa a la praxis vital, como en la lingüística pura".

avances posiblemente eficaces en el terreno de la lingüística aplicada (extraordinariamente basada, en bastantes de sus aspectos, en consideraciones estadísticas), así como algunas puntualizaciones interesantes surgidas en el plano teórico, propician tomar algún partido en el sentido de estudiar cuantitativamente algún corpus determinado del español".

También va quedando claro que dentro de la Lingüística Aplicada se vienen desarrollando estudios teóricos, "hasta poderse hablar, sin paradoja, de una Lingüística Aplicada Teórica" (Marcos Marín, F. y Sánchez Lobato, J., 1988: 11). Otras aproximaciones al concepto de Lingüística Aplicada la consideran una disciplina "orientada hacia la aplicación práctica que contribuye a la solución de tareas, problemas y conflictos en todos los dominios humanos en lo que la lengua está implicada"[14]. Esto se relaciona con otra orientación actual: la preocupación por contactar con la realidad del problema que estudia, con lo cual la Lingüística Aplicada sale también de despachos y aulas (Vez Jeremías, J. M., 1984: 17-18). En resumen, "podría decirse que su objetivo consiste, sobre todo, en recoger o reelaborar, en clasificar y sistematizar por medio de investigaciones científicas especiales los hechos de lengua y los principios de funcionamiento de la lengua que resultan necesarios para fines prácticos: para soluciones prácticas en las respectivas situaciones. En consecuencia, la LA implica hoy, más que nunca, el fomento de estudios especiales basados en los hechos lingüísticos -entre ellos la sistematización de hechos concretos- a fin de poder utilizarlos directamente para la finalidad que se pretenda en tal o en cual estudio, tratando de conseguir que las 'aplicaciones' posean no sólo un alcance práctico sino también humano" (Vez Jeremías, 1984: 18).

---

[14] Spillner, B. "On the theoretical foundations of applied linguistics", *IRAL*, v. 15, nº 2, 1977, p. 155. Citado por Vez Jeremías, J. M. (1984: 16).

Por tanto, se observa una última nota en la Lingüística aplicada: su espíritu no sólo práctico sino humano. En definitiva, se trata de aplicar los conocimientos lingüísticos para la solución de problemas. Uno de ellos es precisamente el de la enseñanza, dominio abordado con gran interés por los lingüistas y en el que se insertaría en última instancia nuestra investigación, si bien ésta además utiliza la lingüística cuantitativa (lexicometría o estadística léxica) y la lexicografía como aplicación lingüística.

1.2. Lingüística estadística

Aunque la estadística comienza siendo una disciplina auxiliar para la lingüística, a la que ofrecía su rigor científico para interpretar sus datos y conclusiones, desde hace décadas es posible hablar de la estadística lingüística como una disciplina científica con un objeto de estudio distinto a la lingüística teórica (Guerra Salas, 1989:12), aunque la relación entre ambas disciplinas es evidente. El uso que la lingüística ha hecho y hace de la estadística es legítimo ya que ésta sólo opera sobre datos cuantificables y las unidades lingüísticas lo son. El interés, no obstante, por aspectos cuantitativos del lenguaje es antiguo, pero los trabajos más o menos sistemáticos sobre tales aspectos, basados sobre la frecuencia de las palabras, surgen a finales del pasado siglo, centrados en el léxico y desde ahí se prolongan hasta nuestros días. Aunque los métodos estadísticos se refinan y perfeccionan, en muchos de estos trabajos, para Guerra Salas (1989: 13) "el objeto de estudio no difiere del de la lingüística, y por tanto la estadística sigue sólo desempeñando el papel de ayudante, si bien muy cualificado". Los pasos que inician la formación de esta nueva disciplina (la estadística lingüística) son el desarrollo de la estadística léxica y los trabajos sociolingüísticos de Labov a principios de los años 70, en donde comienzan a gestarse modelos teóricos acerca de las lenguas basados en la estadística, es decir, los modelos probabilísticos: "Desde el primer modelo probabilístico de Labov hasta los modelos logísticos más recientes, hay unas bases teóricas de

carácter lingüístico comunes: ofrecer un nuevo modelo de la competencia en el que quede inserto un componente probabilístico, que abarcará por igual factores lingüísticos y extralingüísticos. Es por esto por lo que puede hablarse ya con total seguridad de una lingüística estadística (aunque junto a estos estudios se den otros de estadística aplicada a la lingüística): los conceptos probabilísticos se sitúan en una posición central en el estudio de la lengua" (Guerra Salas, 1989: 16-17). No cabe duda que la obra de Charles Müller (1973a) es una contribución importante a dicha consideración independiente de la estadística lingüística.

1.3. Estadística léxica

"La *estadística léxica* es una aplicación de los métodos de la estadística a la descripción del vocabulario" (Dubois y otros, 1992: 243). O dicho de modo distinto, "podemos clasificar bajo este nombre todas las operaciones que toman como unidades las <<palabras>>, a diferencia de las que consideran los fonemas o los hechos morfosintácticos" (Müller, Ch., 1973a: 225). Es precisamente este último lingüista quien, a nuestro juicio, mejor ha desarrollado los postulados, principios y problemas en los que se basa esta disciplina de la lingüística cuantitativa[15] y de los que nosotros trataremos de ofrecer una breve panorámica, ya que nuestro trabajo se servirá de tales postulados y principios.

La estadística léxica comienza verdaderamente, a juicio de Müller, cuando se introduce la noción de *vocablo*, es decir, cuando se empieza a distinguir de un modo claro entre vocablo, como unidad de léxico, y palabra, como unidad de texto: "leemos una palabra en el texto, pero en el diccionario encontramos

---

[15] Tales principios aparecen recogidos en la II parte de su obra (Müller, Ch., 1973a); parte que lleva por título "La estadística lexical", pp. 225-365.

un vocablo" (Müller, Ch., 1973a: 226). También se empiean hoy (y nosotros así los utilizaremos) los términos "lema" y "type" como sinónimos de vocablo, y "ocurrencia", "token", para designar a la palabra. Los términos "type, token, ocurrencia" suelen utilizarse desde perspectivas cuantitativas, y el de "lema", desde perspectivas léxicas y lexicográficas.

Una de las peculiaridades de la estadística léxica le viene dada por el objeto que estudia: el léxico. Quizás sea el léxico el nivel lingüístico más problemático, y ello se refleja en las disciplinas que de él se ocupan. Porque, como hemos dicho, es el que posee mayor número de unidades, pues "sin ser infinito en el sentido matemático del término, no nos da nunca la impresión de ser estrictamente *finito*"[16].

1.3.1. Léxico y vocabulario

Antes de seguir adelante y puesto que trabajaremos, en gran medida, dentro del terreno de la estadística léxica, conviene aclarar entre léxico y vocabulario. Es la propia estadística lingüística la que hace tal distinción, y atribuye el primer término a la lengua y el segundo al discurso, por lo que reserva el término *lexema* a las unidades que componen el léxico, y cuando estas unidades virtuales son realizadas en el discurso se las designa con el término *vocablo*, "designando con el término *palabra* toda ocurrencia de un vocablo cualquiera" (Müller, Ch., 1973a: 230). Desde esta perspectiva, el vocabulario es el conjunto de palabras que un individuo emplea, por lo que será siempre más reducido que el léxico. Éste, como se ha indicado, se refiere a la lengua, y aquél al discurso, al habla, a la actuación, es decir, a manifestaciones concretas de la expresión oral o escrita, por lo que el

---

[16] Müller, Ch. (1973a: 227). La cursiva es nuestra.

vocabulario presupone la existencia de un léxico del que sólo es una representación[17].

Si bien, como hemos visto, la noción de vocabulario, como resultado de la realidad de un texto o discurso, puede ser determinada y definida con cierta exhaustividad, no ocurre lo mismo con la noción de léxico: "El conocimiento del texto es necesario y suficiente para determinar el vocabulario y someterlo al análisis; puede proporcionar indicaciones sobre el léxico del que el texto sólo es una actualización limitada, pero no podría determinarlo" (Müller, Ch., 1973a: 231).

La determinación del léxico pasa por una serie de dificultades de las que trataremos de dar cuenta a continuación. Sabemos que la lengua posee un conjunto de palabras al que podríamos denominar *léxico general* o *léxico global*[18], que a su vez está a disposición de todos los individuos de una comunidad lingüística[19]. Pero la porción de esta globalidad que posee cada

---

[17] Véase Müller, Ch. (1973a: 231); Lamiquiz, V. (1989 : 225). La idea de que el vocabulario está ligado al habla y el léxico a la lengua es algo consabido en lingüística, y no sólo se tiene en cuenta para una investigación estadística, sino didáctica, como recuerda Galisson, R. (1983: 46): "... le choix de **vocabulaire** implique un traitement des mots en *discours*, qui s'oppose à un traitement *en langue*, lequel relèverait d'une étude de *lexique*."

[18] Así lo hace por ejemplo Pastora Herrero, J. F. (1990: 60).

[19] La consideración de que el léxico de un idioma está integrado por la suma de todos los léxicos individuales no es del todo correcta, salvo que tal definición tenga en cuenta consideraciones diacrónicas, pues se sabe que si el léxico de un idioma, por ejemplo el español, es el recogido en el RAE, creemos que la suma de los léxicos de todos los españoles no llegaría al número de vocablos del Diccionario Académico; y lo mismo ocurriría con todos los demás idiomas. Además, si tenemos en cuenta que en el léxico de una lengua entrarían también los tecnicismos, arcaísmos, dialectalismos..., nos moveríamos en cifras alarmantes. Por ejemplo, Marcos Marín, F. y Sánchez Lobato, J. (1988: 125) refieren que existen,

individuo es distinta a la que poseen los demás. Por tanto, el léxico individual estará constituido por las palabras que un hablante determinado posee[20]. La naturaleza de este léxico se puede, teóricamente, establecer con ciertas aproximaciones o consideraciones lingüísticas más o menos limitadas. Por ejemplo, el léxico de un individuo está integrado por lo que se conoce como léxico activo y léxico pasivo, nociones que también entrañan dificultad a la hora de las definiciones. El léxico pasivo se relaciona con el vocabulario disponible -del que hablaremos detenidamente más adelante-, por un lado; es decir, aquellos vocablos que el sujeto conoce y comprende perfectamente pero que tienen una frecuencia muy baja o no aparecen en la lengua usual, porque son vocablos que están unidos a determinados temas; y por otro lado, con aquellos vocablos que el hablante comprende pero que no emplea como locutor. (Müller, Ch., 1973a: 234). A este léxico habría que añadir los "lexemas virtuales", o sea, aquellos que un locutor puede crear en un momento determinado de su discurso utilizando los mecanismos de formación de palabras vigentes en una lengua, pero tales vocablos no están aceptados por la comunidad lingüística y como tal no se recogen en los diccionarios, ya que responden a situaciones comunicativas particulares y se consideran creaciones ocasionales, efímeras e individuales, que si no pasan de ahí y no trascienden al resto de los hablantes, no son adoptadas como componentes del acervo léxico de un idioma, pero que, según Müller (1973a: 234), "estas unidades,

---

según los cálculos de INFOTERM, unos 30 millones de objetos identificados, entendiendo por tales tanto objetos físicos como conceptuales, cifra que supera con creces el caudal de cualquier diccionario.

[20] El léxico es algo que el individuo va ampliando progresivamente a lo largo de su vida. Para De Mauro (1982: 101), la primera fuente de conocimiento de las palabras es la familia, círculo que se va ensanchando a medida que vamos adquiriendo mayor independencia con la edad: "Alrededor de los ocho años, los niños conocen varios miles de palabras. Hacia la adolescencia, la cifra de las palabras conocidas comienza a medirse en decenas de miles. Quienes trabajan en oficios y profesiones que requieren el aprendizaje de diversos lenguajes especiales, típicos de esta o de aquella ciencia o profesión u oficio, y quien conoce varios dialectos y lenguas diferentes, almacena en su memoria más de cien mil palabras diferentes".

que se pueden denominar <<virtuales>>, pertenecen incuestionablemente al léxico". (Nosotros, recordemos, daremos cuenta de estas unidades creadas por los alumnos). Todas estas consideraciones se relacionan con un nuevo concepto: el léxico de situación (Müller, Ch., 1973a: 235-236). Este léxico engloba dos tipos de elementos: unos son de orden estilístico y están ligados al efecto que el hablante quiere producir; y otros son de orden temático y se relacionan con el contenido del mensaje. Estas variantes estilísticas y temáticas excluyen, pues, en un determinado léxico de situación, una parte del léxico individual. El léxico de situación es, por tanto, un subconjunto del léxico ligado a un texto, a un discurso; éste a su vez un subconjunto del léxico individual, el cual es una parte del léxico de un idioma (Müller, Ch., 1975: 5). Para estimar la extensión del léxico individual, habría que partir de textos representativos de todas las situaciones comunicativas, pero hacer un inventario exhaustivo de todas las situaciones sería prácticamente imposible; igualmente para el léxico de un idioma (Müller, Ch, 1975: 5-6). Además, como se ha apuntado más arriba, el vocabulario resultante de la actuación del léxico sólo sería un subconjunto de la totalidad del léxico de un locutor, pues "el vocabulario no cesa nunca de crecer y de hacer aparecer unidades: ningún discurso agota el léxico de su autor" (Müller, Ch., 1973a: 233). Es decir, el léxico es algo potencial, y como tal, no se le pueden asignar límites[21], aunque también es cierto que determinadas palabras son más usadas en el discurso que otras, las cuales a su vez tienen menos posibilidad de aparecer[22].

---

[21] Guiraud, P. (1960: 23) afirma: "Nous entendons par lexique le nombre de mots en puissance, ceux dont l'auteur dispose et que le sujet appelle: c'est une quantité purement hypothétique."; y autores como Michéa, R. (1972: 73) señalan que "un vocabulaire potentiel n'a pas de limites assignables. Continuellement, des mots y entrent e d'autres en sortent."; Müller, Ch. (1975: 3-4) ratifica la idea expresada en otros lugares de su obra: que sólo hipotéticamente se podría agotar el léxico de un sujeto, si se prolongara su discurso indefinidamente, puesto que en la práctica esto no ha sido demostrado.

[22] Müller, Ch. (1975: 4) afirma: "... mais certaines de ces unités ont une probalilité d'emploi très elevée (on

Por todo ello, sólo se puede cuantificar con cierto rigor el vocabulario, campo que, al ser un fenómeno de habla, según la terminología de Saussure, es sobre el que opera el método estadístico (Müller, Ch., 1973a: 238). A pesar de todo, también existen fórmulas que tratan de medir el léxico, pero no pueden tener ni el rigor ni la relativa exactitud que mide el vocabulario o la distribución de frecuencias de un vocabulario[23].

1.3.2. La norma léxica

Para cuantificar un vocabulario se necesita una norma léxica entendida como *norma de despojo* de un texto. Esta norma debe delimitar con precisión dos conceptos: palabra y vocablo, para poder ofrecer con claridad los datos numéricos relativos, en primer lugar, al número de palabras -u ocurrencias, tokens-, representado por N, y que da la medida de la extensión del texto; y, en segundo lugar, al número de vocablos -o lemas, types-, representado por V, que mide la extensión del vocabulario (Müller, Ch., 1973a: 242-243).

Para la determinación de la norma de despojo se deben esgrimir criterios lingüísticos -fonéticos, gramaticales, léxicos- y adoptar una unidad como palabra y otra unidad como vocablo. Uno de los principales obstáculos para la palabra es considerarla como una unidad gráfica o como una unidad de sentido, lo cual supondría contemplar locuciones, perífrasis, refranes,

---

ne peut parler longtemps sans utiliser un verbe comme *faire*, un substantif comme *temps*, un adjectif comme *bon*, un adverbe comme *souvent*, etc.); d'autres ont une probabilité moyenne, faible, très faible, etc.". Véase igualmente Hörmann, H. (1973: 141-145).

[23] Estos asertos se deben a Müller, Ch., (1975: 22-28), quien conocedor de tal problemática, aduce no obstante algunos ejemplos de investigaciones que han tratado de medir una parte del léxico, como por ejemplo el número de verbos "semiauxiliares".

compuestos, etc. Para la delimitación del vocablo, las etapas pasan por la reagrupación de formas flexivas, la separación de los homógrafos (homonimia léxica y sintáctica) y la contemplación o no de las palabras gramaticales (Cfr. 2.11.). Dirimir todos estos aspectos no es tarea fácil, porque hay que enfrentarse a una serie de problemas que no siempre convencen a todo el mundo, ni tan siquiera al propio investigador[24]. Como cualidades de la norma, Charles Müller (1973a: 244-246) aconseja que esta sea simple, sencilla, en el sentido de que recoja aquellos criterios delimitadores que se puedan aplicar con precisión a un texto o a todos los textos de una misma lengua, por lo que propone que se adopte la norma tradicional de los diccionarios: "Los defectos de la norma lexicográfica son ampliamente compensados por su comodidad, su relativa estabilidad, y también por el hecho de que está fuertemente impresa en la memoria de la mayoría de los usuarios" (Müller, Ch., 1973a: 246).

1.3.3. Extensión del vocabulario: ley de crecimiento del vocabulario y riqueza léxica

Cuando se ha definido perfectamente la norma de despojo, se comienza el recuento del total de palabras y vocablos de un texto y se obtiene así un valor numérico para N (extensión del texto) y otro para V (extensión del vocabulario). No existe ninguna fórmula matemática[25] que permita calcular la

---

[24] Müller, Ch. (1973a: 246) afirma: "El carácter más peculiar de una norma es el de no satisfacer a nadie, comenzando por su autor. Pero es necesario franquear el obstáculo de la discusión de detalle".

[25] Müller, Ch. (1973a: 269) ejemplifica este aserto con los datos relativos a una obra de Corneille: "¿qué interpretación se puede obtener de un dato como este?

Fhèdre: N = 14.217

V = 1.653.

Nada nos permite actualmente calcular una esperanza matemática para V en función de N = 14.217".

relación entre N y V, en el sentido de que no se puede afirmar categóricamente que a un texto con una extensión N específica corresponda un vocabulario de una extensión V determinada. No obstante, sí se indican una serie de factores que contribuyen a la extensión de un vocabulario: en primer lugar, el vocabulario -V- está en función de N o extensión de un texto; por lo que cuanto más extenso sea un texto más grande debería ser su vocabulario. En segundo lugar, la extensión de un vocabulario está en función del estilo, es decir, del léxico del locutor en la situación estilística que adopte. En tercer lugar, influye el tema o temas de los que se trate. Y en cuarto lugar, cuando son textos o corpus originados por varios hablantes, influye también el número de sujetos de la muestra[26].

La relación entre N y V conecta a su vez con dos aspectos esenciales de la estadística léxica: la ley de crecimiento del vocabulario y la noción de riqueza léxica.

Es algo consabido que V es función de N, o sea, que el número de vocablos aumenta con el número de ocurrencias, pero V crece menos deprisa que N. También es cierto que llegará un momento en que será cada vez más raro encontrar un vocablo nuevo, quizás porque, en lo tocante a un autor al menos, las palabras se repiten en cierto modo para asegurar la coherencia del

---

[26] Rodríguez Bou (1952: 14) cita una tesis llevada a cabo por Evelyn White en la que la autora se propuso investigar "cuántas palabras de prosa de dificultad promedio tendría que leer un estudiante para encontrar todas o casi todas las primeras 1500 ó 2000 palabras -las básicas- del recuento de Buchanan. La conclusión principal a que llega es que si se leen 20000 de cada uno de tres autores diferentes se podrán encontrar dichas palabras, pero habría que leer muchas palabras más de un solo autor (alrededor de 80000) para encontrar las 1500 ó 2000 palabras aludidas". Dugast, D. (1979: 84) alude así al hecho del que hablamos: "alors que la recherche de la variété des textes du corpus mènera probablement à un excès de richesse lexicale (on expliquera sans doute ainsi l'anormale richesse du dictionnaire de Evy Steinfeldt, qui inclut 75 articles de publicistes...)".

discurso y por el agotamiento e incapacidad del propio locutor para renovar su léxico (Michéa, R., 1972: 75); aunque evidentemente, V no deja de crecer, "pues ningún texto agota el léxico de su autor" (Müller, Ch., 1973a: 268). Y esto debido, quizás, entre otras razones, a que el léxico, al estar ligado al conocimiento del mundo, no deja de ser enriquecido en el transcurso de la vida (Rey-Debove, 1970:4) A medida que va creciendo un texto, que aumenta el número de ocurrencias, podría ser, en un caso hipotético, que los efectivos de frecuencia 1 desaparecieran, pero es una hipótesis, ya que nadie lo ha comprobado todavía (Michéa, R., 1973: 84). O sea, se sabe, en la práctica, que ningún corpus, por muy extenso que sea, (por ejemplo, el *Trésor de la langue française*, de 75 millones de ocurrencias), ha agotado el léxico de sus locutores, pero sí puede establecerse la teoría de que, si el léxico es finito, llegaría un momento en el que V dejaría de crecer, y V1(vocablos de frecuencia 1 ó hápax) se anularía, y los demás efectivos sufrirían este efecto. Pero aunque este punto no se ha observado en la práctica, lo cierto es que debería llegar un momento en que se produjera una estabilización; es decir, un decrecimiento de V1 y, después, de los demás efectivos. Esta experimentación se ha hecho con los verbos "semiauxiliares"[27] y con textos en los que el vocabulario de situación no es muy abundante, como el de los boletines meteorológicos.

El crecimiento del vocabulario en un texto ha generado, desde los inicios de lingüística cuantitativa, intentos por parte de numerosos investigadores para establecer alguna fórmula matemática que diera cuenta de lo que se conoce como la ley de crecimiento del vocabulario, apoyándose para ello sobre N y V, o sobre N, V y V1 o bien teniendo en cuenta la distribución de frecuencias.

---

[27] Véase Müller, Ch. (1975: 23-29).

Por ejemplo, P. Guiraud[28] afirma que V aumenta como la raíz cuadrada de la longitud de N:

$$V = C\sqrt{N}$$

Charles Müller (1973a: 310-322) trata de calcular o diseñar el crecimiento teórico de un texto, "no considerando de ninguna manera N y V como valores fijos, y su rendimiento como una relación estática, sino tratando de determinar cómo V evoluciona cuando se hace crecer a N regularmente, es decir, cuando se lee el texto anotando los vocablos nuevos" (Müller, Ch., 1973a: 310). Es consciente que en el crecimiento del vocabulario intervienen irregularidades, como razones temáticas o estilísticas que pueden hacer crecer en exceso el vocabulario, o momentos en que predominen las repeticiones, lo cual hará disminuir el número de palabras nuevas, sin olvidar la evidencia de que al principio el texto tiene un aumento de vocablos más intenso que al final. Él tratará de establecer un modelo de crecimiento, pero partiendo de datos numéricos conocidos del texto en cuestión, -para contrastar después el crecimiento real- , como el número de vocablos de las partes del texto, o la distribución de frecuencias. Para Müller el problema consiste en saber dónde se halla la primera ocurrencia de cada uno de los vocablos. Müller (1973a: 312) representa el texto por una recta $\alpha \rightarrow \omega$, sobre la que determina el punto *a*; crea así "dos segmentos, $\alpha \rightarrow a$ y $a \rightarrow \omega$; si el lector recorre el texto y se detiene en *a*, el fragmento $\alpha \rightarrow a$ es la parte que ha leído, el fragmento $a \rightarrow \omega$ la que no ha leído todavía. Llamaremos N_a el número de palabras que ha leído, V_a el número de vocablos que ha encontrado hasta ese momento". Su modelo o fórmula calculará los efectivos de los vocablos que no se encontrarán en el fragmento analizado ($\alpha \rightarrow a$):

$$E({}_aV_0 = \sum_f V_f \left(\frac{N - N_a}{N}\right)^f$$

---

[28] Guiraud, P. (1960: 85). Aunque el autor se refiere a textos comprendidos entre 10 y 50 mil palabras.

Tras aplicar esta ecuación a un texto determinado, observa que al principio y al final se producen unas desviaciones excesivas entre el vocabulario teórico y el real. Por ello Ch. Müller deduce que el método sobrestima el crecimiento léxico al principio y lo subestima al final, por lo cual propone aplicar un correctivo. Decide dividir un texto en numerosos trozos de igual longitud, y calcular el crecimiento teórico de uno concreto, denominado trozo *k*. Las probabilidades de aparición de un vocablo cualquiera en el trozo *k* es:

$$P_k = \left(\frac{n - k + 1}{n}\right)^f - \left(\frac{n - k}{n}\right)^f.$$

Si se designa por V_k la extensión del vocabulario al final del trozo k, el crecimiento teórico en este trozo es:

$$E(V_k - V_k - 1) = \Sigma V^f \left(\frac{n - k + 1}{n}\right)^f - \left(\frac{n - k}{n}\right)^f.$$

A Müller, este modelo de crecimiento del vocabulario le sirve, por ejemplo, para detectar las variaciones estilísticas y temáticas en el interior de un texto.

Dugast también propondrá una ecuación para medir la ley de crecimiento de un texto, delimitando las partes de éste o estableciendo una serie de fragmentos de una longitud determinada[29]. Para medir la cantidad de vocablos de estos fragmentos utiliza la medida Uber: $U = n^2/(n-v)$. Considerando $\log V = v$ y $\log N = n$, la fórmula definitiva del crecimiento del vocabulario o "l'apport lexical" que Dugast (1979: 88) postula, de acuerdo con V en función de N y de Uber, es la siguiente:

---

[29] Dugast, D. (1979: xv) afirma: "C'est l'accroissement d'un texte dans ses parties succesives à partir duquel à lui seul peut se dessiner la loi recherchée. C'est dans la comparaison de plusiers tels acrroissements entre eux qu'on trouvera confirmation ou infirmation de cette esquisse".

$$\delta V = \frac{(10^n - n^2}{N}(1 - \frac{2n}{u})(\delta N)$$

Tal fórmula es aplicable al crecimiento de un texto único, que se podrá comparar con el crecimiento de otros textos únicos, siempre y cuando en ellos se haya utilizado la misma norma léxica, pues en aquellos trabajos que tratan de establecer un vocabulario fundamental partiendo de una gran variedad de textos tal fórmula no tendrá validez, pues la variedad de textos llevará a un exceso de riqueza léxica. Dugast (1979: 4) hace observar que contrariamente a lo que se ha dicho -que la relación logV/logN no puede definir una constante- la renovación del vocabulario en un discurso se hace irregularmente, y probablemente de una manera rítmica.

Además de Guiraud, Müller y Dugast, otros autores que han tratado de diseñar sus medidas para dar cuenta de la ley de crecimiento del vocabulario han sido G. Udny Yule (*The Statistical Study of Literary Vocabulary*, 1944), donde expone la fórmula de su famosa "característica *K*"; Gustav Herdan y su índice de Yule-Herdan (*A New Derivation and Interpretation of Yule's Characteristic* K, 1955) y LogV/logN (*Type-token Mathematics. A textbook of Mathematical Linguistics*, 1960); Etienne Brunet (*Le vocabulaire de Jean Giradoux, structure et évolution*, 1978) y su nuevo índice inspirado en Guiraud... (Todas estas medidas se relacionan igualmente con la noción de riqueza léxica de un texto). Para aplicar estos índices, se suele dividir el texto en cuestión en partes de *x* palabras -100, 200, 500-, y se compara el crecimiento del vocabulario real con el vocabulario teórico establecido por tales índices o medidas. Basándose en estas observaciones, se suelen extraer diversas conclusiones de naturaleza estilística, como la relación de los momentos de mayor inflexión o acumulación de palabras con los temas o

personajes esenciales de un texto, o el análisis de las diferencias expresivas de un autor en sus distintas etapas y obras[30], etc.

La relación entre N y V nos lleva, como indicábamos, a otra noción importante para la estadística léxica cuando se trata de extraer apreciaciones cualitativas: la noción de riqueza léxica. Esta noción está en relación con el número de vocablos usados en un texto; es decir, es una noción cuantitativa. Para Cossette (1994: 6-14) la noción de riqueza léxica atraviesa por una serie de problemas como, en primer lugar, la relación que tal noción mantiene con otros conceptos como la de campo léxico-semántico, pues el léxico desplegado en un discurso refleja la extensión relativa (todos los campos semánticos no tienen la misma extensión ni se puede determinar ésta con precisión, ya que muchos están interrelacionados al pertenecer una determinada palabra a varios de ellos) de los campos léxico-semánticos de los que se extraen los vocablos[31], y, en segundo lugar, la relación que mantiene igualmente con el léxico de situación y sus componentes estilístico y temático de los que habla Müller. La distinción de un componente estilístico y de un componente temático obliga a diferenciar entre la riqueza que es propiamente estilística y la propiamente temática. "La richesse stylistique est essentiellement une caractéristique individuelle. Elle réside dans la diversité personnell des choix de vocables effectués par le locuteur dans un lexique considéré comme un invariant fourni par la langue. C'est le contraire de qui se

---

[30] Véase Dugast, D. (1981, 1982, 1985); Rosseli, M. (1992); Müller, Ch. (1967).

[31] Una prueba de ello la aportan también los estudios de disponibilidad léxica, como el llevado a cabo por Etxebarria, M. (1996): el propio autor nos informa de que el número de ocurrencias no está en relación con el número de vocablos. En su trabajo, el centro de interés "el campo" arroja un total de 4.406 palabras y 773 vocablos, mientras que el centro de interés "los colores", con un número casi idéntico de palabras -4224-, tiene un total de vocablos muy inferior: 281. Para Etxebarria, M. (1996: 315) esto "guarda relación con la mayor o menor cohesión semántica de cada uno de los centros".

passe du pont de vue de la richesse thématique, où l'on peut considérer que le nombre et la taille des lexiques de situation effectivement <<en jeu>> dans le texte sont des variables. Quand le locuteur choisit d'aborder tel ou tel sujet, son choix augmente la probabilité d'emploi d'un certain nombre de lexèmes en même temps qu'il diminue celle d'autres lexèmes. Or, ces probabilités d'emploi déterminent la richesse thématique pottentielle dans la mesure où, d'une part, le nombre de lexèmes ayant une forte probabilité d'emploi change selon le sujet et où, d'autre part, le nombre de sujets abordés est grand. Ainsi au départ, le nombre de lexèmes susceptibles d'être utilisés n'est pas le même d'un texte à un autre" (Cossette, A., 1994: 14). Por tanto, la noción de riqueza léxica aparece como una noción compleja, en donde se refleja no sólo el aspecto estilístico o utilización puramente individual de la lengua, sino también el componente temático y la estructuración del léxico en campos semánticos de tamaños desiguales, sin olvidar que el paso de la lengua al discurso facilita sólo una parte del léxico, y que el discurso se realiza en un contexto determinado que lleva al locutor a elegir uno u otro léxico de situación (Cossette, A., 1994: 15). A pesar de que medir la riqueza léxica es algo complicado, varios autores (Yule, Herdan, Guiraud, Müller, Brunet, Dugast...), como indicábamos más arriba, han propuesto sus fórmulas para llevar a cabo tal empresa.

Para Müller (1973a: 2690-272), veíamos, no existe una ley que fije una cantidad de N para una cantidad determinada de V, por lo que sólo tiene sentido la noción de riqueza léxica si se comparan varios textos entre sí, siempre y cuando se haya utilizado en ellos la misma norma de despojo. Se pueden así comparar dos textos A y B para ver cuál de ellos posee un vocabulario más extenso. Müller llama N*a*, N*b* a sus extensiones respectivas y V*a*, V*b* a la extensión de sus respectivos vocabularios, y considera que el vocabulario del texto A es más rico que el del B si se cumplen las siguientes condiciones:

a) N*a* ≤ N*b* y V*a* > V*b* . Es decir, si un texto tiene una extensión (N*a*) menor o igual que otro texto (N*b*), pero el vocabulario del primero (V*a*) es mayor al vocabulario del segundo (V*b*), podremos hablar de que posee una mayor riqueza léxica.

b) N*a* < N*b* y V*a* = V*b*. Es decir, si la extensión de un texto (N*a*) es menor a la de otro (N*b*), pero el vocabulario del primero (V*a*) es igual al del segundo (V*b*).

c) N*a* > N*b* V*a* > V*b* pero (N*a* - N*b*) > (V*a* - V*b*). Este supuesto Müller lo ilustra del siguiente modo:

	A. Médée	B. Phèdre	Desviación
N	14.256	14.217	+ 39
V	1.715	1.653	+ 62

Por tanto, el léxico de *Médée* es más rico que el de *Phèdre* porque la diferencia de los vocabularios es mayor que la diferencia de las extensiones de los textos.

Si no se cumplen estas tres condiciones -a, b, c- se puede recurrir a la frecuencia media: Fm = N/V. En un texto estilísticamente homogéneo, N crece más deprisa que V, de modo que también crece la Fm; por tanto, si el texto más largo tiene una Fm más débil, su vocabulario es más rico:

d) N*a* > N*b* V*a* > V*b* Fm*a* < Fm*b* . Este método de comparación no es útil cuando los textos tienen una extensión muy desigual.

Pero no todos los textos se pueden someter a estos cuatro supuestos, ya que posiblemente en la mayoría de los casos presenten datos cuantitativos muy dispares. Además, casi todos los autores que se han ocupado tanto de la ley de crecimiento del vocabulario como de la riqueza léxica establecen sus fórmulas y cálculos matemático-estadísticos de acuerdo y para ser aplicados a un solo y único texto, concebido como tal, y fruto o producto de un solo y único autor.

Müller (1973a: 290-322) ha desarrollado otro método que permite comparar textos de dimensiones diferentes. El método consiste en calcular el vocabulario teórico de una parte del texto partiendo de todos los datos del vocabulario del texto -o corpus-. En concreto entran en juego los siguientes datos:

N: Número palabras o extensión del texto.

N': Número de palabras o extensión del fragmento.

V: Número de vocablos o extensión del vocabulario del texto.

V': Número de vocablos o extensión del vocabulario teórico del fragmento, que luego se podrá comparar con el V real.

V_i: Número de vocablos de frecuencia 1, 2, 3, etc. del texto.

N'/N: Operación que nos proporciona **dos** cifras de decimales cuya suma es 1 (por ejemplo, 0'513 y la diferencia hasta llegar a 1, 0'487). Estas cifras se relacionan con dos conceptos esenciales que se representan así:

- *p* o posibilidad de que un vocablo determinado se halle en el fragmento del que se está hallando su vocabulario teórico, y cuya cifra resulta directamente de la operación N'/N.
- *q* o posibilidad de que un vocablo de frecuencia 1, 2, 3, etc. no se encuentre en el fragmento que se estudia, y cuya cifra es la diferencia con la anterior hasta llegar a 1.

V'0: Número de vocablos que no tienen ninguna ocurrencia en el fragmento que se estudia.

Conocidos todos estos datos, se puede aplicar la fórmula de la extensión teórica de un vocabulario, que es la siguiente:

$E(v'0) = \Sigma\, q^i V_i$, de donde $E(V') = V - \Sigma\, q^i V_i$

Conviene detener la operación cuando el producto de vocablos sea inferior a 0'5. Esta fórmula permite comparar el vocabulario teórico con el real y extraer las diferencias. Cotejando las diferencias de los fragmentos se pueden sacar conclusiones, más o menos relativas, acerca de la riqueza léxica de cada fragmento. Müller (1973a: 301) resalta que "la experiencia muestra que en el establecimiento de un valor teórico para V', se encuentran frecuentemente valores superiores a los valores reales que a la inversa". El método de Müller es eficaz si no se aplica a fragmentos o textos de extensión muy desigual; en concreto, aconseja no bajar del siguiente valor: $p = 0'1$.

Hay quien llama a los textos con un vocabulario menos rico de lo frecuente *worteng*, "angostos de palabras", y los "dilatados de palabras" o *wortweite* poseen un vocabulario muy rico[32].

1.3.4. Distribución de frecuencias

La distribución de frecuencias se relaciona con la *estructura* de un vocabulario: "La estructura de un vocabulario comprende elementos cuantitativos simples que son el número de los vocablos del texto y la frecuencia de cada uno de ellos; reuniendo los vocablos de idéntica frecuencia, se forman clases de frecuencia, de la que cada una suministra un efectivo. La estructura del vocabulario incluye las relaciones que se establecen

---

[32] Hörmann, H. (1973: 126) alude a tales términos, usados por Meier, y explica: "Los espectros 'angostos de palabras' surgen, sobre todo, como indica el mismo Meier, en momentos de auténtica defensa. Esta regresión al vocabulario más usual debe ponerse en relación con el hecho, conocido por la psicología, de que el estrés provoca un entumecimiento del pensar y una reducción de afluencia de ocurrencias. Los espectros *wortweite* o 'dilatados de palabras' aparecen en cambio en los trabajos científicos, en los escritos sobre materias que tienen su lenguaje especial (por ejemplo, los deportes o la moda) y en las expresiones irónicas...".

entre los efectivos de clases sucesivas de frecuencia. El conjunto de estos datos numéricos es la <<la distribución de frecuencias>>. Se mencionarán a título indicativo los elementos cualitativos que entran igualmente en la estructura, y que son la naturaleza gramatical de los vocablos y las relaciones de asociación (gramaticales o semánticas, paradigmáticas o sintagmáticas, sinonímicas o antonímicas) que existen entre estos vocablos" (Müller, 1973a: 239).

Para apreciar, pues, la estructura cuantitativa de un vocabulario (Müller, 1973a: 273-291), se agrupa cada ocurrencia con su vocablo correspondiente y se reúnen y se cuentan los vocablos de una misma frecuencia para obtener las *clases de frecuencia.* A cada clase de frecuencia se le asigna un *efectivo* (*V1* o número de vocablos de frecuencia 1; *V2* o número de vocablos de frecuencia 2, etc.). Los vocablos se clasifican por frecuencia decreciente, de mayor a menor, y se les atribuye a cada uno de ellos un rango[33] (rango 1 al vocablo más frecuente, rango 2 al siguiente vocablo más frecuente... y así sucesivamente). Se aprecia así el cuadro de distribución de frecuencias y se puede comprobar una serie de rasgos constantes para cualquier distribución de frecuencias[34]:

1. A mayor frecuencia, menor número de efectivos. Los vocablos de mayor rango están integrados por un sólo efectivo, y entre estos vocablos algunas frecuencias no están representadas -es decir, puede haber un vocablo con frecuencia 5400, y el siguiente puede tener 5300, pero no habrá vocablos con frecuencia 5399 o 5398, etc.-. En cambio, por la parte final, no se produce este fenómeno y sí se encuentran vocablos de frecuencia 1, 2, 3, 4, etc.

---

[33] La noción de "rango" es de la ley de Zipf.

[34] Véase Müller, Ch. (1973a: 273-282, 1965a, 1982); Dugast. D. (1981); Michéa, R. (1973).

2. Los vocablos de frecuencia 1 son los más elevados.

3. La importancia de los efectivos varía en sentido inverso a la frecuencia: Vn > Vn +1. Por ejemplo, los 10 últimos reúnen la mayor parte de los vocablos[35].

4. A partir de que los efectivos tienen frecuencias débiles, como alrededor de la frecuencia 20, se producen irregularidades, en el sentido de que puede haber frecuencias con más efectivos que otra frecuencia inferior. Es decir, es posible que los vocablos de frecuencia 19 se compongan de 18 efectivos y los de frecuencia 20 tengan un efectivo de 22, por ejemplo.

5. Hay probabilidades de que los hápax sean el doble de las palabras de frecuencia 2: $V_1 / V_2 \geq 2$

6. Existe la posibilidad de que el número de palabras de frecuencia 2 sea aproximadamente la sexta parte de las palabras diferentes: $V_2 = V/_6$

Si se compara la distribución de frecuencias de textos o fragmentos de distinta extensión, se puede comprobar que a medida que crece N:

a) Crece obviamente el número de frecuencias.
b) La frecuencia máxima aumenta.
c) También crece el efectivo de una frecuencia cualquiera.

Müller (1973a: 282) concluye afirmando que es importante "comprobar que el vocabulario de un texto, que es una muestra de un léxico virtual, obedece en su estructura cuantitativa a impulsos que no son fortuitos, y se construye según leyes complejas y mal conocidas todavía, pero lo bastante imperiosas como para que hayamos podido definir algunos de sus efectos...".

---

[35] Michéa, R. (1973: 82) aporta unos datos que van desde el 79% del francés fundamental -hablado-, hasta sobrepasar el 90% en la obra de ciertos autores.

1.3.4.1. Las leyes de la estructura léxica

La ley de Zipf, o de Estoup-Zipf[36], supone una tentativa de aproximación al conocimiento de las problemáticas que presenta la distribución de frecuencias en un texto, aunque también tiene que ver con la relación que se establece entre N y V, o relación que une la extensión del vocabulario con la extensión de un texto. Si agrupamos los vocablos en rangos, y ordenamos el vocabulario por orden de frecuencia, y se otorga el rango 1 al vocablo con mayor frecuencia, y así sucesivamente; y para los vocablos de igual frecuencia se toma el rango medio, la ley de Zipf establece que *frecuencia x rango es una constante*[37], independientemente del idioma o de la extensión del texto. Esto es así porque tres factores, al menos en lo que atañe a la obra de un autor, contribuyen a asegurar esta constante:

1. La elección de la norma léxica.
2. Razones gramaticales, lógicas y estéticas. Las relaciones que mantienen entre sí las frecuencias se deben, en cierto modo, a la gramática, la lógica y la estética, pues, por un lado, las palabras gramaticales son las de frecuencia más alta; la lógica, porque es necesaria

---

[36] Los estudiosos coinciden en aseverar que fue Jean-Baptiste Estoup quien primero ensayó esta ley, posteriormente desarrollada por Zipf, y mejorada por Mandelbrot. Así lo reconocen, por ejemplo, Guiraud, P., (1960: 75); Dugast, D., (1979: 83-89); Marcus, Nicolau y Stati (1978: 237-240).

[37] Michéa, R. (1973: 69-70) refiere que Zipf extraía una conclusión psicológica de esta constante, pues en ella veía un equilibrio entre la tendencia del emisor a la repetición de palabras, y a la exigencia de precisión del receptor. Tal observación la recoge también Malmberg, B. (1988: 122): "Según Zipf lo que se manifiesta en el efecto de dos tendencias. El hablante tiene una tendencia a repetir lo más posible las mismas palabras (a utilizar pronombres o palabras como *cosa*, *chisme*, *rollo*), por tanto, a salir del apuro con un mínimo de esfuerzo. El oyente por su parte exige un máximo de claridad. Al hablante le interesa satisfacer esa exigencia, y el vocabulario que efectivamente utiliza representa un equilibrio entre estas dos tendencias".

cierta coherencia en el texto, y la estética porque se utiliza el criterio de no repetir muchas palabras seguidas por considerar este hecho pobre y monótono (Michéa, R., 1972: 77).

3. La relación entre la obra y el número de V o voces distintas.

No obstante, esta ley no tiene la generalidad que Zipf le atribuye, como se puede comprobar en numerosos textos en los que no se cumple[38], porque conlleva una serie de irregularidades, relacionadas sobre todo con las frecuencias más altas y las más bajas. Pero, como destaca Guerra Salas (1989: 25), la ley de Zipf también "traduce una tendencia del léxico que explica las teorías de la información, y que ser relaciona también con la estructura de la memoria". Guiraud no está de acuerdo con que mida la distribución de frecuencias de un vocabulario, sino del "léxico" de donde se extrae el vocabulario: "Pour moi la distribution est un caractère du lexique du texte, c'est à dire de l'ensemble des mots du stock mémoriel d'où sont tirés les mots du texte" (Guiraud, P., 1960: 87); por lo que propone una nueva modificación a esta fórmula para un texto dado: $r^{a'} \times f = constante$, en la que a' es una variable que depende de la longitud del texto.

El índice de Yule[39], mejorado por Herdan[40], y la fórmula de Waring-Herdan - o intento de Gustav Herdan de aplicar la fórmula de Waring, matemático inglés del XVIII, al cálculo de efectivos de listas de frecuencias extraídas de

---

[38] Esta afirmación es de Michéa, R. (1973: 70-73). A su juicio, la ley de Zipf se verifica cuando hay un cierto equilibrio entre palabras gramaticales y palabras léxicas.

[39] Yule, G. U. (1968).

[40] Herdan, G., *Language as Choice and Chance*, Groningen, 1956. Citado por Guiraud, P. (1960: 86).

obras literarias[41]-, suponen un paso más para el conocimiento de las leyes de la estructura léxica de un vocabulario.

1.4. Clases de vocabulario

Desde perspectivas tanto lingüísticas como cuantitativas se suelen distinguir varios tipos de vocabulario, aunque nosotros aludiremos a aquellos que entroncan más directamente con nuestra investigación. Por ejemplo, J. F. Pastora Herrero (1990: 62-63) distingue el vocabulario frecuente del vocabulario usual; con el primer concepto se refiere a aquel vocabulario que se cuantifica cuando se analiza un discurso, y llama vocabulario usual a aquel que ponen en práctica la mayoría de los hablantes. Se reconocen igualmente los vocabularios técnicos o específicos, clasificándolos según la materia o disciplina a que hacen referencia, o vocabularios que responden a consideraciones diatópicas porque recogen los términos usados en una determinada zona geográfica, etc.; también se alude al vocabulario activo, pasivo e incluso potencial "compuesto por aquellos vocablos que el alumno no comprende cuando los oye o lee, pero mediante el aprendizaje pueden engrosar su caudal lingüístico" (Sáez Pérez, I. E., 1984: 309); y dentro de esta distinción se hace una segunda matización entre aquellos términos cuyo significado intuyen los alumnos -"les suena"- y las palabras que desconocen totalmente. Pero la terminología lingüística recoge a menudo varios conceptos: vocabulario usual, común, básico y fundamental.

---

[41] Nos referimos a la obra de Herdan, G., *Quantitative Linguistics*, 1964. Esta fórmula ha sido igualmente aplicada por otros investigadores, como Müller, Ch. (1965a) y Ratkowsky, D. A. (1979).

1.4.1. Vocabulario usual, común, básico, fundamental

De Mauro (1982: 103-105) compara el léxico de una lengua con una gran esfera compuesta por distintas capas o estratos de vocabulario. En el nivel más superficial coloca a los hápax o palabras que aparecen una sola vez (vocablos de frecuencia 1) y los términos de los lenguajes especializados que no tienen una transcendencia más allá de sus diferentes dominios. Un estrato más interno está constituido por el "vocabulario común", integrado por aquellas palabras tanto especializadas como regionales e incluso propias de otras lenguas pero que han trascendido de sus dominios y son usadas y conocidas por cualquier hablante. Por ejemplo, palabras técnicas como *ecuación*, *penicilina*, *eclipse*, *preposición*, *afluente*..., o palabras que en castellano se emplean y pertenecen en realidad a otras lenguas de la península, como *morriña* -gallego-, *esquirol* -catalán-. Además, el vocabulario común, comprende, según De Mauro, otros dos estratos más internos: el vocabulario básico y el vocabulario fundamental. El vocabulario básico estaría integrado por aquellos vocablos del vocabulario común "ampliamente conocidos por personas pertenecientes a las más variadas categorías. Hay diferentes criterios para recortar el 'vocabulario básico'. El objetivo es, de todos modos, separar el conjunto de vocablos que resulten conocidos con seguridad por la generalidad de quienes han frecuentado la escuela al menos hasta tercer curso de enseñanza media. Se trata de un conjunto compuesto por unos miles de vocablos" (De Mauro, 1982: 104). El núcleo más interno de la esfera lo forma el vocabulario fundamental. Éste lo componen las palabras de máxima frecuencia en el habla y en la escritura y lo conocen todos los hablantes de un idioma que han cursado estudios elementales. Para nuestro idioma, el vocabulario fundamental "es el conjunto de las palabras que conocen todos los que tienen un conocimiento y una práctica al menos elemental del castellano" (De Mauro, 1982: 105).

Vidal Lamiquiz (1989: 225-226) habla de *vocabulario fundamental* (referido a un grupo de individuos) como la suma de los vocabularios individuales de una muestra de ese grupo, y *vocabulario básico*, a las unidades léxicas *comunes* a esa muestra de individuos, es decir, aquellas unidades léxicas que usan todos los individuos de ese grupo. Bonifacio Jiménez (1986: 79) se sitúa en el nivel de la lengua y alude a los cientos de miles de términos de los que se compone el léxico de un idioma, por lo que surge la necesidad de establecer subconjuntos cuantitativamente reducidos, como el léxico fundamental, compuesto a su juicio entre 7.500 y 15.000 palabras, que son las que realmente se usan. Mesanza López (1990a: 23) distingue entre vocabulario básico pasivo, seleccionado por el sistema de jueces y que debe ser conocido por el destinatario, y el vocabulario básico activo, extraído de una muestra representativa y por tanto es el que usa el destinatario.

Aunque en los ejemplos precedentes existe una diferenciación clara entre léxico y vocabulario, a veces se usan indistintamente uno u otro término para nominar a algunos de los tipos de vocabulario que venimos enumerando, como así lo indican por ejemplo Dubois y otros (1992: 628): "En la terminología lingüística, un vocabulario es una lista exhaustiva de las ocurrencias que figuran en un corpus. A pesar de todo, la oposición entre *léxico* y *vocabulario* no se respeta siempre: en expresiones como vocabulario de base, vocabulario común, vocabulario general, nada indica si las palabras que figuran en la lista figuran en tanto que ocurrencias tomadas de un corpus, o en tanto que unidades de la lengua.

Por otra parte, no todos los lingüistas estructuralistas mantienen esta oposición. L. HJELMSLEV emplea indiferentemente los términos de *léxico* o *vocabulario*. Sin embargo, es conveniente oponer el léxico como conjunto de las unidades de la lengua y el vocabulario como lista de las unidades del habla. Por ejemplo, la estadística léxica, que trabaja sobre las ocurrencias halladas en un corpus y, por lo tanto, sobre el vocabulario de un texto, de un

autor, de una época, busca deducir de ellas las potencialidades léxicas (el léxico)".

Desde perspectivas cuantitativas se utiliza una terminología similar, pero ante todo se mencionan cinco conceptos: vocabulario (o léxico) usual, vocabulario (o léxico) común, vocabulario (o léxico) fundamental, vocabulario (o léxico) básico y vocabulario (o léxico) disponible[42]. Los cuatro primeros son vocabularios frecuentes y usuales, en el sentido de que se extraen según el uso que los hablantes hacen de su léxico y se utiliza la frecuencia como criterio para su delimitación. El vocabulario disponible, incluso basándose también en la frecuencia, tiene matices especiales que comentaremos en otro lugar. Los vocabularios que venimos denominando "usual, común, básico y fundamental" se determinan de un modo más o menos preciso ayudándose de la frecuencia de las palabras en un texto y a veces de otras medidas. López Morales (1987: 57-66) es muy preciso en sus apreciaciones porque define con una gran exactitud los conceptos: utiliza, no obstante, el término de "vocabulario común" para referirse a "los recuentos de vocabulario hechos a base de frecuencia", sin utilizar ninguna otra medida; en cambio, cuando se trata de delimitar el léxico de uso real de una población determinada, utiliza el término "léxico básico", porque en su determinación entran en juego medidas muy precisas: además de la frecuencia, se tiene en cuenta la dispersión -o número de mundos (partes temáticas) en que aparecen los vocablos-, y a través de una fórmula matemática, se pondera frecuencia y dispersión y se obtiene el índice de uso. Por ejemplo, y para el idioma español, López Morales se refiere con este calificativo al trabajo de Juilland y Chang-Rodríguez (1964)

---

[42] Además de estos términos, es habitual encontrarse con otras denominaciones, como "listas de frecuencias, diccionarios de frecuencias", y también hallamos otro concepto: "vocabulario de reconocimiento" (Justicia, F., 1995: 8-9) aunque este término responde más bien a la técnica que se emplea para determinar los niveles de comprensión que presenta una determinada población (normalmente infantil) de un vocabulario o diccionario estándar.

y al de Amparo Morales (1986), cuyo título integra la denominación de "léxico básico". Guerra Salas (1989), siguiendo al lingüista puertorriqueño, utiliza también este término para referirse a estos trabajos. López Morales (1987: 62) habla igualmente de "léxico disponible" para referirse al "caudal léxico utilizable en una situación comunicativa dada". Prefiere el término "léxico" porque para su establecimiento usa, además de la frecuencia, el grado de espontaneidad o disponibilidad -el orden de aparición de las palabras-; teniendo en cuenta frecuencia y grado de espontaneidad calcula el índice de disponibilidad, criterio definitivo para fijar el léxico disponible. La suma de ambos léxicos, básico y disponible, constituye el vocabulario fundamental[43].

Otros autores operan de modo distinto. García Hoz (1953a: 15-16), tras aludir a los conceptos de léxico activo y léxico latente, tal y como los emplea J. Casares, hace la siguiente reflexión: "Léxico activo y léxico latente forman parte del que puede llamarse <<vocabulario usual>>, en contraposición al vocabulario total. Parece que las palabras solamente conocidas no deberían incluirse en el vocabulario usual, puesto que, aparentemente, son las no usadas; mas, para que alguien las comprenda, han de ser previamente pronunciadas o escritas; por tanto, aquellas palabras empleadas en escritos que puedan tenerse por generales han de considerarse dentro del vocabulario usual". Afirma el autor español que el mejor procedimiento para determinar este vocabulario sería ir registrando las palabras corrientes que dice en la vida el hombre corriente, así como las que éste lee o escribe; pero ante la imposibilidad de llevar a cabo tal procedimiento, propone acudir al material escrito para hallar en él el vocabulario usual. Esta premisa guiará la selección de las fuentes para su trabajo[44]. García Hoz (1953a: 22) se plantea también los

---

[43] López Morales, H. (1987: 63) define así este vocabulario: "El vocabulario fundamental de una lengua estará dado por la combinación o la suma del léxico usual y del léxico disponible".

[44] Véase García Hoz (1953a: 16-20).

límites entre palabras usuales y no usuales en los siguientes términos: "Entre los extremos, claramente apreciables, de las palabras muy corrientes y las prácticamente desconocidas hay una amplísima zona intermedia, donde habrá que poner el límite entre las palabras usuales y no usuales; mas el lugar exacto de este límite es imposible de señalar de un modo absolutamente objetivo". Él se basará para ello en la frecuencia media de los vocablos[45]. Siguiendo a García Hoz, Lorenzo Delgado 1981: 37, 53) relaciona igualmente la usualidad con la frecuencia. Para él, un vocablo de frecuencia absoluta 1/100.000 se considera de nivel usual. La usualidad de un vocablo, en su trabajo, viene determinada por la frecuencia o número de veces que aparece. Establecido así el concepto de vocabulario usual, identifica frecuencia media (producto de N/V) con usualidad media. A Fernando Justicia (1985a: 24 y 28; y 1995: 19 y 23) la frecuencia media le sirve también para considerar el grado de usualidad de una palabra[46]. Los conceptos de vocabulario común y

---

[45] García Hoz (1953a: 22-26) va explicando el proceso de selección del número de palabras para su recuento: en un principio registró 60.000 de cada fuente (recordemos que son 4 fuentes), con lo cual obtuvo 240.000, pero como necesitaba un corpus mayor para aproximarse al número de vocablos deseado (aproximadamente 15.000), siguió ampliando 10.000 palabras por cada fuente, hasta que al llegar a 100.000, observó que los vocablos nuevos eran muy escasos: era necesario contabilizar una media de más de cien palabras para hallar un vocablo nuevo. Esto le lleva a la siguiente aseveración: "A mi modo de ver, cuando las palabras tienen una frecuencia media de 100, la que no haya aparecido difícilmente podrá decirse usual". (La cita es de la p. 25). Planteado de esa manera, parece una frecuencia media incluso excesiva -100-, pero en realidad, si operamos matemáticamente, y determinamos la frecuencia media de su trabajo de acuerdo con lo estipulado en estadística léxica, es decir, aplicando la operación N/V, obtendríamos el resultado siguiente: N = 400.000; V = 12.913; frecuencia media = 30'97.

[46] Así lo recoge, por ejemplo, en su primer vocabulario (Justicia, F., 1985a: 24): "... la razón P/V (cociente entre número total de ocurrencias -N- y número total de vocablos -V-) refleja, por un lado, el grado de riqueza o diversidad léxica que contiene un vocabulario, por otro, la podemos considerar como indicador de la usualidad de las palabras ya que la razón P/V lo que nos indica en realidad es la frecuencia media, en términos absolutos, que tienen los vocablos en un vocabulario dado". La aclaración entre paréntesis es

vocabulario fundamental también son definidos con relación a la frecuencia, según la intersección entre varias fuentes o mezclando ambos criterios[47].

1.4.2. Vocabulario disponible

Se ha podido comprobar que además del vocabulario frecuente, usual, es decir, aquel que normalmente se contabiliza y hace acto de presencia en investigaciones cuantitativas, existe otro que es igualmente esencial, y además muy conocido, pero que el hablante no tiene necesidad de utilizar sino en determinadas situaciones comunicativas. Es lo que se llama vocabulario disponible, aquel que está a disposición de los sujetos en circunstancias determinadas porque va asociado a variantes temáticas[48]. Para registrarlo se presenta a los sujetos una serie de temas o centros de interés

---

nuestra.

[47] García Hoz (1953a) selecciona el vocabulario común de acuerdo con el principio de dispersión, es decir, que los vocablos al menos tengan una aparición en cada una de las cuatro fuentes (cartas, periódicos, documentos oficiales y libros); el vocabulario fundamental está formado por los vocablos que tienen sus frecuencias prácticamente distribuidas en partes iguales los distintos tipos de fuentes. Lorenzo Delgado (1981: 37) presenta su vocabulario "común" como aquel que aparece en las distintas fuentes o tiene en alguna de ellas una frecuencia entre 10-100, y si la frecuencia es superior a cien, lo considera vocabulario fundamental. Justicia, F. (1995: 243-255) incluye en esta clase de vocabulario las palabras comunes a los tres niveles de EGB: Inicial, Medio, Superior.

[48] López Morales (1987: 62-63) se refiere a este vocabulario del modo siguiente: "La disponibilidad léxica pasó a entenderse como el caudal léxico utilizable en una situación comunicativa dada; era preciso no confundir las situaciones frecuentes con el léxico frecuente, ni éste con el léxico fundamental. Porque no cabe duda de que mientras algunas palabras aparecen continuamente en cualquier conversación o escrito, no importa su tema, otras sin embargo están muy relacionadas al asunto: es poco probable que, a no ser que se hable de los medios de transporte público, se mencionen palabras como *autobús*, y que salgan en nuestra expresión palabras como *carta* y *sello* si nos referimos específicamente a asuntos del correo".

en torno a los que deben escribir o pronunciar las primeras palabras que acudan a la mente. En estas pruebas, se suele limitar bien el número de palabras (20 por ejemplo) o el tiempo (4 ó 5 minutos). El concepto de disponibilidad léxica se atribuye a los autores de *L'elaboration du français fondamental*[49]. En su trabajo, observaron que algunas palabras, como *metro, autobús*..., alcanzaron una frecuencia tan baja que obligaría a eliminarlas del francés fundamental, algo impensable pues eran palabras muy conocidas, por lo cual decidieron compensar la frecuencia con el vocabulario disponible. Gougenheim, Michéa, Rivenc y Sauvageot (1964: 137-145) señalan que en una lista decreciente de palabras podemos apreciar el siguiente orden: 1. en primer lugar, palabras gramaticales; 2. verbos; 3. adjetivos; 4. algunos nombres de carácter general como *cosa*, *persona*, *niño*... En cambio, no aparecen apenas palabras concretas, cuya frecuencia además es muy inestable, debido a que están ligadas a ciertas circunstancias, a ciertos temas de conversación. Pero de hecho existen voces como *tenedor*, *codo*, *diente*, *falda* que no tienen una frecuencia muy alta pero que son palabras conocidas incluso por niños de tres años. ¿Cómo denominar pues a este tipo de palabras? Deciden denominarlas "palabras disponibles"; es decir, palabras "d'une fréquence faible et peu stable, que son cependant des mots usuels et utiles" (Gougenheim, Michéa, Rivenc y Sauvageot, 1964: 145) Los autores del francés fundamental distinguen pues claramente entre dos tipos de vocabulario: el frecuente y el disponible. Éste consta sobre todo de sustantivos "concretos" y a juicio de los lingüistas franceses mencionados, estos términos son más expresivos e interesantes para los encuestados. Además, realizaron una serie de pruebas en distintos institutos sobre el tema "viaje en tren", y comprobaron que los sustantivos concretos aparecían con bastante regularidad, mientras que los verbos, por ejemplo, eran mucho menos numerosos y mucho más inestables. No aparecían apenas palabras gramaticales, adjetivos o adverbios (Gougenheim, Michéa, Rivenc y Sauvageot,

---

[49] Gougenheim, Michéa, Rivenc y Sauvageot (1964).

1964: 151). Y esto conduce a Michéa a establecer la siguiente conclusión: "En présence d'une *situation donnée*, les mots qui viennent les premiers à l'esprit sont ceux qui sont liés tout spécialement à cette situation et la caractérisent, c'est-à-dire les noms. Les autres parties du discours ne se présentent à l'esprit que dans la mesure où il s'agit d'exprimer les rapports entre les choses. La mémoire, basée sur l'association des idées, est sélective: elle n'obéit pas à una loi probabilitaire simple" (Gougenheim, Michéa, Rivenc y Sauvageot, 1964: 151). Al mismo tiempo, los autores concluyen que mientras en una lista de frecuencias los verbos son a menudo estables y los nombres concretos inestables, las encuestas sobre la asociación de ideas alrededor de un centro de interés establecen sobre todo la estabilidad de los nombres concretos y la inestabilidad de los verbos. Ello los conduce a considerar no sólo la frecuencia sino otros aspectos para seleccionar el vocabulario base que servirá para la enseñanza del francés como lengua extranjera, en concreto, la conjunción o suma del vocabulario frecuente y vocabulario disponible. Para determinar éste último de un modo preciso fijan otra medida: el grado de disponibilidad, que "correspond à la présence plus ou moins immédiate de ces mots dans notre mémoire" (Gougenheim, Michéa, Rivenc y Sauvageot, 1964: 152). Para los autores del francés fundamental, el grado de disponibilidad será el equivalente a la frecuencia (Gougenheim, Michéa, Rivenc y Sauvageot, 1964: 189-194). Posteriormente, además de la frecuencia se han usado otros criterios para determinar el grado de disponibilidad[50].

Otros lingüistas (Mackey y colbs., 1971: 17-23) profundizan igualmente en el concepto de disponibilidad léxica. Son conscientes de que los criterios de frecuencia, dispersión (o número de fuentes donde aparece una palabra) y

---

[50] Otros autores establecen otros índices de disponibilidad. Por ejemplo, Mackey y colbs. (1971: 55-59) acuden al porcentaje del número de sujetos que usan la palabra en Francia y en Acadia; López Morales da, como hemos visto, otro índice: pondera la frecuencia con el orden de aparición de las palabras.

valencia (palabras que incluyen a otras, como *asiento*, *habitación*; que son más generales y por tanto poseen más capacidad de usarse en definiciones, como *persona*, *cosa*; o que tienen mayor número de acepciones, como *hacer*, *tomar*) sirven para instaurar las palabras comunes (léxico común) de una lengua, pero hay otras, también esenciales, que dependen del tema del que se hable (nombres concretos en su mayoría) y que pueden quedar fuera de los estudios de los léxicos básicos. Son las palabras que constituyen el vocabulario disponible: "L'importance des mots concrets ne se révèle point par leur fréquence, leur répartition ou leur valence. Ce sont des mots comme poivre et sel, dont l'occurrence dans une liste de mots fréquentes, comme nous venons de le signaler, est entièrement aléatoire et dépend du sujet ou des textes choisis. Ce sont des mots qui nous viennent à l'esprit quand nous en avons besoin, et c'est pourquoi on les appelle des mots disponibles. Ce sont les mots que nous avons parfois au bout de la langue" (Mackey y colbs., 1971: 23). Las palabras disponibles son, pues, palabras temáticas, ligadas íntimamente al asunto de un discurso. Puede por ello pasar largo tiempo sin que se utilicen pero son también palabras conocidas, a veces muy familiares e integran también el vocabulario común. Estas palabras tienen una frecuencia inestable, como indicábamos, pero una disponibilidad muy estable (Mackey y colbs., 1971: 24-25). Los tres índices mencionados por los citados autores franceses (frecuencia, repartición y valencia), a los que añaden el de disponibilidad, no se reparten por igual en todas las palabras. Una palabra como *hacer* tiene un índice elevado de frecuencia, repartición y valencia, pero su índice de disponibilidad es muy bajo. (Mackey y colbs., 1971: 26).

Quizá el mayor problema que plantea el vocabulario disponible es la elección de los centros de interés, pues difícilmente se pueden incluir todos los que son de interés humano en los temas que se delimitan en las encuestas (Mackey y colbs., 1971: 30). No obstante, siempre se siguen algunos criterios, como partir de un empirismo racional para fijar los centros de interés representantes de los sectores semánticos más universales (como se hizo en

el francés fundamental, donde también se introdujo "la escuela" por familiaridad con los sujetos encuestados), o teniendo en cuenta los temas más interesantes para los escolares[51].

Para algunos autores, no hay que asimilar el vocabulario disponible a vocabulario pasivo ni vocabulario frecuente a vocabulario activo, pues el disponible se puede usar en cualquier momento por el hablante (Pastora Herrero, J. M., 1990: 62). Quizás sea porque cuando se actualiza la lengua, en un discurso determinado, el vocabulario no está dado de antemano, aunque, quizás, se fije el tema que deba desarrollarse, porque las palabras van apareciendo a medida que el autor va componiendo su discurso, sometido este a cambios espontáneos en el escrito[52]. Por otro lado, en varios trabajos sobre la frecuencia de las palabras se alude a este tipo de vocabulario, bien intuitivamente, como, a nuestro juicio, hace Buchanan[53], o indirectamente, como hacen otros autores (Chetouanui, L., 1989: 94-96).

---

[51] López Morales, H. (1978: 77 y1987: 80) alude al trabajo de disponibilidad léxica llevado a cabo por Rodríguez Bou con preescolares, a los que estimuló con láminas cuyos motivos contemplaban los 37 temas preferidos de los alumnos de la escuela primaria. También Díaz Castañón y colbs. (1977) seleccionan centros de interés que resultan familiares a los escolares.

[52] Para Michéa, R. (1972: 73), en oposición al vocabulario disponible, que depende del tema, o "cadre associatif donné", en una obra literaria las cosas son distintas: "Mais une oeuvre littéraire à l'état de project est loin d'être un cadre aussi bien délimité et aussi contraignant. Le vocabulaire utilisable n'est pas donné dès le début de la rédaction. Il se constitue progressivement au cours de celle-ci, un mot actualisé en entraînant d'autres avec lesquels il se trouve lié dans la mémoire, una changement d'orientation dans les associations d'idées faisant surgir du fond de l'inconscient des termes auxquels l'auteur n'avait tout d'abord pas pensé. Nous avons affaire à une réalité dynamique et nos statique, à un pouvoir, une force, variable selon les individus". Aunque se refiere a la obra literaria, se podría relacionar con cualquier composición o actualización en el discurso.

[53] Buchanan, M. (1927: 7) hace la siguiente observación: "In all extensive word-counts, whether English, French, German or Spanish, words of the home, like "spoon", or of the school, like "pencil", have a

Actualmente se admite la distinción entre vocabulario frecuente y vocabulario disponible. Ambos criterios no son opuestos sino complementarios, y a la vez necesarios para la enseñanza de la lengua[54].

1.5. Estudios sobre el vocabulario desde perspectivas cuantitativas. Breve visión histórica

1.5.1. Vocabularios de carácter general

Aunque la afición de componer vocabularios de diversa índole comenzó en la Antigüedad[55] y desde entonces se sintió curiosidad por aspectos

---

surprisingly low frequency, although within their restricted range they are obviously important words".

[54] Véase López Morales (1987: 63, 77); Etxebarria, M. (1996: 310).

[55] Alvar Ezquerra, M. (1983: 153-157) hace un resumen del libro de Gangutia Elicegui, E. *Introducción a la lexicografía griega*, Madrid, CSIC, 1977, donde se expone el nacimiento de la lexicografía en el Próximo Oriente y los primeros trabajos léxicos en Grecia, así como el nacimiento de la lexicografía propiamente dicha en Alejandría, con *glosógrafos* como Filetas de Cos y, algo más tarde, Aristófanes de Bizancio, con quien se eleva la glosografía a la categoría de lexicografía. Coincide con estas apreciaciones Arias, R. (1986: 743-744), quien señala que algunos escritos de los escribas acadios, allá por el año 2600 a. C., "abrigan técnicas de diccionario"; y que en el mundo griego, ya con Platón y Aristóteles, y con Demócrito, se apuntan conatos de saberes lingüísticos que pueden considerarse como "acontecimientos lexicográficos precursores de los glosarios", aunque, a su juicio, la verdadera lexicografía comenzó con los alejandrinos, como Zenódoto de Éfeso, Aristófanes de Bizancio y Aristarco de Samotracia. También habla Luque Durán, J. de D. (1996: 1-18) del nacimiento de los protodiccionarios en Acadia, aunque matiza que en realidad se trata de listas de palabras que los alumnos acadios componían con términos sumerios en una columna y sus correspondientes acadios en otra; Fernández, A. R., Hervás, S. y Báez, V. (1989: 166-177) refieren igualmente la producción de listas de signos gráficos en China o listas de palabras en otras sociedades, ya desde tiempos remotos.

cuantitativos del lenguaje[56], hemos de centrarnos en aquellos trabajos que de alguna manera contemplan la frecuencia de las palabras u otros datos estadísticos en su elaboración. Existen noticias concretas de autores interesados en los aspectos matemáticos del lenguaje, como V. I. Bunjakovskij[57], quien sugirió, ya en el siglo XIX, que la teoría matemática de probabilidades se podría utilizar para destacar matices importantes: conocer la extensión de un vocabulario, su distribución en clases de palabras, la frecuencia de combinaciones de letras...

Se afirma que los estudios cuantitativos sobre vocabulario con fines lingüísticos son muy abundantes[58], aunque sus resultados suelen ser casi siempre insatisfactorios y por ello se suceden con cierta regularidad al cabo de los años[59]. Los trabajos sobre la frecuencia de las palabras o vocabularios de frecuencias siguen, para la selección de su material, técnicas propias de la

---

[56] Guiraud, P. (1960: 5) afirma: "La statistique linguistique es l'héritière d'une tradition très ancienne: les grammairiens alexandrins avaient déjà dressé la liste des hápax legomena d'Homero et les massorètes dénombré tous les mots de la Bible". Y Michéa, R. (1972: 75) recuerda que el análisis cuantitativo trozo a trozo es viejo, y pone como ejemplo una versión alemana del Nuevo Testamento, traducción holandesa de 1698, en la que figuran al final de cada capítulo las palabras nuevas aparecidas y las que aparecen sólo una vez.

[57] Serrano, S. (1975: 285) cita a Bunjakovskij, y su artículo "O Vozmognosti vedenia opredelitelnij mer doveria k pezultatam nekotorij nank nabliudatel nij y preimushetvenno statistiki", en *Sobretennik* III, 1847, otd. II, pp. 36-49.

[58] Véase Morales, A. (1980: 467-474). Rodríguez Bou, I. (1952: 3) cita a Edgar Dale, *Bibliography of Vocabulary Studies*, 1949, donde aparecen, de 1.855 títulos, 84 sobre estudios de frecuencia de palabras. Guiraud, P. (1954) recoge un capítulo dedicado a la "Fréquence des Mots", donde hemos contabilizado 108 títulos, de los cuales sólo 5 se refieren a la lengua española.

[59] Afirmaciones de Morales, A. (1980: 468).

labor lexicográfica[60], pues deben partir de un corpus o selección de fuentes o textos normalmente de diversas características tanto estilísticas como diastráticas y diatópicas, para abarcar todas las posibilidades expresivas de la lengua y por tanto el resultado del trabajo sea representativo. Si bien, otras técnicas son completamente distintas, debido, claro está, a la finalidad de unos trabajos y de otros.

Los estudios estadísticos de vocabulario con propósitos descriptivos han dejado en evidencia el poco caudal léxico que manejan los usuarios de una lengua: se estima que el hombre culto puede manejar entre cuatro mil y cinco mil palabras, y el hombre "común", unas dos mil, cifras que contrastan con otras más optimistas[61].

Tomando como punto de referencia los distintos postulados que se tienen en cuenta en la confección de los recuentos, desde los objetivos, el sistema de muestreo, las fuentes o tipo de material seleccionado, las pruebas de obtención de datos, el "corpus" o número total de palabras, los criterios de clasificación de las palabras, los cálculos para establecer el rango, etc., se podrían hacer múltiples clasificaciones porque casi todos los trabajos, si bien tienen en común establecer un vocabulario, difieren de los demás en algo. Por tanto, y para tener una visión ligera sobre el tema, conviene empezar por una relación histórica y brevemente explicativa de algunos trabajos que creemos

---

[60] Noticias de los pasos que se siguen para seleccionar el material que servirá de base del despojo lo tenemos en varios trabajos. Algunos son de Seco, M. (1979: 2-7; 1987: 221-235;1988: 135-142); véase también el capítulo sexto de Zgusta, L. (1971).

[61] López Morales, H. (1978: 73) aduce que las cifras que se barajaban antes eran mucho más exageradas, por ejemplo, Bloomfield creía que un hombre culto disponía de 25.000, cifra que contrasta con estudios de escritores como Baudelaire, que no usó más de 4000 vocablos en *Les fleurs du mal*.

pueden ser los más representativos a raíz de las coincidencias bibliográficas en que se citan.

El primer recuento de vocabulario de importancia que se cita se realizó en lengua alemana para uso de la taquigrafía y basado ya en la frecuencia de las palabras: nos referimos al monumental trabajo *Häufigkitswörterbuch der Deutschen Sprache* (1898), de Kaeding, con un corpus de 10.910.777 ocurrencias[62]. Este trabajo se destinó, no a la enseñanza de lenguas, sino al establecimiento de una estenografía racional. Kaeding seleccionó su material de distintas fuentes: debates parlamentarios, textos administrativos y comerciales, prensa, escritos de teología y medicina, historia, literatura clásica alemana y extranjera, cartas personales, documentos militares[63]... Las investigaciones cuantitativas de vocabulario con fines pedagógicos se atribuyen a trabajos en lengua inglesa[64], destacando como más influyente *The teacher's Word Book*, (1921), de E. L. Thorndike, donde se recoge una lista de

---

[62] Las referencias de este trabajo las tomamos de Buchanan, M. (1927: 3), Keil, R. D. (1965: 103) y Müller, Ch. (1973a: 281).

[63] Datos aportados por Gougenheim, Michéa, Rivenc y Sauvageot (1964: 31).

[64] Buchanan, M. (1927: 3) cita el trabajo de Leonard P. Ayres, *Measuring Scale for Ability in Spelling*, New York, 1915, como el primero que se confeccionó con fines pedagógicos; en concreto, para intentar establecer las 1.000 palabras más frecuentes, utilizando sobre todo un material que provenía de correspondencia personal y de negocios; Rodríguez Bou (1952: 5) refiere que los trabajos en inglés sirvieron de referencia para otros idiomas, como el francés, para el que Henmon confeccionó uno de los primeros trabajos: *A French Word Book Based on a Count of 400.000 Running Words*, Bureau of Educational Research Bulletin, nº 3, September (1924); trabajo superado por Vander Beke, *French Word Book* (1929), en cuanto al número de palabras base del recuento y al número de fuentes utilizadas.

10.000 vocablos extraídos de un corpus de 4.500.000 ocurrencias, procedentes de textos diversos[65]. Posiblemente en este trabajo se inspiraron otros autores para estudiar otras lenguas. Por ejemplo, V. A. C. Henmon publica en Estados Unidos su *A French Word Book Base on a Count of 400.000 running Words*, 1924, lista de 5.000 palabras que extrae de libros de literatura francesa destinados a la lectura en clases de enseñanza secundaria. Más tarde Vander Beke edita *A French Word Book* (1935), en el que, además del corpus de Henmon, tiene en cuenta revistas y periódicos, textos de historia, biografía, civilización, crítica, ciencias y filosofía[66].

Con la intención de obtener corpus más amplios para seleccionar de una modo más objetivo las palabras, es frecuente fundir varias listas de palabras ya existentes[67], o mezclar el material extraído de letra impresa y de expresión escrita espontánea de una determinada población[68].

---

[65] Rodríguez Bou (1952: 4) nos informa de que se utilizaron 41 fuentes distintas de las que se incluyeron alrededor de 625.000 palabras procedentes de la literatura infantil, 3.000.000 de la Biblia y de los clásicos ingleses, 50.000 de libros de cocina, costura, agricultura, negocios...; 90.000 de periódicos y 500.000 de correspondencia.

[66] Datos aportados por Gougenheim, Michéa, Rivenc y Sauvageot (1964: 32-38).

[67] Rodríguez Bou (1952: 4-5) cita dos trabajos que funden recuentos ya existentes. El primero de ellos pertenece a Faucett y Maki, *A Study of English Word Values Statistically Determined from the Latest Extensive Words-Counts* (1932), donde se combinan las listas de Thorndike y las de Ernest Horn; y el segundo, de Buckingh and Dolch, *A Combyned Word List* (1936): éstos últimos suman a su recuento varios más de múltiples autores. También el *Vocabulaire Fondamental du français* de MM. Robert Dotrens y Dino Massarenti se basan en distintas listas del francés ya existentes (Para más información sobre este trabajo, véase Gougenheim, Michéa, Rivenc y Sauvageot, 1964: 54-56).

[68] Rodríguez Bou (1952: 7) cita el trabajo de T. R. Céspedes, *Investigación acerca de las palabras usadas en castellano*, Panamá, Star and Herald, (1928), en el que se recopila material procedente de

Otras dimensiones que adquieren este tipo de investigaciones son, además de lingüísticas y pedagógicas, de índole sociológico y político. Los países con cierta hegemonía cultural se preocuparon por la difusión de sus lenguas más allá de sus fronteras, por lo que "se arbitraron procedimientos adecuados a las necesidades del mundo moderno para la difusión de sus lenguas respectivas. La mayoría de estos procedimientos radican sobre la idea de reducción y simplificación de los contenidos lingüísticos en general, y de la gramática y el vocabulario en particular. Por consiguiente, los estudios sobre vocabularios básicos o fundamentales aunque en su origen sean directamente dependientes de una necesidad de simplificación y adecuación de las distintas lenguas, indirectamente responden a la necesidad de mantener un predominio no sólo territorial sino también político, económico, comercial y cultural" (Justicia, F. y Palma, A.: 1986: 579). Es decir, surge la necesidad de componer trabajos de esta naturaleza para la enseñanza de lenguas extranjeras, lo cual supuso también un cambio de pedagogía en la didáctica de lenguas, cambio que fue asumido por los organismos oficiales que crearon comisiones específicas para el estudio del idioma. Como los primeros trabajos[69]

---

composiciones escritas de alumnos de Primaria y Secundaria, así como artículos de prensa, informes, conferencias, textos judiciales, etc.; el propio recuento de Rodríguez Bou (1952) incluye, además de textos escritos y orales, el trabajo de Buchanan y el de Casanova y Rodríguez, basado en composiciones de los alumnos de 3° y 8° de primaria. También otros idiomas tienen trabajos de frecuencias elaborados a partir un corpus extraído de textos de adultos y niños, como la lista de M. Aristizabal, establecida sobre un total de 460.727 palabras procedentes de cartas literarias (de autores como Balzac, Racine, Voltaire), correspondencia comercial, y redacciones de alumnos de 2°, 3° y 4° de primaria (Véase Gougenheim, Michéa, Rivenc y Sauvageot, 1964: 51-54).

[69] Gougenheim, Michéa, Rivenc y Sauvageot (1964: 26-29) refieren que el primer trabajo con fines de enseñanza a extranjeros fue el *Basic English*, ideado por Ogdens y Richars en 1920, y elaborado entre 1923 y 1927. En un principio se fijó la cifra de 850 palabras esenciales, pero el número definitivo quedó fijado en 1993. Este método se experimentó en Japón en 1933, y de ahí se extendió a más de 20 países. Aunque este trabajo y otros programas, como el *Defining Vocabulary*, de Michael West, fueron elaborados

seleccionaban el material de un modo más o menos intuitivo, pronto se presenta una nueva alternativa con los estudios de vocabulario basados en la frecuencia de aparición de las palabras. Uno de los primeros trabajos de este tipo fue *Le Français Élementaire*, aparecido en 1954, y gestado por una iniciativa de la UNESCO en 1947 para difundir las lenguas de cultura (Gougenhein, Michéa, Rivenc y Sauvageot, 1964: 9). Este trabajo se reeditó en 1959 con el título de *Français Fondamental (1er degré)*, concebido como un instrumento de difusión rápida de la lengua francesa: recoge un total de 1475 palabras, 1222 léxicas y 253 gramaticales. Entre ambas fechas apareció *L'Élaboration du Français Élementaire* (1956), y posteriormente, una reedición aumentada: *L'Élaboration du Français Fondamental* (1964). Éste ha sido establecido teniendo en cuenta tres aspectos: la frecuencia de las palabras de la lengua oral, las palabras disponibles (la noción de disponibilidad léxica es una innovación en el terreno de los estudios cuantitativos de vocabulario debida, como hemos indicado, a esta obra) y el empirismo racional (Gougenheim, Michéa, Rivenc y Sauvageot, 1964:13). La finalidad del francés fundamental es la enseñanza escolar, la difusión del francés como lengua nacional o extranjera privilegiada y la necesidad de ayudar al extranjero que llega a Francia como turista o como trabajador. El corpus de esta obra fue recogido, en gran parte, de manifestaciones de lengua oral[70], rompiendo así

---

siguiendo "principios lógicos y universalistas", no recogían el material de listas de frecuencias, sino que seleccionaban del diccionario aquellas voces que consideraban más usuales. Por ello recibieron sus críticas, sobre todo el trabajo de West, ya que la enseñanza propugnada por estos métodos conducía a expresarse de un modo no natural.

[70] Gougenheim, Michéa, Rivenc y Sauvageot (1964) seleccionaron una muestra definitiva de 275 sujetos: 138 hombres, 126 mujeres y 11 niños. Parte del material se registró a través del Centro de Estudios Radiofónicos, aunque la mayoría de las entrevistas se realizaron directamente en la zona de París y alrededores. También se recogen testimonios de otras zonas de Francia, e incluso de fuera, como Canadá, Suiza y Guayana. Las personas pertenecían a distintos niveles socioculturales y desempeñaban distintas profesiones, desde profesores, estudiantes, periodistas, escritores, músicos, administrativos..., a obreros, campesinos o amas de casa.

una tradición de trabajos anteriores que tomaban como fuente la lengua escrita.

En lo que atañe al corpus de palabras base de los recuentos, se aprecian también diferencias notables. Resulta curioso observar que los vocabularios más antiguos, que no contaban aún con máquinas informáticas y por tanto los trabajos se hacían de manera "artesana", ocupaban mayor número de ocurrencias[71]. Aunque muchos de estos trabajos no sólo se basaron en la frecuencia de las palabras para establecer las listas definitivas[72], lo cierto es que los corpus van disminuyendo desde que, más recientemente, y para seleccionar los vocablos, se presentan los procedimientos matemático-estadísticos más refinados, aportados por Juilland[73] y otros investigadores que toman como modelo teórico sus trabajos[74], o los procedimientos empleados

---

[71] Por ejemplo, los ya citados trabajos de Kaeding, con casi 11 millones de palabras; o de Thorndike (1921), con 4.565.000; o el de Rinsland, *A Basic Vocabulary of Elementary School Children* (1945), de 6.012.359 palabras; y el propio trabajo de R. Bou (1952), con 7.066.637 palabras.

[72] Buchanan (1927: 8) cita a Thorndike, Henmon y Stanford, quienes en sus trabajos ensayan un método para combinar repartición ("range") (o número de fuentes donde aparece la palabra) y frecuencia, dividiendo esta última por 10 y sumando 1 por cada fuente de aparición o "range". Es el sistema que sigue el propio Buchanan. Vander Beke en su *French Word Book* (1935) tiene en cuenta igualmente la repartición: elimina aquellas palabras que no aparecen en al menos 5 textos (referencias que tomamos de Müller, Ch. (1965b); y Gougenheim, Michéa, Rivenc y Sauvageot (1964) atienden a la repartición y la frecuencia para establecer el orden de importancia de las palabras.

[73] Nos referimos al *Frequency Dictionary of Spanish Word* (1964), de Juilland y Chang-Rodríguez, donde se tienen en cuenta la frecuencia, la repartición y el uso. Además de este trabajo, Juilland es coautor de otros diccionarios de lenguas románicas, como el *Frequency Dictionary of Rumanian Words* (1965), *Frequency Dictionary of French Words* (1971) y *Frequency Dictionary of Italian Words* (1973).

[74] Por ejemplo, y para el español, el *Léxico básico del español de Puerto Rico*, dirigido por Amparo

en estudios más ambiciosos que siguen la norma estadística para componer un diccionario general de lengua[75]. Excepción a esta regla la constituye el hasta hoy trabajo más espectacular en lo que al número de ocurrencias se refiere: el *Trésor de la langue française*, recuento que recoge textos, en su mayoría literarios, de un período que abarca desde 1789 a 1965, con un total de 70.317.234 ocurrencias y que ha arrojado un total de 71.415 vocablos[76]. Si bien estos trabajos extraen sus muestras de diversas fuentes de material escrito ya impreso[77].

Algunas obras de esta categoría, no obstante, también recogen muestras extensas de lengua oral: *L'Élaboration du français fondamental* es una de ellas. Gougenheim, Michéa, Rivenc y Sauvageot (1964), a través de diversas entrevistas, consiguieron un corpus de 312.135 ocurrencias y 7.995 vocablos. De éstos seleccionan los 1.063 que tienen una frecuencia igual o superior a 20

Morales, 1986; y el *Lessico di frequenza della lingua italiana contemporanea*, de Bortolini, Tagliavini y Zampolli (1972).

[75] Nos referimos al trabajo de Lara, L. F. y colbs. (1979: 37-38). Fruto de esta investigación ha sido una primera versión, reducida, publicada con el título de *Diccionario fundamental del español de México* (1982).

[76] Véase Ibms, P. (1971). El *Trésor de la langue française* ha sido elaborado teniendo en cuenta, en un principio, textos literarios -sobre todo en prosa, pues las ocurrencias en verso suman un total de 2.558.586- de los siglos XIX y XX. Exactamente del período comprendido entre 1789 y 1964. Este monumental diccionario, además, será completado con textos no literarios escritos y textos orales.

[77] La mayoría de los trabajos consultados y prácticamente casi todos los trabajos citados, recogen su material de fuentes escritas -aclaremos que el trabajo de Rodríguez Bou recoge también muestras de expresión oral- de muy diversa índole, para tener una muestra de palabras lo más representativa posible de todas las variedades idiomáticas, desde los textos literarios, hasta los jurídico-administrativos, periodísticos, técnico-científicos, hasta los de procedencia más variopinta, como telegramas, tratados de paremiología, libros de costura, cocina, agricultura, etc.

y que constituirán el vocabulario de base, el francés fundamental. Este vocabulario además se verá incrementado por el vocabulario disponible extraído a escolares de distintas regiones de Francia. Otro trabajo más reciente y basado en la lengua oral es el *Lessico di frequenza dell'italiano parlato*, (1994). Sus autores (De Mauro, Mancini, Vedovelli y Voghera) recopilan un corpus de 496.335 ocurrencias de un total de 1653 informantes de distintas ciudades (Nápoles, Roma, Florencia, Milán), edad (entre 6 años y más de 56) y nivel sociocultural. Extraen un total de 15.641 lemas, de los que 7.213 tienen una frecuencia superior a 3, que serán los que integren la lista definitiva de su trabajo.

Prácticamente todas las lenguas han sido objeto de estudios cuantitativos de vocabulario. Por ejemplo, para el ruso se cita el trabajo de Harry H. Josselson, *The Russian Word Count* (1953), en donde se recoge un corpus de 506.044 palabras, de 134 fragmentos literarios de un período de 118 años (1830-1948), repartidos así: 59% prosa, 7% drama, 20% periodismo, 14% de crítica literaria. Steinfeldt reunió 400.000 palabras de 230 muestras de literatura infantil soviética, 75% de novelas y 25% de teatro[78]. La lengua italiana cuenta con algunos trabajos emblemáticos. El primero de ellos es el *Lessico di frequenza della lingua italiana contemporanea* (LIF). Este trabajo sigue los pasos del *Frequency Dictionary of Spanish Words* y del *Frequency Dictionary of Rumanian Words*, de Juilland, tanto en el número de ocurrencias como en los criterios -frecuencia, dispersión y uso- para fijar los 5000 lemas previstos si bien con algunas variantes. Se parte igualmente de un corpus de 500.000 ocurrencias extraídas de cinco mundos distintos, pero éstos difieren en algo de los establecidos por Juilland. Los autores italianos establecen los siguientes mundos: teatro, novela, periodismo, cine y libros de texto[79]. Al final, tras la

---

[78] Los datos de ambos trabajos los aportan Marcus, Nicolau y Stati (1978: 285).

[79] Como señalan los propios autores del LIF (Bortolini, Tagliavini y Zampolli, 1972: 17-18), los cinco

desambiguación del corpus, se consiguieron 15.750 lemas, de los que se seleccionaron 5.356 -de acuerdo con las medidas de frecuencia, dispersión y uso- , que los autores presentan por orden alfabético y por orden decreciente de frecuencia. No sólo las lenguas europeas, sino las del resto del mundo han sido objeto de estudios cuantitativos, como se aprecia repasando las bibliografías de E. Dale[80] y de P. Guiraud (1954).

Otro aspecto se refiere al contenido del vocabulario, pues si bien la mayoría de estos trabajos establecen el vocabulario usual o disponible, también existen otros que tratan de detectar el vocabulario científico de distintas materias. Y más recientemente, se ha establecido un diccionario intermedio, digamos, entre los vocabularios científicos y el usual: nos referimos al "Vocabulario General de Enseñanza Científica" (VGES), que recoge el vocabulario no técnico, sino el que el profesor suele utilizar en la situación de enseñanza-aprendizaje como soporte para transmitir los contenidos científicos[81]. También el lenguaje comercial ha sido objeto de estudios estadísticos con fines didáctico-pragmáticos. Por ejemplo, el trabajo que Lyne (1975) llevó a cabo basándose en la correspondencia francesa de negocios. Logró reunir así un total de 670 documentos suministrados por la

---

grupos o mundos de los que recogen el corpus son "Teatro, Romanzi, Cinema, Periodici, Sussidiari", y muestran la relación de obras seleccionadas.

[80] Dale, E. y Razig, T., *Bibliography of Vocabulary Studies*. Columbus, Ohio, 1963. Esta edición, recogida por Justicia, F. (1995a) en su bibliografía, es más reciente que la citada por Rodríguez Bou (Cfr. nota 58).

[81] Véase Chetouani, L. (1989: 83-103). El autor define este diccionario como el material léxico relevante de la práctica docente que subyace en la enseñanza científica, y lo señala claramente del siguiente modo: "Parallèlement au discours tenu par le professeur en tant que spécialiste de sa discipline (exemple: *différence de potentiel / équation horaire / dérivée de x / oxyde de carbone*...) il existe un autre discours tenu par ce même professeur en tant qu'enseignant (exemple: *nous étudierons d'abord cela / rappelons que / alors vous pouvez noter ça / je récapitule / vous puvez l'appeler /déterminer d'abord / vous pouvez mettre déjà /*...)". La cita corresponde a la p. 85.

Confederación de Industrias Británicas y otros organismos, de donde extrajo un corpus de 78.261 ocurrencias y 3.499 vocablos. La finalidad era proporcionar a los hombres de negocios ingleses un vocabulario adecuado para relacionarse con sus homólogos franceses. Otros trabajos tienen como meta analizar las posibles diferencias lingüísticas existentes entre distintos periódicos, dirigidos éstos a distintos sectores sociales; es decir, ver si la extracción social a la que se dirige la prensa repercute en la variedad lingüística de las palabras utilizadas por determinados periódicos anglosajones[82].

Por tanto, y a lo largo de la historia del lenguaje, los distintos estudios o trabajos acerca del vocabulario han ido ensayando distintas técnicas para establecer sus recuentos, dependiendo sobre todo de los objetivos o fines propuestos y los planteamientos metodológicos que han considerado mejores para la elaboración definitiva, aunque en todos existen algunas constantes, debido a las propias características peculiares de los recuentos, y una de ellas no cabe duda que es la frecuencia de las palabras, condición intrínseca a este tipo de investigaciones.

De alguna manera, muchos de estos trabajos, que toman para la elaboración de sus resultados muestras de lengua impresa, persiguen fijar un vocabulario o léxico básico que sirva de modelo de referencia al resto de la comunidad de hablantes. Posiblemente, partiendo del principio de que el hablante culto usa normalmente entre cuatro y cinco mil vocablos, los léxicos básicos y los léxicos fundamentales de un idioma intentan fijarlos[83], si bien

---

[82] Wolf, Dieter, *Statistiche Untersuchungen Zum Wortschatz Englischer Zeitungen*. Saarbrüken, 1969 . Reseña del autor en *Cahiers de lexicologie*, XXII, 1973, I, pp. 126-127.

[83] López Morales (1987: 59) afirma al respecto: "Este planteamiento plantea dos cuestiones igualmente importantes. De todo el caudal léxico de una lengua, el recogido en los diccionarios, ¿cuáles son esas 4.000

teniendo siempre presente que no se puede hablar de una frecuencia fija de una palabra en lengua ni de una jerarquía definitiva e indiscutible de frecuencia -al menos en lo concerniente a las palabras de contenido semántico-, por otro lado estas investigaciones sí perfeccionan al máximo sus métodos de análisis para situar las palabras en una zona determinada del léxico (Müller, Ch., 1965b: 33-42 y 1976: 92; Michéa, R., 1967: 3-14). Para ello no sólo tendrán en cuenta la frecuencia, sino otros datos más precisos, como la dispersión y el índice de uso. Aunque anteriormente Guiraud (1960: 16 y 74) sí creía que las palabras, además de significante y significado, podrían tener también en la frecuencia una tercera característica estable que debe figurar en nuestra memoria, de ahí surge una de las críticas que se ha hecho a las listas de frecuencia que tratan de establecer el vocabulario fundamental de un idioma, y es que cualquier individuo competente podría establecerlas directamente con muy buena aproximación, pues se intuye más o menos si una palabra es frecuente o no[84]. Haensch (1982b: 401-407) alude a que en la selección de entradas en un diccionario se deberá acudir en primer lugar a la frecuencia, determinada por el análisis estadístico de corpus escritos u orales, pero la frecuencia por sí sola no será suficiente porque siempre hay vocablos con un bajo índice o que no aparezcan documentados y, dada su importancia, deberán ser incluidas en el diccionario. Aporta como testimonio de estos casos la decisión que se adoptó en el *Diccionario del Español de Méjico*: acudir a los miembros de un Consejo Consultivo.

ó 5.000 palabras de uso real entre los hablantes cultos de una determinada comunidad lingüística? ¿Cuáles corresponden a las 2.000 manejadas por los niveles socioculturales más modestos? Las respuestas dadas a preguntas tan fundamentales nos llevan directamente al concepto de *léxico básico*".

[84] Guiraud, P. (1960: 74) expresa así su opinión sobre este punto: "Une des critiques qu'on a pu faire, très justement, aux listes de fréquences lexicologiques établies pour l'élaboration des langues de base, est que tout individu compétent peut les établir directement avec une très bonne approximation. Nous savons tous ce que c'est qu'un mot fréquent ou qu'un mot rare et nous avons, dans l'ensemble, une assez exacte appréciation des hiérarchies quantitatives de notre vocabulaire".

1.5.2. Vocabularios relativos a la población infantil o juvenil

Paralelamente a las investigaciones anteriormente mencionadas, se fueron desarrollando estudios que tenían como motivo el vocabulario de una población infantil o juvenil.

Refiriéndonos, en primer lugar, al vocabulario infantil, e insertos en la descripción y observación del mismo[85], se señalan[86] numerosas investigaciones que tienen como objetivos básicos tratar de establecer por un lado, la adquisición[87] del lenguaje tomando la palabra como unidad de análisis o como componente fundamental en otros índices de medición[88], dentro de las que se insertarían también las investigaciones sobre el bilingüismo o adquisición simultánea de dos lenguas[89]; y, por otro lado, el estudio de la evolución del vocabulario según la edad, que también se lleva a cabo en estudios de bilingüismo. En estos últimos trabajos se han utilizado varios métodos de

---

[85] En este sentido, uno de los primeros trabajos de observación del lenguaje infantil que se nombran corresponde a D. Tiedman, quien en 1782 publica un estudio sobre el desarrollo del lenguaje de su hijo. Citado por Bouton, Ch. (1982: 13-14).

[86] Véanse las referencias que nos ofrece Justicia, F. (1985a: 15-16 y 1995: 7-8).

[87] Por ejemplo, en español existen algunos trabajos recientes, como el de Hernández Pina, F. (1990: 155-321). Aquí nos muestra la autora el análisis pormenorizado del proceso de adquisición de la lengua española de su hijo Rafael, en un período que va de los 0 a los tres años, y el vocabulario obtenido en ese período.

[88] Véase Jiménez, B. (1986: 27-52), quien ofrece una relación de obras interesantes en las que se ha estudiado el tema.

[89] Combarros, S. (1987: 277-295) expone su análisis del lenguaje de un niño bilingüe y cita algunos autores que han llevado a cabo estudios semejantes.

investigación y se han analizado tanto aspectos cuantitativos, o sea, número de vocablos, como cualitativos: por ejemplo, la división de las palabras en clases gramaticales. También se ha estudiado el lenguaje infantil desde perspectivas sociolingüísticas para determinar las diferencias diastráticas de los escolares[90]. F. Justicia (1995: 7) agrupa los métodos de análisis del vocabulario infantil en tres tipos, asociados cada uno de ellos a distintas clases de recuento: "los recuentos absolutos, los vocabularios de reconocimiento y los recuentos basados en cálculos de frecuencias". Los primeros, denominados por B. Jiménez (1986:48) "vocabularios sumarios", tienen por objeto determinar el número total de palabras que emite el niño, por lo que se llevan a cabo con uno o pocos sujetos[91] y en edades muy tempranas -cuatro o cinco años-, ya que después es prácticamente imposible dadas las dimensiones que adquiere el vocabulario[92]. Iniciados a fines del siglo pasado dentro del terreno de la Psicología y continuados durante el siglo XX, tales recuentos absolutos tratan, pues, de medir el desarrollo cuantitativo del vocabulario. De este tipo de estudios son la mayoría, que comenzaron casualmente con la observación directa del lenguaje infantil, y posteriormente se sistematizaron. Estos estudios, afirma F. Justicia (1985a: 15-16 y 1995: 7-8), no ofrecen demasiadas garantías en sus resultados porque son pocos los

---

[90] Por ejemplo, Peronard, M. (1978: 107-125), estudia sobre todo las características fonéticas y morfológicas. La única información cuantitativa relacionada de alguna manera con el vocabulario se refiere a la producción del número de verbos.

[91] Justicia, F. y Palma, A. (1986: 584-585) indican una serie de trabajos que hicieron algunos lingüistas tomando como sujetos experimentales a sus propios hijos, como el alemán W. Leopold; A. Gregoire para el francés; A.F. Chamberlain y J. J. Camberlain y H. V. Velten para el inglés. Para el español podíamos citar el trabajo llevado a cabo por Hernández Pina, F. (1979).

[92] Para Justicia, F. y Palma, A. (1986: 585), estos estudios demuestran la adquisición léxica crece de tal manera y llega a tal volumen de unidades que es prácticamente imposible compilarlas después de los tres años de edad.

sujetos que sirven de muestra, por la subjetividad de los informes y el carácter acientífico de éstos (sobre todo porque fueron recogidos en documentos biográficos por los padres) y además por el escaso rigor con que se tratan las cuestiones lingüísticas y por los fallos metodológicos que presentan. A pesar de todo, tales estudios dejan claro que las primeras palabras aparecen en torno al primer año de existencia y que a partir de ese momento, el vocabulario crece con mucha rapidez. Se centran, pues, en establecer el número total de palabras emitidas o ver cómo crece el vocabulario con la edad, pero no estudian la frecuencia o el número de veces que aparece una palabra.

Según F. Justicia (1995: 8) para subsanar las deficiencias que suponen estas investigaciones, debido a que, a partir de cierta edad, el vocabulario del niño es muy amplio, surgen los vocabularios de reconocimiento, determinados a través de tests léxicos. "El método consiste en deducir de una muestra de palabras la amplitud del vocabulario total que posee el individuo, el empleo de fórmulas matemáticas permite obtener resultados válidos" (Jiménez, B., 1986:50-51). No obstante, las mayores críticas a este procedimiento apuntan al tipo de vocabulario seleccionado, ya que para medir un vocabulario infantil se parte de un léxico de adultos[93]. Es decir, teniendo en cuenta bien un vocabulario estándar[94], bien determinados diccionarios, y por el sistema de catas, se confeccionan unos tests, que previamente se presentan a la muestra de la población pertinente, para ver el número de palabras que identifican y a partir de ahí se extraen las conclusiones. En concreto, se señalan algunos

---

[93] Véanse las observaciones de Justicia, F. (1995: 9) y de Jiménez, B. (1987: 50-51).

[94] Por ejemplo, Fernández Huerta, J. (1995: 11-14) parte de vocabularios de frecuencia del español, de donde selecciona 1020 palabras de mayor frecuencia, de las que suprime las gramaticales, y con el resto se confecciona un test que se pasa a 85 niñas de 5 y 6 años para ver cuántas identifican.

datos cuantitativos del vocabulario, más bien pasivo, y su evolución o aumento con la edad.

Como faceta que no deja de ser curiosa relacionada con esta técnica (mostrar una serie de términos para que sean reconocidos por los alumnos), destacan otros trabajos que tratan de ver la incidencia de los préstamos en el escolar[95]. También se recurre a veces a la utilización de "referentes visuales" como instrumento para estudiar algún aspecto del léxico de una población joven -de 8 a 15 años-, como el proceso de cambio lingüístico o progresiva adaptación a un registro o variedad determinada frente a otras: en concreto, cómo se produce y hasta qué punto el grado de adaptación a la norma montevideana por parte de chicos de origen uruguayo procedentes de países como Suecia, Francia, Venezuela y Méjico[96].

Para F. Justicia (1995: 16) los diccionarios, vocabularios o listas de frecuencias, no tratan de medir las dimensiones absolutas del léxico infantil: son recuentos relativos basados en la relación del número de veces que aparece una palabra con el total de palabras del recuento. Por tanto, tienen en cuenta la frecuencia de uso de las palabras. Estos trabajos participan de los postulados de los recuentos generales y tienen en común muchos aspectos metodológicos, aunque a diferencia de ellos, se analizan muestras de expresión directa y espontánea, por lo cual consideran el léxico activo. Este tipo de trabajos para el análisis de un vocabulario infantil se inicia ya a

---

[95] Alcalá, J., Luengo, M. R. y García, M. A. (1987: 431-448) dan cuenta de una experiencia llevada a cabo con alumnos de EGB, entre 7 y 12 años, para ver el grado de conocimiento de una serie de préstamos actuales, xenismos para ser más exactos, de gran difusión televisiva.

[96] Véase el trabajo de Barrios, G. y Rivero, R. (1994). Es cierto que el número de sujetos encuestados es muy escaso, 40, y también escaso el número de ítems, pero queda patente la elección del vocabulario y de procedimientos cuantitativos para estudiar aspectos lingüísticos.

primeros de siglo[97], si bien y en cuanto a los niños españoles, "son relativamente recientes e insuficientes todavía" (Justicia, F., 1995: 10), y de ellos nos ocuparemos extensamente en páginas posteriores. No obstante, conviene anticipar que los trabajos de frecuencias que se citan en diversas fuentes pertenecen a estudios sobre la escuela primaria y responden más a perspectivas psicológicas o pedagógicas que lingüísticas, quizás porque siempre ha despertado mayor interés el proceso de adquisición de una lengua, propio de edades tempranas, que cuando la lengua está adquirida. Lo mismo ocurre con los trabajos de morfología[98] o sintaxis[99]: son pocos los que analizan una población escolar de Enseñanza Media.

También el vocabulario disponible de los escolares ha sido objeto de estudio. El trabajo de Mackey y colbs. (1971) se centra en una población francófona infantil de 9 a 12 años, sometida a los mismos centros de interés que en el francés fundamental -aunque los autores canadienses añaden 6 temas más- con la finalidad, entre otras pretensiones, de parangonar sus

---

[97] Buchanan, M. (1927: 3) cita la obra *Horn Ashbaugh Speller*, 1921, de Lippincott, quien parte de cartas, apuntes y expresión oral de niños para realizar su investigación; Rodríguez Bou (1952: 5) cita a Henry D. Rinsland, *A Basic Vocabulary of Elementary School Children*, New York, 1945, en donde se recoge el vocabulario de composiciones escritas de alumnos de 1° a 8° de la escuela primaria.

[98] Vila, I. y Muñiz, C. (1986) nos informan de una experiencia llevada a cabo con dos niños, niña y niño, desde los primeros meses hasta los tres años. Se enmarca, pues, dentro de los estudios sobre la adquisición del lenguaje. Urrutia, H., Olaeta, R. y Millá, F. (1987) llevan a cabo un estudio estadístico con los datos obtenidos de 1.212 alumnos de 3° a 8° de EGB, a los que se les pasó dos cuestionarios con 20 ítems cada uno, para estudiar aspectos del género, como la concordancia, su adquisición y grado de dominio a lo largo de la EGB, entre otros...

[99] Véase Salvador Mata, F. (1986); Salvador Mata, F. y Molina, E. (1987); Urrutia, H., Olaeta, R. y Millá, F. (1988), artículo este último en el que se exponen los resultados del análisis de la producción oracional de alumnos de 8 a 13 años; por tanto, del Ciclo Medio y Ciclo Superior de EGB.

resultados con los obtenidos en Francia por Gougenheim, Michéa, Rivenc y Sauvageot (1964).

Se dispone de algunos datos cuantitativos relativos al volumen de vocablos que manejan los niños de distintas edades. Para edades comprendidas entre 4 y 6 años, se aportan unas cifras (relativas a niños estadounidenses) que van desde 1.540 vocablos a los cuatro años, hasta 2.562, con una gradación de entre doscientos o trescientos términos nuevos por cada seis meses de edad[100]. E incluso en otras investigaciones aumentan en 500 palabras las cifras anteriores[101]. También se ha estudiado el vocabulario oral de los párvulos para calcular la media total de palabras, por un lado, y de vocablos, por otro, que utilizan los niños[102]. Ya desde los primeros momentos, las investigaciones pusieron de manifiesto que entre "el vocabulario estimado mediante pruebas y multiplicación por un factor y el utilizado frecuentemente por el niño hay una diferencia extraordinaria. Si nos cabe hablar de millares de palabras poseídas por niños en edad parvular no

---

[100] Fernández Huerta, J. (1955: 5-6) cita los datos del trabajo de M. E. Smith, *An investigation of the development of the sentence and the extent of vocabulary in young children.* University Iowa, 1926, donde se atribuyen 1.540 palabras para los niños de 4 años; 1.870 para los de 4 años y medio; 2.072 para los de 5; 2.289 para los de 5 años y medio y 2.562 para los de 6 años, aunque el estudio se hizo con niños con coeficiente intelectual medio-alto, y las técnicas empleadas para fijar los vocablos, técnicas de reconocimiento, no aseguraban que el niño tuviera realmente un dominio de los vocablos de la escala presentada (203 palabras).

[101] Fernández Huerta, J. (1955: 6-7) se refiere al trabajo de O. J. Grigsby, "An experimental study of the development of concepts of relationship in pre-schoool children as evidence by their expressive ability", *Journal of Experimental Education*, I, 1932, pp. 144-162

[102] Así se recoge en el estudio de A. T. Jersild, and R. Ritzman, "Aspects of Language development: The Growth of Loquacity and Vocabulary", *Child Development*, 9, 3, 1938. Citado por Fernández Huerta (1955: 8).

podemos asegurar lo mismo respecto a su uso. La usualidad por el escolar es muy reducida (sic). Luego el dominio del vocabulario es algo diferente al uso. No se puede encontrar con sólo el uso, aunque se pueda admitir que las usadas por todos los párvulos son conocidas por ellos. El quehacer escolar exige el conocimiento de este grupo reducido de palabras para asegurar el éxito de la labor docente" (Fernández Huerta, J., 1955: 8-9).

1.5.3. Vocabularios en lengua española

Trataremos de hacer en este apartado una breve exposición sobre los recuentos que se han realizado para el idioma español, tanto en países hispanoamericanos como, sobre todo, en España. Comenzaremos con trabajos basados en la expresión de personas adultas y seguiremos con aquellos que analizan el vocabulario de los escolares.

Aunque "el primer diccionario que recogía entre sus páginas el español, el *Vocabulario en latín y en romance* del humanista Alfonso de Palencia"[103], es el primer trabajo lexicográfico completo y de cierta envergadura que nos ha llegado[104]. Amparo Morales cifra el inicio de las investigaciones sobre el

---

[103] Según consideraciones de Alvar Ezquerra (1990: 47).

[104] Ya Menéndez Pidal se refiere a la labor llevada a cabo en las glosas emilianenses y silenses, lo cual presupondría "algún glosario anterior latino-romance dispuesto por orden alfabético" (Menéndez Pidal, R., *El idioma español en sus primeros tiempos*, 6ª ed., Madrid, 1964, pp. 23-24. (La referencia la tomamos de Alvar Ezquerra, 1990: 57); Alvar Ezquerra (1983: 21-24) cita la obra de Américo Castro, *Glosarios latino-españoles de la Edad Media*, 1936, donde se recogen una serie de obras que sirven de precedente a la actividad lexicográfica hispana de importancia que comienza a fines del s. XV; Arias, R. (1986: 744-745) hace un breve recorrido por los primeros testimonios "lexicográficos" de la Península Ibérica, en la que aparece la figura de San Isidoro de Sevilla, quien con "sus *Etimologías -Etymologiarum libri XX-* se erige en el primer diccionarista de la Península Ibérica"; labor que no cesaría ya hasta nuestros días.

español, basadas de algún modo en la frecuencia, tras la obra de A. Merás y S. Roth[105], *Pequeño Vocabulario*, 1915. Buchanan (1927: 4-7) alude a algunos trabajos sobre lengua española, citados también por Rodríguez Bou[106], quien además añade otros que pertenecen más bien al terreno de la psicología[107].

Keniston[108] (1920) realiza un ensayo para establecer una lista de palabras comunes en español. Su intención es que la muestra se aproxime a la lengua coloquial, pero, debido a los problemas que presenta, no contempla la expresión oral sino que recoge obras de teatro, por considerarlas más cercanas a este registro, a las que añade periódicos, revistas, cuentos cortos y novelas. De ahí extraerá la lista de palabras más corrientes, seleccionadas según la frecuencia y la repartición, y las expondrá agrupadas en categorías gramaticales. Aunque es consciente de que su listado no es definitivo, sino

---

[105] Morales, A. (1980: 468) proporciona una breve relación de los recuentos más significativos sobre el español; Buchanan, M. (1927: 7) ofrece la ficha bibliográfica completa del trabajo de Merás y Roth: *Pequeño Vocabulario, a list of two thousand Spanish words arranged in logical groups for sentence building in the first two years*, D. C., Heath and Co. Rodríguez Bou (1952: 7) aporta otros datos: "Merás y Roth han publicado una lista útil de palabras en español escogidas por maestros de experiencia y que lleva por título Pequeño Vocabulario..." alude también a este trabajo de Merás y Roth, P.

[106] Véase Rodríguez Bou, I. (1952: 6-7). Afirma además el autor que los estudios para el español han sido pocos y en cierta medida incompletos.

[107] Rodríguez Bou (1952: 7) se refiere al libro de Teobaldo Casanova, *Educational Psychology and Some Aspects in Latin America*, donde se comentan los trabajos de Alfredo Aguayo, 1920, y de Pino, relacionados con la Escala Binet. También cita a David Pablo Boder, *La Escala Binet-Simon-Terman-Mexico*, donde se adapta para Méjico el examen Binet.

[108] Keniston posee además otros trabajos de naturaleza cuantitativa sobre el vocabulario: *Spanish Idiom List*, New York, MacMillan Co., 1929; y *Basic List of Spanish Words and Idioms*, Chicago, University of Chicago Press, 1933.

una sugerencia de lo que debería hacerse de un modo más extenso, sí destaca la utilidad que puede tener para la enseñanza.

Buchanan (1927) en un principio ideó su trabajo para confeccionar test de vocabulario, aunque después se publica para otros fines. De 40 fuentes, agrupadas en 7 epígrafes distintos, recopila un total de 1.200.000 palabras, 30.000 por cada una de las fuentes[109], y recoge un cómputo final de 6.702 unidades, seleccionadas según una operación matemática que combina frecuencia y repartición[110]. Elige dos tipos de ordenación en la presentación de las listas: por orden de frecuencia decreciente y alfabéticamente. Además, nos ofrece las 189 palabras eliminadas del recuento para una mayor economía, pues son palabras con frecuencia muy alta[111].

T. R. Céspedes (*Investigación acerca de las palabras usadas en castellano*, Panamá, 1928-29) recoge el material de distintas fuentes: mezcla composiciones de los alumnos desde 6º de enseñanza primaria hasta 5º de liceo (equivalente a secundaria) con artículos y secciones de prensa,

---

[109] Buchanan, M. (1927: 5-6) ofrece una relación de las obras que utiliza para su compilación, organizadas en torno a siete núcleos: teatro, novela, poesía, folklore, prosa miscelánea (historia, biografía, viajes...), literatura técnica (ciencia, geografía, filosofía, correspondencia, cocina...) y textos periodísticos. Las fuentes -u obras- seleccionadas van desde un período que comprende el siglo de oro español -*El Quijote; La vida es sueño; La Gitanilla*-, hasta primeros del siglo XX.

[110] Buchanan, M. (1927: divide la frecuencia de una palabra por 10 y le suma la repartición o número de fuentes en la que aparece (f/10 + r). Así obtiene el "credit-number" o "merit-number". Esta operación la realiza con aquellas palabras que tienen una frecuencia 10 como mínimo, o siendo inferior, aparece en 5 fuentes al menos.

[111] Buchanan, M. (1927 10-17) se refiere a las palabras que a su juicio estarían entre las 500 primeras según la frecuencia. Son las llamadas palabras de contenido gramatical, como preposiciones, conjunciones, determinantes, pronombres, pero también algunos sustantivos, adjetivos, verbos y adverbios.

conferencias, telegramas, lenguaje judicial... De un corpus de casi medio millón obtuvo un total de 9.255 vocablos[112].

Aunque no es propiamente un recuento sino una compilación de algunos existentes, hemos de citar aquí el trabajo de Helen S. Eaton (1940,1967). La autora selecciona las 6474 palabras de mayor frecuencia de los recuentos de vocabulario más emblemáticos, en esas fechas, de cuatro idiomas: inglés[113], francés[114], alemán[115] y español[116]. En las distintas ordenaciones que ofrece, siempre aparecen las palabras de un idioma como referencia y, junto a ellas, las palabras equivalentes de los tres idiomas restantes.

Con la idea de hacer el vocabulario español más extenso y representativo y con la finalidad pedagógica de mejorar la tarea de elaboración de libros de lectura para la escuela, surge el *Recuento de Vocabulario Español*, 1952, de Rodríguez Bou. Sus materiales no serán textos clásicos como los de Buchanan, sino completamente actuales a su tiempo, de ahí la determinación de las fuentes y la recopilación. En total se obtuvieron 7.066.637 ocurrencias: 20.542 unidades léxicas; 62.888 formas de inflexión y, por tanto, un total de 83.430. El material se extrae de 10 fuentes distintas, clasificadas en tres grupos: vocabulario de expresión[117], vocabulario basado en el juicio de

---

[112] Información recogida de Rodríguez Bou (1952: 7).

[113] *Teacher's Word Book of 20,000 Words*, de E. L. Thorndike.

[114] *French Word Book*, de Vander Beke.

[115] *Häufigkeitswörterbuch der deutschen Sprache*, de Kaeding.

[116] *A Graded Spanish Word Book*, de Milton A. Buchanan.

[117] Se refiere al confeccionado con muestras de la expresión de los escolares, de 3º a 6º de Primaria, tanto

distintos autores[118] y vocabulario de reconocimiento[119]. Su trabajo mezcla tanto muestras de expresión oral y escrita de escolares, como una selección de textos ya impresos de diversa procedencia, a los que hay que sumar el recuento de Rodríguez y Casanova, por un lado, y el de Buchanan, por otro. Supuso más de 6 años de labor, por parte de un equipo numerosísimo. En el primer volumen, y para facilitar una posible tarea de planificación didáctica del vocabulario, ordena sus listas de palabras de cien en cien unidades, por orden de frecuencia decreciente[120]. También se ofrece una lista con las 20.000 formas flexivas en orden alfabético, con su frecuencia y rango, y otra lista con las palabras de frecuencia menor a 16. En el segundo volumen aparecen en 15 columnas las palabras ordenadas alfabéticamente, con las unidades flexivas agrupadas bajo las unidades léxicas, acompañadas de los datos cuantitativos: las frecuencias calculadas sobre la base de un millón, la frecuencia total de cada uno de los tres tipos de fuentes, la frecuencia absoluta...

---

orales como escritas, y estas últimas sacadas de composiciones y pruebas de asociación libre y controlada de palabras. A este grupo se añade el cómputo de Rodríguez y Casanova, no publicado con anterioridad, que partieron de composiciones escritas de alumnos de 1° hasta 8° de Básica.

[118] Las fuentes son libros de texto básicos y libros escolares de lectura complementaria.

[119] Rodríguez Bou (1952: 11) llama "vocabulario de reconocimiento" al seleccionado de letra impresa, pues sería el que se reconocería, y por tanto, lo considera como parte del léxico pasivo, aunque en este apartado agrupa el material periodístico, religioso, el trabajo de Buchanan, y, curiosamente, programas de radio, como foros, programas de aficionadas, de participación pública..., que más bien apuntan al léxico activo, pues no son de los tipos de programas en los que se hacen guiones para leerlos, sino que se tiene en cuenta la expresión espontánea de los participantes.

[120] Las primeras 105 aparecen sin frecuencia, pues corresponden a las palabras con mayor número de ocurrencias: las llamadas palabras gramaticales. El valor que toma como índice de ordenación es la frecuencia, aunque ponderada por un número recíproco calculado sobre la base de un millón, ya que el número de palabras de las distintas fuentes es muy heterogéneo.

García Hoz (1953a) inicia los estudios cuantitativos del vocabulario en España. Considera el vocabulario usual como aquel que utiliza el hombre de la calle, y es el que se debe enseñar en Primaria. Para determinarlo, obtiene un corpus de 400.000 palabras de textos de distintas fuentes: cartas privadas, periódicos, documentos oficiales -políticos, religiosos, sindicales- y libros[121]. De cada fuente obtiene 100.000 palabras. Al final consigue un vocabulario de 12.913 vocablos, a los que suma otras voces de carácter técnico y regional, hasta conseguir cerca de 15.000, cifra que manejan otros autores de otras lenguas[122]. El "vocabulario común" consta de 1.971 vocablos comunes a las cuatro fuentes, a los que añade una relación adicional de 212 términos de frecuencia muy alta[123]. El vocabulario fundamental está integrado por los 208 vocablos que tienen prácticamente el mismo número de frecuencias en las cuatro fuentes. La frecuencia media de los vocablos del vocabulario usual es de 31, la del vocabulario común de 172 y la del vocabulario fundamental de 1.324. Queda aquí apuntada ya la utilidad de estos trabajos para, por ejemplo, la confección de métodos de enseñanza de la lengua para extranjeros.

---

[121] Las fuentes responden a la selección de cuatro aspectos de la vida que García Hoz (1953a: 18) considera básicos en su investigación: vida familiar, vida social indiferenciada, vida social reglada y vida cultural.

[122] En concreto, García Hoz (1953a: 24-25) cita a A. F. Watts: *The Language and Mental Development of Children*, London (D. C. Heath), 1946. El propio autor nos informa de los resultados obtenidos en su trabajo: " Por otra parte, el recuento de 400.000 palabras (100.000 de cada tipo de fuentes) dió (sic) como resultado la aparición de 12.913 palabras, cifra que, aumentada con los vocablos específicos de la profesión, lugar, etc., que cada hombre posee, viene a coincidir con la aproximativa señalada por Watts, nueva razón para considerar suficiente el recuento de 100.000 palabras de cada tipo de vocabulario". La cita es de la p. 25.

[123] Así lo indica el autor en el apartado correspondiente de "vocabulario común" (García Hoz, 1953a: 387): "De la correspondiente selección de vocablos ha resultado el vocabulario común con 1.917 palabras y la relación adicional con 212; de estas últimas, 97 tienen más de 40 frecuencias y las 115 restantes pasan de esta cifra por acumulación de términos que expresan ideas afines".

Tras esta breve visión de los trabajos cuantitativos de vocabulario de nuestro idioma, llegamos al, quizás, más importante estudio del vocabulario español hasta esos momentos. Nos referimos a la obra de Juilland y Chang-Rodríguez *Frecuency Dictionary of Spanish Words* (a la que a veces aludiremos como *FDSW)*, publicada en 1964. Su finalidad esencial es, además de la enseñanza de lenguas, establecer la base de posteriores investigaciones estructurales, descriptivas, comparativas e históricas. Para ello Juilland, en *The Romance Languages and their Structures* (1964), se propuso confeccionar un diccionario para cada una de las 5 lenguas románicas más importantes: español, francés, italiano, portugués y rumano. El primer volumen en aparecer fue el *Frecuency Dictionary of Spanish Words*, que se puede considerar como representativo del español peninsular estándar, ya que a través de un riguroso sistema de muestreo y de unas medidas muy precisas tratará de determinar el léxico básico; es decir, aquellas palabras usadas por la mayoría de los hablantes de una lengua: unos 5000 vocablos. El punto de partida para su elaboración fue obtener un universo léxico -corpus- de 500.000 palabras. Para reunirlas, se dividió el material en cinco subconjuntos o mundos: teatro, novela, ensayo, prensa y escritos técnico-científicos. Se hizo una selección de las mejores obras de cada mundo, y se recogieron, dentro de ellas y aleatoriamente, una serie de oraciones hasta conseguir 100.000 palabras para cada mundo. Para establecer los vocablos, primero se eliminan las palabras con frecuencia inferior a 5. Pero como la frecuencia por sí sola no era garantía de selección en un trabajo con esas pretensiones, se aportaron otros medidas. Primero, la dispersión, que hace referencia, en principio, a la distribución de las palabras en los mundos; pero la dispersión que idea Juilland va más lejos, pues para su cálculo se necesitan operaciones matemáticas tales como la media aritmética, la variancia[124], la desviación típica, coeficiente de variación,

---

[124] Preferimos utilizar este término, recogido en el DRAE, en vez de "varianza", término éste quizás más común entre la terminología matemática que se emplea, al menos, en Enseñanza Media.

etc. Así se extrae el índice de dispersión, representado por un número fraccionario de 0 a 1. Con la frecuencia y la dispersión se calcula otra medida: el índice de uso, que será el factor determinante en la selección de los vocablos. Clasificados éstos según el uso, se fija un límite para aquellos que presenten un índice de uso de 3'08, que es el que poseen los que ocupaban el lugar desde 4.869 a 5.024. Razón por la que la lista final no es de 5.000 sino de 5.024.

El *Frecuency Dictionary of Spanish Words* es, pues, una obra ambiciosa y muy elaborada, que ha sido motivo de estudio por muchos autores[125], analizada, ponderada y elogiada hasta considerarla como, posiblemente, la mejor obra de recuento de vocabulario hecha sobre el español peninsular, que a su vez ha servido y sirve de modelo teórico de diversas investigaciones realizadas en diferentes países. Pero también ha sido criticado en algunos aspectos, como en la norma léxica, en la que se han primado sobre todo aspectos sintácticos sobre léxico-semánticos[126], o los criterios de segmentación, o en la elección de los mundos léxicos, sobre todo por dejar fuera a la poesía o el lenguaje jurídico administrativo, o en la delimitación espacial y temporal, e incluso en el sistema aleatorio y objetivo a la hora de seleccionar las obras y las oraciones de las obras, por lo que "algunas decisiones indudablemente han tenido que ejercer notable influencia en los resultados que se nos ofrecen" (Bustos Tovar, 1966-7: 174). No obstante, creemos que es el mejor trabajo acerca del vocabulario peninsular y su utilidad

---

[125] Entre los muchos trabajos que se ocupan de analizar este trabajo, se pueden destacar los de Müller, Ch. (1965b y 1965c); Morales, A. (1980); Ortega, J. (1985); Huynh-Armanet (1977); Bustos Tovar (1966-7)

[126] Bustos Tovar (1966-7: 183-184) censura la nula atención que Juilland ha prestado a las acepciones de las palabras.

sigue vigente, si bien en algunos aspectos ha podido quedar anticuado[127], ya que el vocabulario se renueva con el paso del tiempo, y en este sentido es posible que en el trabajo de Juilland no se recojan determinados neologismos que formen parte, en la actualidad, del léxico básico del español.

En Puerto Rico se ha confeccionado un léxico básico siguiendo los presupuestos metodológicos de Juilland: el *Léxico básico del español de Puerto Rico* (Morales, A., 1986). Amparo Morales eligió un material escrito, constituido por las obras publicadas por puertorriqueños entre los años 1948 y 1970. También subdivide el material en 5 mundos: drama, narrativa, ensayo, literatura técnica y periodismo, y elige 100.000 palabras de cada uno de ellos. Y utilizando los criterios de frecuencia, dispersión y uso, selecciona un total de 4.456 vocablos, de frecuencia al menos de 4, repartición de al menos en tres mundos e índice de uso igual o mayor a 3'55. Por tanto, no coincide exactamente con los valores del trabajo de Juilland, tampoco en los criterios de segmentación lexicográficos. La idea la autora de aumentar el índice de uso obedece al deseo de eliminar en lo posible la intervención del azar en la fijación de lo que sería un léxico con las palabras básicas de un idioma, sobre todo cuando uno de los fines del *Léxico básico del español de Puerto Rico* es "favorecer una mejor utilidad del léxico en su propósito más inmediato que es la gradación léxica en la labor educativa" (Morales, A., 1986: 30).

Tenemos noticias de que en la República Dominicana se ha llevado a cabo una investigación más o menos reciente sobre el léxico básico[128].

---

[127] Coincidimos en este punto con Guerra Salas (1989: 2) y señalamos con él la necesidad de un nuevo trabajo de estas características para nuestro idioma.

[128] La referencia la tomamos de Morales, A. (1986: 36), quien alude la trabajo de A. González, S. Cabanes y F. García: *Léxico básico de la lengua escrita en la República Dominicana*, Santo Domingo, Universidad Nacional P. Henríquez Ureña, 1982.

También la lengua oral usada en TV ha sido objeto de estudio en nuestro país. Lorenzo Delgado (1981), inspirándose en el estudio de García Hoz[129] (1953a) y con una finalidad didáctica destacada[130], elabora su investigación. Para ello, el autor justifica, en primer lugar, tanto las posibilidades educativas de la televisión como la clasificación de los programas -información, instrucción, distracción[131]- y sobre la base de ésta última selecciona 300.000

---

[129] Lorenzo Delgado (1981: 11-12) se refiere al *Vocabulario Usual, Común y Fundamental* de García Hoz como la obra que constituyó " la primera curiosidad intelectual que llevó al planteamiento de la presente investigación: ver la posible evolución del vocabulario durante este largo período de tiempo, utilizando para ello las mismas fuentes y procedimientos del original. Se trataba, pues, de repetir el trabajo veinte años más tarde para estudiar la evolución observable en el vocabulario a través de ese tiempo con vistas a determinar si resultaba o no un documento válido todavía para servir de fuente de extracción de contenidos por parte del maestro en su cotidiana tarea. Una segunda tentación intelectual hizo dar al proyecto una nueva dimensión para que resultara más útil y creativo: ¿Qué aportaciones podría hacer la televisión, sin duda el más importante de los mass media, a dicho estudio del vocabulario? ¿En qué medida contribuirá este nuevo trabajo a enriquecer la ardua tarea de determinar los contenidos de la didáctica del vocabulario?"

[130] El propio Lorenzo Delgado (1981: 13) así lo manifiesta: "Por ello, la pretensión última del Vocabulario que aquí se presenta es su aplicabilidad a la escuela -de ahí la doble adjetivación de <<orientación>> y <<didáctica>>-, ofreciendo al maestro un cuerpo de palabras que, dada la amplitud de fuentes o parcelas de la lengua de las que se ha hecho su extracción y el correspondiente nivel de usualidad, garanticen la racionalidad de su trabajo en la enseñanza del vocabulario, ortografía y, en general, las actividades de expresión y comprensión verbal".

[131] Lorenzo Delgado (1981: 32) delimita los programas del siguiente modo:
"1.- Información: recoge los programas informativos sobre la actualidad, la publicidad y la propaganda (telediarios, crónicas, comentarios políticos, encuestas, información deportiva, anuncios...).
2.- Instrucción: recoge los programas de intencionalidad instructiva, tanto los propios de la <<televisión didáctica o escolar>>, como aquellos que buscan una confrontación del telespectador con los bienes y valores culturales (documentales sobre las diversas ciencias y costumbres, descubrimientos, mesas redondas <<temáticas>>, arte, música, historia...).

palabras (100.000 de los espacios informativos, 100.000 de espacios instructivos y 100.000 de espacios de ocio) durante algunas semanas de enero y febrero de 1975. Con este corpus, confecciona un "Vocabulario Televisivo" en tres niveles: usual, común y fundamental. Además, y de acuerdo con el principio de que la frecuencia de una palabra es un dato importante porque nos da una idea de la necesidad relativa de conocerla (cuanto más frecuente sea, más necesidad hay de conocerla); propone la hipótesis siguiente: "La necesidad de introducir en los contenidos de la didáctica del vocabulario las palabras aparecidas en el Vocabulario Televisivo, estará en proporción directa a su aparición en otras fuentes o vocabularios y al nivel de usualidad en que lo haga" (Lorenzo Delgado, 1981: 34). Como consecuencia de estos postulados, contrasta su Vocabulario Televisivo con dos trabajos de García Hoz (1953a y 1976) y, fruto de la intersección de los tres, compone el Vocabulario de Orientación Didáctica (VOD), también en los niveles usual, común y fundamental. A nuestro juicio, este es un trabajo serio cuyos resultados son perfectamente válidos para la enseñanza del vocabulario en el instituto. Nosotros completaremos las palabras seleccionadas (Cfr. 4.8.) del trabajo de Juilland y Chang-Rodríguez con el Vocabulario de Orientación D[illegible]tica en su nivel usual, integrado por 3257 vocablos, los cuales son comunes a varias fuentes: el Vocabulario Usual y el de "Mayor frecuencia" de García Hoz (1953a y 1976), y el Vocabulario Televisivo en su nivel usual de Lorenzo Delgado (1981). Este vocabulario, a juicio de su autor (Lorenzo Delgado, 1981: 54) debería ser conocido en su totalidad por los alumnos al terminar la EGB. Nosotros hemos podido comprobar, al menos, que gran parte de él no lo usan, lo cual confirma la falta de planificación del vocabulario en el recinto escolar o -y esto es más seguro- la no contemplación del criterio sociológico (que atiende a la usualidad de una población) en la selección de los contenidos (Cfr. 4.5.).

3.- Distracción: recoge -a nivel de hogar- aquellos espectáculos que el hombre busca para su relax y esparcimiento. Su montaje es más de espectáculo que de contenido (telefilmes, concursos, variedades, musicales, cine...)".

El trabajo de Jenaro Ortega Olivares (1983) se enmarca también dentro del apartado del léxico televisivo. El autor recoge un corpus representativo de los telediarios emitidos en febrero de 1975. En total graba 40 en cinta magnetofónica: 20 de tarde y 20 de noche, de lunes a viernes. Después de las oportunas transcripciones, se consiguieron 1016 páginas, de las que se seleccionaron definitivamente 180 para conseguir un corpus entre 30.000 y 35.000 ocurrencias. El total, tras el despojo, fue de 32.488, de donde se obtuvieron 4.342 vocablos (2.038 de frecuencia 1). El corpus es relativamente pequeño si lo comparamos con otros trabajos de frecuencia, pero en compensación, Jenaro Ortega contempla también las llamadas lexías complejas o en vías de lexicalización, registradas como subentradas bajo el lema considerado como básico[132].

1.5.4. Vocabularios escolares en España

Nos referiremos a aquellos trabajos sobre el vocabulario de los escolares que de alguna manera utilizan datos cuantitativos. Por tanto, daremos cuenta tanto de los trabajos de los llamados por Justicia (1985a: 16) "vocabularios de reconocimiento", como aquellos trabajos basados en listas de frecuencias que tienden a recoger el vocabulario disponible o el usual. Éstos son relativamente recientes y escasos, surgidos tras la aparición del trabajo de García Hoz y con un enfoque más pedagógico que lingüístico[133].

---

[132] Ortega, J. (1983) adopta como unidad de texto la palabra, entendida como unidad gráfica, pero aporta también la frecuencia relativa de las lexías. Así, por ejemplo, bajo el lema *acto*, ofrece en primer lugar la frecuencia absoluta, y luego, registra como subentrada *acto académico, acto de clausura, hacer acto de presencia*, etc., o bajo *agencia: agencia de noticias, agencia de información*, etc.

[133] Justicia, F. y Palma, A. (1986: 586) citan unos pocos: la investigación de Fernández Huerta, el

1.5.4.1. Vocabularios de reconocimiento

En España , uno de los autores pioneros en estas investigaciones fue García Hoz. Con un grupo de colaboradores emprende sus investigaciones en 1944, las continúa hasta 1976 y las publica en *Estudios Experimentales sobre el Vocabulario* (1977). Se trata, en realidad, de un vocabulario de reconocimiento, pues para detectar el nivel de comprensión que del vocabulario tiene el escolar, no extrae un vocabulario de los escolares, sino que somete a éstos a una serie de pruebas. El autor confiesa que como no tenían ningún trabajo de referencia en nuestro país por esas fechas, debieron acudir al sistema de catas, partiendo del DRAE, 16ª ed., Madrid, 1939. Seleccionaron en un principio la primera palabra de cada página del diccionario y rechazaron aquéllas que no consideraron de uso general[134]. Para los tecnicismos, y dado que algunos son de dominio general y otros no, siguieron el siguiente criterio: considerar tecnicismos de uso general cuando su significado era conocido por todos los alumnos becarios del Instituto "San José de Calasanz". Igualmente, eliminaron los nombres de letra, artículos, pronombres, preposiciones, conjunciones e interjecciones. Quedaron 2.714 palabras, de las que se eliminaron al azar 14 para constituir 27 series de 100. Las palabras de cada serie se eligieron empezando por la palabra que figuraba en los lugares 1, 21, 41, 61, etc. La segunda, por la 2, 22, 42, etc., y así sucesivamente. En la investigación se utilizó la primera serie, dispuestas en cinco tipos de pruebas distintas. Cada prueba constaba de 20 vocablos. Para hacerla más cómoda, no se tenía que escribir la palabra, sino simplemente subrayarla. Además, se colocaron al principio las palabras más fáciles, para

---

vocabulario básico de Díaz Castañón y colbs., los trabajos de Armayor González, el vocabulario catalán-español de Mata, M. y el del propio Justicia.

[134] García Hoz (1977: 15) rechaza los regionalismos, arcaísmos y términos jergales.

asegurar que los escolares entraran en la prueba sin dificultad. A partir de aquí, García Hoz extrae una serie de conclusiones. Entre ellas, intentar determinar la evolución cuantitativa del vocabulario de los escolares entre 10 y 18 años.

Márquez Villegas sigue una estrategia similar en la confección de su trabajo: *Vocabulario del español hablado* (1975). Aunque se denomina "español hablado", es en realidad un vocabulario de reconocimiento. Todo surgió porque le encargaron en 1964 confeccionar un método de español como lengua extranjera. Para seleccionar la base léxica del trabajo, acudió al concepto de frecuencia. Pero el material lo recogió basándose en encuestas de reconocimiento. Estas encuestas la efectuó con 200 alumnos de 15-18 años de centros de Jaén. Partiendo de un diccionario de 20.000 palabras, los alumnos debían señalar en fichas cuáles usaban en su lengua coloquial. Tras el despojo y recuento de frecuencias, quedaron 1.647 palabras, guardadas y no publicadas ni utilizadas. Diez años más tarde, y para elaborar un vocabulario fundamental del español en tres niveles, repiten la experiencia en Madrid, con 200 alumnos desde 5º de Bachillerato hasta COU. Para efectuar la encuesta partieron de un diccionario con 9.000 palabras más que anteriormente (29.000, pues), del que los alumnos debían seleccionar tres tipos de palabras, para los tres niveles de enseñanza previstos: "a) palabras usadas muy frecuentemente en la lengua familiar; b) palabras usadas bastante en las mismas circunstancias; c) palabras conocidas, pero no usadas con mucha frecuencia en la lengua hablada, o pertenecientes a áreas semánticas restringidas" (Márquez Villegas, 1975: 6).

Se fijaron los tres niveles partiendo de las frecuencias: Mayor frecuencia, nivel I, y así sucesivamente. El trabajo presenta en primer lugar las listas de palabras por niveles, y posteriormente, basándose en el total de estas palabras, confeccionan una guía gramatical, con información morfológica, sobre todo.

Guerra Salas (1989) realiza una investigación interesante para determinar el grado de conocimiento que del léxico básico poseen los estudiantes madrileños de EE.MM. (en concreto, de 1º de BUP y de COU). Su punto de partida es el *Frequency Dictionary of Spanish Words*. De este diccionario extrae el material[135] para elaborar dos pruebas de 100 palabras cada una (una de producción y otra de comprensión léxica), divididas en 10 categorías según el índice de uso de los vocablos seleccionados, que pasará a los alumnos para comprobar, pues, dos aspectos: si se comprende correctamente el léxico básico y si se utiliza con propiedad. Estos dos aspectos permiten al autor sacar conclusiones acerca del léxico pasivo -el que se comprende- y del léxico activo -el que se usa en la expresión oral o escrita-. Para las pruebas de expresión se ofrece la lista de palabras a los alumnos y se les pide que compongan una frase con cada palabra. Para la prueba de comprensión decidieron elegir la modalidad de respuesta verbal con sinónimos. Para seleccionar los sinónimos acudieron al *Diccionario ideológico de la lengua española*, de Julio Casares, y al DRAE. Tras la aplicación de las pruebas, se recogen los datos, se tabulan y se someten a un test estadístico para apreciar la probabilidad de las desviaciones. Guerra Salas hace un análisis pormenorizado de los resultados y expone sus conclusiones. En la prueba de producción y referente a las 10 categorías, constata que el mayor número de aciertos -el 92'16%- lo obtienen las cinco primeras categorías que pertenecen a los vocablos que en el *FDSW* tienen un índice de uso mayor. El número de aciertos -el 72'12%- disminuye con las otras cinco categorías, que

---

[135] El procedimiento que sigue Guerra Salas (1989: 71-71) es el siguiente: elimina las 504 primeras palabras del Diccionario de Juilland y Chang-Rodríguez por tratarse de palabras gramaticales o muy usuales, y las 4520 restantes son divididas en 10 categorías de 452 vocablos -por tanto, según un orden de índice de uso correlativo en el diccionario de Juilland y Chang-Rodríguez- de donde se extraen en un principio 25 vocablos al azar, con lo cual se obtiene un excedente de 50 para poder elegir los vocablos más idóneos para las pruebas.

corresponden a vocablos que en el diccionario de Juilland y Chang-Rodríguez presentan un índice de uso menor: "esta primera percepción indica una adecuación entre los índices de uso que ofrece el FDSW y el comportamiento de los alumnos en las pruebas: las palabras teóricamente más usadas son conocidas por más alumnos que las palabras que tienen un índice de uso más bajo" (Guerra Salas, 1989: 90). Esto refuerza el valor del citado diccionario, y el autor pone en evidencia además que el porcentaje mínimo de aciertos se da en la categoría o serie X (el 66'2%), por lo que un 33'8% de los alumnos encuestados -recordemos, de 1º BUP y de COU-, no sabe construir correctamente una oración con los vocablos que en el *Frequency Dictionary of Spanish Words* están comprendidos entre los índices de uso 4573 y 5024. En cuanto a las diferencias por estratos, el autor señala que es el nivel académico el que marca las auténticas diferencias, ya que COU (86'31%) supera en aciertos a BUP (78'94%), mientras que las diferencias entre la enseñanza pública y la privada (levemente a favor de ésta última) y entre alumnas y alumnos -levemente a favor de éstos, aunque COU femenino privado es el que obtiene mejores resultados- no son significativas. Las pruebas de producción arrojan otros datos, pues en lo que atañe a las categorías, son las cinco últimas, aquellas que corresponden a los vocablos de un índice de uso más bajo, las que obtienen globalmente mejores resultados, es decir, mayor número de aciertos, lo cual indica que las palabras que las integran son más conocidas por los alumnos. Además, el porcentaje de aciertos de la prueba de comprensión es inferior al de la prueba de producción. Esto demuestra la inadecuación entre el conocimiento de los alumnos del léxico básico que se recoge en el *FDSW* y el índice de uso que este diccionario presenta. Para el autor, esto quizás se deba a que si bien los alumnos madrileños de EE.MM. "utilizan el léxico básico en sus producciones lingüísticas en correspondencia aproximada con los índices de uso del FDSW, cuando se trata de precisar el significado de dicho léxico, muestran un conocimiento anárquico del mismo, fruto de una inexistente planificación dirigida a la enseñanza de los significados precisos de las palabras integrantes del léxico básico del español"

(Guerra Salas, 1989: 152). Los datos revelan que los alumnos de EE.MM. poseen un importante desconocimiento del significado preciso de los vocablos más usuales del español peninsular; en concreto, y para el conjunto de la prueba, el porcentaje de desconocimiento es del 37'24% de los vocablos comprendidos en el *FDSW* entre los índices de uso 505 y 5024; o lo que es lo mismo, el 37'24% de los alumnos desconoce el significado exacto de tales vocablos. Por lo demás, las diferencias según la enseñanza pública o privada (levemente a favor de ésta) y entre alumnos y alumnas -levemente ahora a favor de éstas- no son significativas. De nuevo es el nivel académico el que marca realmente las diferencias: los alumnos de COU obtienen mayores porcentajes de aciertos en todas las categorías. No obstante, el 32'8% de los alumnos de COU desconoce el significado preciso del léxico básico español. Guerra Salas concluye al final de su trabajo que se pone en evidencia en su investigación que no existe planificación de la enseñanza y adquisición del léxico en EE.MM., lo cual origina que un porcentaje importante de alumnos llegue a la Universidad sin manejar adecuadamente el léxico básico de su comunidad lingüística, por lo que es necesaria una planificación seria que contemple el material que le suministran los léxicos básicos y los trabajos de disponibilidad léxica (Guerra Salas, 1989: 153).

1.5.4.2. Los vocabularios basados en el cálculo de frecuencias

A este apartado pertenecerían los vocabularios disponibles y los vocabularios propiamente basados en listas de frecuencia que tratan de determinar el vocabulario usual.

1.5.4.2.1. Vocabularios escolares: disponibilidad léxica

De entre los trabajos españoles que estudian el vocabulario de una población escolar, utilizando para ello muestras de expresión oral o escrita, conviene destacar algunos, como el trabajo de Díaz Castañón y colaboradores (1977 Y 1978). Su *Vocabulario básico del español y sus aplicaciones a la enseñanza* estaría cerca de lo que sería un vocabulario disponible (Cfr. 1.4.2.), pero conviene introducir algunos matices. Veíamos que en el léxico disponible, según López Morales (1987: 62-66), se conjugan varios factores. En primer lugar, delimitar una serie de temas o centros de interés que será en torno a los que los hablantes deberán escribir aquellas palabras que recuerden. Estas palabras se recogen y se les otorga una frecuencia, pero la frecuencia es matizada por "el grado de disponibilidad o espontaneidad" (López Morales, H., 1987: 64) que tiene en cuenta el orden de aparición de las palabras. Combinados ambos índices, frecuencia y grado de disponibilidad, se obtiene el rango real de una palabra. En el trabajo de Díaz Castañón se utilizan unas pruebas que tienden a extraer el léxico disponible, pero no efectúa las operaciones reseñadas por el lingüista puertorriqueño. No obstante, confecciona un vocabulario importante extraído de palabras escritas por alumnos de EGB y BUP, tanto de enseñanza pública como privada. Utiliza para su investigación una muestra de 3.150 escolares, de los que 1.890 son de EGB y 1.260 de BUP. Corresponden a la enseñanza pública 2.520, y 630 a la privada. *A priori* contempla la variable zona, pero no la de sexo. Realiza dos modalidades de pruebas para la obtención del material de trabajo:

a) Encuesta libre, que consiste en escribir durante cinco minutos las palabras que primero acudan a la mente.

b) Encuesta controlada: escribir los 20 primeros términos que acudan a la mente sobre una serie de temas o centros de interés[136].

Así se obtuvo un corpus de 850.000 palabras: 170.000 de la encuesta libre, que produjo 9.782 vocablos, lo cual supone una frecuencia media de 17'13%; y 680.000 palabras de la encuesta controlada, que arrojó 16.761 vocablos, con una frecuencia media, pues, del 40'57%. Con todo este material se compusieron varias listas, por orden alfabético y de frecuencia.

En el 2º volumen se presenta a su vez un vocabulario básico, compuesto por 2.773 vocablos -la mayoría sustantivos-, fruto de los dos tipos de encuesta, libre y controlada, con una frecuencia igual o mayor que 20. De este total, 939 vocablos coinciden en ambas encuestas. Al mismo tiempo, realizan otro trabajo con un corpus de 32.017 palabras de las obras literarias más leídas o próximas a los alumnos. Se extraen 6.957 vocablos, de los que se forma un vocabulario básico con términos de frecuencia 3 como mínimo. Este vocabulario lo componen 2.021 voces, que unidas al básico de los escolares, 2.773, supone 4.344 vocablos en total, porque 630 son comunes a ambos. En sus observaciones finales, constatan que con el tipo de prueba diseñado, no se iba a determinar el léxico activo, "sino ese otro, el formado por vocablos a los que el alumno asocia ya conceptos, pero quizá no de forma muy exacta" (Díaz Castañón y colbs., 1978: 408).

Tal vez la obra más ambiciosa relacionada con el estudio del vocabulario escolar sea el *Vocabulario básico en la E.G.B.*[137], fruto del esfuerzo y

---

[136] Díaz Castañón (1977: 15) selecciona los temas siguientes: Los animales, el campo, la casa y los muebles, comidas y bebidas, los diferentes oficios, la escuela y el material escolar, juegos y deportes, los medios de transporte, las partes del cuerpo, los vestidos.

[137] En la bibliografía lo hemos registrado bajo el nombre de sus coordinadoras: Casanova Rodríguez, Mª

coordinación de un equipo de investigadores y colaboradores de todas las Comunidades Autónomas del Estado, y con el apoyo del Ministerio de Educación y Ciencia. Si lo hemos incluido dentro de los trabajos de disponibilidad léxica ha sido porque una parte de él ha sido extraído siguiendo una técnica de test de asociación controlada de palabras. Pero su contenido es mucho más amplio, por lo que trataremos de ofrecer una ligera visión de todos los aspectos que recoge. Engloba en realidad varios vocabularios: vocabulario "tipo" y vocabulario "activo". El primero se creó por el sistema de jueces: varios especialistas hicieron un despojo del DRAE (edición 1970) y la selección de vocablos la cotejaron con varios vocabularios existentes (Juilland y Chang-Rodríguez; Rodríguez Bou; García Hoz; Díaz Castañón; Márquez Villegas). Posteriormente analizan el contenido semántico de las palabras para, de acuerdo con ese criterio, incluirlas en los distintos ciclos de EGB, para lo cual se consultó también a diversos profesores de la primera enseñanza. Para recoger el vocabulario "activo" se prepararon dos tipos de pruebas, una sugerida y otra espontánea. En la prueba espontánea se pidió a los alumnos que escribieran durante 5 ó 10 minutos las palabras que quisieran; en la prueba sugerida, debían escribir palabras relacionadas con 20 temas concretos.

Para comprobar el grado de adecuación del vocabulario "tipo" y para recopilar el corpus del vocabulario "activo", se aplicaron las pruebas pertinentes a una muestra representativa de alumnos de 2º, 5º y 8º de EGB de todas las Comunidades Autónomas de nuestro país. Con la corrección de las pruebas del vocabulario "tipo", determinaron el índice de comprensión por parte del alumnado de los vocablos sometidos a encuestas. Estas pruebas pusieron de manifiesto, entre otros aspectos, "las carencias expresivas de nuestros escolares, las cuales (y aquí se abren interesantes campos de investigación) pueden deberse no sólo a una pobreza de vocabulario activo, ya que no se ha manifestado así en las encuestas aplicadas, sino más bien a

---

Antonia y Rivera González, Melitina (coords.) (1989).

una pobreza de estructuras sintácticas" (Casanova, Mª A. y Rivera, M., 1989: X). Asimismo, el vocabulario "tipo" se organizó en campos nocionales, muy similares a los temas propuestos en la recogida de datos del vocabulario "activo".

El *Vocabulario básico en la E.G.B* recoge en el tomo I y en primer lugar todo el vocabulario "tipo", compuesto de 10.017 vocablos[138] ordenados alfabéticamente, con indicación de los ciclos y el número de acepciones recomendadas para éstos; en segundo lugar, registra un tratamiento específico de cada ciclo. Dentro de cada ciclo se ofrece el vocabulario correspondiente agrupado en campos nocionales, las pruebas de evaluación del vocabulario y los resultados obtenidos. El volumen II consta esencialmente de tres partes. La primera contiene todo el vocabulario "activo", extraído de un corpus de 266.729 ocurrencias[139], con un total de 18.881 vocablos[140] (nombres propios incluidos) ordenados alfabéticamente, distribuidos por ciclos y con indicación de la frecuencia relativa y frecuencia total; la segunda presenta el vocabulario según un orden decreciente de frecuencias y distribuido por Comunidades Autónomas; en la tercera parte se compara el vocabulario "tipo" y el vocabulario "activo", en orden alfabético y teniendo en cuenta los tres ciclos. También se recoge un cuadro informativo con los alumnos y los centros donde se aplicaron las pruebas. El volumen III se dedica a la didáctica del vocabulario: se compone de tres apartados extensos en los que se plasma la

---

[138] El número de vocablos que corresponde a cada ciclo es: 3.340 para el Ciclo Inicial, 3.490 para el Ciclo Medio y 5.185 para el Ciclo Superior.

[139] Se consiguieron 78.967 ocurrencias en el ciclo Inicial, 95.599 en el Ciclo Medio y 92.163 en el Ciclo Superior.

[140] 8.185 vocablos en el Ciclo Inicial, 11.329 en el Ciclo Medio y 15.016 en el Ciclo Superior.

experiencia en el terreno del campo léxico de varios profesionales de la educación.

Trabajos de disponibilidad léxica también se han llevado a cabo en Almería[141], Madrid[142], País Vasco[143], Gran Canaria[144]..., y en varios países hispanoamericanos[145].

---

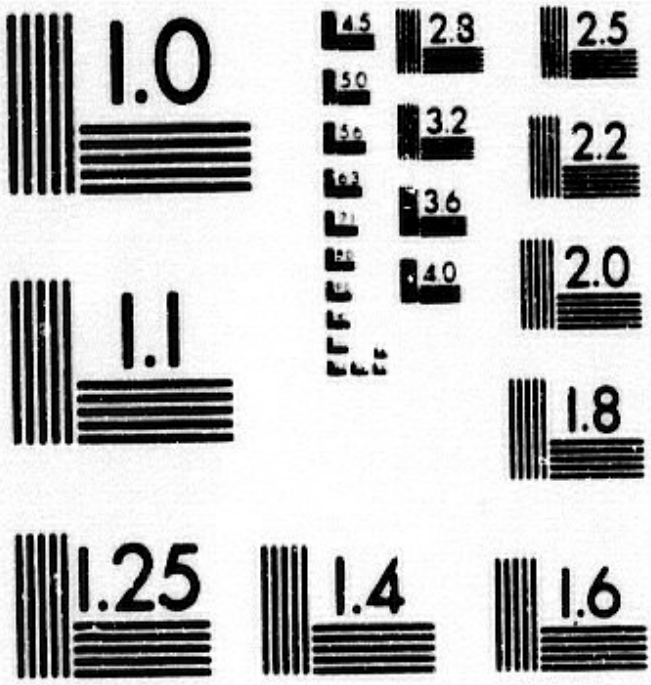
[141] Mateo García, M. V. (1995) realiza una investigación centrada en un principio sobre los centros de interés "el campo" y "la alimentación". Sus datos los contrasta con otros obtenidos en España y en Hispanoamérica. La autora nos indica también que sus estudios de disponibilidad léxica se inscriben en un proyecto inter-hispánico dirigido por Humberto López Morales.

[142] Benítez Pérez (1992) estudia el centro de interés de los medios de transporte en alumnos de COU de Madrid. La muestra la constituyen 257 estudiantes, de los que recoge 2401 N y 111 V. No hay diferencias sustanciales entre las distintas variantes de los sujetos encuestados (variante sexo, variante clase social y variante tipo de educación -privada/pública-), en cuanto a extensión de N y de V, ni con relación al índice de disponibilidad de los vocablos.

[143] Etxebarria, M. (1996) realiza su investigación con 245 alumnos de tres institutos -donde la enseñanza se imparte en castellano y el euskera se imparte como asignatura- y tres ikastolas -en las que la enseñanza se imparte en euskera y el castellano como asignatura- de Bilbao. Estableció 17 centros de interés y obtuvo un total de 72.891 palabras y 7.432 vocablos. Contrasta los resultados obtenidos por los alumnos de instituto y de las ikastolas entre sí y con otros trabajos de disponibilidad de Madrid y Gran Canaria, y concluye que no hay diferencias cuantitativas importantes entre los alumnos que han cursado en castellano, tanto en Bilbao como en las otras zonas españolas indicadas, y los alumnos de las ikastolas.

[144] Samper, J. A. y Hernández, C. E. (1997) llevan a cabo los estudios de disponibilidad léxica en Gran Canaria. Tienen en cuenta 17 centros de interés. Para recoger los datos seleccionan una muestra de 540 alumnos de COU. Sus objetivos son, por un lado, contrastar los resultados con otras investigaciones similares desarrolladas en el mundo hispano y, por otro, contribuir al establecimiento del léxico fundamental de estas comunidades.

[145] Además de los trabajos de López Morales en Puerto Rico, Max S. Echevarría ha llevado a cabo



MICROCOPY RESOLUTION TEST CHART
NATIONAL BUREAU OF STANDARDS
STANDARD REFERENCE MATERIAL 1010a
(ANSI and ISO TEST CHART No. 2)

experiencia en el terreno del campo léxico de varios profesionales de la educación.

Trabajos de disponibilidad léxica también se han llevado a cabo en Almería[141], Madrid[142], País Vasco[143], Gran Canaria[144]..., y en varios países hispanoamericanos[145].

---

[141] Mateo García, M. V. (1995) realiza una investigación centrada en un principio sobre los centros de interés "el campo" y "la alimentación". Sus datos los contrasta con otros obtenidos en España y en Hispanoamérica. La autora nos indica también que sus estudios de disponibilidad léxica se inscriben en un proyecto inter-hispánico dirigido por Humberto López Morales.

[142] Benítez Pérez (1992) estudia el centro de interés de los medios de transporte en alumnos de COU de Madrid. La muestra la constituyen 257 estudiantes, de los que recoge 2401 N y 111 V. No hay diferencias sustanciales entre las distintas variantes de los sujetos encuestados (variante sexo, variante clase social y variante tipo de educación -privada/pública-), en cuanto a extensión de N y de V, ni con relación al índice de disponibilidad de los vocablos.

[143] Etxebarria, M. (1996) realiza su investigación con 245 alumnos de tres institutos -donde la enseñanza se imparte en castellano y el euskera se imparte como asignatura- y tres ikastolas -en las que la enseñanza se imparte en euskera y el castellano como asignatura- de Bilbao. Estableció 17 centros de interés y obtuvo un total de 72.891 palabras y 7.432 vocablos. Contrasta los resultados obtenidos por los alumnos de instituto y de las ikastolas entre sí y con otros trabajos de disponibilidad de Madrid y Gran Canaria, y concluye que no hay diferencias cuantitativas importantes entre los alumnos que han cursado en castellano, tanto en Bilbao como en las otras zonas españolas indicadas, y los alumnos de las ikastolas.

[144] Samper, J. A. y Hernández, C. E. (1997) llevan a cabo los estudios de disponibilidad léxica en Gran Canaria. Tienen en cuenta 17 centros de interés. Para recoger los datos seleccionan una muestra de 540 alumnos de COU. Sus objetivos son, por un lado, contrastar los resultados con otras investigaciones similares desarrolladas en el mundo hispano y, por otro, contribuir al establecimiento del léxico fundamental de estas comunidades.

[145] Además de los trabajos de López Morales en Puerto Rico, Max S. Echevarría ha llevado a cabo

1.5.4.2.2. Vocabularios extraídos de libros de lectura o de libros de texto escolares

Aunque para su elaboración no se utilizan directamente las palabras que usan los estudiantes, sí al menos se relacionan con el ambiente escolar, pues se extraen con fines didácticos y de fuentes cercanas a los centros de enseñanza. En tal sentido, habría que citar el propio trabajo de Díaz Castañón y colbs. (1978), quienes ofrecen, como se ha indicado, un vocabulario de obras de lectura del Bachillerato.

El vocabulario técnico también ha sido objeto de estudio. En este caso, extraído de los libros de texto, bien sean éstos de enseñanza secundaria (antiguo bachillerato y preuniversitario), como el trabajo de García Hoz (1976), bien de enseñanza primaria, como el de Armayor (1983b). Otras investigaciones del mismo género son las de Pérez Vázquez (1958), Bellanato (1958) y Tamajón (1978)[146].

1.5.4.2.3. Listas de frecuencias: vocabulario usual

Armayor González (1975, 1983a, 1983b, 1985) realiza una serie de trabajos sobre el vocabulario con escolares de Córdoba. Los inicia (1975) con alumnos de seis años, de donde extrae un total de 2.964 vocablos, y sigue una metodología parecida al vocabulario de alumnos de 8º de EGB. En esta investigación (1983a), con alumnos de 8º, utiliza una muestra de 201 escolares, seleccionados de 9 centros, tres en cada una de las siguientes

---

estudios de disponibilidad léxica en Chile, según información bibliográfica recogida en el nº 6 de *Lingüística*, revista de la Asociación de Lingüística y Filología de la América Latina (ALFAL).

[146] Las tres últimas citadas por Justicia, F. y Palma, A. (1986: 586).

zonas cordobesas: sierra, campiña y ciudad. Para la obtención de datos emplea un doble método:

a) Fase espontánea, expresión oral. Consta de 420 conversaciones, de 15 minutos, de tema libre.

b) Fase sugerida, expresión escrita, en torno a 8 temas o núcleos de motivación. Así extrae un vocabulario de 4.286 vocablos, que presenta en dos ordenaciones: alfabética y por frecuencias.

Moral Santaella (1983) realiza un trabajo con una muestra de alumnos de un solo centro educativo. Surge esta investigación ante la necesidad de mejorar la ortografía de los alumnos, pero siguiendo las orientaciones didácticas en este terreno, que promulgan la enseñanza de la ortografía en estos niveles a través de un léxico básico, realizan un estudio de vocabulario de su centro con una muestra de 100 alumnos de 3º de EGB. Tras recoger varias redacciones y palabras sobre diversos temas, confeccionan el vocabulario básico compuesto de 2.558 vocablos.

Otro trabajo sobre el vocabulario infantil del que tenemos referencias es el de Bonifacio Jiménez, quien lleva a cabo una investigación basada en composiciones escritas de alumnos de 5º de EGB y dos cursos de 7º. Entre otras cosas, calcula el índice "razón tipo/figuración" o TTR (Type/Token Ratio), analiza las curvas acumulativas del vocabulario, y ofrece una lista de más de 7.000 vocablos[147].

En español también existen trabajos que funden o, mejor, contrastan varias listas de frecuencias. Nos referimos al *Vocabulario básico para la EGB* (1989), compuesto a través de un procedimiento mixto: por el criterio de jueces se seleccionaban las palabras apropiadas para los distintos niveles de

---

[147] Citado por el propio autor, Jiménez, B. (1986: 109-114).

EGB y se contrastaban éstas con los diccionarios de frecuencias del español de García Hoz, de Juilland y Chang-Rodríguez, de Rodríguez Bou, de Díaz Castañón y de Márquez Villegas (Cfr. 1.5.4.2.1.).

Desde perspectivas diferentes pero basándose también en la frecuencia de las palabras, han sido elaborados algunos vocabularios ortográficos. Por ejemplo, Mesanza López (1990a), tomando como base el vocabulario precedente, ha seleccionado aquellos vocablos con dificultades ortográficas. Estos vocablos han sido dictados a una muestra de 120 alumnos para cada nivel de EGB. Tras los correspondientes análisis, el autor ha podido establecer los distintos vocabularios ortográficos básicos por niveles, con el índice de dificultad ortográfica de cada palabra. El mismo autor (Mesanza López, 1990b) compone un inventario cacográfico pero esta vez basándose en la expresión de los escolares; en concreto, recoge redacciones de 1016 alumnos, de 34 colegios de 8º de EGB. Este trabajo recoge tanto los fallos de acento, signos de puntuación, como los vocablos con errores de letras. Las cacografías se ordenan según la frecuencia, de mayor a menor.

Como se puede apreciar, en España también el vocabulario de los alumnos ha despertado interés y se ha abordado su estudio con unas u otras finalidades y con más o menos coincidencia en la metodología, aunque con grandes diferencias si nos centramos en puntos particulares.

A nuestro juicio, quien hace un primer planteamiento serio a la hora de emprender un trabajo de estas características, sin descuidar prácticamente ningún aspecto, es Fernando Justicia[148]. Conocedor de los problemas que

---

[148] Justicia, F. ha confeccionado dos vocabularios (1985a y 1995) con materiales de los escolares de EGB, desde 1º a 8º. El primero (1985a) recoge muestras de expresión escrita de alumnos del Ciclo Inicial y Medio. El segundo (1995), supone una compilación del primer trabajo y de una investigación posterior llevada a cabo en el Ciclo Superior de EGB.

presentan estas investigaciones, aborda cada uno de ellos con rigor y exhaustividad. Fija con profundidad su metodología, en la cual no escapa casi ningún detalle, y presenta sus postulados perfectamente engarzados e interrelacionados, con coherencia, desde el inicio, con el planteamiento de los objetivos de investigación hasta la propia confección de las listas. Los estratos están perfectamente definidos y aporta referencias claras acerca de su consideración. Se basa en procedimientos científicos como la estadística y la estadística lingüística, somete sus resultados a determinadas leyes del vocabulario e incluso propone referencias para detener un recuento.

En su obra *El desarrollo del vocabulario. Diccionario de frecuencias*, compilación de su anterior trabajo (1985a) junto con otro del Ciclo Superior de al Enseñanza Básica, Fernando Justicia estudia una muestra del vocabulario de alumnos de todos los cursos de la EGB, desde 1º hasta 8º. Su objetivo primordial es analizar, desde perspectivas de la Psicología Evolutiva, el vocabulario escolar de estos alumnos, pertenecientes a la zona geográfica de Andalucía Oriental: Almería, Granada, Jaén y Málaga. Parte de un tipo de muestreo de afijación proporcional, y se obtiene una muestra total de 3.042 alumnos, seleccionados de 198 cursos de 26 Centros de EGB. Contempla las variables de zona, sexo y clase social. Utiliza dos pruebas para recoger los datos: una composición escrita de tema libre y un test de asociación libre de palabras. Obtiene un corpus de 528.544 palabras y 8.937 vocablos, cifra cercana a las 8.500 palabras fijadas por Averril[149] (a quien F. Justicia toma como referencia) de promedio de uso para los alumnos comprendidos entre los 6 y 13 años de edad[150]. Con los resultados organiza varias listas: de todos

---

[149] Averril, M.: *La vida psíquica del escolar*. Buenos Aires, 1965. (Citado por Justicia, F., 1985a),

[150] En su primera investigación (Justicia, F., 1985a) la muestra es de 2.166 sujetos, comprendidos entre 6 y 10 años de edad. Aquí obtiene un vocabulario usual de 5.750 unidades, extraído de un *corpus* de 258.983 palabras.

los vocablos, una por orden alfabético y otra por orden decreciente de frecuencias; un vocabulario común a todos los ciclos de EGB, compuesto por 2.123 unidades, y varios vocabularios diferenciales, que recogen las palabras que aparecen en un ciclo (por ejemplo, el Ciclo Inicial) y no aparecen en los restantes (Ciclo Medio, Ciclo Superior). Además, en la parte final, ofrece un diccionario silábico de frecuencias.

Fernando Justicia (1995: 10) define los trabajos como el suyo de la siguiente manera: "Los recuentos de frecuencias son cálculos inferenciales del vocabulario basados en la relación type/token (tipo/figuración; vocablo/palabra), es decir, en la cantidad de veces que aparece una misma palabra en proporción al número total de palabras emitidas por una muestra de sujetos". Por tanto, estos vocabularios tienen en cuenta el vocabulario activo, y frente a otro tipo de trabajos, en estos últimos, "las palabras que sirven como base al recuento no son independientes de las muestras de sujetos que las producen" (Justicia, F., 1995: 17).

Nuestra investigación se inscribe dentro del último apartado de la clasificación establecida; por tanto, responde en cierta medida a los presupuestos de los recuentos basados en listas de frecuencias que tratan de estudiar el vocabulario de una población escolar. Es decir, tratará, como se ha indicado (Cfr. 1.), de determinar el vocabulario usual de los alumnos de primer curso de Enseñanza Media, cuya edad por tanto oscilará entre los 14 y 15 años. Para ello, trabajamos con muestras de expresión escrita, recogidas en composiciones de producción espontánea de los escolares. Al mismo tiempo que cumplimos con este objetivo, ofrecemos los vocablos y sus frecuencias; es decir, la estructura del vocabulario responde a la de los vocabularios o diccionarios de frecuencias.

La metodología empleada toma como punto de referencia la seguida por F. Justicia en sus investigaciones, si bien con matices distintos en algunos

aspectos concretos de los objetivos, el sistema de muestreo, la estratificación o la presentación del vocabulario.

Creemos, por tanto, que los trabajos de vocabulario basados en cálculos de frecuencia tienen vigencia, y son necesarios para remozar consideraciones léxicas y para cuestiones pedagógicas. Es más, los estudios cuantitativos del vocabulario están plenamente integrados dentro del marco de los estudios lingüísticos[151], ya que nos aportan otra visión importante para obtener un conocimiento total de los hechos de lengua y sus resultados son útiles para muchos campos de la Lingüística Aplicada y la didáctica de la lengua. Prueba de la integración dentro de la Lingüística es la inclusión en obras de Lingüística General de capítulos dedicados a la lingüística cuantitativa, tanto de aspectos de sonidos o letras, como de gramática o léxico. Así lo demuestra la consideración de tales aspectos en algunas obras de autores tanto extranjeros (Lyons, J., 1975; Malmberg, B., 1988) como nacionales (Jauralde Pou, P., 1973; Fernández, A. R., Hervás, S. y Báez, V., 1989); y los muchos artículos publicados en numerosas revistas, tanto nacionales como internacionales, además de existir una extensa bibliografía acerca de lingüística cuantitativa que se edita aparte como tal especialización, sin olvidar el uso que de los recuentos se hace para otras disciplinas como la Psicología, Pedagogía o Psicolingüística.

Como colofón, recordemos la siguiente cita: "Una última cuestión fundamental, para que el lector tenga una información completa de los problemas semánticos y lexicológicos, es el aspecto cada vez más importante de la lingüística cuantitativa. Es un *segundo plano* básico actualmente. Frente a la mera consideración *cualitativa* de los hechos de lengua, que busca la descripción más o menos exhaustiva de formas, diferencias, clases,

---

[151] Ortega, J. (1983: 5) reconoce la importancia de estas investigaciones y así lo refiere en la justificación de su trabajo.

estructuras, subsistemas, etc., se va extendiendo la convicción cada vez mayor de que también lo *cuantitativo*, la frecuencia mayor o menor de los distintos elementos de la lengua necesita un estudio especializado, *válido* siempre desde el punto de vista exclusivamente científico -del conocimiento *total* de los hechos de lengua-, y *útil* para las aplicaciones prácticas en muchos campos concretos de la *lingüística aplicada*" (Fernández, A. R., Hervás, S. y Báez, V., 1989: 172).

1.6. Conclusión

1.6.1. La lingüística aplicada, que en principio tenía un campo de estudio restringido (la aplicación de la lingüística general y descriptiva a la enseñanza de idiomas) ha ido ampliando progresivamente sus dominios. En la actualidad, acoge un gran número de disciplinas, muy variadas entre sí, pero que tienen en común aplicar sus conocimientos para la solución de aquellos problemas en los que la lengua está implicada. Como más recientes incorporaciones al terreno de la LA hemos de destacar la lingüística cuantitativa. Por otro lado, una nueva dimensión acompaña a la lingüística aplicada: su cariz humano.

1.6.2. La estadística léxica resulta una disciplina indispensable para abordar una investigación de índole cuantitativa. Esta disciplina opera con criterios metodológicos importantes como la distinción entre vocablo y palabra, por un lado, y entre léxico y vocabulario, por otro.

Ante los problemas para determinar el léxico de un hablante o de un grupo, la estadística léxica actúa con manifestaciones concretas de la lengua, los textos, para cuantificar el vocabulario, ya que éste es un fenómeno del habla. Para cuantificar un vocabulario se necesita una norma léxica o norma de despojo. El mayor escollo -aún no superado- con el que se enfrenta la norma es, para la palabra, considerarla como unidad gráfica o como unidad de sentido, y para el vocablo, la homonimia sintáctica y la homonimia léxica.

Aunque se suelen dar orientaciones generales, cada recuento justifica aquella norma que considera más oportuna o coherente, por lo que los datos tanto cuantitativos como cualitativos que arrojan investigaciones de este tipo deben ser debidamente ponderados.

Al aplicar la norma léxica se obtienen dos valores numéricos: N (cantidad de palabras) o extensión del texto y V (cantidad de vocablos) o extensión del vocabulario. No existe ninguna fórmula matemática que permita calcular la relación entre N y V, en el sentido de que no se puede establecer con precisión que a un texto con una extensión N específica corresponda una extensión V determinada. Alrededor de estos valores numéricos (N y V) giran dos conceptos esenciales: la ley de crecimiento del vocabulario y la noción de riqueza léxica. La ley de crecimiento del vocabulario se basa en la idea, consabida, de que a medida que aumenta N aumenta V, aunque el crecimiento de V no es homogéneo: crece más rápidamente al principio que al final. Numerosos investigadores han ensayado sus fórmulas para dar cuenta de cómo se produce el crecimiento del vocabulario a lo largo de un texto. La relación entre N y V nos conduce también a la noción de riqueza léxica. Ésta, para la estadística léxica, es una noción cuantitativa que depende de:

a) extensión de los campos semánticos;
b) léxico de situación: componente estilístico, de carácter individual, y componente temático, ligado al tema del discurso;
c) número de sujetos de la muestra.

Debido a que no hay una ley que fije una relación numérica determinada para N y V, Charles Müller establece cuatro supuestos para poder comparar la riqueza léxica de dos textos de datos cuantitativos homogéneos. Para textos heterogéneos o de dimensiones diferentes propone la siguiente ecuación: $E(V^i) = V - \sum q^i V^i$. Conviene detener la aplicación de esta fórmula cuando el producto sea inferior a 0'5.

Todo recuento presenta una distribución de frecuencias, formada por los vocablos ordenados de mayor a menor frecuencia y agrupados en clases de frecuencias con sus efectivos correspondientes. La distribución de frecuencias permite apreciar la estructura cuantitativa de un vocabulario, cuyos rasgos más característicos son:

1. A mayor frecuencia, menor número de efectivos.
2. Los vocablos de frecuencia 1 son los más numerosos.
3. La importancia de los efectivos varía en sentido inverso a la frecuencia.

La ley de Zipf estudia la estructura cuantitativa de un vocabulario. Postula que la frecuencia por rango es una constante (F x R = C). No obstante, presenta una serie de irregularidades relacionadas con las frecuencias más altas y más bajas.

1.6.3. Desde perspectivas lingüísticas se suele aludir a cuatro tipos de vocabulario: usual, común, básico y fundamental. Autores como De Mauro consideran que el vocabulario común está integrado por una serie de palabras de diversa índole que suelen ser conocidas por los hablantes. Este vocabulario comprende a su vez el vocabulario básico, integrado por vocablos que son ampliamente conocidos, y el vocabulario fundamental, constituido por las palabras de máxima frecuencia.

Aunque los conceptos de léxico y vocabulario se relacionan con los respectivos de lengua y habla, no todos los lingüistas hacen esta distinción, debido a la interrelación que presentan.

Desde perspectivas cuantitativas se menciona (además de los cuatro vocabularios anteriores) el léxico disponible. Los recuentos suelen fijar el

vocabulario usual tomando como referencia la frecuencia (por ejemplo, 1/100.000 para Lorenzo Delgado) y la frecuencia media. El vocabulario común se determina de acuerdo con la intersección entre varias fuentes y el vocabulario fundamental con relación a las palabras de mayor frecuencia. Para seleccionar el léxico básico se emplean, además de la frecuencia, medidas muy precisas. El vocabulario disponible es algo peculiar, pues es conocido, pero no se emplea sino en determinadas situaciones comunicativas y va asociado a variantes temáticas. Si en las listas de frecuencia, que recogen el vocabulario usual, aparecen palabras gramaticales, sustantivos de carácter general, verbos, adjetivos..., en el vocabulario disponible se registran sobre todo sustantivos concretos. La distinción entre ambos tipos de vocabulario está plenamente aceptada. Son dos vocabularios compatibles y necesarios para programar debidamente la enseñanza sistemática del léxico.

1.6.4. Los estudios sobre aspectos cuantitativos de la lengua no son nuevos. Hay noticias de que los gramáticos alejandrinos dejaron constancia de esa curiosidad. Desde finales del siglo XIX comienzan a elaborarse de manera continuada recuentos de vocabulario. Estos primeros trabajos comenzaron recopilando extensiones muy amplias de palabras, pero los corpus fueron progresivamente acortándose a medida que se perfeccionaban los criterios de selección de las fuentes y los cálculos para fijar los vocablos.

Todos estos estudios se inician con alguna finalidad didáctica destacada: bien para mejorar la taquigrafía, bien para componer tests estandarizados, bien para facilitar las relaciones comerciales..., y, sobre todo, para la enseñanza de lenguas.

Las fuentes de selección son muy diversas y variados son también los registros. Aunque se trabaja sobre muestras de lengua escrita, algunos estudios atienden a la lengua oral. No faltan compilaciones en las que se funden diversas listas o donde se mezclan composiciones de los escolares con

la expresión de los adultos. Por otro lado, casi todas las lenguas cuentan con estudios cuantitativos de vocabulario.

Paralelamente a estas investigaciones se fueron desarrollando otras que estudian el vocabulario infantil, insertas unas en el estudio de la adquisición del lenguaje y en la evolución o aumento del vocabulario con la edad, otras.

1.6.5. De nuestro idioma también se han elaborado algunos recuentos (los trabajos de Merás y Roth, Keniston, Buchanan, T. R. Céspedes, Rodríguez Bou...) con similares características a los de otras lenguas. En España los estudios cuantitativos de vocabulario los inicia García Hoz, con su *Vocabulario usual, vocabulario común y vocabulario fundamental.* Pero será otro estudio sobre el español, el *Frequency Dictionary of Spanish Words*, el que otorgará una nueva dimensión a estos trabajos. Su metodología contempla aspectos novedosos y exhaustivos que han sido aplicados en investigaciones similares de otros idiomas.

La lengua oral usada en TV nos ha proporcionado algunos trabajos interesantes. Uno es de Lorenzo Delgado, elaborado con una finalidad pedagógica esencial: su aplicación en la escuela; otro, de J. Ortega, que aporta la peculiaridad de registrar no sólo palabras sino también unidades mayores.

1.6.6. En España también se han desarrollado estudios sobre el vocabulario de los escolares. Varios de ellos consisten en presentar una lista de vocablos para que sean reconocidos: son los llamados vocabularios de reconocimiento. García Hoz lleva a cabo una investigación de estas características con alumnos de distintas edades para medir la evolución cuantitativa del vocabulario. El trabajo de Márquez Villegas se vale de esta técnica para determinar un vocabulario básico destinado a enseñanza de español para extranjeros. La investigación de Guerra Salas difiere de las

anteriores ya que tratará de ver el nivel de comprensión y de producción que los alumnos poseen del léxico básico del español peninsular.

Dentro del terreno de los léxicos disponibles podría enmarcarse el trabajo de Díaz Castañón y colbs., puesto que recoge las palabras de acuerdo con encuestas libres y encuestas controladas alrededor de ciertos centros de interés. El *Vocabulario básico en la E.G.B.* participa, en parte, de las características de la disponibilidad léxica, pero su contenido es mucho más amplio, ya que también gran parte de éste se creó por el sistema de jueces. La investigación precedente pone de manifiesto las carencias expresivas de los alumnos, las cuales podrían deberse no sólo a la pobreza léxica sino también a la pobreza sintáctica de los escolares. Actualmente los estudios de disponibilidad léxica se extienden por numerosos puntos de nuestra península y por varios países hispanoamericanos.

Algunos vocabularios han sido extraídos también de los libros de texto y de lectura de los alumnos.

Otros recuentos escolares tratan de recoger el vocabulario usual. Para ello recopilan el material de muestras de expresión espontánea. Los trabajos que al respecto hemos hallado se centran en la EGB. Quizás el trabajo más significativo sea el de Fernando Justicia, elaborado con muestras de expresión escrita de todos los cursos de EGB y dentro del marco de la Psicología Evolutiva.

Nuestra investigación se sitúa en este último apartado (los recuentos de vocabulario escolares que tratan de fijar el vocabulario usual), se centra en la E. M. y responde a consideraciones lingüísticas y educativas.

CAPÍTULO II: METODOLOGÍA

2. Metodología

Los trabajos sobre el vocabulario de una población escolar realizados en España para determinar el léxico usual se realizan en el nivel de EGB, por un lado, y desde perspectivas más psicológicas y pedagógicas que lingüísticas[152], por otro. Los estudios sobre el vocabulario referentes a la Enseñanza Media utilizan técnicas para extraer el vocabulario disponible (Díaz Castañón y colbs., 1977 y 1978) o el vocabulario de reconocimiento (García Hoz, 1977; Márquez Villegas, 1975; Guerra Salas, 1989), pero no se basan en la expresión espontánea de los escolares. En ese sentido, nuestra investigación aporta la novedad de recoger muestras de expresión escrita espontánea, usada en composiciones. Además, obedece a consideraciones lingüísticas, si bien el campo de la pedagogía, en el sentido de la práctica docente, es la finalidad que lo alienta.

Como trabajo de investigación científica, el nuestro pasa por una serie de etapas u operaciones metodológicas indispensables. Fijamos así, en primer lugar, el objeto de estudio. Como éste se centra en una determinada población, emitimos algunas hipótesis[153] relacionadas con ciertos aspectos de

---

[152] Véanse las indicaciones de Justicia, F. y Palma, A. (1986: 580-581).

[153] Carrión, J. J. (1985: 92) entiende por hipótesis "una afirmación provisional que hay que verificar mediante el recurso a un procedimiento de análisis. Se trata de una suposición basada en la intuición, teorías establecidas, etc., que permanece en cuestión hasta que no se somete a la prueba de los datos".

sus componentes, como la riqueza léxica o locuacidad[154], para seleccionar. Teniendo en cuenta los puntos precedentes, debemos seleccionar la muestra y el material -o corpus- indispensable para el análisis. Hemos de recurrir a la estadística y diseñar un sistema de muestreo apropiado, en este caso, de los sujetos que producirán los textos. Los textos se someten a un despojo informático, previo establecimiento de una norma léxica y una unidad textual clave -la palabra, en este caso-. Unidad que posteriormente habrá que clasificar de acuerdo con unos criterios, lingüísticos en nuestro caso, para finalizar con un tratamiento estadístico mínimo de algunos de los datos del trabajo.

Para una investigación como la que nosotros nos proponemos, la metodología atraviesa, pues, por las siguientes etapas:

"Las cuestiones metodológicas que presentan estos trabajos radican principalmente en torno a los siguientes aspectos básicos: el establecimiento de los límites del vocabulario, la fijación del universo léxico de procedencia, las técnicas de registro de datos, la consistencia de las frecuencias observadas, la representatividad de las muestras de sujetos, la definición de las unidades de análisis, el proceso de cuantificación de los datos y, por último, las formas de codificación, clasificación y presentación de las muestras de palabras obtenidas en los recuentos" (Justicia, F., 1995: 10). Aquí se recogen las claves que habrá que ir desvelando con cierto rigor para conseguir un resultado óptimo.

---

[154] Utilizamos este término tal y como lo emplea Justicia, F. (1985a y 1995) para referirse al número de palabras emitidas por un sujeto o la media de palabras de un grupo determinado de sujetos encuestados.

2.1. Determinación del tema de trabajo

La investigación versa sobre el siguiente punto: la determinación del vocabulario usual, activo, de los alumnos de 1º de EE.MM. (BUP y FP) y del nivel equivalente de Reforma o experimentación del nuevo sistema educativo (LOGSE), es decir, 3º de Enseñanza Secundaria. Los alumnos pertenecen a una zona geográfica concreta: Granada, capital y provincia. Hasta aquí la parte más amplia del estudio. Trabajamos sobre muestras de expresión escrita. La expresión escrita tiene ciertas ventajas sobre la oral para ser cuantificada, ya que no debe ser transcrita. También supone un registro más rico que la expresión oral. Sabido es que la comunicación lingüística se lleva a cabo en múltiples situaciones, y no todas ellas exigen el mismo esfuerzo ni imponen las mismas exigencias idiomáticas. Por ejemplo, no se realiza el mismo esfuerzo ni el grado de exigencia es el mismo cuando hablamos en la calle o en público, ni cuando escribimos una carta o un artículo de una revista. "Evidentemente, el idioma de un individuo se hace más rico, es decir, se hace un instrumento más poderoso y más capaz si el sujeto en cuestión ejerce de manera regular su actividad lingüística en contextos exigentes" (Núñez, R. y del Teso, E., 1996: 18). Y es la lengua escrita la situación comunicativa donde más se exige a nuestro dominio del idioma, por sus características especiales, pues es un tipo de comunicación en la que emisor y receptor no están presentes simultáneamente, no comparten el contexto, por lo que no se podrán pedir ni ofrecer aclaraciones, y en consecuencia se deben formular las ideas con mayor claridad y, por tanto, donde el nivel de exigencia es mayor. Esto comporta que la explicitud del mensaje se realice aumentando el número de palabras semánticamente más llenas o con mayores componentes semánticos y utilizando mayores recursos gramaticales. Por todo ello, para algunos autores, estos son los motivos de que sea la lengua escrita, en su faceta de comprensión y expresión, el objetivo fundamental de la enseñanza

del idioma[155]. Y esto a pesar de que la Lingüística moderna ha subrayado la prioridad de la lengua oral o hablada[156], como la más completa y esencial, sobre la lengua escrita, lo cual ha repercutido sobre todo en la enseñanza de una segunda lengua. En la lengua materna, debido a que el escolar ya tiene incorporada en el nivel de enseñanza media al menos, la competencia fonológica y domina la sintáctica, la enseñanza del Bachillerato suele centrarse, es cierto, sobre la lengua escrita, como lo demuestra también el hecho de que en los libros de texto destinados a tal fin la casi totalidad de los ejercicios propuestos apuntan a esta modalidad de expresión[157].

---

[155] Asertos que exponen Núñez, R. y del Teso, E. (1996: 19).

[156] Ya el mismo Saussure en su *Curso de lingüística general* (15ª ed. Buenos Aires, Losada, 1976), p. 72, afirmaba: "Lengua y escritura son dos sistemas de signos distintos; la única razón de ser del segundo es la de representar al primero; el objeto lingüístico no queda definido por la combinación de la palabra escrita y la palabra hablada; esta última es la que constituye por sí sola el objeto de la lingüística". Véase también, por ejemplo, Alarcos, E. (1965: 10). Quilis, A. (1977: 16) destaca varios principios esenciales en los que se asienta la lingüística moderna, y uno de ellos es precisamente que el lenguaje es un fenómeno ante todo oral: "La lingüística actual no desdeña la lengua escrita, como la tradicional arrinconaba la hablada, pero da más importancia a esta última, porque: a) genéticamente, la lengua hablada es anterior a la escrita; b) la lengua escrita es una confirmación de la lengua hablada, que se configura artísticamente en la lengua escrita literaria; c) la lengua escrita pierde elementos que posee la lengua hablada: las variaciones significativas de tono, de acento, los valores expresivos, que difícilmente se pueden transcribir con los signos ortográficos, y todo el lenguaje gestual que acompaña al acto de habla.

Las diferencias entre lengua hablada y lengua escrita son muy importantes. El estudio de las dos es necesario. Por otra parte, la forma escrita es un fuerte nexo de unión en una lengua, nexo que evita su desmembración".

[157] Poch Olive, M. (1984: 294-308) afirma y pondera la primacía de la lengua oral sobre la escrita, pero reconoce que en la enseñanza es la escrita la que predomina, y como prueba alude a que en los libros de texto más usados tan sólo ha encontrado un ejercicio de lengua oral, y ése no estaba dedicado a fomentar la producción oral, sino a analizar las desviaciones con respecto a la norma.

Por otro lado, el vocabulario obtenido lo sometemos a un criterio mínimo de dispersión y frecuencia (teniendo en cuenta para ello las diferencias más significativas entre los distintos niveles de alumnos que constituyen la muestra) para contrastarlo con el léxico básico de nuestra comunidad lingüística -en este caso, el recogido en el *FDSW*- y otro vocabulario usual de mayor actualidad, el Vocabulario de Orientación Didáctica (VOD), de Lorenzo Delgado (Cfr. 1.5.3. y 4.8.), y de acuerdo con las diferencias, sugerimos la parte del léxico básico peninsular que debería ser objeto de planificación en la Enseñanza Secundaria. Pero desde el principio advertimos que en absoluto pretendemos fijar un vocabulario cerrado para el proceso de enseñanza-aprendizaje, sino precisar, de acuerdo con unos criterios rigurosos (Cfr. 4.4., 4.5. y 4.8.) y al mismo tiempo siguiendo la opinión de lingüistas ilustres (Humberto López Morales, Antonio Quilis, Gregorio Salvador...), aquel vocabulario que consideramos esencial e imprescindible para una programación sistemática, sin que esto entre en contradicción con otros tipos de léxico, como el disponible o el técnico, que igualmente deben ser materia de aprendizaje.

Somos conscientes de que la enseñanza del léxico es algo que preocupa a los docentes, pero lo cierto es que a menudo y con más frecuencia de lo habitual, el nivel gramatical ocupa casi todo el programa de nuestra asignatura, y aunque se dedican algunas unidades al nivel léxico-semántico, su enfoque se efectúa desde perspectivas más o menos teóricas, pues, en la práctica, se descuida la enseñanza sistemática del léxico, entendida como enseñanza instrumental, que sirva para aumentar la competencia lingüística y comunicativa del escolar. Algunos autores corroboran con sus estudios estas afirmaciones, y llegan incluso a proponer una reforma de los programas de la asignatura de Lengua Española, en los que deban incluirse mayor parte de unidades didácticas dedicadas al léxico[158].

---

[158] Salvador, G. (1985: 72) propone 12 temas para aumentar el estudio del léxico y ponerlo así en el lugar

Dentro de estos planteamientos, pues, se inscribe nuestra investigación: hay que conceder más importancia al vocabulario en la enseñanza de la lengua y se ha de partir del que dominan los alumnos para hacer una programación cuidadosa.

2.2. Objetivos

De acuerdo con el tema de trabajo, nos hemos fijado una serie de objetivos generales. En primer lugar:

a) Determinar el vocabulario usual del alumnado de 1º de Enseñanza Media y Reforma (3º de Enseñanza Secundaria Obligatoria), cuya edad oscila entre 14 y 15 años. (Cfr. 2.1.). El grueso de nuestra investigación gira en torno a este objetivo y la mayor parte de las páginas lo reflejan, algo obvio si consideramos que tanto el marco teórico, la metodología y la presentación del vocabulario han ocupado y exigido casi todo nuestro tiempo. Pero este ha sido el primer paso importante para cumplir con otros dos objetivos igualmente esenciales e interrelacionados con el primero, sin el cual no habríamos podido desarrollarlos. Los dos objetivos, y en orden de importancia en nuestro trabajo, son los siguientes:

b) Proponer aplicaciones didácticas del vocabulario extraído y diseñar algunas orientaciones para la planificación y adquisición del léxico en el entorno escolar, partiendo de los diccionarios de frecuencias.

---

que a su juicio le corresponde (Cfr. nota 357).

c) Proporcionar información lingüística sobre el léxico real que poseen los alumnos, a través de la presentación y análisis del primer objetivo (a).

La primera aplicación, como hemos indicado más arriba, es servir de punto de referencia para fijar el vocabulario que debe ser, a nuestro juicio, objeto de enseñanza-aprendizaje. En función de esta finalidad hemos realizado un breve análisis del léxico de nuestros alumnos, porque creemos que conocer algunos aspectos de él puede contribuir a una mejor planificación de la enseñanza.

2.3. Población

En estadística, se denomina "población" a todo conjunto de cualesquier elementos sometidos a su estudio, e "individuo" a cada uno de esos elementos del conjunto. Por tanto, una población la constituiría un grupo de personas, pero también, en el caso del campo en el que nos movemos, un conjunto de palabras[159]. Nosotros emplearemos el término en el primer sentido: conjunto de individuos. Esta es, por otra parte, la acepción general que se recoge en los diccionarios.

Para determinar el vocabulario activo, de uso, de la población escolar que estudiamos, es necesario conocer dos datos esenciales: el número de individuos que integran esa población, y, al mismo tiempo, la cantidad de palabras que deberán registrarse como base del recuento. Ya se han señalado las cuestiones metodológicas esenciales para establecer un vocabulario; se ha indicado también que el vocabulario es la realización, la "actuación" del léxico, por lo que el vocabulario refleja el léxico mental de donde ha sido extraído, y

---

[159] Así lo afirma Müller, Ch. (1973a: 21).

desde esta perspectiva, cualquier texto es una muestra de ese léxico[160]. Además, aunque el léxico como tal se considera un conjunto finito de unidades, no se tiene en realidad esa consciencia debido sobre todo a su naturaleza de "carácter abierto", en el sentido de que está continuamente en movimiento, introduciendo palabras nuevas, apartando otras que quedan en desuso, o "resucitando" arcaísmos que habían caído en el más profundo de los olvidos[161]. Se sabe también que el léxico de un individuo varía en algo con respecto a otro, e incluso que las probabilidades de aparición son distintas de unas palabras a otras, según que sean palabras gramaticales o palabras de contenido semántico, las cuales están muchas veces unidas al tema de elocución (Cfr. 1.3.3.)

Así, pues, queda claro, primero, que para estudiar el léxico de una población se debe partir del vocabulario, que, al ser la realización del léxico, es lo que podemos computar, y segundo, que el léxico de una determinada población lo constituiría la suma de todos los léxicos individuales de sus componentes. Éstos, a su vez, forman un conjunto tan numeroso que resulta imposible aprehender sus léxicos. Por tanto, hay que seleccionar una muestra de sujetos y del vocabulario que usan. Pero de tal manera que el vocabulario final que se extraiga sea realmente representativo de la totalidad de la población. Así, se impone analizar con detenimiento los criterios de selección de la muestra para que los condicionamientos que se han presentado más arriba queden atenuados en la medida de lo posible.

---

[160] Guiraud, P. (1960: 84) afirma: "C'est que le vocabulaire d'un texte est le reflet du lexique mental d'où il est tiré, de l'emsemble des mots dont l'auteur dispose pour la composition du texte, mots dont chacun est affecté d'une probabilité qui est déterminée à la fois par le lexique total de l'auteur et par les limites sémantiques et stylistiques de son sujet. Tout text donc est un échantillon de ce lexique...".

[161] Véase el artículo de Seco, M. (1985).

Todo esto nos conduce al primer gran problema que se plantea y que surge en trabajos sobre el vocabulario: el de la "representatividad". Es decir, primero: ¿qué número de alumnos será necesario seleccionar para que sea representativo de esa población total? Y segundo: ¿qué número de palabras se deberán recoger para que el vocabulario sea representativo? Ambas cuestiones están vinculadas, a su vez, con el sistema de muestreo. Es decir, y como hemos indicado, como la población es muy numerosa, se debe seleccionar una porción de ella que la represente, y al mismo tiempo, se deberá recoger un número de palabras suficiente para fijar el *corpus* del que se extraiga el vocabulario.

2.5. Determinación de la muestra

Para determinar el tamaño de la muestra y repartir ésta según el sistema de muestreo, hemos seguido una serie de pasos para dar respuesta a las interrogantes apuntadas en el epígrafe anterior. Es decir, ha sido imprescindible establecer una serie de principios:

1. Acudir a disciplinas auxiliares, como la estadística y la estadística léxica, para dar solución a problemas que la Lingüística no puede resolver por sí misma, como el caso de la representatividad de la muestra.

2. Claridad para determinar la parte de la población que entra en nuestra investigación y la parte que se excluye de acuerdo con el criterio de la edad.

3. Objetividad. Utilizar un método de selección de la muestra que garantizará la objetividad y eliminará en lo posible la mediatización y el subjetivismo. Por ello nos hemos servido de tablas de números aleatorios para seleccionar la muestra.

4. Por supuesto, *sincronía*[162], por lo que se ha fijado un período temporal para las muestras seleccionadas. En concreto, 2 cursos escolares: 91/92 y 92/93. Durante este período, obtuvimos el 90% del material. Con posterioridad, el 10% restante. Hemos de reconocer que la faceta de recogida de datos ha sido la más problemática, debido a que nuestra investigación no dispone de un equipo investigador ni de ningún presupuesto, como muchos trabajos de este tipo, y hemos tenido que contar con la colaboración del profesorado de los distintos centros. Desde el primer momento en que se inició el trabajo de campo hemos tenido una respuesta positiva en casi todos los Institutos, pero tuvimos que repetir los contactos e instrucciones en años sucesivos, hasta completar definitivamente toda la muestra. También trabajos precedentes al nuestro dan noticias del problema de recogida de datos[163].

5. Dispersión y diversidad, para recoger una proporción significativa de los puntos heterogéneos que caracterizan los niveles educativos que estudiamos. A ello responde el tipo de estratificación fijada.

---

[162] Nos referimos a uno de los componentes de la dicotomía de Saussure, y por tanto *sincronía* hemos de entenderla como el estado de lengua considerado en su funcionamiento en un momento determinado del tiempo. La idea de enumerar este criterio (que pudiera resultar obvio) responde, en primer lugar, al rigor que queremos otorgar a nuestra metodología, y en segundo lugar, pretendemos seguir así a los autores de trabajos de frecuencia anteriores, en los cuales se tuvo que fijar un período histórico para seleccionar las fuentes.

[163] Díaz Castañón y colbs. (1977: 10) exponen sus pretensiones de recoger muestras de 4000 alumnos, y a pesar del equipo de investigación con el que contaba, no pudieron obtener pruebas en determinadas zonas de Asturias, por lo que "al no haber tenido respuesta de algunos centros, ha quedado reducida a una muestra real de 3.150". Nosotros hemos sido un poco más pertinaces y no hemos cejado hasta completar la estadística. También Guerra Salas (1989: 77) tuvo que recoger el material de sus encuestas durante dos cursos académicos.

6. Acumulación. Por lo que a cada alumno se le ha pedido dos composiciones, en vez de una composición al doble de alumnos. Esto ha sido así porque se trata de buscar una muestra ni escueta ni amplia de cada alumno. Como los objetivos apuntan a un léxico usual, no conviene multiplicar las posibilidades de vocablos de frecuencia 1, que responden a variedades estilísticas o individuales. Aunque, como en cualquier trabajo de frecuencias, los hápax constituyen el mayor número de efectivos.

7. Delimitación espacial. Nos hemos circunscrito a una zona geográfica específica: la provincia de Granada. Creemos que a pesar de esta delimitación, los resultados pueden trascenderla y, con las debidas precauciones y cautelas, relacionarse con los mismos niveles de los alumnos con característica similares de, al menos, toda Andalucía.

8. Y por último, el de economía. Es decir, acudir a sistemas que, sin perjuicio de los postulados anteriores y ofreciendo garantía en los resultados, facilitaran en algo la tarea. Nos hemos servido de tablas estadísticas y del conglomerado como unidad muestral.

De acuerdo con estos principios y para fijar la muestra, hemos tenido que conocer en primer lugar el censo real de población escolar que estudiamos. Por ello, nos ha sido necesario acudir a fuentes solventes de la administración educativa; en concreto, las fuentes de las que partimos fueron las correspondientes Inspecciones de BUP y FP de la Delegación Provincial de Educación y Ciencia de Granada.

Decidimos eliminar del censo el número de alumnos de 1º de EE.MM. matriculados en Estudios Nocturnos y Bachillerato a Distancia, porque para poder acceder a estos niveles se requieren una serie de circunstancias

especiales, entre otras, tener 18 años[164]. La idea de excluir esos dos tipos de enseñanza está relacionada, pues, con la edad: nos interesaba que la población estudiada fuese homogénea en este aspecto, por lo que nos centramos en los estudios diurnos, cuyos alumnos tienen unas edades comprendidas entre 14 y 15 años aproximadamente. No olvidemos que la edad es uno de los factores que influyen en la adquisición del lenguaje y por supuesto en la riqueza y cantidad de palabras que se utilizan[165].

Restando pues las cifras correspondientes a los alumnos del Nocturno y educación a distancia, partimos de un censo que arrojaba las cifras siguientes: 8554 alumnos en BUP, 4476 en FP y 940 en Reforma (anticipo de la LOGSE). Nos hemos servido de tablas estadísticas para determinar el número representativo de alumnos en cada nivel; en concreto, hemos tomado como base las tablas de Arkin y Colton "Tables for statistians", en las que se recogen distintas muestras correspondientes a diferentes márgenes de errores y para distintos tamaños de la población, con un nivel de confianza del 95%[166].

En nuestro caso, con la selección de la muestra, teníamos la respuesta a la primera pregunta, es decir, determinar el número representativo de la población, pero no respondíamos a la segunda cuestión: número de palabras

---

[164] En el caso de los Estudios Nocturnos, la ordenación que rige es la siguiente: Orden de 1 de agosto de 1978 (B.O.E. de 15 de septiembre), Resolución del de 12 de noviembre de 1985 (B.O.E. del 30 noviembre), y Orden de 1 de julio de 1988 (B.O.E. del 6 de julio). Aquí se especifica que para poder optar a estos estudios se debe tener 18 años, o 16, y en este último caso, acreditar su condición de trabajador. Para el IBAD, pueden verse las Resoluciones de la Junta de Andalucía, donde se especifica la condición de tener 18 años para poder cursar, o estar en una serie de circunstancias peculiares.

[165] Recuérdense los trabajos de Justicia, F. (1985a y 1995), en los que el nivel de instrucción es influyente, pero también la edad; García Hoz (1977); Almela Pérez (1986)

[166] Recogida por Pulido San Román (1984: 183).

necesario par efectuar el recuento y establecer el vocabulario. Para ello tomamos como referencia algunos trabajos publicados en los que se determina el vocabulario de alumnos de EGB. Aunque decidimos que quien con mayor rigor había aplicado una metodología para la determinación del vocabulario era F. Justicia[167]. Con esta referencia, nuestro propósito fue aproximarnos al número de palabras base del recuento de esos trabajos o incluso superarlo.

Para conocer la media aproximada de palabras que emiten los alumnos de este nivel, realizamos una prueba con un curso del instituto "Ulyssea" de Ugíjar y otra con un curso del "Padre Suárez". El primer instituto es de Reforma y de zona rural, y el segundo de BUP y de zona urbana. Ambos centros podrían, a nuestro juicio, extrapolar muy bien las diferentes características de la población que estudiamos.

Cada uno de los alumnos de los cursos seleccionados efectuó dos composiciones escritas: una narración y una descripción, ambas de tema libre. La media de palabras por alumnos fue la siguiente:

- I. "Ulyssea":.....413
- I. "Padre Suárez":.....510
- Media de ambos centros:.....461'5

De acuerdo con estos datos y con la muestra seleccionada en principio según las tablas estadísticas mencionadas -muestra: 844 sujetos-, se superarían las 350.000 palabras. Pero nuestra intención era alcanzar al menos las 400.000. Por lo que decidimos ampliar proporcionalmente la muestra de cada uno de los niveles, hasta quedar definitivamente como sigue:

---

[167] El recuento de Justicia, F. (1985a) asciende a 258.983 ocurrencias y 5.750 vocablos.

- 1º de B.U.P. .....419 alumnos.
- 1º de F.P. .....406 alumnos.
- 1º de Reforma (3º E.S.O.).....107 alumnos.
- Total .....932 alumnos.

2.6. Método de muestreo

Determinado definitivamente el tamaño de la muestra, necesitábamos un método garantizará la representatividad de la población. Elegimos así un *muestreo aleatorio estratificado de afijación proporcional y por conglomerados*[168]. Recurrimos, primero, a un procedimiento de selección aleatorio, muy utilizado, basado en el *aleas* o azar. Profundizando y matizando un poco más, elegimos la estratificación porque la población de la que haremos las estimaciones la consideramos en cierto modo heterogénea. Por lo que hemos tenido que dividirla previamente en estratos.

Con el término *afijación* designamos la operación matemático-estadística que consiste en repartir la muestra entre los diferentes estratos. La afijación proporcional es un criterio sencillo, pero presenta ciertas ventajas. Consiste en repartir la muestra total en partes proporcionales a la población de cada estrato. Para ello hay que conocer los siguientes datos:

m: muestra total
p: población total
p1: tamaño de la población en el primer estrato
p2: tamaño de la población en el segundo estrato
p3, p4, p5, et..

---

[168] Véase Pulido San Román (1984: 165-174); Azorín, F. y Sánchez-Crespo. J. L. (1986: 83-89).

Para averiguar *n1* o tamaño de la muestra en el primer estrato, *n2* o tamaño de la muestra en el segundo estrato, *n3* o tamaño de la muestra en el tercer estrato... y así sucesivamente, habría que aplicar la siguiente fórmula para estimar los correspondientes tamaños de la muestra en cada uno de los estratos:

$n_1 = p1 \times m / p$

$n_2 = p2 \times m / p$

etc.

Como unidad muestral elegimos el conglomerado. El conglomerado la ventaja que presenta es que, al ser la unidad muestral no un elemento de la población, sino un grupo de ellos, disminuye el número de unidades muestrales. Pero presenta el inconveniente de que los elementos de cada conglomerado tengan características comunes, con lo cual disminuiría el grado de representatividad de la muestra.

En nuestro caso, los "conglomerados" con los que hemos operado han sido grupos de alumnos. Pero consideramos que el inconveniente anterior queda matizado e incluso superado con el criterio de dispersión que hemos introducido para determinar, por un lado, el número de centros, y por otro, la estratificación que hemos llevado a cabo. (Cfr. 2.6.).

Para seleccionar la muestra, y para garantizar el principio de objetividad, se ha asignado un número a cada grupo tomando como referencia el código del Instituto al que pertenece. Enumerados así los cursos, hemos utilizado posteriormente una tabla de números aleatorios para seleccionar los grupos necesarios y completar la muestra de los distintos estratos.

Para intentar cumplir con el principio de dispersión, diversidad y representatividad, hemos dividido la población en una serie de estratos de los que se ha extraído la parte proporcional de acuerdo con la fórmula de afijación que se ha indicado. Para la estratificación, hemos aplicado un diseño factorial que incluye, pues, el nivel y tipo de enseñanza -1º de EE.MM.: BUP, FP, Reforma-, el tipo de centro educativo -privado y público- y la zona -urbana y rural-. Los estratos que hemos determinado y la muestra en cada uno de ellos se especifica a continuación:

2.6.1. Tipo de enseñanza: BUP, FP, Reforma (LOGSE)

Según este criterio, hay que establecer una primera consideración atendiendo a tres modelos de enseñanza que integraban las EE.MM. y existentes en el momento de iniciar la investigación. Estos tres niveles (BUP, FP, Reforma) tienen características peculiares en cuanto a los elementos integrantes de las planificaciones y programaciones que orientan el quehacer educativo. Es decir, estos tres modelos no coinciden exactamente ni en los objetivos, ni en los contenidos, ni en la metodología, ni en los criterios de evaluación ni, por supuesto, en los requisitos que deben reunir los alumnos para acceder a ellos. Por tanto, hemos considerado imprescindible seleccionar, en primer lugar, una muestra proporcional de estos tres niveles. No olvidemos que estos alumnos, pertenecientes a lo que recientemente era la Enseñanza Media, poseían características similares a todos los que ahora integran la Enseñanza Secundaria (LOGSE).

2.6.2. Tipo de centro educativo

La decisión de recoger una muestra de alumnos que cursan estudios tanto en centros públicos como privados responde a la idea de "representatividad". No queremos pasar por alto ningún elemento que integre el sistema

educativo, por lo que hemos elegido pues una muestra proporcional de alumnos de enseñanza pública y privada.

2.6.3. Variante zona: urbana y rural

Consideramos esencial obtener muestras de alumnos tanto de la capital como de las distintas zonas de la provincia: interior y costa.

2.6.4. Nivel educativo, tipo de centro y zona

Combinando simultáneamente los tres factores de la estratificación, la muestra queda repartida como sigue:

2.6.4.1. BUP

a) Centro público y zona urbana.

Tamaño de la muestra.....132 alumnos.

Los alumnos fueron seleccionados de los siguientes institutos:

- I.B. "Albayzín"
- I.B. "Fray Luis de Granada"
- I.B. "Padre Manjón"
- I.B. "Padre Suárez"

b) Centro público y zona rural

Tamaño de la muestra.....139 alumnos

Institutos:

- I.B. "Hiponova" .Montefrío
- I.B. "Jiménez Montoya". Baza
- I.B. "Jiménez Quesada". Santa Fe
- I.B. "Julio Rodríguez". Motril
- I.B. "Padre Poveda". Guadix

Total BUP y centro público.....271 alumnos

c) Centro privado y zona urbana

Tamaño de la muestra.....126 alumnos

Centros:

- "Dulce Nombre de María"
- "Juan XXIII". Chana
- "Progreso". Zaidín
- "Sagrado Corazón"
- "Virgen de Gracia"

d) Centro privado y zona rural

Tamaño de la muestra.....22 alumnos

Centros:

- "Nuestra Señora del Rosario". Atarfe

Total BUP y centro privado..... 148 alumnos

e) Total BUP..... 419 alumnos

2.6.4.2. FP

La Formación Profesional poseé una serie de características especiales, debido sobre todo a la diversidad de ramas que la integran. En Granada, capital y provincia, y en los distintos centros públicos y privados, se impartían las ramas siguientes: administrativa y comercial, agraria, artes gráficas, automoción, delineación, electricidad y electrónica, hostelería y turismo, madera, metal, moda y confección, peluquería y estética, química, sanitaria, servicio a la comunidad... Estas ramas a su vez se subdividen en diversas especialidades: informática de empresas, explotación agropecuaria, impresión, mecánica del automóvil, cocina, operador de laboratorio, jardín de infancia... Ante un panorama tan amplio, decidimos que en la selección de la muestra también se recogiera una representación proporcional de las múltiples ramas que integran este tipo de enseñanza. Por tal motivo, las agrupamos por bloques atendiendo al número de alumnos y al perfil académico de cada especialidad. De esta manera se conseguía que las ramas con mayor número de alumnos estuvieran bien representadas y garantizábamos al mismo tiempo que la muestra reflejara la realidad heterogénea de este nivel de enseñanza, pues es conocido entre los estamentos educativos de la FP que no todo el alumnado presenta las mismas características y que el hecho de elegir una rama y otra está en función de las aptitudes de los escolares. En lo que concierne al dominio y uso de la lengua, los alumnos que cursan determinadas Ramas poseen una mayor competencia lingüística, como en la Rama Administrativa, Sanitaria, Delineación, etc., frente a los alumnos que cursan otras Ramas como Automoción, Electricidad...

Para la selección de la muestra en FP tuvimos también en cuenta, pues, estos planteamientos. Por estos motivos, indicaremos no sólo los centros donde se llevó a cabo, sino también la rama o especialidad:

a) Centro público y zona urbana

Tamaño de la muestra.....118

Institutos:

- I.F.P. " Luis el Crespo" de Armilla. 1º de Administrativo
- I.F.P. "Hermenegildo Lanz". 1º de Electricidad
- I.F.P. "Virgen de las Nieves". 1º de Madera
- I.F.P. "Hurtado de Mendoza". 1º de Cocina

b) Centro público y zona rural

Tamaño de la muestra.....149

Institutos:

- I.F.P. "Pedro J. Montoya" de Baza. 1º de Administrativo
- I.F.P. "ILYWRA" de Illora. 1º de Administrativo
- I.F.P. Alhama de Granada. 1º de Administrativo
- I.F.P. "La Zafra" de Motril. 1º de Electricidad
- I.F.P. de Iznalloz. 1º de Automoción
- I.F.P. "Acci" de Guadix. 1º de Peluquería

Total FP y centro público.....267 alumnos

c) Centro privado y zona urbana

Tamaño de la muestra.....105

Centros:

- "Caja General de Ahorros". 1º de Administrativo
- "San Juan Bosco". 1º de Administrativo
- "Ave María". 1º de Artes Gráficas

d) Centro privado y zona rural

Tamaño de la muestra.....34

Centros:

- "Virgen de la Cabeza" de Motril. 1º de Sanitaria

Total FP y Centro privado.....139

e) Total FP.....406

2.6.4.3. REFORMA (LOGSE)

En Reforma tan sólo contemplamos la variante zona, debido a que la población de la que seleccionamos la muestra era muy pequeña y prácticamente toda pertenecía a centros públicos.

a) Zona urbana

Tamaño de la muestra.....49 alumnos

Institutos:

- I.E.S. "Aynadamar"
- I.E.S. "Hermenegildo Lanz"

b) Zona rural

Tamaño de la muestra.....58 alumnos

Institutos:

- I.E.S. "Al-Andalus" de Almuñécar
- I.E.S. "La Sagra" de Huéscar
- I.E.S. "Ulyssea" de Ugíjar

- Total Reforma (LOGSE).....107

c) Total FP y Reforma (LOGSE).....523

Con este sistema de muestreo se observa la diversidad en lo que a la ubicación geográfica de los centros se refiere. En concreto, y para Granada capital se han seleccionado centros que pertenecen tanto al centro de la ciudad como a la periferia; en cuanto a la provincia, los centros están situados en todas las zonas de la provincia, desde la costa -Motril, Almuñécar-, como las zonas de interior, y de comarcas tan diversas como La Alpujarra -Ugíjar-, Huéscar, Los Montes Orientales -Iznalloz-, Alhama, Montefrío y las cercanías de Granada, como Santa Fe, Atarfe y Armilla.

2.7. Consideraciones previas sobre la riqueza léxica y la locuacidad de los alumnos que integran los distintos estratos

Pensamos que la estratificación que hemos realizado en nuestra investigación no es gratuita y se hizo de acuerdo con unos planteamientos previos acerca de la competencia lingüística de los escolares. En esos planteamientos tuvimos en cuenta tanto trabajos anteriores (García Hoz, 1977; Justicia, F., 1985a; Díaz Castañón, 1977 y 1978) como consideraciones de orden académico y de experiencia docente. Estas consideraciones, expuestas como hipótesis parciales, las contrastamos después con los resultados obtenidos (Cfr. 3.1.).

2.7.1. BUP, FP, Reforma (LOGSE)

Evidentemente, se puede aventurar que, en lo que concierne a la competencia lingüística, y más en concreto, a la competencia léxica, los alumnos de BUP deberán ofrecer una mayor producción de palabras y mayor riqueza léxica que los alumnos de FP y Reforma. Y digo Reforma, porque en los años en que se inició la investigación, al menos en Granada, casi todos los Centros que estaban experimentando el nuevo sistema educativo y que, por tanto, anticiparon la LOGSE, eran Centros de FP y muchos de los alumnos que a ellos asistían venían con retraso de EGB. Además, este aserto lo corrobora un dato objetivo, pues para acceder a Reforma y a FP, no se necesitaba 8º de EGB aprobado, es decir, el Graduado Escolar.

2.7.2. Zona urbana, zona rural

La segunda estratificación, es decir, la distinción entre zona urbana y zona rural, así como las demás distinciones, se deben sobre todo al principio de

representatividad, variedad y dispersión que desde el principio ha guiado nuestro trabajo.

Además, en varios trabajos anteriores (Justicia, F., 1985a; Armayor, H. (1975, 1983a, 1983b) se contempla esta distinción, porque el hábitat, evidentemente, produce diferencias a todos los niveles: no sólo en número de palabras y riqueza léxica de los escolares, sino en cuanto a variedad de vocabulario.

Creíamos que, en principio, los alumnos de zona urbana, es decir, de Granada capital, deberían presentar una mayor competencia léxica, que se traduciría en un mayor número de ocurrencias y de vocablos. Además, en todos los trabajos consultados, los resultados indican esta tendencia. Pero matizamos que los institutos de Enseñanza Media se encuentran en poblaciones grandes, que se erigen en centros culturales y económicos de su comarca, por lo cual este factor contribuiría a paliar las diferencias.

2.7.3. Centro público y centro privado

La distinción en BUP y FP entre ambos tipo de centros -y por tanto, entre lo que se conoce como enseñanza pública y privada- se debe, más que nada, a lo dicho anteriormente: contemplar en nuestra investigación toda la población de Enseñanza Media.

En este estrato, creemos que la diferencia en lo concerniente a competencia lingüística no debe ser muy significativa[169], por lo que no lo tendremos en cuenta en la presentación definitiva de los datos cuantitativos. Evitamos así, al mismo tiempo, mostrar unas listas de palabras con datos demasiado dispersos.

---

[169] Así queda constancia en los trabajos de Guerra Salas (1989) y de Benítez Pérez (1992).

2.7.4. Variante de sexo: alumnos, alumnas

Aquí hemos de hacer una aclaración: En el sistema de muestreo que hemos seguido no tuvimos en cuenta, al principio, los datos cuantitativos referidos al número de alumnos y alumnas, pero creemos que el hecho de haber aplicado la estratificación proporcional es suficiente garantía para asegurar un porcentaje fiable y representativo de alumnos y alumnas. En ese sentido, se han contrastado, posteriormente, los datos que ofrece el MEC relativos al número de alumnos y alumnas matriculados en Granada en los niveles que estudiamos y durante los cursos 91/92 y 92/93, en los que recogimos la casi totalidad del material, y hemos obtenido los resultados que a continuación exponemos:

En Primero de BUP, la muestra de alumnas seleccionada por nosotros es de 195, lo que supone un porcentaje del 46'53% , y la muestra de alumnos es de 224, lo cual supone el 53'47%. Los datos de la Estadística del MEC acerca del número de matrículas en 1º de BUP de Granada y Provincia arrojan un total de 19.063 alumnos, de los que 9.754 son alumnas. Estas cifras, traducidas en porcentajes, suponen que el 51'16% son alumnas y el 48'84% alumnos. Si estas últimas cifras las contratamos con los porcentajes ofrecidos en primer lugar, y que corresponden como se ha dicho a nuestra muestra, podremos apreciar una pequeña diferencia de menos de 5 puntos aproximadamente. Consideramos que no es significativa, y se debe sobre todo a la selección de la muestra en centros de enseñanza privada, pues la equivalencia de porcentajes en la enseñanza pública es casi perfecta, debido tal vez a que en ésta los grupos de alumnos son mixtos, es decir, los cursos están integrados por porcentajes más o menos proporcionales de alumnas y alumnos. No ocurre esto en la enseñanza privada, donde existen, al menos en Granada, varios centros donde la enseñanza se dirige a alumnas o alumnos exclusivamente.

En los otros niveles de enseñanza -FP y Reforma (LOGSE)- que hemos agrupado en un solo estrato, las diferencias son aún más reducidas. Porque nosotros seleccionamos un total de 235 alumnas, que representa el 45'80% del total de la muestra en este estrato, y un total de 278 alumnos, lo que supone el 54'20%. En la Estadística del MEC se ofrece una cifra de 8.513 alumnos, de los que son alumnas 3.775, lo cual supone un porcentaje de 44'34. Aquí la diferencia es más pequeña y muy poco significativa. Esto creemos que se debe a que en FP el número de centros de enseñanza privada destinado con exclusividad a un alumnado femenino es menor e igualmente es menor la selección de alumnos en este nivel del muestreo: 139 de un total de 513.

El número definitivo de alumnas de nuestra muestra asciende a 430 de un total de 932, lo cual supone un porcentaje de 46'14%. Los alumnos seleccionados han sido pues 502, cuyo porcentaje es de 53'86%. La Estadística del MEC arroja un total de 27.576 alumnos, de los que 13.529 son alumnas. Por tanto, el porcentaje de alumnas matriculadas es de 49'06% y el de alumnos de 50'94%.

A tenor de estos datos no cabe sino concluir que la muestra seleccionada por nosotros es lo suficientemente representativa del número de alumnos y de alumnas, con lo cual podremos afirmar que el sistema de muestreo seguido en nuestra investigación es válido y garantiza los resultados. Por ello, emitimos también unas hipótesis que posteriormente contrastaremos con los datos que se obtengan. Pensamos, pues, de acuerdo con investigaciones precedentes (Justicia, F., 1985a y 1995) que las alumnas deberán emitir un mayor porcentaje de palabras pero la extensión de su vocabulario será algo menor. No obstante, las diferencias no deberán ser muy significativas, como así se aprecia en estudios como los de García Hoz (1977) y de Guerra Salas (1989).

Para terminar este apartado hemos de hacernos una autocrítica. Quizás se eche de menos en esta estratificación el aspecto social. No vamos a discutir la importancia de tal factor en el uso del idioma ni la relevancia que están adquiriendo los enfoques sociológicos en la Lingüística. No obstante y como justificación a tal decisión ha sido el considerar dos sistemas de enseñanza no obligatoria: el BUP y la FP, que ya marcan lo suficiente las diferencias de competencia entre los alumnos que eligen o van a un tipo de enseñanza u otro. Además, y dentro de la FP, como hemos indicado, hay determinadas ramas que presentan niveles más bajos en la utilización de la lengua materna -y otros aspectos-. No olvidemos que la educación -o mejor, la ausencia de ella- es uno de los factores que marca las diferencias de las variedades diastráticas, y todos los alumnos seleccionados en nuestra muestra han recibido al menos ocho años de enseñanza. Algunos testimonios de trabajos precedentes refuerzan esta decisión: el estudio de Díaz Castañón y colbs. (1977:11) y el artículo de M. V. Mateo García (1995). En este último se alude al trabajo de P. Benítez Pérez sobre disponibilidad léxica, realizado con estudiantes de Madrid (Nivel COU), en el que se vio que no hay diferencias significativas entre los tres niveles socioculturales estudiados, quizás porque los alumnos de COU tienen ya formado su léxico. También es cierto que en nuestra investigación, y teniendo en cuenta los problemas que comporta y en los que nos hemos visto inmersos, es más cómodo no hacer tal distinción, pues hacerla supone tomar algún o algunos criterios como base -los estudios de los padres, su nivel de ingresos, etc.-, ya que de hecho no hay indicadores comúnmente aceptados que indiquen claramente las variedades sociales[170],

---

[170] Peronard, M. (1978: 112) se queja de que no hay estudios sociológicos que determinen la situación estratificada de la sociedad de una ciudad, ni existen datos que permitan seleccionar los indicadores de grupo social: "Aparte de ciertas investigaciones a nivel latinoamericano que señalan que el factor entrada económica no es muy buen indicador y que educación sistemática parece ser más adecuada, poco o nada se sabe al respecto". Ella tomará como criterios los factores barrio en que habita el niño y ocupación del padre. Para López Morales (1978: 74) es el factor educación el que impone la diferencia de los distintos

como se observa en los criterios de los distintos trabajos de esta índole que se han llevado a cabo[171]. Adoptar alguno de estos criterios hubiera supuesto manejar datos amplísimos de los centros en los que hemos seleccionado una muestra, y como hemos indicado al principio, no tenemos en nuestra investigación un grupo de personas que se dividan el trabajo de encuesta. Además, creemos que con el sistema de muestreo elegido habrá en la muestra una representación de todos los estratos socioculturales, que es lo que nos interesa, aunque su estudio no podremos realizarlo a posteriori como hemos hecho con la variante sexo. Pero insistimos que lo importante para nosotros era la representatividad de la muestra, porque se trata de poner de manifiesto el nivel de vocabulario que presentan unos alumnos determinados para contribuir así a la enseñanza de la lengua materna. Por lo que los resultados se podrán extrapolar y no quedarán circunscritos a un determinado grupo, sino a un nivel educativo determinado y a una edad concreta, si bien considerando que la nota característica de cualquier nivel educativo es la heterogeneidad.

---

registros; Morales, A. (1978: 91) considera para su investigación -con niños de 10 años- como indicador de variedad socioeconómica el tipo de escuela: privada/pública; Lope Blanch (1978: 44-45) tiene en cuenta el grado de instrucción de los encuestados -que son adultos- y su profesión. En definitiva cada investigador de acuerdo con las características esenciales de su trabajo y del contexto en que se lleva a cabo, elige aquellos factores que cree mejores y más justificativos de su decisión.

[171] García Hoz (1977: 13) expone: "Ante la dificultad que presenta la determinación del ambiente social en que viven los escolares en este primer estudio nos hemos limitado a diferenciar la clase popular, por un lado, y la clase media, por otro, considerando la clase popular a los niños de las escuelas suburbanas y gratuitas de Madrid, y como clase media, a los que van a las escuelas que exigen contribución económica de sus alumnos". Armayor González, H. (1983a: 9) contempla como variable la situación socioeconómica (alta, media, baja), pero no indica los criterios en que se basa para hacer tal delimitación; Justica, F. (1995: 29) señala que la "variable social se estableció con posterioridad a la estratificación de la muestra. Dicha variable se cuantificó con los datos de un cuestionario en el que se solicitaba a los alumnos información sobre el estatus cultural y socioeconómico de la familia de procedencia. Con estos datos se establecieron tres niveles en la variable, correspondientes a tres clases sociales: alta, media y baja".

2.8. Corpus o universo léxico

El problema de la determinación del corpus o universo léxico, es decir, la cantidad de palabras necesaria para hacer un recuento con garantía se ha planteado desde diferentes posiciones. Dentro de la propia lingüística, P. Guiraud (1960) cuestiona la pobreza numérica de los corpus de donde se extraen los vocabularios básicos, es decir, aquellos vocabularios que tratan de fijar las palabras más frecuentes de un idioma a partir de observaciones estadísticas. Guiraud funda sus críticas basándose en primer lugar en la teoría de que la frecuencia de las palabras constituye un atributo esencial de los signos lingüísticos, por lo que éstos reposan en nuestra memoria con su probabilidad, y de ahí que se repitan con una frecuencia fija en un estado de lengua dado[172]. Por ello, para el lingüista francés la extensión de las compilaciones tiene que ser de tal manera que permita desgajar una lista de base útil en la práctica, teniendo en cuenta que la "fréquence des mots dans les textes est le reflet de leur probabilité dans la mémoire des auteurs; la fréquence relative du mot tend vers sa probabilité à mesure que la compilation augmente, elle présent donc un écart aléatoire qui diminue progressivement à mesure que la fréquence augmente et il est donc possible de déterminer à partir de quel moment le rang du mot dans la liste correspond avec une approximation suffisante à son rang réel dans le dictionnaire de la langue" (Guiraud, P., 1960: 95-96). Partiendo de la fórmula $r^{1}.^{35} \times f = constante$, podría calcular el rango verdadero de una palabra según la extensión del corpus: por ejemplo, el vocablo de rango 500 estaría entre el 420 y 635 rango de un corpus de 300.000 palabras, o entre el rango 450 y el 560 para un corpus de un millón, o entre el rango 480 y 530 para un corpus de cinco millones; el

---

[172] Véase Guiraud, P. (1960: 16, 74). Esta opinión categórica no la comparten Müller, Ch. (1965b: 33-34) ni Michéa, R. (1967: 8-9)

rango teórico 2000, estaría entre el rango 1350 y 6000 para un corpus de 300.000, o entre el rango 1750 y 5500 para un corpus de un millón, o entre 1900 y 2700 para cinco millones de ocurrencias. Por lo que una lista de 300.000 palabras no ofrece garantía más allá del vocablo 500, y en una lista de un millón de ocurrencias no se podría obtener un vocabulario con garantía de más de 1000 vocablos, etc. Algo parecido opina Keil (1965: 103), para quien, siguiendo a Frumkina, se necesitarían 10 millones de palabras para alcanzar un margen de error de un 1 ó un 2% a partir de la palabra mil.

No obstante y a pesar de teorías tan enjundiosas, investigadores sesudos como Juilland, conocedor del problema, sopesa previamente la cantidad de palabras del corpus, la metodología, el sistema de muestreo y los criterios para seleccionar el listado de vocablos definitivo, y, así, opera con una muestra de palabras menos extensa. Autores como Charles Müller o Amparo Morales, parecen aceptar la esta decisión, como se deduce del análisis que el primero realiza del *Frequency Dictionary of Spanish Words*, y de la adaptación de los presupuestos de Juilland por parte de la lingüista puertorriqueña en el recuento que efectúa[173].

En el terreno de la Psicología Evolutiva, Fernando Justicia (1985a: 23-24) hace unas reflexiones interesantes, y ofrece otra perspectiva para abordar un asunto tan espinoso. El autor granadino propone una serie de características que sirven de indicadores para detener un recuento:

a) Cuando se obtiene un número de vocablos deseado: 1000, 2000, etc. Las referencias para fijar el número de vocablos de una determinada población la proporcionarían estudios anteriores. Por ejemplo, él toma como referencia a Averril.

---

[173] Nos referimos a los trabajos ya citados de Müller, Ch. (1965c) y de Morales, A. (1986).

b) Estructura estable de la distribución de frecuencias. Es decir, cuando el recuento no modifica esencialmente el número de vocablos que integran las distintas clases de frecuencia. (Clase de frecuencia: conjunto de vocablos que tiene la misma frecuencia).

c) Cuando un vocabulario presenta un orden decreciente de frecuencias parecido a otro vocabulario. Es decir, cuando las palabras de uno y otro ocupan más o menos el mismo orden.

d) Proporción en palabras contenido: nombres, adjetivos, verbos..., y en palabras función: artículos, preposiciones...

e) Ratio P/V: Cociente entre palabras y vocablos. Cociente que crece, evidentemente, a medida que aumentamos el corpus. Pero para F. Justicia no es una medida estable a la hora de contrastar trabajos con muestras de palabras desiguales.

El autor aclara que estos criterios o características no son excluyentes, pueden ser complementarios.

Algunas de estas consideraciones presuponen que el proceso de codificación de palabras, de lematización, etc. ya se haya terminado para poder contar con todos los datos precisos y establecer las comparaciones con otros trabajos. Por lo que en principio, decidimos, siguiendo a F. Justicia, fijar un corpus lo suficientemente amplio para conseguir un número de vocablos determinado. Tuvimos también en cuenta estudios en los que se establecen unos límites más o menos indicativos del vocabulario que utilizan los escolares de una edad concreta. Recordamos así, de nuevo, la diferencia entre léxico y vocabulario, y que los diccionarios de frecuencias registran el vocabulario usual, activo. No sabemos, pues, el número de lexemas que integra el léxico de un individuo, ni de una población. No obstante, hay autores que ofrecen cifras al respecto. En concreto, hay quien piensa que serían 24.000 palabras las que constituirían el léxico de un adulto, distribuidas como sigue: 12.000 sustantivos, 5.500 adjetivos, 4.500 verbos y

1.000 adverbios[174]. A pesar de estas aproximaciones numéricas, se ha de hacer la siguiente observación: dentro de estas cifras entraría lo que conocemos hoy con el nombre de léxico pasivo y léxico activo. Éste seguramente está integrado por una cantidad menor. En tales términos se mueve F. Justicia (1985a: 27), para quien "los vocabularios muestra por el contrario presentan una perspectiva diferente. Su finalidad es ofrecer una muestra del lenguaje del niño. [...] Son muestras del léxico infantil. Ofrecen una menor precisión respecto al número total de palabras conocidas (amplitud del vocabulario), sin embargo, proporcionan una mayor información respecto a cuáles son las palabras más utilizadas, con qué frecuencia aparecen y el contexto en que se dan. De aquí que se les denomine vocabularios fundamentales, básicos o usuales. El promedio de palabras que presentan por edad es sensiblemente inferior al que proporcionan los vocabularios de reconocimiento. [...] Por otra parte, las diferencias también resultan obvias si consideramos el tipo de competencia lingüística que unos y otros analizan. Los vocabularios muestra utilizan como base del recuento palabras producidas por el niño. Muestrean, por tanto, sólo una parte de su léxico activo. Los vocabularios de reconocimiento muestrean un léxico mucho más amplio: palabras no sólo producidas por el sujeto sino también otras que el niño puede reconocer y comprender aunque no emplee normalmente y que formen parte, por tanto de su léxico pasivo". F. Justicia inscribe su trabajo dentro de los vocabularios muestra, y se propone conseguir alrededor de 5.500 vocablos, tomando como referencia a Averril, quien fija la amplitud media del vocabulario de niños de 6 a 10 entre 2.500 y 5.500 vocablos. Para conseguir esta cifra, F. Justicia partió de un corpus de 258.983 ocurrencias, de donde extrajo un vocabulario de 5.750 vocablos.

---

[174] Pastora Herrero, J. F. (1990: 58) recoge los datos de Peytard y Genouvrier, *Linguistique et enseignement du français*, Larousse, 1970

Nosotros hemos seguido un criterio semejante, tan sólo hemos cambiado las cifras, ya que la amplitud del vocabulario fijado para los jóvenes de 14 años es de 9.500 vocablos[175]. Por ello nos propusimos obtener un corpus de más de 400.000 ocurrencias. Para tal fin, hemos recogido 1.864 composiciones escritas de los escolares, que han arrojado un total de 412.433 palabras u ocurrencias. El total de vocablos es de 10.353. Esta cifra puede parecer excesiva, por lo que hemos de introducir alguna aclaración. Para establecer un vocabulario muestra, basado en la frecuencia de las palabras, se tiene que usar una norma léxica o norma de despojo (Cfr. 2.11.) De ahí que las cifras de un recuento sean válidas para ese recuento, y no se puedan comparar con otros que hayan adoptado una norma diferente, ya que, posiblemente, existan tantas normas como recuentos. La nuestra tiene un alcance muy amplio, pues contempla tanto aspectos formales (morfológicos y sintácticos) como léxicos -homonimia- El resultado es una extensa información lingüística llena de matices.

El corpus de nuestro trabajo creemos que es lo suficientemente representativo para establecer un vocabulario usual y un vocabulario básico. El primero lo constituye la suma total de vocablos (10353) y el segundo se podría establecer utilizando como criterios una determinada frecuencia mínima y la dispersión según los estratos que fijamos para determinar el tamaño de la muestra. Este vocabulario básico se podría utilizar para seleccionar las palabras esenciales que integren textos de lectura para escolares con retraso, o para planificar la enseñanza del vocabulario en niveles educativos inferiores al que nosotros estudiamos, o para adaptaciones curriculares en la Enseñanza Secundaria Obligatoria[176].

---

[175] García Hoz (1977: 31) ofrece una lista de palabras conocidas para distintas edades, según Averril en su obra *La vida psíquica del escolar*, Buenos Aires, 1956, p. 176.

[176] Aclaramos en este momento que, para no aumentar en exceso los listados, no ofreceremos al final una

2.9. Recogida de datos

Somos conscientes de que el procedimiento de obtención de datos es importantísimo y puede condicionar en mucho los resultados finales. Cuando se trata de establecer el "vocabulario usual" de una comunidad de hablantes, se debe acudir al concepto de *frecuencia*, y tomar muestras de textos orales o escritos de esa comunidad. Por tanto, nuestra investigación no recoge el léxico disponible, por lo cual no contemplamos las pruebas de test de asociación libre y controlada de palabras. El test de asociación controlada, como dijimos, consiste en pedir al alumno que escriba un número determinado de palabras en torno a unos centros de interés. Este tipo de prueba apunta más, en rigor, a un léxico pasivo o "latente", en el que predominan sobre todo sustantivos, después algunos verbos, y por último, pocos adjetivos y adverbios; no suelen aparecer palabras "útiles" o gramaticales.

Nosotros, sin dejar de reconocer la importancia del léxico disponible, decidimos registrar el léxico activo, el que usan normalmente los escolares, para ofrecer así una proporción real de las clases de palabras que integran su vocabulario. Igualmente elegimos la lengua escrita, modalidad que tiene sus peculiaridades y ofrece ciertas ventajas para ser cuantificada. Por tanto, para nuestra investigación, nosotros propusimos las siguientes pruebas:

a) Narración de tema libre.
b) Descripción de tema libre.

---

nueva relación con los vocablos que podrían integrar un vocabulario básico, pero se puede fijar sin dificultad consultando el vocabulario usual y seleccionando las palabras de una determinada dispersión y frecuencia.

La elección del tema libre responde a varios presupuestos. El primero fue tratar de evitar la mediatización que supondría fijar un tema o unos temas determinados, lo cual comportaría que determinadas palabras de contenido semántico obtuvieran una frecuencia que no correspondiera a su uso real, pues, como sabemos, la frecuencia de una palabra está también en función con el tema del discurso. El segundo es que, dada la edad y el grado o nivel académico de los alumnos, sabíamos que habían superado ya el segundo estadio del aprendizaje de la composición. Es decir, tradicionalmente se vienen señalando varias etapas en el proceso de enseñanza-aprendizaje de la redacción: período del pensamiento dirigido y período de la espontaneidad, o redacción dirigida y libre[177]. La primera fase, de aprendizaje dirigido, llegaría hasta los diez años y constaría de redacciones de tema obligatorio, y el segundo período, para el que se requiere el adiestramiento del período anterior, se iniciaría a partir de esa edad, con lo cual los escolares de nuestra muestra tienen superadas las fases y pueden realizar perfectamente una redacción de tema libre.

Dentro de la redacción de tema libre elegimos, en primer lugar, la narración, forma de expresión más usual en los escolares cuando no se exige un modelo de composición determinado. La otra forma de expresión elegida ha sido la descripción, pues creemos que ambas, narración y descripción, son las más practicadas por los alumnos. Si bien la descripción ofrece mayores dificultades a los escolares que la narración[178], por lo que su enseñanza se

---

[177] Véase Sepúlveda, F. (1992: 19); Iglesias, J. (1971a: 212); Romero López, A. (1986: 88).

[178] Gili Gaya hace la siguiente observación: "En los ejercicios de composición escrita realizados por nuestros discípulos, obtenemos siempre los mejores resultados en los temas narrativos, y los mayores fracasos en los descriptivos, tanto en las edades que abarca la escuela primaria como en los primeros años de la enseñanza media. La causa reside, además del interés mayor que un relato suele suscitar, en que la narración da hecho al niño el orden expositivo que debe seguir, sin más que dejarse llevar por el

suele posponer, lo cierto es que ambas formas son conocidas y practicadas por los alumnos del nivel que nos ocupa. Ahora, somos conscientes de que la descripción producirá un número menor de vocablos que la narración, pero quizás aporte un mayor número de adjetivos, cuestión importante porque el adjetivo es una de las clases gramaticales más significativas. Además, en testimonios de trabajos sobre el lenguaje infantil queda constancia de la escasez y pobreza de que hacen gala los escolares, tanto en el período infantil[179], como adolescente y juvenil[180]. El adjetivo además es síntoma de otros aspectos sociales, como el grado de cultura, según Gili Gaya (1976: 216): "Algo parecido puede observarse en el habla coloquial de los adultos poco instruidos. El uso abundante y preciso de adjetivos está en razón directa del grado de cultura, y constituye (al lado de las conjunciones) un criterio diferenciador muy importante entre los planos sociales de las tablas sincrónicas".

---

encadenamiento cronológico de los hechos que se hallan presentes en el recuerdo. (...)... raras veces he conseguido composiciones organizadas de asunto descriptivo antes de la pubertad" (Gili Gaya, S., *Estudios de lenguaje infantil*, Barcelona, Vox-Biblograf, 1972, p. 171). La cita la recogemos de Sepúlveda, F. (1992: 105-106). Algo parecido opina Iglesias, J. (1971a: 220): "La descripción es un modo expositivo más difícil y delicado que la narración; mientras los niños narran sucesos, aventuras, fantasías y hechos reales desde muy pequeños, la descripción de personas, paisajes, ambientes, situaciones, requiere mayor finura de observación y una riqueza de vocabulario que no podemos exigir en los escolares de los primeros cursos".

[179] Gili Gaya (1976: 215) alude a un trabajo efectuado con niños de cuatro a diez años en el que se refleja el poco uso de adjetivos, a pesar de que, al menos a los diez años, los alumnos están en contacto, por la influencia de la escuela, con la escritura y la lectura. Y afirma: "Calificar supone en el hablante una actitud contemplativa o descriptiva poco común entre niños, cuya expresión es dinámica y salta del substantivo al verbo sin detenerse en las cualidades de las cosas".

[180] Zabal Lázaro, M. (1984: 325-352), tras analizar algunos aspectos del vocabulario de los alumnos de 1º, 2º, 3º de BUP y de COU, llega a varias conclusiones, y una de ellas es precisamente la escasa adjetivación que utilizan.

Hemos de resaltar, finalmente, que, a pesar de nuestra precaución para no mediatizar en lo posible las composiciones y facilitar así la naturalidad del alumno, la espontaneidad absoluta no existe y creemos que cualquier tipo de prueba puede influir en los resultados. Por ejemplo, hemos observado al introducir los datos en el ordenador que muchas composiciones tienen influencia de la realidad histórica del momento, o que al proponer a los alumnos una redacción, casi todos acuden al tema de la "excursión"; en fin, al delimitar el género de expresión (la descripción), es posible que inconscientemente se fomentara la aparición de determinadas palabras, como las concernientes a la vivienda o a cualidades de las personas[181].

No obstante estas observaciones, otra razón por la que hemos optado por este tipo de prueba es contar con un material vivo del alumno para poder utilizarlo en ulteriores trabajos y determinar la competencia sintáctica, ortográfica, etc.

2.10. Introducción de datos en el proceso informático

El material recogido -las 1864 composiciones escritas de los escolares- se introdujo en un ordenador utilizando para ello el programa WP51. En principio respetamos la expresión del alumnado, con lo cual nos limitamos a reproducir textualmente sus escritos. (Después, copiábamos todo en un nuevo fichero en donde aplicábamos la corrección ortográfica). A medida que

---

[181] Los autores del francés fundamental (Gougenheim, Michéa, Rivenc y Sauvageot, 1964:137-145) señalaban ya que la frecuencia de muchas palabras dependía de ciertas circunstancias y del tema del discurso. Y ponen como ejemplo el trabajo sobre el alemán de M. Auber *Guide de fréquence*, 1953, basado sobre todo en textos descriptivos, por lo que palabras como *montaña, sol, árbol, nieve, valle, campanario*..., tienen una alta frecuencia.

realizábamos la introducción de datos, íbamos colocando alguna clave a las palabras que posteriormente presentarían algún tipo de ambigüedad lingüística, como la homonimia sintáctica y morfológica. No obstante, en aquel momento, esa operación no pudo ser ni mucho menos exhaustiva, por lo que posteriormente, cuando realizamos un primer listado de palabras, fuimos desambiguando todos aquellos casos complejos de acuerdo con la norma léxica que consideramos más idónea para nuestro trabajo.

2.11. La norma léxica de nuestra investigación

En todo trabajo cuantitativo sobre el vocabulario es necesaria una norma de despojo. Para ello hay que recurrir a diversas disciplinas como la lexicología, la lexicografía, la gramática... al objeto de cumplir con una tarea esencial: ofrecer un número preciso de V; es decir, se han de aplicar una serie de criterios lingüísticos para agrupar las palabras en vocablos y estructurar éstos de acuerdo con algún tipo de ordenación. La ordenación que se sigue es la que suelen recoger los diccionarios.

Nuestro trabajo adopta la estructura de un diccionario de frecuencias. Éstos suelen presentar de dos maneras los datos: por orden alfabético y por orden decreciente de frecuencias. El orden alfabético como criterio formal[182] de

---

[182] Alvar Ezquerra, M. (1994: 3-13) reconoce que casi todos los diccionarios -alfabéticos, inversos, de rimas, etc.- siguen criterios formales para clasificar sus materiales, aunque algunos no lo hacen. Para superar la distinción tradicional entre diccionarios formales y diccionarios conceptuales, el autor establecerá una amplia clasificación partiendo de la noción de signo lingüístico. Esta manera de proceder se ajusta a la tipología "monista", según Gak, V.G. (1991: 106): " ...tipología monista, que pretende explicar la variedad de diccionarios y de sus caracteres partiendo de un solo criterio".

clasificación es el más usado por todo tipo de diccionarios[183]. Esto se debe, posiblemente, a que quizás sea el más cómodo ya que permite identificar o localizar fácilmente una palabra en la obra, o porque es la ordenación arbitraria a la que más habituados están los miembros de las comunidades culturales que conocen la escritura, y ello incluso a pesar de que el origen de este orden permanece inexplicado, lo cual asegura su carácter convencional[184]. Un diccionario normal de lengua está integrado por dos componentes esenciales denominados macroestructura y microestructura (Haensch, G, 1982b: 452-480). La macroestructura está constituida no sólo por el conjunto de entradas de un diccionario, sino también por la ordenación general, es decir, la introducción, los anexos, los suplementos... La microestructura en cada caso concreto está formada por los artículos, llamados también entradas[185], que son las unidades más pequeñas del diccionario y se componen de lema y el cuerpo del artículo, que incluye la parte definitoria y otros tipos de información[186]. Ambas partes están relacionadas y afectan a los registros de

---

[183] Alvar Ezquerra, M. (1994: 4) afirma que son muchos los autores que prefieren el orden alfabético y son los diccionarios alfabéticos los más frecuentes y conocidos, aunque no los únicos. Estas afirmaciones son fácilmente comprobables si revisamos cualquiera de los muchos diccionarios existentes en el mercado o en las bibliotecas.

[184] Dubois, J. et C (1971: 57) afirman: "La suite des lettres qui constitue l'alphabet est une convention universelle des communautés culturelles qui connaissent l'escriture: son caractère conventionnel est accentué encore par le fait que l'origine de cet ordre reste inexpliqué".

[185] Así lo hace Haensch, G. (1982b: 462)

[186] En términos parecidos hallamos la siguiente consideración de Gak, V. G. (1991: 109): "La microestructura del diccionario se define por la organización del artículo que, además de la información principal, que constituye la razón de ser del tipo dado del diccionario, puede contener varias informaciones complementarias". Y Martínez de Sousa (1995: 275) afirma: "La microestructura afecta al artículo como unidad de estructuración del contenido léxico y a la descripción lingüística, la colocación, disposición y

los diccionarios generales de la lengua. En un diccionario de frecuencias, la macroestructura estaría constituida por las distintas ordenaciones del vocabulario -en este caso, alfabética y por orden decreciente de frecuencias, como se ha indicado-; y la microestructura estaría formada por el lema, las formas flexivas y los datos cualitativos y cuantitativos. Se suelen distinguir dos tipos de trabajos de frecuencias: unos registran simples listas de palabras distintas, entendidas gráficamente, y otros tienen como unidad el lema, el cual representa todas las variantes flexivas de una palabra. Si afinamos un poco, con el primero nos estamos refiriendo a aquellos diccionarios, como el de Rodríguez Bou (1952), que ofrecen listas alfabéticas con todas las variantes flexivas sin agrupar; y dentro del segundo tipo, el de los diccionarios lematizados, se encuentran los que dan cuenta sólo del lema, como los trabajos de Fernando Justicia (1985a y 1995), y aquellos otros que recogen no sólo la unidad que sirve de lema sino también las formas flexivas que la representan, como los trabajo de Juilland y Chang-Rodríguez (1964) y de Amparo Morales (1986): éste será el criterio que sigamos nosotros.

Cualquier confección de un vocabulario determinado -sobre todo si se parte de un corpus compuesto de realizaciones de lengua, de textos, de "datos primarios directos" (Quemada, B., 1990: 174), y no de una selección de palabras recogidas en diccionarios ya existentes, de "datos primarios indirectos" (Quemada, B., 1990: 175)-, supone entrar de lleno en labores lexicográficas[187] y, si además se analizan otros aspectos, en tareas propias de

---

separación de las acepciones, la disposición de los sintagmas y de la fraseología, las subentradas, la separación entre los artículos, etc.".

[187] Entendiendo lexicografía no sólo como "arte o técnica de componer diccionarios", según el DRAE, el diccionario de Casares o el de María Moliner, sino además "la ciencia lingüística que trata de fijar las metodologías aplicables a su elaboración" (Arias, R., 1986: 743). Esta consideración es más actual, como se señalan en otros artículos, como Anglada Arboix, E. (1987: 11), quien señala que "la lexicografía, tal como se viene entendiendo y practicando, es un técnica de confección de repertorios léxicos, y un análisis

la metalexicografía[188], desde el momento en que se tienen que ordenar los vocablos y, por tanto, someterlos a unos criterios de lematización determinados. En la lexicografía confluyen varias disciplinas lingüísticas: la gramática, la lexicología, la semántica... Si dentro de tales disciplinas no hay una homogeneidad absoluta ni en terminología ni en las interpretaciones de un mismo fenómeno, etc., difícilmente se podrá hallar una identidad de criterios en la confección de diccionarios, sean éstos de la índole que sea. Y en lexicometría, en la estadística lingüística, se deben utilizar criterios propios de la lexicografía. Son varios los problemas que atraviesa la labor lexicográfica, y en cada caso particular se han aportado soluciones distintas. Para ello basta con echar una ojeada bien directamente a los distintos diccionarios, bien a los numerosos estudios que analizan un trabajo lexicográfico concreto o plantean postulados generales. Todos los caminos llevan a la conclusión de que no hay uniformidad en la lexicografía ni en la lexicometría[189]. Pero también hay muchos puntos en común que se adoptan

---

lingüístico de esta técnica".

[188] Hausmann, F. J. (1988: 80) cita como labores de la metalexicografía, por ejemplo, el estudio de la historia de la lexicografía, la crítica de diccionarios actuales, la tipología de diccionarios, la recolección y tratamiento de datos en lexicografía... Para Marcos Marín, F. y Sánchez Lobato, J. (1988: 121-122) el metalenguaje incluye tanto la terminología para la presentación de las distintas categorías gramaticales, como el criterio de ordenación de las acepciones y el léxico interno utilizado para su definición.

[189] Alvar Ezquerra, M. (1982: 159-164) habla de que hay distintas maneras de registrar y ordenar las entradas y, tras apuntar numerosos problemas, admite que aunque se dieran orientaciones teóricas, no se adoptarían por todo el mundo; Morales, A. (1986: 22) reconoce que ningún diccionario ha cumplido con éxito el problema de la asignación de las palabras a las distintas clases gramaticales. Cito textualmente un apunte de la autora: " Indudablemente, cualquier decisión que se tome ofrecerá reparos debido a la asistematicidad que aún presenta la teoría en muchos de estos casos, lo cual impide poder llegar a acuerdos generales en cada una de las lenguas." La cita la tomamos de la p. 24. Otros testimonios para otras lenguas nos los proporciona Michéa, R. (1973: 79): "Le statisticien est condamné à se débattre perpétuellment au milieu de critères contradictoires d'ordre formel, grammatical, sémantique ou étymologique. On est

sin problemas y que hacen que la lexicografía -y la lexicometría- tenga su interés y eficacia. Además, los casos polémicos sirven para poner sobre la mesa los problemas más acuciantes, a los que se les intenta dar una solución ahondando en las distintas disciplinas lingüísticas. Posturas atrevidas en un tiempo concreto se convierten en algo superado en otra época. En definitiva, los retos sirven para que las disciplinas avancen.

Los trabajos de frecuencias que analizan el vocabulario escolar, aun concediendo importancia capital a estos asuntos, no se detienen en exceso en informar ampliamente acerca de los criterios lingüísticos seguidos para organizar la macroestructura de su diccionario; sí dedican mayor atención aquellos trabajos que tratan de establecer el léxico fundamental de nuestro idioma. Nos estamos refiriendo a dos obras esenciales, una sobre el español de la Península: el *Frequency Dictionary of Spanish Words*, y otra sobre el español de un país hispanoamericano: el *Léxico básico del español de Puerto de Rico*. Nosotros nos inspiramos en ellos para la presentación del vocabulario y seguimos en gran medida muchos de sus postulados, aunque con alguna que otra diferencia.

Los diccionarios de frecuencias son especiales por algunos motivos. Si bien es cierto que gozan de algunos privilegios, como el hecho de no contemplar definiciones obviando así el cúmulo de problemas que esto pueda comportar, en otros muchos aspectos están en franca desventaja con respecto a los diccionarios generales. Tomar cualquier decisión en lexicografía pasa por uno o varios postulados de la gramática o la lingüística teórica. Por lo que una decisión complicada para un diccionario general se duplica para un diccionario de frecuencias, en el que no es posible registrar varios casos dentro de un mismo artículo. De ahí lo difícil de algunas decisiones. Quizás por ello sea preciso adoptar un criterio claro y ser coherente con él.

---

finalment obligé de reconnaitre qu'aucune norme de dépouillement n'est entièrement satisfaisante". Véase igualmente Lyne, A. (1973: 84).

Recordemos que a la hora de ofrecer las listas de palabras o determinar cualquier vocabulario influyen varios puntos importantes: las propias decisiones y orientaciones de la lexicología y lexicografía; las distintas concepciones gramaticales y lingüísticas sobre un mismo hecho de lengua o de habla; y las opiniones que, basándose en lo anterior, vierten los especialistas o estudiosos al enjuiciar obras concretas, a lo que hay que sumar las decisiones que, de acuerdo con estos puntos, adopta la estadística lingüística. Teniendo en cuenta todo lo precedente acerca del problema general que inunda las listas de frecuencia, trataremos de establecer nuestra norma léxica. Explicitar una norma léxica supone intentar superar una serie de problemas de índole metodológica, conceptual y terminológica, para lo cual se deben especificar con precisión los criterios que se siguen a la hora de consignar los lemas y las ocurrencias. Pero sabemos que, aunque operemos en ese sentido y tratemos de ser coherentes, las decisiones adoptadas no serán satisfactorias para todo el mundo, pues la lingüística dista mucho de ofrecer terminologías homogéneas, coherentes e idénticas[190]. Seguiremos ante todo criterios formales, morfológicos y sintácticos, sin descuidar los léxicos, aunque con predominio de los primeros, quizás más claros o patentes a la hora de esclarecer dudas, considerando que contamos con un contexto lingüístico que ayuda a dirimir casos complejos. Al decantarnos por criterios formales no hacemos sino seguir, como apuntábamos más arriba, las directrices de los trabajos de Juilland y Chang-Rodríquez y de Amparo Morales[191].

---

[190] Véanse las distintas gramáticas que sobre el español existen o artículos donde se tratan directamente problemas de terminología, como Calvillo, M. y Pulgarín, A. (1987: 153-167); Manjón-Cabeza, A. (1987: 230-244).

[191] Nos referimos al *Frequency Dictionary of Spanish Word* y al *Léxico básico del español de Puerto Rico.* La autora puertorriqueña aclara que en su diccionario (Morales, A., 1986: 18) se busca "establecer el

Aunque la finalidad esencial es mostrar el vocabulario usual de una población escolar determinada, queremos que nuestras listas den además información gramatical acerca de los tiempos o personas verbales preferidas por los alumnos o en qué medida emplean las distintas categorías o clases gramaticales[192].

2.11.1. Consideraciones previas: Palabras gramaticales y palabras léxicas

La primera consideración lingüística que se ha de tener en cuenta es la del tipo de palabras que deberán registrarse; es decir, si habrán de ser sólo las propiamente léxicas o también las denominadas ampliamente como gramaticales, dentro de las que se incluirían las relacionantes[193].

---

debido equilibrio entre las distinciones gramaticales y las semánticas, por lo menos, hasta donde ello fuera posible".

[192] Cierto es que la forma de expresión sugerida en la recogida de datos (Cfr. 2.9.) no es gratuita para influir en la aparición de unas formas u otras, sobre todo en el verbo, pronombre... Pero si bien la descripción apunta sobre todo al presente y al imperfecto, y al adjetivo; la narración, aunque apunta al perfecto simple, puede dar entrada al diálogo, con lo que la variedad verbal se enriquece y las composiciones se llenan de matices.

[193] Aludimos así a la dicotomía léxico/gramática y sus repercusiones en la lexicografía. Recordemos que Martinet, A. (1971: 148-151) distingue entre monemas gramaticales y monemas lexicales, o con otra terminología (Alvar Ezquerra, M., 1982: 155-156), morfemas gramaticales o gramemas y morfemas léxicos o lexemas, y que Coseriu, E. (1987: 133) matiza con aguda sutileza: "... hay que distinguir tres clases de palabras: 1) *palabras lexemáticas*, que estructuran la realidad extralingüística, como, por ejemplo, *hombre*, *bosque*, *blanco*, *correr*, etc.; 2) *palabras categoremáticas* (<<pronombres>>) que presentan sólo la forma de estructuración de lo extralingüístico (que funcionan, por tanto, como sustantivos, adjetivos, etc.), pero que no representan ninguna materia extralingüística determinada, como, por ejemplo, *yo*, *éste*, *aquí*, *ahora*; 3) *palabras morfemáticas* (o <<instrumentales>>) que no funcionan de

Las posturas sobre la inclusión o no de elementos gramaticales en los diccionarios no son unánimes. En algunas obras de cierta proximidad temporal se suelen descartar las palabras de mayor uso, aunque tal decisión se adopte por razones no puramente lingüísticas, sino precisamente por su elevada frecuencia[194]. No obstante, la tradición lexicográfica, no sólo la española, siempre ha apostado por la inscripción de tales unidades[195]. De hecho, la gramática siempre está presente tanto en la macroestructura como en la microestructura[196], porque, aparte de razones didácticas e imperativos impuestos por usuarios y editores, la lexicografía no suele operar con monemas o morfemas, sino con palabras[197].

---

modo inmediato como configuradoras del <<mundo>>, sino sólo, en relación con otras palabras, en la estructuración del hablar, como, por ejemplo, *y*, *o*, *sobre*, *en*, *sí*, *no*, etc. Sólo las palabras lexemáticas pertenecen con pleno derecho al léxico y, en consecuencia, al objeto propio de la lexicología". A esta clasificación habríamos de añadir además un cuarto apartado: el de las palabras diacríticas, que no tienen significado léxico, ni categorial ni relacional, como *lirondo*, *vilo*, *bruces*, *bemoles*, *bledo*, *troche*, *moche*, *patatín*, *patatán* (Zuluaga, A., 1980: 102-103).

[194] Alvar Ezquerra, M. (1982: 156) cita dos obras: 1ª. E. Sarmiento: *Concordancias de las obras poéticas en castellano de Garcilaso de la Vega*, Madrid, 1970, donde se eliminan las palabras más frecuentes, muchas de ellas son relacionantes; y 2ª. "Le Dictionary of the Old Spanish Language (*DOSL*)", alude al trabajo sobre el español medieval desarrollado en EE.UU., donde provisionalmente se ha adoptado el criterio anterior.

[195] Así queda constatado en el amplio artículo de Alvar Ezquerra, M. (1982); directamente lo afirman Dubois, J. et C. (1971: 33), cuando aseguran que el acopio léxico del francés contiene palabras gramaticales y léxicas, y que además a algunas palabras funcionales, como la preposición *à*, se dedican un número importante de columnas en la microestructura.

[196] Alvar Ezquerra, M. (1982) analiza en profundidad este aspecto y deja clara constancia de esta presencia gramatical en el diccionario.

[197] Véase Werner, R. (1982b: 194). Alvar Ezquerra, M. (1982: 155-156) aduce que si se registran en los

Los diccionarios y listas de frecuencias sobre el español, quizás haciéndose eco de la problemática en diferentes épocas, han dado respuestas distintas a lo largo de su historia. Así, por ejemplo, en los primeros trabajos al respecto que se conocen no suelen registrar en su inventario las palabras de mayor frecuencia -aunque las ofrecen en apéndices aparte-, muchas de las cuales son elementos de relación[198]. Pero los demás recuentos del español que hemos citado a lo largo de la investigación no hacen distinciones e incluyen tales palabras en sus repertorios. Además, añaden información gramatical al registrarlas, pues suelen ir acompañadas de una abreviatura[199] o de un código numérico[200] especificando la clase gramatical a la que pertenecen. Incluso algunos diccionarios han sido sometidos a análisis gramaticales[201], los cuales a su vez han sido cuestionados[202].

---

diccionarios entradas de afijos, con menos razón se podrán excluir los elementos de relación.

[198] Buchanan, M., (1927: 11-17) ofrece una lista aparte de palabras eliminadas -no sólo conjunciones y preposiciones, sino numerales, demostrativos, posesivos, relativos, artículos, pronombres personales y algunos sustantivos, adjetivos, verbos y adverbios- de su recuento. Aparecen sin datos numéricos, porque el propio autor aclara que presentan una frecuencia muy alta, y por tanto estarían entre las 500 palabras más usuales. Rodríguez Bou, I. (1952: 645 y 654) expone una lista de 105 palabras más frecuentes, sin el dato numérico de la frecuencia, (entre las que se hallan *a*, *si*, *por*, *de*...), ya que llevaban contabilizadas un millón de ocurrencias, y éstas palabras suponían el 50% de los tokens, por lo que decidieron actuar así, ya que el recuento lo efectuaban "a mano".

[199] Juilland, A. y Chang-Rodríguez, E. (1964) y Morales, A. (1986) acompañan las entradas con abreviaturas como *aj.*, *n.*, *v.*, *c.*, etc.

[200] Justicia, F. (1985a y 1995) asigna un número del 1 al 9 a las distintas clases gramaticales.

[201] Es lo que hace Huynh-Armanet, V. (1977) con el diccionario de Juilland y Chang-Rodríguez (1964).

[202] Morales, A. (1980: 467-474) cuestiona el trabajo de Huynh-Armanet, V. (1977) porque no existen

Se puede comprender que tales palabras gramaticales, muchas de ellas funcionales, se hayan excluido de las listas de frecuencias de cualquier idioma por los numerosos problemas que presentan a la hora de inventariarlas, o se hayan tomado con ellas medidas particulares[203]. Para Müller (1963: 168-169), aunque esta decisión elimina problemas, genera otros, pues partiendo de la distinción entre palabras de significación y palabras de relación, hay casos delicados, como ciertos adverbios, o incluso los verbos auxiliares, que podrían ser tanto contados como eliminados, además de que suprimir estas palabras enmascara hechos lingüísticos y estilísticos interesantes, por lo que, afirma, hay que registrarlas, pero merecen un tratamiento distinto a las otras unidades léxicas.

Nosotros seguiremos lo comúnmente aceptado por la tradición lexicográfica y por la mayoría de los diccionarios de frecuencias e introduciremos tanto las palabras relacionantes como el resto de palabras gramaticales en nuestro vocabulario. Con ello, pues, no hacemos sino adaptarnos a un hábito lexicográfico arraigado y que parece a todas luces ventajoso.

---

criterios convincentes ni claros que sustenten aún los recuentos de vocabulario.

[203] Por ejemplo, los autores del Trésor de la langue française (Imbs, P., 1971: XIV, Vol. I), que han seguido el criterio de recoger bajo el lema o "vedette" las variantes flexivas de las palabras de contenido léxico, no actúan así con las palabras gramaticales: "Les variantes fléchies des mots grammaticaux (lequel, laquelle, lesquels, lesquelles...) ne sont pas regroupées; il n'est pas sans intérêt de connaître les fréquences d'emploi de chacune d'elles, et l'on évite de cette manière le problème du choix de la vedette: por une catégorie comme les pronoms personnels, par exemple, les possibilités seraient diverses mais également discutables (*moi* sous *me*? ou sous *je*? *nous* sous *je*?...)".

2.11.2. Problemas de N y de V

Todo recuento de palabras supone previamente, como indicábamos más arriba, el establecimiento de unos criterios lingüísticos que constituyen la norma léxica. Ésta se relaciona con la problemática que presentan tanto los diccionarios generales de la lengua como los diccionarios de frecuencias. Para éstos, que manejan datos numéricos y están sometidos a análisis matemático-estadísticos continuos, se pide que haya una uniformidad en la norma, sobre todo si se someten a contrastes y comparaciones. Por ello, y por otros motivos, son varios los autores (Müller, Ch., 1963; Lyne, A., 1973) que han ensayado o propuesto los principales escollos por los que atraviesan estas listas, con el propósito de conseguir una cierta homogeneidad. Ciertamente a la hora de aportar los datos cuantitativos de un vocabulario, nos encontramos, en primer lugar, con un problema esencial, que consiste en adoptar una unidad lingüística como unidad de tratamiento lexicográfico o lexicométrico. Este proceso forma parte de la norma léxica -que debe ser lo más coherente posible[204]- de un recuento determinado, para establecer correctamente los tokens y los types. En lexicografía, la diatriba entre monema y palabra siempre se ha dirimido en favor de ésta, por una serie de ventajas. Pero ello acarrea también una serie de inconvenientes. En lo que atañe a los datos cuantitativos de un diccionario de frecuencias, el problema de elegir la palabra como unidad de tratamiento léxico comporta, en primer lugar, hacer una distinción clara entre palabra como unidad de texto (Müller, Ch., 1963: 155-173), que es lo mismo, como decíamos en otro lugar, que ocurrencia o token -y a veces se utiliza el propio término palabra en este sentido-, es decir, alude a la cantidad de N o conjunto total de palabras, repetidas o no, que tiene un corpus; y

---

[204] Michéa, R. (1973: 76) define "la norme lexicologique" de la siguiente manera: "Entendons por là un ensemble, aussi logique et aussi cohérent que possible, de conventions destinées à definir ce que l'on se propose de compter".

palabra como unidad de léxico (Müller, Ch., 1963: 155-173) que es lo mismo, recordamos, que vocablo, lema o type, es decir, palabras distintas y no repetidas, y que se suele representar como V[205]. Según que se adopten unas determinadas convenciones u otras, tales decisiones repercutirán de varias maneras en los datos tanto cuantitativos como cualitativos; es decir, la admisión de ciertos criterios, bien léxicos (considerar o no la homonimia), bien gramaticales (distinguir, por ejemplo, entre una misma forma según su función sustantiva, adjetiva o adverbial...), afectará a los datos esenciales de los diccionarios de frecuencias:

1. En el número total de ocurrencias o N.
2. En el número total de vocablos (V) que ofrecen los distintos diccionarios.
3. En el cuadro de distribución de frecuencias, o vocablos con frecuencia 1, 2, 3, etc.
4. En el número total de ocurrencias y vocablos para las distintas clases gramaticales[206].

2.11.3. Palabra como unidad de texto: palabra, ocurrencia, token

Con respecto a la palabra como unidad de texto, cuya consideración repercutirá directamente sobre el total de N de un recuento, podemos distinguir, siguiendo a Müller (1963), los siguientes casos: primero, aquellos

---

[205] Queda, pues, claro, que las variedades flexivas -por ejemplo, *tenía*, *tenemos*, *tuvisteis*...; *hermosa*, *hermosas*...- no constituyen type sino token.

[206] Según la consideración gramatical de tal o cual unidad, así se verá afectada su frecuencia. Por ejemplo, el caso de las construcciones *el que*, *la que*..., las formas *el*, *la*..., según se consideren artículos o pronombres, así incrementarán el valor numérico de una u otra clase gramatical.

en los que una forma gráfica representa varias palabras y segundo, aquellos otros en los que varias unidades gráficas representan una palabra. Para el español, en el primer supuesto, no encontramos excesivos casos. Tan sólo los artículos contractos *al*, *del*, que no representan ningún problema de desambiguación, y las formas verbales que llevan varios pronombres enclíticos. Aquí los problemas, en rigor, estriban en contabilizar como una sola palabra o como más de una las formas de los verbos reflexivos. Los diccionarios de frecuencia del español han dado distintas respuestas a este problema. Keniston incluye "los verbos reflexivos, los regímenes preposicionales y ciertos usos de clíticos. Así, cada entrada aparece con sus <<idioms y derivados>>: por ejemplo, *agradecer*, incluye entre ellos: *me lo agradece*" (Morales, A., 1986: 20). Juilland y Chang-Rodríguez (1964) y A. Morales (1986) separan los pronombres átonos, que pasan a engrosar sus categorías correspondientes. La única diferencia entre ellos es que mientras que Juilland hace una notación para los verbos con reflexivos, A. Morales no la hace.

Nosotros, en este caso, separamos también los clíticos y ofrecemos además una marca de reflexividad en aquellas formas verbales que la presenten, pues esta actuación está más acorde con la práctica que seguiremos de distinción entre los distintos tipos de reflexivos (medial, propiamente dicho, e incluso recíproco).

El segundo supuesto, más complicado, se presenta cuando varias unidades gráficas equivalen a una sola palabra (entendida como unidad de sentido), bien por razones morfosintácticas o por motivos léxicos. Este punto constituye uno de los mayores problemas en lexicografía y en lexicometría, no sólo por la abundancia de casos sino por lo difícil de su tratamiento. En español se pueden encontrar desde las formas compuestas de los tiempos, las

perífrasis verbales, verbos reflexivos o con uso obligatorio de preposición[207], pasando por las locuciones preposicionales, conjuntivas, adverbiales..., hasta las palabras compuestas con o sin guión, las colocaciones[208], modismos[209], refranes... Toda esta amplia enumeración se relaciona y afecta simultáneamente al segundo problema en lexicografía y lexicometría, es decir, la palabra como unidad de léxico, o lo que es lo mismo, type[210].

Aunque tales unidades, generadas en el habla y en una perspectiva diacrónica, posean el estatus de palabra, en cuanto a "elemento virtual anterior a la actividad lingüística que lo realiza..." (Zuluaga, A., 1980: 23), no

---

[207] Morales, A. (1986: 21) cita ejemplos polémicos como *tratar de*, *insistir en*.

[208] Haensch, G. (1982b: 415) aboga por la inclusión en los diccionarios generales colocaciones como *un éxito clamoroso*; *un soltero empedernido*; *aburrirse como una ostra*...

[209] Alonso, M. (1982: 232-233, Vol. I) remonta el origen del término "modismo" al siglo XVIII, término semánticamente muy cercano a "idiotismo", usado por Covarrubias para definir " ciertas frases y modos de hablar particulares a la lengua de cada nación". El propio Ortega y Gasset organizó un seminario sobre el modismo en su Instituto de Humanidades. Alonso, M. distingue entre locución y modismo: "Entendemos por modismo una locución popularizada, peculiar de la lengua respectiva, intraducible, inalterable y con cierto sentido metafórico. Ejemplos: *A ojos vista. Acostarse con las gallinas. Clarearse de hambre. Partirse de risa. Charlar por los codos*". Otros autores, como Domínguez, P., Morera, M. y Ortega, G. (1995: 7), obvian la definición de modismo: "Algo que sí nos gustaría reseñar es que, dada la índole eminentemente práctica de este trabajo, hemos orillado en él todas las cuestiones que pudiera suscitar la noción de modismo, aunque admitimos que tal concepto, y en consecuencia el material que se pueda seleccionar a partir de su aplicación, no siempre presenta unos perfiles inequívocos respecto de sus vecinos (la frase hecha, la locución, el refrán, etc.).

[210] Creemos que el criterio que se elija para contabilizar el número total de ocurrencias de un vocabulario deberá ser idéntico al proceso de lematización. No sería coherente, suponemos, dar una cifra para los tokens, considerando la unidad gráfica, y registrar varios lemas con más de una unidad. De ahí la interrelación.

pueden equipararse sin más a palabras, como unidades simples. Todos estos problemas presuponen que los rasgos que suelen manejarse para la delimitación de palabra no sean satisfactorios en la práctica para determinadas unidades[211], por lo que se propone "trascender a unidad superior: la lexía"[212].

Por razones gramaticales, tenemos en español los tiempos compuestos de la activa, lo cual comporta unir en una sola unidad dos componentes verbales que si bien es verdad que se sienten plenamente como forma única, pueden aparecer alejados en el discurso , por la intercalación de elementos, como *ella ha, de alguna manera, insultado a...*, "con lo cual se viola uno de los requisitos que exigen las formas para su consideración de independientes" (Morales, A., 1986: 21), o casos que suponen incertidumbres como *ha llegado pronto y comido...*, en los que se elide el verbo auxiliar. Las perífrasis presentan la dificultad de que su delimitación y casuística sea completa, rigurosa, clara y unánime. Los lingüistas suelen coincidir en la definición o en los rasgos esenciales que permiten interpretar una estructura verbal como tal perífrasis[213], en el sentido de secuencia formada por un verbo en forma

---

[211] Urrutia Cárdenas (1978: 195) alude a tres rasgos que suelen ser los más utilizados para referirse a la palabra como unidad: Su carácter reemplazable y separable en el sintagma, su límite gráfico y su designación de un determinado significado.

[212] Urrutia Cárdenas (1978: 195-196) distingue tres clases de lexías: 1, simple: *casa, flor, candil*; 2, compuesta: *casaquinta, caballo de vapor, bronquitis, hiperfunción*; 3, textual: *A quien madruga, Dios le ayuda*. Por tanto, considera lexías las palabras definidas tradicionalmente como simples, derivadas, parasintéticas, y las locuciones y refranes.

[213] Véase por ejemplo: RAE., *Esbozo de una nueva gramática de la lengua española*, Madrid, Espasa-Calpe, 1983, 9ª reimpresión, pp. 444-453; Gili Gaya, S., *Curso superior de sintaxis española*, Barcelona, Vox-Biblograf, 1976, 11ª ed., pp. 103-119; Sánchez Márquez, M. J., *Gramática moderna del español. Teoría y norma*, Buenos Aires, Ediar, 1972, pp. 386-392; Marcos Marín, F., *Curso de gramática española*,

personal, que suele perder su significación esencial y se gramaticaliza, actuando en consecuencia como auxiliar, y otro en forma no personal (entre ambos, cuando la perífrasis es con infinitivo, a veces, pueden aparecer nexos) que aporta el significado esencial y selecciona el sujeto y los complementos, y que toda la perífrasis posee unidad funcional y significativa. Normalmente se utiliza el criterio semántico, o la pérdida de significación del auxiliar, para identificar una secuencia como perífrasis de otra que no lo es, pero también se citan otros procedimientos de índole funcional o gramatical, como la exigencia del infinitivo por parte de determinados verbos, la inadmisión de la transformación pasiva e interrogativa con *¿qué?* o con *¿cómo?*, o la anteposición o posposición de los clíticos. Aunque estos recursos son de cierta consistencia, no se cumplen al cien por cien en todos los casos. Además, como las perífrasis son fruto de una evolución, de un fenómeno diacrónico, se señalan ciertos grados de lexicalización, desde casos fronterizos o en vía de lexicalización[214], pasando por un estrato intermedio entre la perífrasis y la lexicalización total[215], hasta la lexicalización total, en donde estarían aquellos casos que no podrían considerarse como tales perífrasis, sino como locuciones verbales[216]. Pero incluso reconociendo con

---

Madrid, Cincel-Kapelusz, 1980, pp. 273-288; Fente Gómez, R., Fernández Álvarez J. y Feijóo, L. G., *Perífrasis verbales*, Madrid, Edi-6, 1983, pp. 11-13; Marsá, F., *Diccionario normativo y guía práctica de la lengua española*, Barcelona, Ariel, 1986, pp. 203-231; Lamiquiz, V., *Lengua Española. Métodos y estructuras lingüísticas*, Barcelona, Ariel, 1989, 2ª ed., pp. 176-179; Alcina y Blecua, *Gramática española*, Barcelona, Ariel, 1991, 8ª ed., pp. 777-783; Alarcos, E., *Gramática de la lengua española*, Madrid: Espasa Calpe, 1994, 3ª reimpresión, pp. 259-265; Gómez Torrego, L., *Perífrasis verbales*, Madrid, Arco/Libros, 1988, y *Manual de español correcto*, Madrid, Arco/Libros, 1994, 5ª ed., 2º Vol., pp. 188-195.

[214] Gómez Torrego, L. (1988: 53-63) se refiere a casos como *querer*, *pensar*, *dejar*, *osar*, *saber*+ infinitivo.

[215] Gómez Torrego, L. (1988: 26-27) nos habla de expresiones que casi pueden considerarse ya como locuciones verbales, como *dar que hablar*, *dar que pensar*, *dar de mamar*, *dar de comer*, *dar de beber*.

[216] Gómez Torrego, L. (1988: 23-25) cita casos como *echar a perder*, *dar a conocer*, *dar a entender*.

más o menos exactitud todas las características mencionadas, las posturas se diversifican a la hora de señalar como perífrasis construcciones concretas. Por ejemplo, la Academia (1983: 450) y Gili Gaya (1976: 118-119) no las reconocen cuando el primer componente es un verbo "modal", como *poder*, *soler*, *deber*, criterio que no comparten Alarcos (1994: 260), Gómez Torrego (1988: 34), Marcos Marín(1980: 273); o una perífrasis comúnmente aceptada, como *estar+gerundio*, no lo es para Alarcos (1994: 264), etc. Se podrían enumerar otros casos, porque como hemos afirmado, difícilmente existe coincidencia absoluta en las listas que ofrecen los lingüistas en sus estudios, pero la proximidad es igualmente estrecha[217]. Mención especial merecen las perífrasis con participio, porque dentro de ellas habría que registrar la llamada voz pasiva o conjugación pasiva[218] con *ser*[219] y con *estar*[220], a pesar de que no hay consenso en su consideración o descripción.

Además, aporta una serie de recursos lingüísticos rigurosos para distinguir entre perífrasis y locución. Tales recursos son la segmentación sintáctica, la conmutación léxica o estructural, etc.

[217] Puede comprobarse el aserto consultando la bibliografía dada en este apartado para las perífrasis.

[218] Alarcos, E. (1974: 98-100; 1982: 90-94) niega la voz pasiva. Lamiquiz, V. (1989: 172) adopta una postura ecléctica, ya que reconoce como Alarcos que no se puede hablar de voz, en el sentido de que no hay ningún morfema que la exprese, pero sí se puede hablar de construcción pasiva.

[219] Hablan de la pasiva con *ser* o perifrástica, la RAE (1983: 378-379 y 450-452; Gili Gaya (1976: 121-129); Marsá, F. (1986: 221-229); Marcos Marín (1980: 280-288); Alcina y Blecua (1991: 782-783).

[220] La RAE (1983: 452-453) reconoce además la pasiva con el verbo *estar*; también Gili Gaya (1976: 124-125) y Hadlich, R. L. (1973: 305-309). Marsá, F. (1986: 226) se refiere a la construcción con *estar* más participio como voz "estativa".

Las expresiones fijas o unidades fraseológicas[221] constituyen un problema mucho mayor, debido sobre todo a su abundancia (por lo que se debería tener una gran memoria lingüística para no omitir ninguna cuando se procede a las labores de despojo de los textos), a su continua y progresiva lexicalización[222] y a la dificultad, por tanto, de identificar de un modo exhaustivo todos los casos[223]. Entre las locuciones se suelen distinguir aquellas que tienen función meramente gramatical, como las prepositivas y conjuntivas, y las que presentan valor semántico, que serían las nominales, verbales y adverbiales[224]. En general, el rasgo esencial de la fraseología es para Zuluaga (1980: 15) "la fijación", pues "no son expresiones formadas libremente en el hablar sino repetidas como estructuras prefabricadas, combinaciones <<ya hechas>>...", que presentan además otras notas distintivas[225] como la inalterabilidad del

---

[221] Zuluaga, A. (1980: 15) afirma: "Las llamamos <<expresiones fijas>> porque su rasgo constitutivo es la fijación, o, también <<unidades fraseológicas>> porque funcionan como unidades en diferentes niveles gramaticales y porque, con muy pocas excepciones perfectamente identificables, están formadas por combinaciones de palabras".

[222] Valga como testimonio Zuluaga, A. (1980: 105-109), quien pone varios ejemplos de locuciones que presentan un grado de lexicalización no absoluto, bien porque se intercalan otras palabras entre sus elementos, como *todo queda en familia = todo queda, como quien dice, en familia*; bien porque se altera el orden de sus componentes: *no dejar titere con cabeza = ni titere que dejen con cabeza*; bien por la modificación o sustitución de algunos de sus componentes: *echar una cana al aire = echar una canita al aire, a mis anchas, a tus anchas, a sus anchas, pasar una noche en vela, pasar dos noches en vela...*

[223] Martínez Marín, J. (1991: 121), en relación al tratamiento en los diccionarios, afirma que "un tratamiento objetivo y sistemático de la fraseología sigue siendo prácticamente imposible todavía hoy por no haberse realizado los estudios descriptivos previos".

[224] Alonso, M. (1982: 232, Vol. I) distingue también las locuciones "interjectivas".

[225] Ettinger, S. (198a: 233-258) recoge un resumen de las características que aplican a estas unidades

orden de sus componentes: *corriente y moliente*, y no **moliente y corriente*; o la invariabilidad de alguna categoría gramatical (género, número, tiempo, etc.): *pagar los platos rotos*, y no **pagar el plato roto*, etc. Normalmente, y para los diccionarios generales de la lengua, los lingüistas están de acuerdo en que se recojan tales unidades, a pesar de los inconvenientes que pueden suponer[226]. La lexicografía española suele actuar de esta manera, siguiendo con una larga tradición lexicográfica[227], aunque no se esgriman criterios uniformes para inscribirlas[228]. La tendencia actual, no obstante, ha optado por acoplar las expresiones fijas como subentrada al final del artículo[229]. Para los

---

lingüísticas varios autores como W. L. Chafe, J. Häusermann, Zuluaga...

[226] Así opinan, por ejemplo, Werner, R. (1982b: 230); Haensch, G. (1982b: 416). Ettinger, S. (1982a: 258) considera que se deberían tener en cuenta, mucho más que en el pasado, las unidades fraseológicas, lo cual contribuiría a mejorar la calidad de los diccionarios, "aunque no exista una clasificación satisfactoria de ellas".

[227] Martínez Marín, J. (1991: 117-118) cita trabajos de fraseología de nuestra lengua en los siglos XVII y XVIII, y destaca la importancia que adquieren estas expresiones para la lexicografía actual.

[228] Alvar Ezquerra, M. (1982: 161) pone varios ejemplos de "lexías complejas", -como las llama- recogidos del diccionario de la Academia: mientras *hurtadillas* (*a*), *rondón* (*de*), se recogen en la macroestructura, directamente como entrada, *a golpes*, *a pies juntillas*, se registran en la microestructura, dentro de golpe y de pie respectivamente. Martínez Marín, J. (1991: 120) alude a los muchos "puntos débiles" en el tratamiento de estas unidades en los diccionarios, y pone como ejemplo *de bote en bote*, recogida, a juicio del autor, con buen criterio como subentrada por el Diccionario actual de la lengua española, frente a "prácticas superadas" de lematizar, como entrada, *bote en bote* (*de*) en el Diccionario moderno del español usual.

[229] Valga como testimonio Zgusta, L. (1971: 269), quien aconseja que se registre la frase dentro de cada palabra de contenido léxico que integre la locución. Haensch, G. (1982b: 507-508) ofrece tres posibles criterios para registrarlas bajo los siguientes lemas: 1. O el orden de aparición. 2. O el orden alfabético. 3. O el orden de categorías: sustantivo, verbo, adjetivo, etc. No se atribuyen o no se recogen bajo palabras funcionales. Alvar Ezquerra, M. (1982: 164) afirma que las locuciones, por varios motivos, se suelen

diccionarios de frecuencias muchas opiniones consideran importante que las unidades fraseológicas se registren, aunque presenten distintos grados de lexicalización, para poder constatar así el movimiento del lenguaje[230]. No olvidemos que la frecuencia de uso y la disponibilidad en distintos contextos son características de estas secuencias, convertidas así en unidades léxicas "duraderas" frente a otras estructuras, formalmente similares, pero que bien por ser creaciones coyunturales o bien por carecer de una frecuencia de uso y de una disponibilidad importantes, están determinadas para ser creaciones efímeras[231]. En este último caso, no serían por tanto auténticas frases hechas "que deben repetirse en su literalidad, sino que se trata de una vida oral en variantes, una parte de la cual vida llega a la escritura" (Abad, F., 1994: 447). Por lo que puede ser un riesgo registrar determinadas unidades que estén en vías de lexicalización, sobre todo si tenemos en cuenta que no es posible tener dotes proféticas para determinados fenómenos del idioma, según asevera Humberto Hernández (1989: 45).

Los refranes y las frases proverbiales son otro caso especial que entrarían dentro de lo que Urrutia Cárdenas (1978: 204) llama "lexías textuales", y la distinción entre frase proverbial y refrán pasa por considerar la primera como una especie de cita que recuerda algo que se dijo en una situación histórica

---

recoger como subentrada o en apéndices.

[230] Por ejemplo, consúltese Ortega, J. (1985: 437).

[231] Urrutia Cárdenas (1978: 209-210) indica un coeficiente que alude al grado de lexicalización: por ejemplo, *caballo de vapor* presenta un coeficiente del 100 por 100, frente a otras expresiones como *el caballo de carrera*, que posee un coeficiente mucho menor, o *el caballo de Juan*, que es una creación coyuntural. Al mismo tiempo, se refiere a expresiones como *cosacas negras*, *política del ping-pong*, que "si es que no han logrado una frecuencia y una disponibilidad independientes a una determinada situación o contexto, están expuestas a vivir lo que dura una moda".

determinada y que refuerza el mensaje. La frase proverbial tiene carácter autónomo, por lo que no se considera *en menos que canta un gallo* y sí *otro gallo me cantara*[232]. El concepto de refrán también presenta problemas, pues mientras algunos, siguiendo a Casares, lo definen como "el refrán es un texto con valor de cláusula independiente, en al que se relacionan por lo menos dos ideas, que expresan una verdad con carácter de enunciado moral y con valor atemporal" (Urrutia Cárdenas, H., 1978: 205), otros consideran la dificultad de delimitar "las zonas de los refranes y de las frases proverbiales, debido a que estas, en muchos casos, son restos que han sobrevivido a la desaparición de antiguos proverbios, y también porque a veces el refrán presenta la estructura de una oración simple cuando solo se expone uno de los términos del juicio, quedando el otro sobrentendido" (Campos, J. y Barela, A., 1975: VIII). No tenemos noticias de que los diccionarios de frecuencias recojan refranes o aludan a su inclusión o no. Las pocas referencias al respecto de otros autores apuntan la tendencia a eliminar los refranes de los diccionarios generales, y si se mantienen algunos, es por influencia del DRAE, donde a su vez también se van eliminando en las ediciones más noveles, pues parece que su intención es recogerlos fuera del diccionario (Alvar Ezquerra, M., 1982: 62).

Por tanto y teniendo en cuenta que no se podría ser exhaustivo en el cómputo de todos estos casos -nos referimos a todos aquellos que suponen más de una unidad gráfica- unas veces porque no se consigue unanimidad en su consideración, otras porque son difíciles de distinguir debido a la intercalación de elementos incómodos entre sus componentes; bien porque se presenta el problema adicional de que habría que separar las locuciones de los compuestos, secuencias entre las que existen rasgos comunes[233], bien

---

[232] Urrutia Cárdenas (1978: 205) cita el ejemplo *Si yo fuera bachiller, otro gallo me cantara*, considerada por él frase proverbial, frente a la lexía compuesta, porque en este caso funciona como oración principal.

[233] Zuluaga, A. (1980: 142-144) cita algunos compuestos como *falda pantalón, ciudad jardín, ave del paraíso, cabello de ángel, lengua de gato, hombre rana, perro lobo, coche cama, café teatro, sofá cama,*

porque su grado de lexicalización es imperfecto, bien porque carecemos de un estudio descriptivo categórico..., cabrían varias posturas: registrarlas (lo cual añadiría el problema adicional de ¿en dónde?: ¿cómo subentrada de un elemento o de todos los elementos, o en un apéndice aparte?), utilizando para su delimitación muchos de los recursos lingüísticos que se han enumerado, aun sabiendo que hay casos limítrofes, o adoptar, precisamente por la falta de exhaustividad, la decisión de no contemplarlas, al menos en una primera instancia.

Ante la perspectiva de no actuar con rigurosidad[234], conviene, pues, tomar la determinación de no registrarlas, desglosando los compuestos en unidades gráficas[235], porque "las distinciones que por la propia naturaleza de los fenómenos no permiten el apoyo de metodología rigurosa, no hacen sino entorpecer la labor de los análisis posteriores" (Morales, A., 1986: 22). De este modo, se toma la palabra (entendida como unidad gráfica) como token, por lo que los compuestos no influyen en el número total de ocurrencias que ofrecemos. No obstante, nosotros ofrecemos aparte, aun con el riesgo de no ser exhaustivo, una relación de estas unidades. Tal decisión, adoptada por interés didáctico de ofrecer las preferencias de los alumnos, no repercute en las operaciones cuantitativas de un diccionario de frecuencias (Cfr. 3.4.4.).

---

máquina de escribir, capillo de dientes, sala de espera...

[234] Morales, A. (1986: 21) censura que en el *FDSW* no se actúe con rigor, cuando se registran las locuciones *en seguida*, *no obstante*, y, en cambio, no se encuentran *sin embargo*, *sin duda*; pero *embargo*, *duda*, sí tienen sus entradas respectivas.

[235] Es la solución que se adopta en el Léxico básico del español de Puerto Rico, y en definitiva, es la propuesta que hace Müller, Ch. (1963: 161).

2.11.4. La palabra como unidad de léxico: Lema, vocablo, type

Con este apartado nos referimos a la tarea de encontrar la unidad léxica que debe tomarse como entrada en un diccionario y que recibe el nombre de lema[236]. La disyuntiva que a menudo se ha planteado entre "monema" (o más frecuentemente "morfema"[237]), y "palabra", casi siempre se ha dirimido a favor de la segunda, a pesar de que el monema presenta menos problemas de identificación[238]. Sobre el concepto de palabra existe una abundante

---

[236] El lema, por tanto, es la palabra que sirve como entrada de un diccionario y que reúne todas sus variantes flexivas; es decir, como afirma Alvar Ezquerra, M. (1982: 166): "La reducción del paradigma se hace a una forma canónica, el lema, representante de todas las variantes de la palabra". Así lo confirma Haensch, G. (1982b: 463): "Desde el punto de vista formal, el lema es [...] el representante de una serie de formas que se pueden obtener a partir de él". En este sentido, es comprensible la idea de Dubois, J. et C. (1971: 62) de que las entradas de un diccionario son unidades teóricas las cuales a su vez son "abstracciones" de palabras en el discurso, en donde se actualizan todas sus variables flexivas, por lo que se puede considerar el lema como una forma arbitraria y teórica despojada de marcas gramaticales. El Trésor de la langue française prefiere denominar vedettes a estas palabras que sirven de entrada en los diccionarios: "On appellera "vedette" la forme de regroupement des variantes fléchies d'un même vocable..." (Imbs, P., 1971, p. XIII, Vol. I).

[237] Hörmann, H. (1973: 62) afirma: "A las unidades obtenidas de esta manera se les llama, con Martinet [...], unidades de la primera articulación o monemas; más frecuente el término morfemas". Es el término que utilizan, por ejemplo, Morales, A. (1986: 16); Alvar Ezquerra, M. (1982: 155); Lyons, J. (1975: 175-200); Collado, J. M. (1974: 175-226).

[238] Como este tema es consabido, valgan como referencias las siguientes: Alvar Ezquerra, M. (1982: 155); Werner, R. (1982b: 219) alude al mismo caso y recuerda que todas las lenguas europeas toman como unidad la "palabra"; Morales, A. (1986: 16) también nos ofrece su mirada sobre la disyuntiva y afirma que los autores ponderan una unidad u otra según sus preferencias; asimismo informa de que el estructuralismo lingüístico impulsó la consideración lexicográfica del morfema, pero casi nadie siguió en la práctica tal consideración.

bibliografía, y muchas veces no coinciden todas las caracterizaciones[239]. A pesar de todo, la consideración teórica ha influido poco en la tradición lexicográfica, ya que, como afirma Hausman, "para el lexicógrafo la palabra es ante todo una unidad gráfica en el marco de la convención ortográfica"[240]. Desde este punto de vista, la palabra se ha definido como aquella unidad gráfica entre dos espacios en blanco o algún signo de puntuación[241]. La elección de la palabra como unidad de tratamiento en lexicografía tiene una serie de ventajas y varios inconvenientes. Las ventajas que ofrece son su fácil reconocimiento como unidad gráfica, el arraigo que la palabra tiene dentro de una comunidad lingüística -mucho más que conceptos como lexema, monema o morfema-, y la influencia que ejerce en el proceso de enseñanza aprendizaje, dominado éste, pese a que el registro oral se considere más importante, por

---

[239] Ya Casares, J. (1950: 51) aludía al desacuerdo existente entre los gramáticos a la hora de definir la palabra; Werner, R. (1982b: 219-221) ofrece una abundante bibliografía que han tratado el tema; Jiménez, B. (1986: 70-71) ofrece diversas definiciones de palabra desde distintas perspectivas; lo mismo hace Lyons, J. (1975: 201-213), quien además acompaña su estudio con algunas referencias bibliográficas.

[240] La cita, recogida por Werner, R. (1982b: 222), es de Hausmann, J.: *Einführung in die Benutzung der neufranzösischen Wörterbücher*, Tubinga, Romanistiche Arbeitshefte, 19, 1977, p. 9.

[241] Justicia, F. (1995: 13) remite a Hockket, C. F.: *A Course in Modern Lingüistics*, New York, The MacMillan Co. 1958. Werner, R. (1982b: 222) la define así: "por 'palabra', en el modo de ser de la lengua llamado 'discurso', se entiende una unidad gráfica entre dos espacios en blanco"; Müller, Ch. (1963: 158) afirma: " Si l'on appelle unité graphique un groupe de signes alphabétiques séparé des autres signes ou groupes par des blancs ou par des signes de ponctuation, on pourra presque toujours admettre l'equivalence: unité graphique = mot"; Dubois, J. y C. (1971: 62) exponen: "...les entrées lexicographiques doivent être des 'mots' compris entre deux blancs typographiques au sens le plus strict", también consideran como entradas las palabras compuestas con guión. Michéa, R. (1973: 77), en cambio, aclara que para evitar la vacilación que supone el guión en determinadas formas (pues sí es una unidad *arc-en-ciel*, pero no *dites-le-moi*) se acude a la matización de palabra, a la que se define también como una unidad gráfica no comprendida entre blancos internos.

la expresión escrita, en la que la división gráfica es decisiva[242]. Además, la lexicografía se elabora a partir del modelo de "perfomance", de habla o realizaciones del discurso, y no a partir del modelo de competencia, de lengua (Dubois, J. et C., 1971: 61).

Los inconvenientes, no obstante, son muchos: "Del hecho de que la unidad 'palabra', como unidad lemática del diccionario, no esté definida unívocamente se deriva una serie de problemas para el lexicógrafo" (Haensch, G., 1982b: 463). Sobre todo si tenemos en cuenta que "el lema no se funda en una definición creada por la lingüística moderna, sino en una tradición plurisecular de lexicografía" (Werner, R., 1982b: 218). Los criterios seguidos a la hora de establecer el lema tampoco son unánimes. Prueba de ello lo tenemos en las entradas que ofrecen los distintos diccionarios, tanto los generales[243] como los de frecuencias[244]. Algunos autores (Haensch, G., 1982b: 463-487; Morales, A., 1986: 21) suelen señalar los casos más polémicos para la lexicografía, aquellos en los que no hay acuerdo, y que suelen coincidir con los propios de los diccionarios de frecuencia. La diferencia entre ambos estriba en que la lexicografía general , como no entra en cuestiones cuantitativas, puede tomar decisiones diferentes en casos polémicos, como registrar una entrada con varias subentradas en el mismo artículo, o dedicar

---

[242] Véase Werner, R. (1982b). Recordemos también las palabras de Galisson, R. (1983: 4): "Ensuite la nécessité de reconnaître que la notion de mot a une grande consistance psychologique chez le plupart des apprenants. Et ce, grâce à une représentation graphique précoce du mot en langue maternelle (par le canal de l'école: apprentissage de la lecture, de l'écriture, pratique des exercices à trous, etc.),...".; consideración similar observamos en Lyons, J. (1975: 206).

[243] Por ejemplo, el DRAE recoge este, aquel como representantes de todas las formas de estos demostrativos, y en cambio dedica tres entradas para las formas del artículo *el*, *la*, *lo*. La observación, corroborada, la tomamos de Alvar Ezquerra, M. (1982: 166).

[244] Véanse, por ejemplo, los artículos de Ortega, J. (1985) y de Morales, A. (1980).

dos entradas diferentes, etc., pero la estadística léxica debe eliminar estos matices, porque al final se deben ofrecer unas cifras para N y para V. Según Müller (1963: 171-173) no hay norma léxica perfecta porque el carácter complejo y movible del lenguaje no obedece a la cuantificación; pero, a pesar de todo, recomienda que sea explicitada y aunque su establecimiento sea laborioso, su aplicación debería ser simple, para ello conviene utilizar ante todo criterios formales, sintácticos.

En cuanto a la palabra, pues, como unidad de léxico o lema, se contemplan algunos casos polémicos:

A) Una palabra equivale a numerosas formas.

En este apartado habrían de tenerse en cuenta tanto las palabras que modifican su grafía según el contexto lingüístico[245], como las formas flexivas de los verbos, adjetivos, sustantivos, pronombres y artículo[246]. Es decir, nos referimos a las variaciones que producen en una misma palabra las categorías gramaticales como género, número, persona, modo, tiempo...

---

[245] Müller, Ch. (1963: 163) pone los ejemplos para el francés de algunos adjetivos polimorfos, *beau/bel*, o variaciones gráficas del tipo *encor/encore*. En español algo similar puede ocurrir con algunas formas apocopadas, como *san*, frente a *santo*.

[246] Müller, Ch. (1963: 164-165) recoge aquí también casos de sustantivos que forman el masculino y femenino con lexemas o sufijos distintos, como *coq/poule*, *menteur/menteuse*. (Recordemos que en español tenemos casos como *caballo/yegua*, *actor/actriz*, *tigre/tigresa*, *gallo/gallina*...). Müller, Ch. (1963: 165) propone que se registren por separado, alegando que "les deux substantifs opposés par leur genre, au contraire, peuvent évoluer séparément; l'un des deux genres peut prendre des acceptions sans les communiquer à lautre".

B) Una forma equivale a numerosas palabras.

1. Homografía y polisemia.
2. Homonimia sintáctica[247]. Con esta designación se recogen aquellas palabras que cambian de categoría[248] y adquieren, por tanto, valores sintácticos diferentes, como *fuerte*, *salida* (sustantivo/adjetivo), o *deber* (sustantivo/verbo), y sobre todo, palabras de frecuencia muy alta, como ciertos indefinidos que pueden funcionar como sustantivos, adjetivos o adverbios.
3. Homonimia morfológica. Nos referimos, siguiendo a Juilland y Chang-Rodríguez (1964: XXXI y XXXV) a aquellas formas flexivas idénticas pero que pertenecen a distintas categorías de persona (*amaba* 1ª y 3ª persona), tiempo (*llegamos* presente y pasado), etc.

Si se decide valorar determinadas medidas gramaticales, funcionales frente a otras de contenido semántico, no se dedicarían entradas distintas a la polisemia y homonimia léxica, y por tanto se obviarían casos como *bote*, 'botella', *bote*, de remos, o *bote*, de 'botar', y *marcha*, de 'marchar', o *marcha*, voz juvenil para designar el ambiente y la diversión; pero en cambio, sí se registrarían, por ejemplo, los gentilicios, como *granadino*, *andaluz*..., u otras palabras como *gitano*, cuando funcionan como adjetivo y como sustantivo[249].

---

[247] Este término lo usan Juilland y Chang-Rodríguez (1964: XXXI-XXXV) y Morales, A. (1986: 22-25). Refiriéndose a los mismos casos, Müller, Ch. (1963: 165-171) los registra bajo las denominaciones de "polivalencia sintáctica" o "derivación impropia", y "polivalencia funcional".

[248] Müller, Ch. (1963: 165-171) se refiere a los casos de participio de presente que se han sustantivado o adjetivado; los participios de pasado que funcionan como sustantivo o adjetivo, etc.

[249] En otros diccionarios de frecuencias se hacen, obviamente, estos planteamientos. Por ejemplo, Morales, A. (1986: 19) es consciente de este problema, al que se refiere cuando analiza la manera de actuar del

Lo mismo ocurrirá con los pronombres, a los que habrá registrar según su función sustantiva, adjetiva o adverbial. Todo esto repercute en el recuento final y en las diferencias que separan unos trabajos de otros. Por ello hay quien propone[250] hablar sólo de nombres, que según las circunstancias contextuales, funcionarían como términos primarios (función sustantiva) o términos secundarios (función adjetiva), y actuar de modo idéntico para los pronombres, registrándolos una sola vez, informando, además de su función primaria y secundaria, de su función como términos terciarios (función adverbial).

2.11.4. 1. Consideraciones de naturaleza semántica

2.11.4.1.1. Homonimia y polisemia

Este es uno de los mayores problemas con los que se encuentra la lexicografía. Problema que permanece sin resolver, y al que además los

---

FDSW, en donde se otorgan dos entradas para los gentilicios (En el propio *Léxico básico del español de Puerto Rico* también se encuentran dos entradas para americano, puertorriqueño...). Y concluye la autora que con "este estado de cosas, las distinciones que se establezcan en cada diccionario deberán ser cuidadosamente ponderadas si queremos evitar las irregularidades. Por ahora nos movemos en la situación de que, por lo menos para los recuentos del español, no tenemos ninguna pauta teórica general que haga más fácil la tarea". Lyne, A. (1973: 103-105) propone la idea contraria, es decir, en tales casos conceder una sola entrada en la macroestructura, y en la microestructura, en cambio, especificar la frecuencia como sustantivos y como adjetivos. Aunque es consciente que al no distinguir entre adjetivo y sustantivo en la macroestructura se presentan casos problemáticos: "par exemple on combinera *trayeuse* (machine) avec *trayeur*, *-euse* (personne), et *bonne* (personne) avec *bon*, *-nne* (adjetif). Cela semble inévitable à moins de ouvrir le débat sur l'homonymie".

[250] Ortega, J. (1985: 432-433) sigue de este modo la terminología propuesta por Alcina y Blecua. Alcina y Blecua (1991: 499 y 636) se refieren a casos como *francés*, *verde*, *sabio*, *filósofo*, *físico*, *viajero*..., y, por otro lado, a casos como *mucho*, bastante...

diccionarios dan soluciones distintas[251]. Humberto Hernández (1989: 49) propone que se adopten, para evitar la falta de uniformidad, alguna de las siguientes medidas cuando son homógrafos: o dar una sola entrada, o "entender que cuando no aparecen rasgos significativos comunes a dos significados de lo que parece una misma palabra se produce una situación de homonimia, o lo que es lo mismo, de palabras distintas con idéntico soporte sonoro. De estas dos soluciones quizás sea la segunda la de mayor utilidad desde el punto de vista didáctico, aunque también es la que ofrece mayor cantidad de dificultades". Pero, como afirma Ch. Müller, estos problemas son conocidos por los lexicógrafos, que al no considerar aspectos cuantitativos, pueden proponer una o dos entradas, pero la estadística léxica debe excluir estas disyuntivas, pues al final establecerá unas cifras determinadas para los vocablos y las ocurrencias[252]. Y no olvidemos que incluso los criterios esgrimidos para la identificación de los homónimos no son ni homogéneos ni concluyentes, pues, por ejemplo, el comúnmente usado criterio etimológico es puesto en tela de juicio[253], aunque las alternativas propuestas no sean en

---

[251] Calvo Pérez, J. (1986: 738-739), tras enumerar, siguiendo a Coseriu, una serie de precauciones que deben tomarse en la confección de diccionarios, reconoce que hoy en día el problema de la homonimia y la polisemia dista mucho de ser resuelto; y cita como ejemplo de comportamiento peculiar el diccionario de Casares, que da una sola entrada para *gato*, *hoja*, tomando el significante como referencia, y no el significado.

[252] Müller, Ch (1963: 172) expone: "La statistique lexicale, à un moment de son oeuvre, doit exclure ces nuances. Après avoir dressé l'inventaire des formes, il arrive un instant où elle doit fournir des chiffres en face des symboles N et V...".

[253] Dubois, J. y C. (1971: 66) no consideran totalmente rigurosos ni el criterio gramatical ni el etimológico para distinguir entre palabras distintas; Werner, R. (1982c: 300) afirma que "una distinción entre homonimia y polisemia según el criterio de la etimología ya no se puede admitir para la descripción sincrónica de la lengua desde el punto de vista de la lingüística moderna. Sin embargo, en la práctica lexicográfica, especialmente en los países hispanohablantes, es aún muy corriente".

absoluto clarificadoras, como el criterio de la conciencia lingüística de los hablantes[254] o el de que las palabras se relacionen no sólo con el significado nominativo, sino con el metafórico o metonímico (Werner, R., 1982c: 297-313).

En cuanto a la polisemia, el principal obstáculo que presentan los diccionarios de frecuencias es que al dar cuenta sólo del significante de las palabras, junto a una escueta información gramatical, no se registran, normalmente, los diferentes significados de una palabra. Esto no cabe duda que supone pérdida de información al no saber de qué sentido se trata[255], pero registrar todas las acepciones supondría ser detallista en exceso, propio de otros diccionarios distintos al nuestro.

Quizás, por tanto, la decisión más acertada hubiese sido no sólo desechar la polisemia, sino no hacer distinciones entre homógrafos léxicos, y tan sólo aceptar entradas diferentes cuando los homónimos pertenecen a clases

---

[254] Según el cual habría homonimia cuando no hubiese parecido en el significado de las palabras. Pero recordemos las afirmaciones de Werner, R. (1982c: 303): "...los hablantes de una lengua pueden (...) tener ideas totalmente equivocadas sobre el uso de los elementos de su lengua. De todas formas, en la mayoría de los casos, el hablante, al usar unidades léxicas, no reflexiona sobre relaciones entre los contenidos que se dan para él...", y cita los ejemplos de *gato*, 'animal' y *gato*, 'herramienta'. A éste caso se podría unir un término como *perras*, que designa 'animal' y 'dinero', debido al dibujo que presentaban antaño determinadas monedas, lo cual ya no está en la mente de los hablantes.

[255] Véase Bustos Tovar (1966-7: 187); Morales, A. (1986: 19). Ortega, J. (1985: 438-439) expone: "Resolver el problema de la polisemia es tarea verdaderamente incómoda, tanto que es uno de los grandes que tienen planteados los estudios de semántica actuales. Lo cierto es que si no se tiene en cuenta (como hace el *Frequency*) esta realidad, podríamos llegar a considerar que *punto*, por ejemplo, significa siempre lo mismo, lo que es falso. Si, por el contrario, se asigna una entrada a cada uno de los "sentidos" que encierra un determinado lexema, la representación del léxico se atomizaría hasta el infinito y se daría pábulo, así, a sustentar que entre esos sentidos no existen lazos".

morfológicas distintas,[256] No obstante, nosotros hemos preferido recoger la homonimia léxica, siguiendo para ello criterios etimológicos y basándonos en el *Diccionario de la lengua española*, 1992, de la Real Academia.

2.11.4.2. Consideraciones de naturaleza gramatical. Homonimia sintáctica y morfológica

2.11.4.2.1. El sustantivo y el adjetivo. Género y número. Otras consideraciones

Se concede un valor distinto al género según se trate de adjetivos o sustantivos. Sirve para otorgar dos entradas en los sustantivos, una para el masculino y otra para el femenino, pero una sola para los adjetivos. Lo cual está perfectamente acorde con la lingüística, que considera el género en el sustantivo como un hecho de lengua[257]. Además, el género debe servir para separar aquellos casos en los que se produce cambio de significado, dando lugar por tanto a signos distintos, como *el/la editorial*, *el/la pez*, *el/la cometa*, *el/la radio*, *el/la clave*, *el/la doblez*, *el/la orden*, *el/la margen*, *el/la frente*, *el/la cura*, *el/la cólera*, e incluso *el/la guardia*, *el/la policía*.

Asimismo, el número deberá distinguir entre *la corte/ las Cortes*, y casos similares.

Cuando palabras se distinguen sólo por la tilde, se han de poner entradas distintas si tienen cambio de significado, si no, sólo una (una entrada:

---

[256] Véase Müller, Ch. (1963: 165-166); Haensch, G. (1982b: 467-468). Así se actúa en el *Frequency*... de Juilland y Chang-Rodríguez y en el *Léxico básico*... de Morales, A.

[257] Morales, A. (1986: 25) cita textualmente las palabras de Rey-Debove, J. en *Étude linguistique et sémiotique des dictionnaires français contemporains*: "...contrario a los adjetivos el género en los nombres es un hecho de lengua, ya que no depende gramaticalmente de otra palabra en el discurso".

dínamo/dinamo; dos entradas: *cártel* 'convenio de empresas', *cartel* 'papel impreso para difundir noticias' (Haensch, G., 1982b: 464).

2.11.4.2.2. El pronombre: forma (género, número) y función

La mayoría de las formas pronominales, por la propia naturaleza del pronombre[258], presentan algunas peculiaridades que afectan a los diccionarios. En nuestro caso, y si se mantiene la coherencia de actuar según criterios funcionales, se han de conceder varias entradas: según su función de adjetivo, adverbio o pronombre, y en esta última, distinguir según el género, concediendo una entrada para el masculino y otra para el femenino. Recuérdese que el género se considera propio de la categoría nominal, y no de la adjetiva (Morales, A., 1980: 473; y 1986: 25). Por lo que el no actuar así supondría someterse a críticas justificadas[259], a pesar de que esta medida es en exceso prolija, ya que algunas formas pueden aparecer cuatro veces: por ejemplo, *mucho*, como sustantivo, adjetivo y adverbio, desdoblándose al mismo tiempo en *mucha*, femenino. En lo referente a otras formas pronominales, como *mi*, *tu*, *su*, *le*, *se*, *sí*..., hay coincidencia en señalar que no se hagan distinciones de género, pues no se hallan marcados[260].

---

[258] Mondéjar, J. (1977: 54-55) afirma la entidad del pronombre y lo define como "morfema funcional referencial. En cuanto a morfema, tiene significación gramatical; en cuanto funcional, es capaz de contraer relaciones y desempeña oficios; en cuanto referencial, en el discurso se carga de contenido lexemático sobre la base de los morfemas que lo caracterizan".

[259] Morales, A. (1980: 473) aprueba que Juilland y Chang-Rodríguez, en el *Frequency*..., utilicen el género para conceder dos entradas distintas a *ella*, *nuestra*, *vuestra* y *él*, *nuestro*, *vuestro*, como pronombres, pero no que se actúe así para el artículo *el*, *un* y *la*, *una*.

[260] Morales, A. (1980: 473; y 1986: 25) y Ortega, J. (1985: 435) aconsejan que no se hagan distinciones ya que no están marcadas respecto al género. De la misma opinión es Müller, Ch. (1963: 171), quien

El número no ofrece prácticamente casos de ambigüedad, salvo *sí*, *se*, y quedaría a juicio del investigador dirimirlo o no, aunque se podrían tomar criterios generales y uniformes (Morales, A., 1980: 474).

2.11.4.2.3. Pronombres personales

Las formas flexivas ocasionadas por el "caso" -*me*, *te*, *mí*, *ti*...- se agruparán bajo el lema de *yo*, *tú*, *él*, *ello*. En cuanto a la persona, la única observación hasta el momento se refiere a *usted*, lo cual exige una tercera persona del verbo, aun refiriéndose a la segunda. Amparo Morales (1986:24) critica que no se dé noticia de las formas verbales de tercera y segunda persona con *usted*, lo cual es importante, sobre todo para Sudamérica. Nosotros seguiremos las indicaciones de la lingüista puertorriqueña, aunque su contemplación no sirva como referencia general de uso, sino sólo aplicable a una población escolar de las características de la que analizamos. Otras observaciones, como la crítica de Bustos Tovar (1966-7: 188) al *Frequency*... de Juilland, de que no haya distinguido entre *le*, *les* tanto función de directo como indirecto, e incluso que no se registren las repeticiones del tipo *a mí me*, casi gramaticalizadas, salen fuera de las características de nuestro recuento, que contempla caracteres morfosintácticos, pero sólo para la distinción de clases y categorías gramaticales, no de funciones sintácticas.

2.11.4.2.4. Reflexivos

Las construcciones con *se* son muy variadas y frecuentes en español, por lo que han llamado poderosamente la atención a estudiosos y son objeto de

---

propone " neutraliser les oppositions qui ne sont pas manifestées morphologiquement, mais de compter celles qui le sont...".

numerosos artículos y libros[261]. Sobre sus características y usos, los autores son más o menos claros a la hora de considerar tales construcciones, aunque hay sus diferencias, y las justificaciones son igualmente rigurosas. Bobes Naves (1974: 87-127) hace una clasificación entre usos pronominales o pleremáticos: reflexivo (incluido el valor medial), recíproco y personal (alomorfo de *le, les*); y usos sintácticos o morfemáticos: el *se* de la pasiva refleja y el *se* impersonal, a los cuales la autora no los considera pronombres[262], si bien para ella el nombre que figura en la pasiva refleja no es sujeto, sino objeto, por lo que este hecho no distingue a la pasiva con *se* y a la impersonal con *se*, sino la presencia o ausencia, pues, del objeto directo (1974, IV, 2: 313-325). Desde otras perspectivas se aprecian igualmente

---

[261] Véanse por ejemplo Anderson, J. O. y Hills, E. C., "Relative frequency of Spanish Personal Pronouns", *Hispania*, XIV, 5, 1931, pp. 335-337; Bobes Naves, M. C., "Construcciones castellanas con "Se". Análisis transformacional", *Revista Española de Lingüística*, 1974, IV, 1, pp. 87-127 y IV, 2, pp. 301-325; Molina Redondo, J. A., *Usos de "se"*, Madrid, S.G.E.L., 1974; Alarcos Llorach, E., "Valores de /se/", *Estudios de Gramática Funcional del Español*, Madrid, Gredos, 1982, pp. 213-222; Alcina J. y Blecua, J. M., *Gramática española*, Barcelona, Ariel, 1991, pp. 907-923.

[262] Bobes Naves (1974: 93) afirma: "En los usos sintácticos o morfemáticos, el SE pierde su categoría propia, como unidad léxica y morfológica y queda reducido a índice categorial de verbo. No tiene significado, ni es susceptible de soportar accidentes gramaticales (género, número, caso, persona)". En la segunda parte de su artículo (*R.S.E.L.*, IV, 2, pp. 301-325), Bobes Naves desarrolla sus teorías acerca del se impersonal y el se pasivo, a los que se refiere como usos "no-pronominales". Para el se pasivo aduce que no tiene significado ocasional, ni es capaz de recibir morfemas, ni alternar con otros pronombres personales como *me, te, le*, ni desempeñar función sintáctica propia de las partes nominales. Si bien la autora considera que el nombre que acompaña a las pasivas funciona no como sujeto sino complemento directo. Además estima que la construcción impersonal es históricamente una derivación de la pasiva refleja. Ambas estructuras, impersonal y pasiva, no deben oponerse por el rasgo de ser "impersonales", pues las dos lo comparten: lo que las opone es que la pasiva lleva objeto directo y la impersonal carece de él. Alarcos, E. (1982: 218-221) defiende el valor reflexivo de las llamadas construcciones de pasiva refleja, con lo cual coincide con la Academia, que considera la forma se como un pronombre.

opiniones coincidentes, cuando se pide[263] que en los diccionarios de frecuencia se registren con entradas distintas el *se* reflexivo y el *se* impersonal, o se aconseja[264] que, aunque el reflexivo se incluya dentro del pronombre personal, no obstante se recojan todos los valores del *se*: el propiamente reflexivo, el medio, etc.

Por tanto, la postura más coherente será registrar bajo la forma del pronombre personal *él*, los valores del *se* personal, el propiamente reflexivo, el medial y el recíproco (y actuar así con las formas de 1ª y 2ª persona, a pesar de que existan posiciones encontradas)[265], y otorgar además una entrada para el *se* impersonal y otra para el *se* pasivo, siguiendo nuestra consideración sintáctica y morfológica de conceder entradas distintas cuando hay cambio de categoría.

---

[263] Morales, A. (1986: 24; y 1980: 471) censura que el *Frequency*... registre dentro de la entrada de *él*, el *se* reflexivo y el *se* impersonal. A su juicio, se deberían haber separado, dedicando una entrada para el *se* impersonal, porque en una consideración sincrónica, argumenta, así debe procederse. En cambio, y a pesar de que en el *Léxico básico del español de Puerto Rico*, (1986: 23) se afirma: "Los autores del *FDSW*, contrario a nosotros, no separan en la forma *se* los casos que son pronombre de los que son partícula impersonal...", no se aprecia en el diccionario puertorriqueño una entrada distinta para *se*, pues se pasa del lema *savia* al lema *secar*. Aunque es posible que se actúe así porque el *se* impersonal, al someterse al índice de dispersión, frecuencia y uso, no alcanzará la medida exigida.

[264] Véase Ortega, J. (1985: 434); Bustos Tovar, (1966-7: 188) habla de funciones recíproca, impersonal y signo de pasiva.

[265] Bobes Naves (1974: 87) considera arbitrario estudiar el valor reflexivo sólo para la 3ª persona, *se*, obviando la 1ª y la 2ª; mientras que Alarcos, E. (1982: 213-222) concede valor de auténtico reflexivo sólo a la 3ª persona.

2.11.4.2.5. Numerales

A nuestro juicio, en este apartado se presenta el problema de *uno* y sus variantes; es decir, si cabría registrarlo como pronombre indefinido o artículo, frente a otros usos como numeral. Los diccionarios de frecuencias no se plantean este problema, y por tanto no aparece la distinción entre *uno* numeral y *uno* indefinido[266]. Por tanto, hemos de tomar como referencia la opinión que vierten algunas gramáticas al respecto. La Academia (1983: 237) distingue dos aspectos en los numerales: uno de carácter semántico, vinculado a un lenguaje formal, y otro de naturaleza gramatical, pues muchos poseen género y número, y además pueden funcionar como pronombres o adjetivos; por lo que poseen la peculiaridad de presentarse en la escritura de dos formas distintas, con signo lingüístico (*uno*, *dos*, *tres*, *cuatro*, *cinco*...) y con signo matemático (1, 2, 3, 4, 5...). Los numerales cardinales del 1 al 0 actúan como sustantivos: a) "cuando designan cada uno de los diez guarismos 1,2,3...0" (RAE: 1983: 238). o b) "cuando designan los números enteros positivos de la serie ilimitada 1, 2, 3 ... 15 ... 20 ... 100 ... 1000"[267]. Refiriéndose al caso concreto de *uno*, la Academia (1983: 242-243) distingue el numeral cuando expresa un valor numérico, implícito o explícito, en ejemplos como *A ti te daré una peseta y a él le daré dos* (que llamaremos criterio c); en plural, *unos*, *unas*, siempre son indefinidos, salvo en los dos casos a) y b), y si designan a sustantivos que siempre se usan en plural, como *unas castañuelas*, *unos prismáticos*, etc., habrá que aplicar el último criterio.

---

[265] Consúltese, por ejemplo, Morales, A. (1986); Justicia, F. (1985a y 1995).

[267] RAE (1983: 238). Por lo cual en este caso no sólo designan números, como *el 30* (el número 30), *el 50* (el número 50), sino otros conceptos, como *La generación del 27* (año 27), *la 8* (la fila ocho), *el XVI* (el siglo XVI), etc.

Otros gramáticos coinciden con la RAE y conceden a *uno* una serie de valores, como el indiferenciador y numérico, los cuales se neutralizan en algunos contextos, pero en otros predomina el valor numeral, como en el caso de las enumeraciones de objetos, o con el refuerzo de *solo*, *siquiera* y otras expresiones[268]. Alarcos (1994: 114-117) coincide con la Academia en el tipo de referencia imprecisa que realizan numerales e indefinidos, pero por las peculiaridades que presentan ciertos indefinidos y los numerales cardinales, propone que se incluyan dentro de los cuantificadores. Además, para el autor no tiene justificación hablar de artículo indeterminado ni la distinción que tradicionalmente se ha hecho entre *uno* numeral y *uno* indefinido: "No puede ser artículo por cuanto este (*el*, *la*, etc.) carece de acento y no es palabra independiente. Separar, de otra parte, el sentido numeral respecto del indefinido es innecesario, ya que de todas las maneras se trata de un cuantificador" (Alarcos, E., 1994: 122). Alarcos incluso deja fuera de los cuantificadores numerales las formas que designan los guarismos o los números enteros positivos, ya que para él son sustantivos[269].

Con respecto a la disyuntiva de si *uno* es artículo o pronombre, además de Alarcos no hablan de artículo indeterminado la Academia (1983: 213-217), ni Alcina y Blecua[270]. En cambio, otros autores sí distinguen entre artículo

---

[268] Alcina y Blecua (1991: 671) citan algunos ejemplos, como "cuatro sillas, un brasero, un sillón y un bufete componían su mueblaje (P. A. Alarcón, *El Escándalo*, 19)".

[269] Alarcos, E. (1994: 120-121) afirma: "Estas unidades comportan género masculino y admiten número plural: *el uno*, *el dos*, *el tres*, *el cuatro*..., *el diez*, *el ciento*, *el mil*..., *el cero*; *los unos*, *los doses*, *los treses*, *los cuatros*, *los dieces*, *los cientos*, *los miles*, *los ceros*. Como auténticos sustantivos pueden recibir términos adyacentes: *este dos*, *algún cuatro*, *tres cincos amarillos*, etc.".

[270] Alcina y Blecua (1991: 670), dentro del apartado del indefinido *uno*, exponen: "Tradicionalmente, los gramáticos presentan a las formas singulares en función adjunta como gramaticalización del numeral paralela a la de los derivados del demostrativo latino *ille*, como artículos indeterminados".

determinado e indeterminado (Sánchez Márquez, 1972: 278; Gili Gaya, 1976: 242-243; Marcos Marín, 1980: 235).

Nosotros, en el caso de *uno*, creemos que el contexto permite distinguir entre su función como numeral y como indefinido, según la explicación de la Academia y Alcina y Blecua. Al fraccionario *medio*, además, lo contabilizaremos cuando cumpla una función adverbial[271].

2.11.4.2.6. Los indefinidos.

Dentro de este tipo de elementos, lo peculiar es destacar la doble o triple función de algunas formas. Alarcos (1994: 115-116) señala que los indefinidos *alguien*, *algo*, *nadie*, *nada*, *quienquiera* funcionan exclusivamente como sustantivos, pero con respecto a las formas *algo*, *nada*, considera que su función sustantiva no impide que, en combinación con un adjetivo, "se comporten como adyacentes de este, por tanto, como si fuesen adverbios". Así por ejemplo, se distingue función sustantiva en *Dame algo frío*, *No dijo nada importante*, pero adverbial en *El café está algo frío*, *Eso es cuestión nada importante*. También se concede valor adverbial a *algo*, *nada* cuando son adyacentes circunstanciales, como en *Algo se acordaba de nosotros*, *No le molestan nada esas críticas*. Esta doble función la recoge también Alcina y Blecua (1991: 650-651). Alarcos (1994: 120) reconoce función adverbial para los demás cuantificadores si se inmovilizan con valor neutro cuando están en la forma de masculino singular. Alcina y Blecua (1991: 656-663) otorga valor de adverbio cuando van con adjetivo, adverbio o verbo en función circunstancial. La RAE (1983: 226-236) también reconoce la función adverbial de algunos indefinidos.

---

[271] La RAE (1983: 246) concede un valor adverbial a *medio*, y se citan como ejemplos *medio muerto*, *a medio vestir*.

Nosotros, siguiendo las indicaciones de los gramáticos, haremos tales distinciones, por lo que a algunos indefinidos los registraremos como pronombres, adjetivos y adverbios.

2.11.4.2.7. Relativos

A nuestro juicio, las formas *el cual, la cual, el que, la que, lo que...*, son las que constituyen los casos polémicos. Las formas con *cual* se pueden considerar como una unidad o como unión de artículo más pronombre; y las formas *el que, la que...*, se pueden interpretar también de varias maneras: a) constituyen una unidad; b) están integradas por pronombre más *que*; c) la forma *el, la, lo...* no es pronombre sino artículo + relativo; d) la secuencia se interpreta como artículo más conjunción[272]. En el primer caso, si hemos decidido adoptar la unidad gráfica como lema, debemos registrar separadamente ambas formas: como artículo, *el, la, los...*, y como pronombre, *cual, cuales*; en el segundo caso, creemos que aunque *el* sea un resto del demostrativo, en una consideración sincrónica funciona como artículo, ya que carece de acento y por tanto no tiene independencia, y *que* funciona como pronombre relativo.

En las formas *como, donde, cuando*, que se interpretan como adverbios relativos, como conjunciones, o como pronombre (Ortega, J., 1985: 434), conviene destacar que seguimos los pasos de Juilland y de Amparo Morales[273],

---

[272] La interpretación d) la recoge Morales, A. (1986: 24)

[273] Morales, A. (1980: 471, y 1986: 24) expone la dificultad que supone la distinción de estas formas en algunos contextos, pero, a pesar de todo, concede entradas distintas cuando funcionan como conjunción y como adverbio.

y distinguiremos entre adverbio y conjunción. En el *Léxico básico del español de Puerto Rico*, se registra *como* adv. y conj. y en cambio se recoge *cuando* sólo como conj. Si la consideración de adverbio o conjunción es que lleven antecedente o no, *cuando* creemos que tiene una frecuencia mayor con antecedente que *como*.

2.11.4.2.8. El verbo. Tiempo, modo, persona.

Los morfemas gramaticales peculiares de tiempo y modo a veces no son suficientes para distinguir determinadas formas en pasado y presente, como *llegamos*, *paramos*, *estudiamos*..., y algunas formas del imperativo que coinciden con el indicativo, como *toma*, *bebe*... Y el morfema de persona tampoco sirve para distinguir aquellas formas que pueden ser 1ª y 3ª (*había*, *temía*, *llegaba*...) y cuando la tercera persona se refiere en realidad a la segunda según impone el tono cortés con *usted*. Todos estos fenómenos se agrupan bajo el nombre de homonimia morfológica (Juilland y Chang-Rodríguez, 1964: XXXI y XXXV; Morales, A., 1986: 23-25). En estos casos colocamos alguna notación que permita distinguir unas palabras de otras (Cfr. 6.).

Por otro lado, las distintas formas verbales han ocasionado numerosos casos de homonimia sintáctica o cambio de categoría, como infinitivos sustantivados, participios de presente sustantivados, participios pasados sustantivados..., como *poder*, *deber*, *estudiante*, *salida*, *llegada*..., cuya categoría se reconoce muy bien en el contexto.

2.11.4.2.9. Participios de pasado adjetivados

No ocurre lo mismo con los participios pasados en función adjetiva. Estos presentan numerosos problemas a la hora de registrarlos en los diccionarios

debido a que nadie ha fijado el criterio definitivo que permita identificar sin ambigüedad cuándo son adjetivos o cuándo no lo son. Si acudimos a criterios sintácticos, los lingüistas suelen señalar una serie de funciones propias de los participios: como atributos, como predicativos referidos al sujeto o al complemento directo, como adjetivos adyacentes al sustantivo -caso que puede equivaler a una proposición adjetiva[274]-, construcciones de participio en cláusula absoluta, y formando los tiempos compuestos de la activa, pasiva y algunas perífrasis verbales. Cuando no forman perífrasis, las gramáticas no hablan, en general, claramente de adjetivos, aunque en los apartados y capítulos destinados a la descripción y características del adjetivo, ponen ejemplos de adjetivos que son participios[275]. Los libros de texto destinados a los estudiantes de C.O.U.[276] tienen la necesidad de manejar criterios más o menos claros, y a veces pueden servir para dirimir algunos problemas, pero en este caso es un reflejo de la indistinción que se establece en la lingüística teórica. Lázaro Carreter y Vicente Tusón (1978: 123; y 1992: 104), cuando hablan del participio señalan que puede completar a un sustantivo, exactamente igual que un adjetivo en casos como *Los periódicos publicados hoy no traen la noticia*, *Sólo se alimentaba de patatas cocidas*, aunque hacen notar que a pesar de su carácter adjetivo, conserva su capacidad verbal de llevar complementos. Añaden que el participio puede realizar además todas las funciones del adjetivo, como las de atributo o complemento predicativo. Y conceden carácter preferentemente verbal cuando el participio va en cláusula

---

[274] La RAE. (1983: 496) pone como ejemplo *Hojas de árbol caídas*, que equivale a *hojas que han caído...*

[275] Gili Gaya (1976: 219) aporta el siguiente ejemplo: *La niña lo miraba todo con una cara asustada.* y la RAE. (1983: 408): *Los excursionistas volvían cansados.*

[276] Creemos que es pertinente en este apartado -o en otros- citar estudios destinados a una población escolar, sobre todo si están confeccionados por lingüistas de categoría incontestada, y porque nuestro trabajo considera al estudiante como el protagonista, ya que él es la fuente y la finalidad.

absoluta, bien referida al sujeto de la proposición principal: *Juan deambulaba, el gesto cansado, por las calles*, bien referidos a un sujeto propio que no figura en la proposición principal: *Acabada la función, cerraron el cine*. Para J. A. de Molina Redondo y M. Rodríguez (1993: 10-105), el participio, conservando su valor verbal, tiene varios usos: como adjetivo (*Recuerden las notas tomadas ayer*), en construcciones absolutas y en perífrasis verbales (*El problema está resuelto, Llevo cinco folios pasados a limpio*). Gómez Torrego (1988: 16) reconoce la doble naturaleza lingüística de las formas no personales del verbo, pero es tajante cuando afirma que "cuando estas formas no personales constituyen el verbo principal o auxiliado de una perífrasis verbal, lo que domina en ellas es el valor verbal, de forma que lo nominal del infinitivo, lo adverbial del gerundio y lo adjetival del participio desaparecen". El tono conciliador de F. Marsá (1986: 201-204) cuando con una enorme sencillez no exenta de profundidad trata de aclarar que el participio es verbo y es adjetivo -y pone para ello varios ejemplos, entre ellos, el caso de María Moliner, que recoge *perseguido* como "participio adjetivo"-, y por tanto, no tiene sentido hacer distingos, tampoco nos soluciona si los participios pasivos deben registrarse una o dos veces en los diccionarios de frecuencia. Para Ch. Müller (1963: 167) son muy difíciles de separar el sentido adjetivo del verbal, puesto que por naturaleza, los participios tienen algo de adjetivos, y funcionan como tales.

Se pueden utilizar criterios semánticos, como sugerir que se aproximan a su carácter verbal cuando permanece en los participios la noción de resultado de un proceso, y se aproximan a su carácter adjetivo cuando no es así[277]; o acudir a criterios de índole sintáctica y semántica, simultáneamente, como hace la Academia (1983: 495) en aquellos casos en que el participio equivale a

---

[277] Müller, Ch. (1963: 167) no obstante, aconseja dejar la mayor parte de los participios dentro de sus verbos correspondientes: "On sera donc tenté de laisser toutes ces formes -ou la plupart d'entre elles, au prix d'une liste d'exceptions...- à l'intérieur des formes verbales".

un adjetivo de significación activa: "Cuando los verbos transitivos tienen además uso pronominal, a este corresponde un participio o adjetivo de significación activa; p. ej.: de ***resolver un problema*** nace un participio pasivo (***problema resuelto***); pero de ***resolverse*** sale un adjetivo activo (***un hombre resuelto***). Por analogía se propaga este doble valor a otros participios de verbos transitivos que, aunque no tengan uso pronominal, expresan acciones producidas por el hombre y cuyo participio adjetivo denota costumbre, hábito de realizar determinados actos. Así, ***leído*** es participio pasivo en ***un libro leído***, y adjetivo activo en ***una persona leída***...". De las mismas características que ***leído*** cita otros casos, como ***disimulado***, ***sabida***, e incluso afirma que va aumentando el número de estos participios con este doble valor, y ofrece una lista considerable[278]. Gili Gaya (1976:200), refiriéndose a la misma circunstancia, y utilizando los mismos ejemplos, no habla de "adjetivo activo", sino en todo momento de participios: "El participio de los verbos intransitivos y reflexivos tiene significación activa". Y enumera una serie de participios con similares características a los de la Academia[279]. Con todo, la propia Academia, al actuar según estos criterios, no convence de un modo absoluto cuando no registra en su diccionario algunos participios, claramente adjetivados[280], con

---

[278] La RAE (1983: 495) ofrece la lista siguiente: *agradecido, almorzado, bebido, callado, cansado, cenado, comido, considerado, descreído, desesperado, desprendido, disimulado, encogido, entendido, esforzado, fingido, leído, medido, mirado, moderado, precavido, resuelto, sabido, sacudido, sentido.*

[279] Gili Gaya (1976: 200) cita en concreto *acostumbrado, arrepentido, atrevido, comedido, osado, parecido, porfiado, preciado, presumido, recatado, sentido, valido.*

[280] Alvar Ezquerra, M. (1982: 168-169) afirma que *atrevido, sosegado, exagerado* o *corriente, durmiente, vituperante* son adjetivos, aunque el DRAE no los recoja como tales, y todavía acompañe para los segundos la información de "participio activo", hoy en desuso. Tampoco diccionarios como el de María Moliner o el de Casares recogen todas estas formas como adjetivos.

lo cual no es válido, por ejemplo, considerar adjetivos sólo los casos que recoja el DRAE[281].

Por tanto, parece que si utilizamos sólo criterios sintácticos o simultáneamente, por así decirlo, sintácticos y semánticos, las siguientes preguntas quedan sin respuesta: 1. ¿Hay que dar cuenta de todas las veces que el participio desempeña una función de adjetivo? ¿Y cuándo se producen esos casos? 2. ¿Sólo hay que registrar los participios que se han lexicalizado como adjetivos? ¿Cuáles son esos casos o cómo distinguirlos?

La clave hemos de buscarla en consideraciones morfológicas y sintácticas: Alcina y Blecua (1991:753) reconocen que el participio admite las mismas construcciones que el adjetivo, pero aconseja que se conserve la denominación de participio, si bien distinguiendo entre participio y adjetivo verbal: "Aquí se llamará participio siempre a las formas inmovilizadas con neutralización de género y número, y se llamará adjetivo verbal y, a veces, participio adjetivo a las realizaciones concordadas". Esta parece ser la solución más sencilla, es cierto, pero también la más rigurosa, y por ello es la adoptada, creemos, por Juilland y Amparo Morales. Porque está plenamente de acuerdo con la postura inicial de conceder mayor valor a los criterios morfológicos y sintácticas que a los semánticos. No obstante, esta decisión entra en fricción con otras opiniones, relativas a los auxiliares y las formas compuestas. En el caso de los verbos auxiliares, como *haber*, *ser*... hay quien propone que cuando son tales y por tanto forman los tiempos compuestos, se registren bajo el lexema verbal no auxiliar (Ortega, J., 1985: 436). Porque, por

---

[281] García Hoz (1953a: 39) actuó según esta línea, como él mismo recuerda: "he de advertir que para que un participio pudiera ser registrado como adjetivo necesariamente ha de ser consignado como tal en el Diccionario de la Real Academia; así, por ejemplo, *sombreado* ha de registrarse siempre como *sombrear*, porque no aparece en el Diccionario independientemente del verbo; en cambio, *tirado* puede registrarse como *tirar* y como *tirado*, porque en esta forma figura también en el Diccionario."

ejemplo, en el *Frequency...*, a una forma verbal como *he cantado* se le dan dos entradas distintas: una como auxiliar, bajo *haber* y otra bajo el lema *cantar*. Para J. Ortega (1985: 436), una forma como *he cantado* debería recibir el mismo tratamiento que otra como *canté*. Algunos autores llegan más lejos, y piden que se especifiquen los usos auxiliares de *ser*, *estar*, en la pasiva, junto a sus valores copulativos y predicativos, y que no sólo se registren los usos modales de *ir*, *tener*, *deber...*, sino que además en *haber* se distingan los tiempos compuestos de su uso en construcciones perifrásticas (Bustos Tovar, 1966-7: 187-188).

Actitud totalmente contraria a ésta, pero igualmente efectiva y rigurosa es la de contabilizar todas las formas del participio, incluso las de función adjetiva, dentro del verbo correspondiente[282].

Nosotros actuaremos de acuerdo con criterios formales, contabilizando las formas como participio cuando están inmovilizadas en género y número, por tanto, y en rigor, cuando forman los tiempos compuestos de la llamada voz activa.

---

[282] Para Lyne, A. (1973: 89), "tout participe passé en fonction adjectivale sera donc affecté au verbe". Este criterio es el que parecen seguir los autores del *Trésor de la langue française* cuando afirman: "On en fait jamais le départ entre l'emploi adjectif y l'emploi participial, même comme distingué, pourtant sémantiquement éloignés du participe lorsqu'iis sont pris adjectivement" (Imbs, P., 1971: XIIX, Vol. I). Quede pues constancia del problema que presentan estas palabras incluso para otras lenguas.

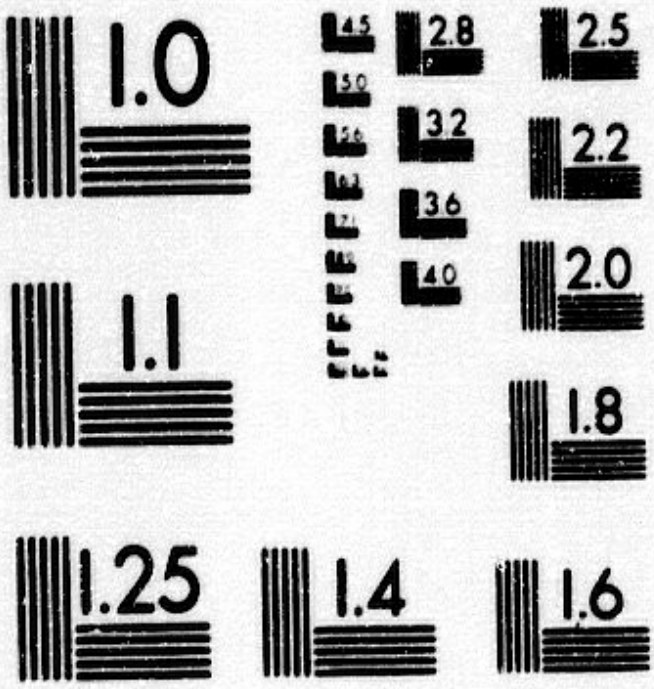
2.11.4.3. Consideraciones de naturaleza léxica

2.11.4.3.1. Palabras con guión. Derivación

Las unidades que lleven en su interior un guión, bien uniendo diversos lexemas, bien uniendo un lexema con prefijo (*ex-amiga*, *ex-marido*...) los contabilizamos como un solo vocablo. Actuamos así porque, en el caso de los lexemas, el guión es quizás la primera fase para una lexicalización definitiva, y en lo tocante a los prefijos, porque no son elementos autónomos, sino dependientes.

Algunas observaciones que retomamos relativas a la derivación se refieren a las propuestas de que se incluyan como subentradas o en listas aparte los morfemas derivativos (Ortega, J., 1985: 438; Werner, R., 1982b: 230-231), o llamadas de atención sobre los adverbios en *-mente*, a los que diccionarios normativos como el DRAE prefieren introducir cuando su "existencia no ofrezca dudas" (Alvar Ezquerra, M., 1982: 168).

Los elementos derivativos nos servirán para conceder entradas distintas cuando supongan un cambio de categoría en la palabra, para distinciones de género en el sustantivo o cuando remitan a un referente distinto y por tanto con un significado diferente. A los diminutivos, aumentativos, despectivos y superlativos no se concederán entradas distintas, sino que se registraran dentro del lema que les corresponda, salvo en los casos precedentes. Obviamente, los pocos comparativos sintéticos que existen en español como *mejor*, *peor*, *superior*, *mayor*, *menor*, *peor*... serán consignados con diferente cabecera al adjetivo positivo, pues son lexemas distintos.



MICROCOPY RESOLUTION TEST CHART
NATIONAL BUREAU OF STANDARDS
STANDARD REFERENCE MATERIAL 1010a
(ANSI and ISO TEST CHART No. 2)

2.11.4.3. Consideraciones de naturaleza léxica

2.11.4.3.1. Palabras con guión. Derivación

Las unidades que lleven en su interior un guión, bien uniendo diversos lexemas, bien uniendo un lexema con prefijo (*ex-amiga*, *ex-marido*...) los contabilizamos como un solo vocablo. Actuamos así porque, en el caso de los lexemas, el guión es quizás la primera fase para una lexicalización definitiva, y en lo tocante a los prefijos, porque no son elementos autónomos, sino dependientes.

Algunas observaciones que retomamos relativas a la derivación se refieren a las propuestas de que se incluyan como subentradas o en listas aparte los morfemas derivativos (Ortega, J., 1985: 438; Werner, R., 1982b: 230-231), o llamadas de atención sobre los adverbios en *-mente*, a los que diccionarios normativos como el DRAE prefieren introducir cuando su "existencia no ofrezca dudas" (Alvar Ezquerra, M., 1982: 168).

Los elementos derivativos nos servirán para conceder entradas distintas cuando supongan un cambio de categoría en la palabra, para distinciones de género en el sustantivo o cuando remitan a un referente distinto y por tanto con un significado diferente. A los diminutivos, aumentativos, despectivos y superlativos no se concederán entradas distintas, sino que se registraran dentro del lema que les corresponda, salvo en los casos precedentes. Obviamente, los pocos comparativos sintéticos que existen en español como *mejor*, *peor*, *superior*, *mayor*, *menor*, *peor*... serán consignados con diferente cabecera al adjetivo positivo, pues son lexemas distintos.

2.11.4.3.2. Criterio purista, aperturista: neologismos, préstamos, formas apocopadas, siglas...

Aunque hay voces en el terreno de la lexicografía general que llegan muy lejos a la hora de proponer la inclusión de unidades en los diccionarios generales[283], nosotros no podemos abarcar tanto, porque nuestro vocabulario tiene unas limitaciones específicas. Pero en nuestra intención está el deseo de aportar un cierto espíritu aperturista. Por otro lado, este espíritu se observa en obras lexicográficas generales actuales, que recogen voces no aparecidas antes en diccionarios (aunque éstas se utilizaran desde hace tiempo), tratando así de reflejar el verdadero uso del idioma[284]. Por tanto, los neologismos de creación propia de nuestro idioma, y teniendo en cuenta las características de nuestro vocabulario, que recoge la lengua viva del escolar, conviene registrarlos; igualmente los xenismos, sobre todo aquellos que tienen cierto uso entre la población adulta, y las palabras tabuizadas, para dar así un testimonio real del uso de la población cuyo léxico analizamos[285]. Al mismo

---

[283] Haensch, G. (1982b: 415-422, y 1982c: 248-252), partiendo de que los llamados por él diccionarios descriptivos deben reflejar el uso y por tanto deben suponer una puesta al día, llega a proponer que se incluyan expresiones y colocaciones como *puerta de la calle, lavado en seco, en el acto, un éxito clamoroso, un soltero empedernido*, modismos como *¿qué tripa se te ha roto?*, palabras marca, fórmulas de la vida social, formas muy usuales pero contrarias a la norma, como *mesmo, arbañil...*, pero con la indicación de la palabra considerada correcta, etc.

[284] Jiménez, J. L. (1994: 496) se refiere al *Diccionario General Ilustrado de la Lengua Española*, de Alvar Ezquerra, que recoge términos como *biodegradable, strip-tease, pizza, cojonudo, gilipollas, movida, pasota, software....*

[285] Haensch, G. (1982b: 406-413) aconseja que un diccionario normativo debe sopesar la inclusión tanto de neologismos y extranjerismos como de palabras tabuizadas, pero un diccionario descriptivo que tenga que describir la realidad de la lengua y dar el máximo de información, deberá incluirlos.

tiempo, las formas apocopadas irán como subentrada, bajo el lema correspondiente, lo cual está acorde con nuestro criterio de no conceder cabeceras distintas a éstas ya que conservan la misma clase gramatical y el mismo significado básico.

Con respecto a las siglas, no se puede negar que constituyen hoy en día uno de los procedimientos más característicos de creación léxica. En España, su implantación es relativamente tardía y de menor vitalidad que en otras lenguas, a pesar de que en las últimas décadas su auge es algo evidente (Rodríguez González, 1993: 9-10). No cabe duda, pues, de que las siglas representan un papel cada vez más destacado en todas las lenguas actuales de cultura, en las que muchas son propias creaciones de la lengua materna, o siguen, otras veces, la vía del préstamo (Santoyo, J., 1980: 17), y así, en español, tenemos siglas propias como BUP, COU, EGB, y extranjeras, como OTAN, IBM, IRA, NASA... A estas notas externas hay que sumar otras de carácter lingüístico, pues se pueden considerar palabras que poseen unas determinadas características fonológicas, gramaticales y léxico-semánticas. Las siglas van sufriendo un proceso de lexicalización hasta convertirse en un nuevo vocablo, que no difiere ni en apariencia ni en comportamiento de otras unidades léxicas (Rodríguez González, 1993: 9-24; Santoyo, J., 1980: 17-19; Mejía, J., 1980: 30-31). Además, la mayoría de las siglas se suelen ajustar al sistema fonológico del español, si bien también son numerosas las que no (CAMPSA, PYME, MPAIAC...). En el plano gramatical, "la referencia a organizaciones e instituciones de variado orden determina el carácter sustantivo de las siglas, la mayoría de las cuales constituyen nombres propios. En menor número se siglan designaciones de nombres comunes y concretos referidos a personas (PNN, ATS) u objetos (TV, OVNI)" (Rodríguez González, 1993: 13). Como sustantivos, admiten género y número; el género se refleja en el determinante[286], y para el número reduplican sus componentes (I.C.E.,

---

[286] Santoyo, J. (1980: 18) pone ejemplos como: " la BBC, la CIA, la FAO, el IRA....".

II.CC.EE., EE.UU) (Rodríguez González, 1993: 13; Marcos Marín, 1984: 27), o adoptan cualquiera de los morfemas que a tal efecto existen en español (*los* LP, *los* LPs, *los ovni*, *los ovnis*), con la consiguiente vacilación propia de los términos que terminan en consonante final inusual en el sistema fonológico español[287]. Como argumento definitivo para incluir las siglas en los diccionarios, o al menos, en los diccionarios de frecuencias, citemos de nuevo las palabras de Rodríguez González (1993: 14): "Los rasgos examinados hasta aquí constituyen de por sí una buena muestra de la integración progresiva de las siglas en el sistema lingüístico. Ahora bien, más importantes a la hora de medir el grado de lexicalización alcanzado son dos aspectos directamente relacionados con la formación de las palabras a partir del lexema siglar, a saber, la composición y la derivación". Y así se observan, sobre todo en la prensa, términos como *la ucedé-Madrid*, UNICEF-*España*, *peceprofesor*, *Ucedepeneuve*, o derivados, ya desde la II República, como *faísta* (FAI), *cedista* (CEDA)..., o ya durante los años setenta, años de aluvión para las siglas, se aprecian términos como *peceros* (PCE), *peneuvistas* (PNV), *ucedización*... (Rodríguez González, 1993: 14-20). Este último aspecto, referente la formación de nuevas unidades léxicas tomando a la sigla como elemento lexicogenésico, es uno de los más valorados por el autor[288]. Otros

---

[287] Rodríguez González (1993: 13-14) pone como ejemplos *los LP*, *los LPs*, *los LP's*, *los LPS*, *los elepés*. Para Santoyo, J., (1980: 19) las terminaciones que no se adaptan a nuestra fonética son la permanencia del morfema inglés, como en *vips*, *vip's*. Véase también para el plural de los sustantivos: Lorenzo, E. (1980: 80-95); RAE (1983: 180-189); Marsá, F. (1986: 124).

[288] Véase Rodríguez González (1993: 10): "...con ello, la puerta queda abierta para la creación de nuevas unidades léxicas a partir de ellas. Precisamente este aspecto es lo que hace pertinente su inclusión en los estudios sobre formación de palabras". Y en otro lugar de su artículo (1993: 20): "A pesar del rico y variopinto caudal de usos y formas que acompañan a los derivados de siglas en su conjunto, la formación y empleo de éstos obedece a reglas fonológicas, morfológicas y semánticas muy precisas, análogas a las seguidas por otras voces de la lengua. Desde el punto de vista grafemático, la adopción de la minúscula constituye un fiel exponente del grado de lexicalización alcanzado por la sigla base, en virtud del cual el

investigadores opinan que hay que ser cautos a la hora de considerar estas unidades como auténticos vocablos, integrados en nuestra lengua, si bien reconocen que cuentan con numerosas características a su favor para reconocerlos como tales (Santoyo, J., 1980: 19). Lingüistas reconocidos de la talla de Haensch en varios lugares de su obra (1982b: 420-421 y 1982c: 241) abogan porque se recojan.

2.11.4.3.3. Nombres propios

Los nombres propios se suelen suprimir atendiendo a razones semánticas, pero por razones didácticas algunos diccionarios los incluyen, aunque manejen para ello criterios no totalmente coherentes[289]. La lexicografía moderna aboga porque no se registren sino en las enciclopedias.

No obstante, G. Haensch (1982b: 416-417) matiza que aunque los nombres propios no deben recogerse en diccionarios, sí cuando formen parte de una locución, como *más feo que Picio*, *saber más que Lepe*, o los propios que han pasado a comunes, como *jerez*, *celestina*, *rebeca*... En ese sentido, en nuestro trabajo se encontrarán varios nombres propios: *Picio*, usado en la expresión citada; *aladino*, utilizado como sustantivo común en la frase particular de un alumno: *...se les concede por medio de un aladino, de un hada*; *valentina*, en la frase de otro: *...despedirme a la valentina*; *alsina*, muy usado en vez de *autobús* en algunas zonas de Granada y así queda constancia en los textos de los escolares (*Todos los días subían alsinas. Allí muchas plazas: la*

---

lexema derivativo puede considerarse como una verdadera nueva unidad léxica de la lengua".

[289] Seco, M. (1987: 181) cita los diccionarios de M. Moliner, el Diccionario Vox, el Petit Robert, el Collins, como los que recogen nombres propios, normalmente en apéndices o en la nomenclatura general, y destaca la crítica al DRAE por aceptar nombres propios de astros y algunos religiosos y rechazar los de otro tipo, como continentes o países, etc.

del Ayuntamiento, la plaza de la iglesia, la de la alsina... Cogimos por la mañana temprano la alsina que nos llevaría hasta allí), y *Pedro*, nombre integrado en la frase por todos conocida, *como Pedro por su casa*, reproducida literalmente por un escolar de nuestra investigación.

2.11.4.3.4. Palabras únicas o palabras idiomáticas

Nos referimos a aquellas palabras que suelen formar parte de las frases hechas o expresiones fijas, como *bruces*, *lirondo*..., que Zuluaga (1980: 18-19) llama "palabras únicas" o "palabras diacríticas" y otros autores, como Reichling, "idiomatic words"[290]. Para Zuluaga (1980: 18) son palabras, pues gozan de unidad fonológica y su función semántica se caracteriza por "la constitución y distinción de signos lingüísticos". Tales palabras carecen de significado léxico, categorial y relacional. Pero también deberán consignarse en el recuento.

2.12. Proceso de recuento de palabras y vocablos

No cabe duda de que la informática facilita la labor del recuento. Los años en que el trabajo de despojo de los textos se hacía de un modo artesano, a mano, como la ingente obra de Rodríguez Bou (1952), han quedado como reminiscencias del pasado. La Lingüística se ayuda ahora de la tecnología para acciones cuantificativas necesarias en este tipo de búsquedas. En varias investigaciones se citan programas y bases de datos para efectuar los recuentos y la presentación del vocabulario[291], o programas de análisis de

---

[290] Citado por Zuluaga, A. (1980: 18).

[291] Justicia, F. utiliza el DATA PERFECT (1985), y el DBASE PLUS IV (1995);

contenido por ordenador que, aunque con otros fines, pueden efectuar dicha labor[292]. Si bien la informática facilita el cómputo, la labor lexicográfica previa debe realizarla el investigador. Aunque el programa ALCESTE parece que ofrece algunos pasos más avanzados: divide los textos automáticamente en enunciados, que constituyen las filas, y en las columnas correspondientes se aprecian los vocablos. El ordenador separa las palabras función de las palabras contenido o con sentido pleno, y estas se agrupan tras un proceso de lematización. E incluso otros programas tras sufrir varios procesos de innovación, podrían conseguir desambiguaciones importantes en la polisemia y homonimia, por ejemplo[293]. La mayoría de estos programas suelen operar sobre la base de la palabra como unidad de información, por lo que pueden generar un listado de todos los vocablos o palabras distintas y su frecuencia, así como otra serie de cálculos.

A pesar de esta sofisticación en los programas informáticos, nosotros hemos utilizado otros más sencillos. Nos servimos del programa WP51 para la introducción de los textos. Para convertir éstos en un listado de palabras por orden alfabético, usamos la base de datos LEXI (Sabido, V. y Guirao, J. M., 1993). Posteriormente, los ficheros se asociaron a un programa de DBASE III PLUS, donde se sumaron las frecuencias parciales y se creó una nueva variable con la frecuencia total. Por último, utilizamos el programa Word para configurar la estructura definitiva del vocabulario de los escolares.

---

[292] García, E., Gil, J. y Rodríguez, G. (1994: 107-124) citan el programa GENERAL INQUIRER como inicio del análisis automático de contenido aplicado a los textos, además del propio programa ALCESTE, que es el que ellos manejan. Mochmann, E. (1985: 37) cita el programa WORDS, capaz no sólo de listar las palabras y su frecuencia, sino de suprimir las palabras funcionales y lematizar las palabras de contenido semántico, tras pasarlas a su forma radical.

[293] Mochmann, E. (1985: 18-24) se refiere al programa GENERAL INQUIRER, citado anteriormente.

2.13. Codificación

Tarea ingrata, aunque imprescindible, es la desambiguación de un texto para su posterior codificación: asignar a cada vocablo su clase gramatical correspondiente. Para ello se tienen que tener en cuenta previamente una serie de preceptos lingüísticos que constituyen la norma léxica. Teniendo ésta claramente definida, comienza la labor de búsqueda y demarcación de las ambigüedades: se deben localizar uno a uno los casos dudosos en su contexto y ponerles una marca que permita posteriormente distinguirlos cuando a través del proceso informático aparezca el listado de palabras. Posteriormente, se deberán reunir bajo el mismo lema todas aquellas formas flexivas que le correspondan.

Esta tarea, necesaria e imprescindible, es en exceso problemática y agotadora porque al no haber (o al menos nosotros no hemos encontrado) un programa informático de desambiguación gramatical y léxica, se debe acudir a modelos lingüísticos variados[294]. Además, son precisamente las palabras de mayor frecuencia, es decir, las llamadas gramaticales, las que presentan los panoramas más desoladores debido, entre otras cosas, a la multiplicidad de funciones que pueden desempeñar y la poca uniformidad que existe a la hora de su determinación. Por ejemplo, (y de acuerdo con nuestro criterio de distinguir sus distintas funciones) la forma *se* suma en total de más de seis mil ocurrencias, por lo que ha sido necesario en primer lugar delimitar muy bien sus funciones e ir buscando una por una en los textos para colocarle una

---

[294] Recordemos, por ejemplo, las palabras de Tournier, M. (1989:16), cuando refiriéndose a la tarea del recuento de las palabras de un texto afirma: "La première option ne permet pas l'étude de gros corpus, car elle exige comme préalable à tout traitement et en l'absence d'analyseurs morpho-syntaxiques automatiques, un énorme investissement du chercheur en pré-analyses de toutes sortes, décryptant pas à pas les énoncés et faisant appel pour chaque desambiguïsation à une multiplicité de modèles linguistiques et situationnels, dont certains d'ailleurs son loin d'être établis...".

marca que permitiera posteriormente ser identificada; lo mismo ocurre con algunas de las formas flexivas, como *lo, los, la, las*, muy numerosas, y que pueden pertenecer tanto al artículo como al pronombre.

2.14. Conclusión

2.14.1. Para establecer el vocabulario usual de los escolares hemos seguido una serie de etapas metodológicas indispensables: a) determinación del tema de nuestra investigación, b) objetivos, c) población y muestra, d) método de muestreo, e) consideraciones previas sobre la riqueza y locuacidad de los alumnos que integran los distintos estratos, f) corpus o universo léxico, g) recogida de datos, h) la norma léxica, i) codificación.

2.14.2. El tema de nuestra investigación es, pues, la determinación del vocabulario usual de los alumnos de primer curso de EE.MM. (BUP y FP) y del nivel equivalente a 3º de Enseñanza Secundaria Obligatoria (1º de Reforma). Trabajamos sobre muestras de expresión escrita. El vocabulario obtenido lo sometemos a un criterio mínimo de dispersión y frecuencia según las diferencias más significativas de los alumnos de los distintos estratos, para contrastarlo con el *FDSW* y el VOD y sugerir la lista de palabras que deben ser objeto de enseñanza-aprendizaje en el 2º Ciclo de la ESO. No pretendemos con ello fijar una lista cerrada de vocablos, sino precisar, de acuerdo con un criterio didáctico y sociológico, y siguiendo la opinión de lingüistas ilustres (López Morales, A. Quilis, G. Salvador...), aquel vocabulario que consideramos imprescindible para programar sistemáticamente la enseñanza de la lengua, sin que esto entre en contradicción con otros tipos de léxico, como el disponible o el técnico, que igualmente deben ser materia de aprendizaje.

2.14.3. Como primer objetivo, señalamos la determinación del vocabulario usual de los alumnos, paso importante para cumplir con el segundo: su aplicación para la enseñanza del español. En función de éste, nos hemos

propuesto también presentar un breve análisis de algunos aspectos del léxico de nuestros escolares.

2.14.4. La población la constituye el número de alumnos del nivel que estudiamos. Como es muy numerosa, hemos seleccionado una muestra representativa. Además, hemos registrado un número significativo de palabras para fijar el corpus. Conocida la población real, acudimos a tablas estadísticas para seleccionar la muestra: ésta quedó definitivamente fijada en 932 alumnos. Para seleccionar los alumnos que compondrían la muestra elegimos un muestreo estratificado de afijación proporcional y por conglomerados. Para aplicar este método elegimos el conglomerado o grupo como unidad muestral y para fijar la estratificación utilizamos un diseño factorial que incluye el nivel y tipo de enseñanza (1º de EE.MM.: BUP, FP y Reforma), el tipo de centro educativo (privado y público) y la zona (rural y urbana). En FP procuramos también que en la selección de la muestra hubiera una representación de las distintas ramas y especialidades. Esta estratificación responde al principio de representatividad, pero al mismo tiempo partíamos de consideraciones previas acerca de la locuacidad y riqueza léxica de los alumnos que integran los distintos estratos. Así, consideramos que los alumnos de BUP deberían tener una mayor competencia lingüística que los de otros niveles educativos, y que los de zona urbana superarían igualmente a los de zona rural. Entre los de centro público y los de centro privado pensamos que no deben existir diferencias significativas. Aunque en el sistema de muestreo no tuvimos en cuenta la variable sexo, alumnos y alumnas, al pensar que al aplicarlo seleccionaríamos una porción representativa de los dos, sí analizamos posteriormente los resultados obtenidos por unos y por otras. Por ello, establecimos la hipótesis siguiente: las alumnas deberían emitir mayor número de palabras pero la extensión de su vocabulario sería algo menor que la extensión del vocabulario de los alumnos, aunque las diferencias no deberían ser significativas.

2.14.5. Quizás uno de los mayores problemas por los que atraviesa una investigación de naturaleza cuantitativa sea el de la determinación del corpus o cantidad de palabras necesarias para efectuar un recuento con garantía. En el terreno de la Lingüística, a las críticas de Pierre Guiraud acerca de la extensión de las compilaciones los investigadores han respondido tan sólo parcialmente, proponiendo nuevos sistemas de muestreo y reuniendo compilaciones más pequeñas, pero con criterios de selección matemático-estadísticos muy precisos tanto para fijar N como V. En el terreno de la Psicología Evolutiva, Fernando Justicia propone, entre otras, una referencia interesante para fijar el corpus: que éste sea tal que con él se consiga un número de vocablos determinado. Nosotros seguimos la vía sugerida por el investigador granadino y decidimos fijar un corpus superior a 400.000 palabras para obtener así alrededor de 9.500 vocablos, que es la amplitud del vocabulario fijado por algunos investigadores (Averril, García Hoz) para jóvenes de 14 años. En total, el corpus real que hemos recopilado asciende a 412.433 palabras y 10.533 vocablos, cifra esta última quizás elevada debido a nuestra norma de despojo.

2.14.6. Para la recogida de datos, pedimos a cada alumno de la muestra dos composiciones: una descripción y una narración. Obtuvimos así 1864 composiciones escritas, que fuimos introduciendo en el ordenador para el posterior tratamiento informático de los datos y para establecer el vocabulario de los escolares, de acuerdo con la norma léxica de nuestra investigación.

2.14.7. La norma léxica de nuestra investigación quedó definida en síntesis de la siguiente manera: elegimos la palabra entendida como unidad gráfica para establecer tanto N como V. Incluimos las palabras léxicas y las palabras gramaticales. Resolvimos los problemas que presenta la palabra como unidad de texto separando los contractos y los pronombres enclíticos, que pasaron a engrosar sus clases gramaticales correspondientes. No contemplamos unidades fraseológicas en la confección del vocabulario, por lo que no influyen

para el establecimiento final de N y de V (sí ofrecemos una lista de estas unidades en otro apartado del trabajo). Para fijar los lemas tuvimos en cuenta consideraciones de naturaleza fonética, gramatical y léxica. El nivel fonético nos sirvió para conceder entradas distintas cuando las palabras se distinguen por la tilde y pertenecen a clases gramaticales diferentes. Las decisiones de naturaleza léxica no fueron en exceso complicadas, salvo el de la homonimia, que optamos por resolver acudiendo al criterio etimológico. Por lo demás, consideramos los compuestos con guión como un solo vocablo, porque, en el caso de los lexemas, suponen una primera fase de lexicalización y, en el caso de los prefijos, porque no son elementos autónomos. Los diminutivos, aumentativos y despectivos se incluyen como variante flexiva bajo el lema correspondiente, salvo que remitan a una realidad distinta a la de la base. De acuerdo con un criterio descriptivo, incluimos neologismos, préstamos, siglas. Las formas apocopadas figuran bajo el lema. Suprimimos de nuestro recuento los nombres propios, salvo que formen parte de una locución o los escolares los usen como nombres comunes. Fenómenos de orden semántico, como la polisemia, no tienen cabida en un recuento de estas características ya que se multiplicarían las entradas y las explicaciones. Frente a esta ausencia de información semántica contrasta la abundancia de información gramatical, ocasionada por nuestra decisión de atender a criterios formales y sintácticos, por lo que algunos vocablos pueden registrarse tres veces de acuerdo con su función de pronombres, adjetivos o adverbios. Seguimos así a dos de los trabajos más emblemáticos sobre el español: el *FDSW* y el *Léxico básico del español de Puerto Rico*. No obstante, a veces hacemos distinciones donde otros estudios no las han hecho, como en el caso de *uno*, numeral e indefinido. Con respecto a las clases nominales, el género sirve para registrar aparte a los sustantivos, no a los adjetivos, de acuerdo con la consideración lingüística de que el género es atributo del sustantivo. Además, este morfema nos sirve para separar aquellos nombres que tienen distinto significado según sean masculinos o femeninos. De la misma manera actuamos con el número. El grado superlativo de los adjetivos no sirve para otorgar nuevo lema, sino

que figura como variante flexiva del adjetivo positivo pertinente. Entre las clases pronominales conviene hacer algunas aclaraciones. Los pronombres personales presentan, a nuestro juicio, un solo punto conflictivo: el género y la persona. Si actuásemos como en el resto de los pronombres, tendríamos que registrar aparte *nosotros* y *nosotras*, *vosotros* y *vosotras*, ya que poseen género distinto, pero eso supondría atender al número plural y dar dos entradas a una misma persona gramatical. La decisión adoptada por nosotros en este caso ha sido actuar como los dos diccionarios anteriormente citados, y distinguir solamente entradas diferentes en los casos en los que la persona en singular presenta significantes distintos para el masculino, femenino y neutro, es decir, la tercera (*él*, *ella*, *ello*). En el caso del pronombre reflexivo, incluimos bajo el lema *él* el *se* personal, el propiamente reflexivo, el reflexivo medial y el recíproco. En cambio, y siguiendo opiniones de algunos lingüistas, registramos aparte el *se* impersonal y el pasivo. De los demás pronombres, recordaremos que algunas formas de los indefinidos se recogen varias veces de acuerdo con su función. El verbo presenta diversos casos de homonimia sintáctica que el contexto permite distinguir con claridad. No ocurre lo mismo con los participios, registrados como verbos sólo en aquellos casos en que se produce invariabilidad de género y número.

CAPÍTULO III: DATOS CUANTITATIVOS Y CUALITATIVOS

3. Datos cuantitativos y cualitativos de nuestra investigación

En primer lugar hemos de recordar que el número total de ocurrencias o N es de 412433, y los vocablos o V ascienden a un total de 10353.

A continuación aportamos los datos cuantitativos, N y V, correspondientes a los distintos estratos. Ofrecemos los datos cuantitativos correspondientes a agrupaciones realizadas en función de tres criterios:

a) Tipo de enseñanza, que recoge un bloque integrado por los alumnos de BUP -tanto de la enseñanza pública como de la enseñanza privada- y otro bloque integrado por los alumnos de FP, Reforma y LOGSE, igualmente de enseñanza pública y privada. En este primer apartado ofrecemos dos tablas. La primera contiene los datos referidos a los ocho estratos que sirven de base para el vocabulario usual, y que corresponden simultáneamente a las variables de nivel educativo, zona y sexo. El segundo cuadro ofrece los datos de los alumnos de BUP y de FP-Reforma-LOGSE, según las variables zona y sexo. Esta manera de operar permitir contrastar en primer lugar los datos cuantitativos de cada nivel de enseñanza, en sí mismos, comparando todas las variables de zona y sexo; y en segundo lugar, se podrán apreciar con claridad las diferencias entre los dos niveles de enseñanza.

b) Zona. El segundo criterio muestra los resultados de los alumnos pertenecientes a la zona urbana frente a los pertenecientes a la zona rural, según la variable sexo. Ofreceremos una tabla que sirve para completar los datos anteriores.

c) Totales. El tercer criterio recoge los datos totales en este caso correspondientes a las variables nivel educativo, zona y sexo. Ofrecemos una tabla con estos totales que servirán para asentar conclusiones más o menos definitivas según la variable, ya que los datos parciales se han podido apreciar en los dos apartados anteriores.

En todos los casos aludiremos al número de alumnos, al total de palabras, representado como N; al número de vocablos, representado como V, a la media (M) de palabras emitidas por el alumno (media que resulta de la división del número total de ocurrencias del estrato correspondiente con el número de alumnos), la frecuencia media (Fm) o resultado de la operación N/V, el vocabulario teórico (V'), calculado según la fórmula de Müller y la diferencia entre el vocabulario real y el teórico (V - V').

3.1. Locuacidad y riqueza léxica

Los datos anteriores los utilizaremos para contrastar los resultados obtenidos por los alumnos pertenecientes a cada estrato y comentar las diferencias con respecto a la locuacidad (entendida, como hemos dicho en otro lugar, como el número de palabras emitidas por un alumno o la media de palabras por grupo de alumnos) y a la riqueza léxica que existen entre ellos. Tomamos como base para el primer punto la media de palabras por alumno, y para la riqueza léxica, y debido a que no hay igual número de alumnos por estrato, aplicamos la fórmula de Ch. Müller (Cfr. 1.3.3.). Hemos de aclarar que de los condicionantes que influyen en la riqueza léxica (variedades estilísticas y temáticas, número de sujetos de la muestra) sólo puede influir realmente en

los resultados el mayor o menor número de alumnos de cada estrato, pues pensamos que los escolares han partido de las mismas circunstancias para componer sus textos, y la elección de tema libre tanto para la descripción como para la narración debe influir o no por igual en todos. Debido precisamente a que las muestras de los estratos no son iguales, las conclusiones no serán categóricas, sino relativas, pero pueden servir en algunos casos como indicadores significativos. De todas maneras, las mayores diferencias observadas entre los diferentes estratos nos servirán para, de acuerdo con la dispersión y la frecuencia, seleccionar y fijar una lista de palabras apropiadas para la enseñanza sistemática del vocabulario.

Antes de ofrecer los cuadros y gráficos con los datos reseñados más arriba, explicamos las abreviaturas correspondientes a cada estrato:

E1: Alumnas, nivel educativo FP-Reforma-LOGSE; zona urbana.
E2: Alumnas, nivel educativo FP-Reforma-LOGSE; zona rural.
E3: Alumnos, nivel educativo FP-Reforma-LOGSE; zona urbana.
E4: Alumnos, nivel educativo FP-Reforma-LOGSE; zona rural.
E5: Alumnas, nivel educativo BUP, zona urbana.
E6: Alumnas, nivel educativo BUP, zona rural.
E7: Alumnos, nivel educativo BUP, zona urbana.
E8: Alumnos, nivel educativo BUP, zona rural.

Estos estratos a veces los combinamos según las tres variables que analizamos: el nivel educativo, la zona y el sexo, por lo cual resultarán combinaciones como E1+E2, es decir, alumnas de FP-Reforma, o E4+E8, es decir, alumnos de zona rural, etc.

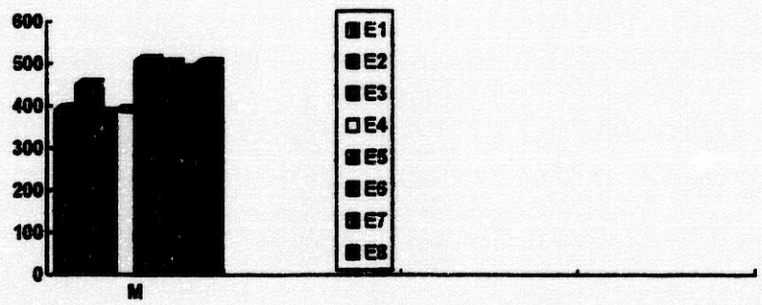
A) Tipo de enseñanza, zona y sexo: estratos individuales

Tabla 1

	Nº AL.	N	V	M	Fm	V'	V - V'
E1	106	41363	3348	390'21	12'35	3987	-639
E2	129	57969	3904	449'37	14'84	4686	-782
E3	166	63007	4742	379'56	13'28	4872	-130
E4	112	43290	3833	386'51	11'29	4077	-212
E5	96	48519	3865	505'40	12'55	4308	-433
E6	99	49284	3680	497'81	13'39	4340	-660
E7	162	78129	5786	482'27	13'50	5372	+418
E8	62	30872	3255	497'93	9'48	3446	-191

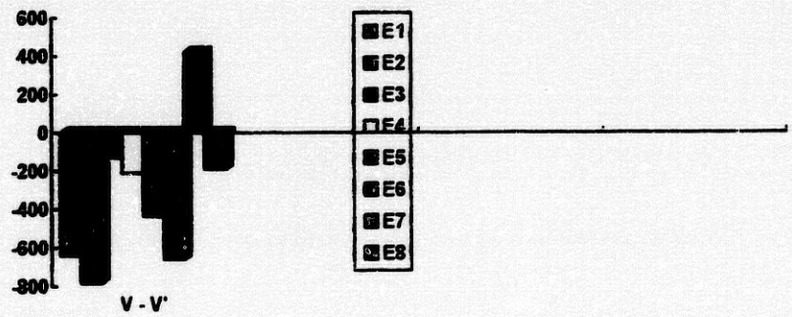
En los gráficos representamos, pues, la locuacidad o número de palabras emitidas por alumno (M) y la riqueza léxica, entendida como extensión del vocabulario. Para este último punto destacamos la diferencia (V - V') entre el vocabulario real y el vocabulario teórico calculado según la fórmula de Müller: $E(V') = V - \Sigma q^{i} V_{i}$.

Gráfico 1



Con respecto a la media de palabras por alumno, lo primero que apreciamos con claridad es que todos los alumnos de los distintos estratos de BUP emiten mayor número de palabras que los alumnos de FP-Reforma(LOGSE). En BUP, son las alumnas de la capital (E5) las que ofrecen una mayor locuacidad, pero todos los estratos están muy igualados, si bien los alumnos de zona urbana (E7) emiten el menor número de palabras por sujeto. En el otro nivel educativo, son las alumnas de zona rural las que obtienen el mayor número de palabras por sujeto encuestado, con una diferencia significativa con el resto de estratos de este nivel. Aquí también son los alumnos de zona urbana (E3) los que ofrecen una media (M) menor.

Gráfico 2



Para la riqueza léxica, entendida como mayor extensión de un vocabulario, hemos recurrido, como decíamos, a la fórmula de Müller: $(E(V') = V - \Sigma q V_1$, para calcular el vocabulario teórico (V') y contrastarlo con el vocabulario real (V), pues ninguno de los cuatro supuestos -a, b, c, d- (Cfr. 1.3.3.) aportados por el lingüista es válido para todos los vocabularios, debido a que los estratos no poseen el mismo número de alumnos y los datos son heterogéneos.

No obstante, acudiendo a la frecuencia media (Fm) se pueden anticipar algunas evidencias. Por ejemplo, en FP-Reforma, los alumnos de zona rural (E4) poseen un vocabulario más extenso que las alumnas de zona urbana (E1), ya que con un número mayor de N, la frecuencia media es más baja. Igual ocurre con los alumnos de zona urbana (E3) con respecto a las alumnas de zona rural (E2): aquellos tienen una frecuencia media más baja a pesar de que el número de N es mayor. Todas estas apreciaciones se confirman con la diferencia entre V - V'.

Si comparamos V y V' de todos los estratos y observamos las desviaciones, se aprecia, en primer lugar, que los alumnos de BUP y zona urbana (E7) poseen el vocabulario más rico (único estrato que en el gráfico figura sobre la abscisa), seguidos por los alumnos de FP-Reforma y zona urbana (E3). Es cierto que el mayor número de alumnos de zona urbana en ambos niveles puede ser un factor determinante para la riqueza léxica, pero este factor no influye para el resto de los estratos. Es más, los alumnos de BUP y zona rural (E8), en menor número, poseen un vocabulario más rico que los 5 estratos restantes. En cuarto lugar se sitúan los alumnos de FP-Reforma de zona rural (E4); le siguen las alumnas de BUP y zona urbana (E5), en quinto lugar, y las alumnas de FP-Reforma de zona urbana (E1), en sexto. Las alumnas de BUP y zona rural (E6) irían posteriormente y en último lugar, las alumnas de FP-Reforma de zona rural (E2). Por tanto, y a tenor de nuestros datos, el número de alumnos influye relativamente, porque E7 posee, como hemos indicado, el vocabulario más rico, a pesar de que, por ejemplo, posee un número de alumnos algo inferior a E3; e incluso E8, con un número de alumnos muy inferior (62) a E3 (166), tiene una diferencia de V-V' muy poco significativa, lo que nos conduce a pensar que los alumnos de BUP de zona rural poseen un vocabulario también más rico que los alumnos de FP-Reforma de zona urbana (E3). La variante que de momento tiene más influencia en lo concerniente a la extensión del vocabulario es el sexo: los alumnos presentan una mayor riqueza léxica que las alumnas, pues los cuatro estratos de

alumnos superan a los cuatro de las alumnas. También influyen el nivel educativo y la zona, pero en menor proporción. No obstante, en las tablas y gráficos siguientes, donde ofrecemos las cifras de los estratos agrupados por nivel, zona y sexo, podremos apreciar mejor estas diferencias.

Si contrastamos la riqueza léxica con la locuacidad o media de palabras por alumnos, se observa que los dos estratos que obtienen la mayor riqueza de vocabulario (E7 y E3) son los que poseen una media (M) más baja. Además, las alumnas, emitiendo mayor número de palabras, poseen un vocabulario menos extenso que los alumnos. Estos datos coinciden con estudios precedentes. (Justicia, F.: 1985a y 1995).

B) BUP y FP-Reforma-LOGSE: variables zona y sexo

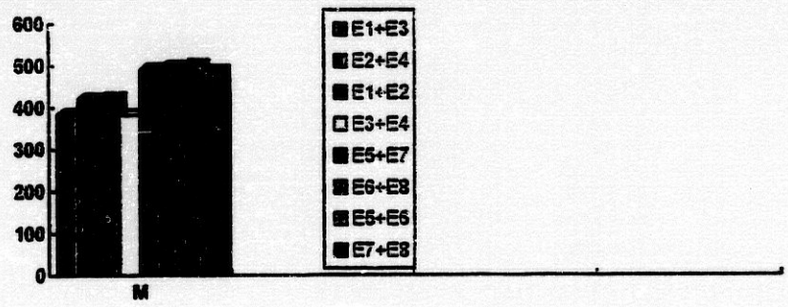
La tabla siguiente ofrece los datos agrupados en cada nivel educativo por las variables zona y sexo.

Tabla 2

		Nº AL.	N	V	Media	Fm	V'	V - V'
BUP	Z.urbana (E5+E7)	258	126648	6904	490'88	18'34	6617	+287
	Z.rural (E6+E8)	161	80156	4898	497'86	16'36	5675	-677
	Alumnas (E5+E6)	195	97803	5209	501'55	18'77	5930	-739
	Alumnos (E7+E8)	224	109001	6639	486'61	16'41	6212	+427

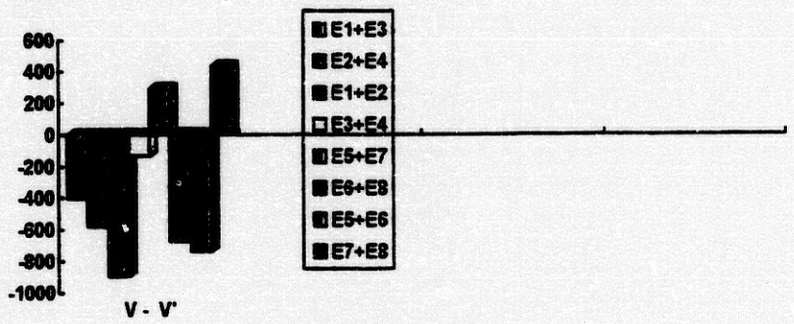
FP,RE,LO	Z.urbana (E1+E3)	272	104370	5694	383'71	18'32	6098	-404
	Z.rural (E2+E4)	241	101259	5440	420'16	18'61	6019	-579
	Alumnas (E1+E2)	235	99332	5073	422'68	19'58	5970	-897
	Alumnos (E3+E4)	278	106297	6011	382'36	17'68	6146	-135

Gráfico 3



En cuanto a la media de palabras por alumno, observamos que las características anteriores se mantienen, en el sentido de que los alumnos de BUP poseen una mayor locuacidad que los alumnos de FP-Reforma. Tanto en un nivel como en otro, son las alumnas (E1+E2 y E5+E6) las que producen mayor número de palabras, y los alumnos (E3+E4 y E7+E8) los que menos. Estos datos contrastan de nuevo con la riqueza léxica.

Gráfico 4



Si observamos en primer lugar los resultados obtenidos en FP-Reforma, apreciamos que los alumnos (E3+E4) son los que poseen mayor riqueza léxica, seguidos por el alumnado de zona urbana (E1+E3), si bien puede contribuir a ello el mayor número de sujetos de la muestra en estos estratos. En tercer lugar se sitúa el alumnado de zona rural (E2+E4) y en último lugar las alumnas (E1+E2).

En BUP existe una correspondencia idéntica que en el nivel educativo anterior: son los alumnos (E7+E8) quienes poseen un vocabulario más extenso, a pesar de que, por ejemplo, el número de sujetos es menor que en la zona urbana (E5+E7), estrato que se sitúa en segundo lugar. En tercer lugar, el alumnado de zona rural y en último lugar, las alumnas.

Al contrastar los datos de ambos niveles educativos, destacamos en primer lugar lo siguiente: si establecemos una relación paralela por variante zona y variante sexo, apreciamos que el alumnado de BUP, a pesar de tener un número menor de sujetos, supera al alumnado correspondiente de FP-Reforma; es decir, el alumnado de BUP de zona urbana (E5+E7), con 258 sujetos, supera al alumnado de FP-Reforma de zona urbana (E1+E3) -que cuenta con 272-; el alumnado de BUP y zona rural supera al alumnado de FP-Reforma de zona

rural..., y así sucesivamente. En segundo lugar, llama la atención que son los alumnos de BUP (E7+E8) los que tienen mayor riqueza léxica, a pesar de que el número de sujetos es menor que el estrato de zona urbana del mismo nivel y que todos los estratos de FP-Reforma. A éste sigue el alumnado de BUP y zona urbana. Ambos estratos son los únicos que presentan la diferencia de V-V' por encima de la abscisa, lo cual indica que poseen un vocabulario real más rico que su vocabulario teórico correspondiente. A estos dos estratos siguen en riqueza léxica los alumnos de FP-Reforma (E3+E4) y el alumnado de zona urbana del mismo nivel educativo (E1+E3), pero las diferencias de V-V' es negativa, lo cual indica que el vocabulario teórico es mayor que su vocabulario real, a pesar de que en estos estratos (E3+E4 y E1+E3) poseen mayor sujetos que sus equivalentes en BUP. En penúltimo y último lugar se sitúan las alumnas de BUP y de FP-Reforma, respectivamente.

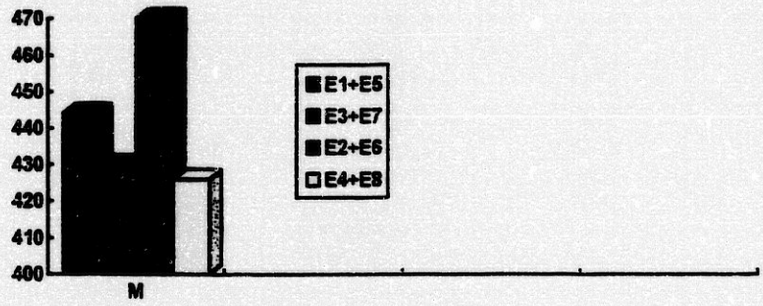
Total zona: variante sexo

En la tabla siguiente agrupamos los alumnos por niveles educativos para ver la incidencia que la variable zona ejerce sobre alumnos y alumnas. De este modo obtenemos un panorama más amplio.

Tabla 3

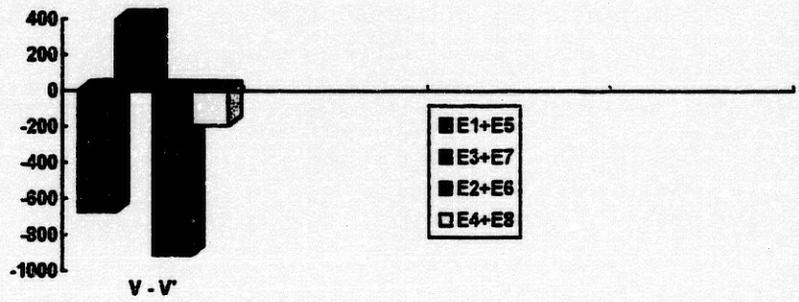
		Nº AL.	N	V	M	Fm	V'	V - V'
ZONA URBANA	Alumnas (E1+E5)	202	89882	5043	444'96	17'82	5716	-673
	Alumnos (E3+E7)	328	141136	7311	430'29	19'30	6919	+392
ZONA RURAL	Alumnas (E2+E6)	228	107253	5255	470'40	20'40	6169	-914
	Alumnos (E4+E8)	174	74162	5054	426'21	14'67	5247	-193

Gráfico 5



En cuanto a la media de palabras por alumnos (M), se observa que son las alumnas de zona rural (E2+E6) las que muestran mayor locuacidad, seguidas por las alumnas de zona urbana (E1+E5). Los alumnos emiten menor número de palabras, sobre todo los de zona rural, (E4+E8) que se sitúan en último lugar.

Gráfico 6



Si miramos el gráfico 6, vemos que se invierte el orden con respecto a la locuacidad. Las alumnas de zona rural (E2+E6) son las que muestran una diferencia negativa mayor entre V-V', lo cual indica que su vocabulario es el

menos extenso. A éstas les siguen en cuanto a pobreza léxica las alumnas de zona urbana (E1+E5), a pesar de contar con menor número de sujetos. De nuevo son los alumnos los que muestran una mayor riqueza de vocabulario, sobre todo los de zona urbana (E3+E7), único estrato cuyo vocabulario se sitúa sobre la abscisa, es decir, que supera el vocabulario teórico, aunque también aquí podría influir el número de componentes de la muestra. Los alumnos de zona rural (E4+E8), con menor cantidad de sujetos, poseen igualmente un vocabulario más extenso que las alumnas.

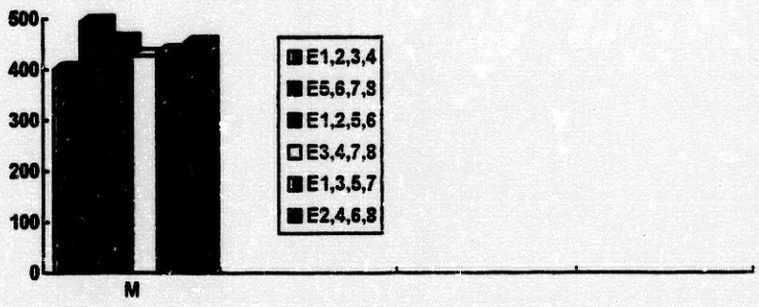
C)Total

A continuación ofrecemos una tabla con la suma total de todas las variantes. Podremos apreciar así la influencia que ejercen los estratos que venimos analizando en el vocabulario determinado en nuestra investigación. Con estos datos completamos de un modo exhaustivo todas las combinaciones que se producen de acuerdo con el nivel educativo, la zona y el sexo.

Tabla 4

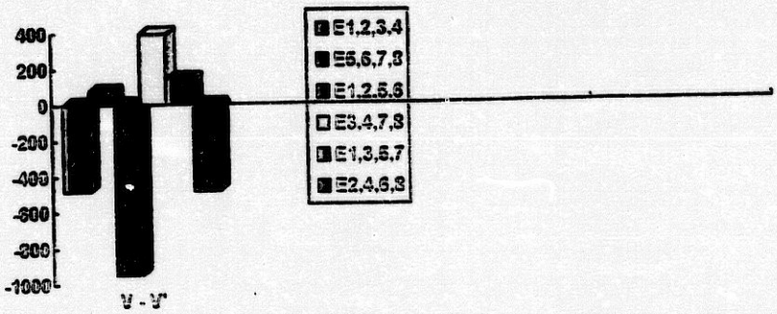
		Nº AL.	N	V	M	Fm	V'	V - V'
V. sexo	Alumnas (E1,2,5,6)	430	197135	6956	458'45	28'34	7905	-949
V. sexo	Alumnos (E3,4,7,8)	502	215298	8570	428'88	25'12	8178	+392
Z.urbana	(E1,3,5,7)	530	231018	8543	435'88	27'03	8400	+143
Z. rural	(E2,4,6,8)	402	181415	7051	451'28	25'72	7652	-601
FP,RE,LO	(E1,2,3,4,)	513	205629	7551	400'83	27'23	8035	-484
BUP	Total (E5,6,7,8)	419	206804	8138	493'56	25'41	8138	+76

Gráfico 7



En cuanto a la locuacidad, observamos que la variante con mayor influencia es el nivel educativo. El alumnado de BUP (E5,6,7,8) es el que emite mayor número de palabras por sujeto. En el polo opuesto se sitúa el alumnado de FP-Reforma (E1,2,3,4), con la media (M) más baja. Combinados ambos niveles educativos, la variable que tiene una incidencia mayor es el sexo: las alumnas (E1,2,5,6) emiten mayor número de palabras que los alumnos. Y según la zona, apreciamos que el alumnado de zona rural posee una locuacidad algo mayor que el alumnado de zona urbana.

Gráfico 8



Antes de comentar la riqueza léxica quisiéramos hacer algunas aclaraciones. El único factor que influye en la extensión del vocabulario y que puede desequilibrar los datos es la desigualdad de alumnos de los estratos. Pero al ofrecer los totales, el número de alumnos y de N es lo suficientemente elevado como para que todos los vocabularios reales fuesen superiores a los vocabularios teóricos correspondientes, pues hemos visto en tablas anteriores que un estrato individual (E7) e incluso combinaciones como E5+E7, E7+E8, tienen un vocabulario más rico que el teórico.

Según la variable sexo, se aprecia claramente que los alumnos, a pesar de emitir una media de palabras por alumno menor, posee un vocabulario más extenso que las alumnas. Éstas, a pesar de contar con un número elevado de sujetos, poseen un vocabulario real mucho menor que el vocabulario teórico correspondiente.

Según la variable zona, el alumnado de zona urbana maneja un vocabulario más extenso que el de zona rural. Es cierto que la muestra de sujetos es mayor en la zona urbana, pero el alumnado de zona rural también se mueve con una muestra de sujetos lo suficientemente elevada como para haber obtenido un vocabulario real más alto que el teórico, pero no ha ocurrido así.

Según la variable nivel educativo los datos son categóricos: el alumnado de BUP, a pesar de contar con una muestra muy inferior -419- al alumnado de FP-Reforma -513-, supera al otro nivel educativo tanto en locuacidad como en riqueza léxica.

Si comparamos todos los totales, apreciamos en primer lugar que es la variable sexo la que tiene mayor incidencia en el vocabulario: los alumnos (E3,4,7,8) poseen el vocabulario más rico, a pesar de contar con menor número de sujetos que, por ejemplo, el alumnado de zona urbana (E1,3,5,7,), que se sitúa en segundo lugar. En tercer lugar se sitúa el nivel educativo, BUP, si bien es cierto que el número de sujetos de la muestra es muy inferior a las dos combinaciones precedentes. Los vocabularios de esas tres combinaciones se sitúan todos por encima de la abscisa, lo cual demuestra que poseen un vocabulario más extenso que el teórico. Las otras combinaciones restantes poseen un vocabulario menos rico que el teórico correspondiente, si bien son las alumnas (E1,2,5,6) las que muestran una desviación entre V-V' más profunda, a pesar de que en sujetos superan al alumnado de zona rural (E2,4,6,8).

3.1.1. Conclusión

El objeto esencial de esta investigación no es establecer un método infalible para contrastar, definir y establecer la riqueza léxica o la locuacidad de los alumnos. Si ofrecemos estos datos es porque nos interesa mostrar también características cuantitativas del vocabulario de los escolares así como la heterogeneidad de los distintos estratos que configuran el alumnado de la EE.MM., del que hemos obtenido una muestra representativa.

Por ello no hemos seleccionado un número igual de alumnos en cada caso, lo cual supone que no podamos ser categóricos en el comentario de los

datos, al menos en lo que a la riqueza léxica se refiere. Pero creemos que se observan, a tenor de los resultados, algunas tendencias interesantes que confirman de alguna manera las hipótesis que emitimos al seleccionar la muestra.

A pesar de estas aclaraciones, sí quisiéramos hacer una pequeña síntesis de aquellos puntos más significativos. Por ejemplo, y en lo que atañe a la media de palabras por alumno, las conclusiones pueden ser firmes porque en este aspecto no afecta nada el número de alumnos por estrato. En este sentido, hemos podido comprobar que el factor que más influye es el tipo de enseñanza: los alumnos de BUP emiten mayor número de palabras que los de FP-Reforma, y esta superioridad es bastante significativa. Por otro lado, esto no hace sino confirmar nuestras hipótesis, lógicas, porque como indicamos, existe una característica académica importante que distingue los dos tipos de enseñanza: para FP y Reforma no se necesitaba el Graduado Escolar. En cuanto a la variante zona, no existen diferencias significativas, pero, en general, son los alumnos de zona rural los que emiten mayor número de palabras por sujeto. Salvo en el caso de las alumnas de BUP y zona urbana (ES), que son las que muestran mayor locuacidad, el alumnado de zona rural supera siempre al de zona urbana en este punto. Esto no se corresponde con nuestra hipótesis inicial, pero ya advertimos que el alumnado de zona rural de EE.MM. vive en localidades grandes y muchas de ellas son auténticos enclaves culturales y económicos. Según la variable sexo, las alumnas emiten mayor número de palabras que los alumnos: en todos los casos en que hemos contrastado estos datos según el sexo, las alumnas superan en locuacidad a los alumnos, si bien las diferencias no son tan significativas como las que se producen entre el nivel educativo. Este dato coincide con estudios precedentes, donde se pone de manifiesto una mayor locuacidad de las niñas (Justicia, F., 1985a y 1995).

Los comentarios que hemos venido haciendo acerca de la riqueza léxica se basan, y lo volvemos a repetir, en la aplicación de la fórmula de Ch. Müller

con la que se obtiene un vocabulario teórico por lo general más extenso que el real. Comparando ambos vocabularios, real y teórico, se obtiene una desviación que ha sido la que nosotros hemos tenido en cuenta para formular nuestras observaciones. De nuevo queremos aludir a que puede influir en la riqueza léxica el mayor número de sujetos de los estratos, ya que la variedad de locutores es un factor determinante en la extensión de un vocabulario. Pero a veces hay estratos con menor número de sujetos que muestran una desviación entre V-V' más positiva que otros estratos con mayor número de alumnos. No obstante, y a pesar de ello, queremos ser cautos en las conclusiones, que en absoluto pretenden ser definitivas, sino que responden al deseo de informar de todos los aspectos posibles que rodean el vocabulario de la población escolar que estamos analizando. Desde esta perspectiva, sólo podemos ser rotundos en una conclusión: el alumnado de BUP posee una mayor riqueza léxica que el alumnado de FP-Reforma. A pesar de contar, como indicábamos, con una muestra de sujetos inferior, en BUP se obtiene un total de N mayor y un vocabulario mucho más rico que en FP-Reforma. Esta conclusión es cierto que no aporta nada nuevo acerca de lo que ya se conoce entre las diferencias de ambos niveles educativos. Pero confirma de un modo riguroso lo que muchas veces se afirma intuitivamente.

Si bien sólo en esta conclusión podemos ser rotundos, también podemos comentar otras tendencias que se esbozan en las tablas analizadas, sin llegar a la exactitud del primer aserto. Por ejemplo, y según la variable zona, desde los estratos individuales se aprecia que el alumnado de zona urbana posee un vocabulario más rico que el de zona rural. Es cierto que, en total, la muestra de sujetos en la zona urbana es superior, pero en algunas ocasiones, hemos contrastado estratos individuales, como E1 y E2, y combinados, como E1+E5 y E2+E6, en donde se aprecia que a pesar de que los estratos de zona urbana (E1 y E1+E5) cuentan con una muestra de sujetos menor que sus equivalentes en zona rural, la desviación de V-V' es más pequeña. Además, al final, en la tabla 4, que muestra los totales, se aprecia que el vocabulario real de la zona

rural es muy inferior a su teórico correspondiente, a pesar de que el total de sujetos de la muestra es bastante elevado, lo cual tendría que contribuir a la extensión del vocabulario. Este dato acerca de la zona rural resulta llamativo si establecemos otra comparación: el total de zona rural (E2,4,6,8), con 402 sujetos, tiene un vocabulario real de 7051 vocablos, mientras que , según la variante sexo, los alumnos de zona urbana (E3+E7), con 328 sujetos, y un total de N menor, tienen un vocabulario de 7311 vocablos, por tanto, más rico. Todo ello nos mueve a concluir que el alumnado de zona urbana posee un vocabulario más variado y extenso que el de zona rural, a pesar de que éste último ha sido seleccionado de todas las zonas de la provincia de Granada, zonas tan distantes como Huéscar y Almuñécar, Ugíjar y Alhama, etc.

La variable sexo es la que ofrece, en cuanto a la riqueza léxica, las diferencias más significativas. Los datos siempre muestran que los alumnos manejan un vocabulario más amplio que las alumnas. Es cierto que en total, el número de alumnas es menor, pero a pesar de ello, hemos de hacer algunas observaciones. Desde la primera tabla en la que se ofrecen los estratos individuales, es la variable sexo la que muestra las mayores diferencias: las alumnas, en general, obtienen un vocabulario real mucho más bajo que el teórico correspondiente. Esa tendencia ha ido aumentando a medida que íbamos combinando los estratos hasta llegar a la mayor desviación de V-V' para las alumnas: -949. En todos los demás estratos, las diferencias se han ido más o menos neutralizando. En el caso de las alumnas, el número de sujetos no ha servido para amortiguar la poca extensión que desde el principio mostraba su vocabulario. Es más, el total de alumnas (E1,2,5,6), a pesar de contar con mayor número de sujetos que el total de zona rural (E2,4,6,8) y mayor número de N, tiene un vocabulario inferior: 6956 de las alumnas, frente a 7051 de zona rural. Es decir, que a pesar de su mayor locuacidad, las alumnas presentan un vocabulario menos extenso. Por contra, el total de alumnos (E3,4,7,8) es el que muestra el vocabulario más amplio; más incluso que el total de zona urbana, a pesar contar con menor número de sujetos. A

tenor pues de todo lo que venimos comentando, hemos de afirmar que los alumnos poseen un vocabulario más rico que las alumnas, a pesar de que éstas emiten mayor número de palabras. Cuando hablamos de vocabulario más rico nos referimos a que es más extenso y variado. A nuestro juicio esta diferencia tan significativo a favor de los alumnos se debe a que éstos posiblemente utilicen más temas y más variados en sus conversaciones, ya que, para nosotros, son las variedades temáticas las únicas que influyen en las diferencias entre el vocabulario de los alumnos y el de las alumnas. En estos datos no se reflejan aspectos tales como la capacidad de comprensión, inteligencia, etc., e incluso aspectos cualitativos importantes en lo que al uso de la lengua se refiere, como riqueza gramatical, corrección ortográfica, coherencia en el discurso, etc. Los datos que aportamos sólo dan constancia de una cosa: la extensión del vocabulario; es decir, que unos estratos poseen un vocabulario más amplio que otros.

3.2. La distribución de frecuencias de nuestro recuento. Ley de Zipf

La distribución de frecuencias sirve para conocer la estructura cuantitativa de un vocabulario. Está constituida por los efectivos que integran las distintas clases de frecuencia, ordenadas éstas habitualmente de forma decreciente[295] (Cfr. 1.3.4.). Teniendo en cuenta que las clases de frecuencias son muy numerosas, ofreceremos tres cuadros. El primero representa los 10 vocablos de mayor rango, con su frecuencia relativa -la obtenida en cada uno de los 8 estratos- y la frecuencia absoluta (ET), por lo que esta ordenación responde más a la conocida ley de Zipf. El segundo cuadro sí es propiamente de

---

[295] Dugast, D. (1981: 101) define así la distribución de frecuencias: "On appelle *distributions des fréquences* la suite des nombres qui, pour un texte donné, présente les effectifs des vocables pour chacune des *classes de fréquences* successives, habituellement rangées dans l'ordre décroissant, de la plus élevée à la plus basse".

distribución de frecuencias y en él ofreceremos los datos de las últimas clases de frecuencias, es decir, el número de vocablos de frecuencia 1 hasta los de frecuencia 20, con el número de efectivos, para observar así ciertas características que presentan los recuentos. Esta manera de proceder es habitual cuando se ofrecen datos cuantitativos[296]. El tercer cuadro representa todas las clases de frecuencias agrupadas en 10 grupos de 1000 vocablos y uno de 353.

Los dos cuadros nos sirven para corroborar una serie de constantes que se producen en cualquier distribución de frecuencias.

Cuadro 1: Rango, type, frecuencia relativa y frecuencia absoluta

R.	TYPE	E1	E2	E3	E4	E5	E6	E7	E8	ET
1	EL ar.	3301	4559	5304	3904	3771	3731	6725	2449	33744
2	DE	1841	2604	3092	2159	2177	1982	3824	1346	19025
3	Y	1768	2287	2624	1661	1791	1930	2918	1335	16314
4	A	1370	1988	2110	1342	1536	1661	2314	1106	13427
5	UN	1225	1622	2130	1227	1257	1260	2427	998	12146
6	SER	1152	1498	1497	1052	1327	1311	2045	747	10629
7	ÉL pr.	1067	1566	1306	1180	1162	1287	1622	751	9941
8	EN	767	1171	1287	1061	1003	896	1804	600	8598
9	QUE c.	659	1042	1069	745	992	1032	1345	515	7399
10	QUE p.	631	853	949	699	786	766	1280	433	6367

En primer lugar, si atribuimos el rango 1 al vocablo más frecuente (en este caso, el artículo *EL*, con un total de 33744 ocurrencias), el rango 2 al

[296] Véase el cuadro de distribución de frecuencias que ofrece Müller, Ch. (1973a: 276-277).

siguiente y así sucesivamente, se observa que estos vocablos que tienen la mayor frecuencia están integrados por un sólo efectivo: es decir, el número de los efectivos disminuye cuando crece la frecuencia. Además, entre estos vocablos hay frecuencias que no están representadas (si observamos la frecuencia absoluta, se aprecia que de la frecuencia 33744 se pasa a la frecuencia 19025, etc.). Esto no ocurre en la parte final, donde es normal encontrar vocablos de frecuencia 1, 2, 3, 4, 5, etc.

Contrastando los dos cuadros, se puede apreciar también que el número de efectivos varía en sentido inverso de la frecuencia; es decir, a menor frecuencia, más efectivos.

Cuadro 2: distribución de frecuencias : vocablos de menor frecuencia

Además del número de vocablos de las últimas 20 clases de frecuencias, recogemos en este cuadro el número total de palabras o N de cada estrato y el número de vocablos de vocablos (V).

	E1	E2	E3	E4	E5	E6	E7	E8	ET
N	41363	57969	63007	43290	48519	49284	78129	30872	412433
V	3348	3904	4742	3833	3865	3680	5786	3255	10353
V1	1477	1654	2070	1743	1780	1601	2561	1591	3564
V2	547	643	789	650	559	613	907	516	1447
V3	294	307	401	329	320	299	496	279	876
V4	175	211	240	191	207	213	303	159	574
V5	118	171	186	147	155	135	223	113	435
V6	86	92	145	106	109	86	176	71	280
V7	60	90	85	81	93	87	133	58	266
V8	57	67	83	48	73	66	95	52	213

V9	52	63	68	54	50	64	76	33	180
V10	40	60	55	34	40	41	69	32	166
V11	28	32	48	32	31	36	53	33	138
V12	27	35	42	34	30	24	45	19	129
V13	29	31	32	34	22	23	61	21	110
V14	19	30	34	16	17	25	40	23	84
V15	23	22	24	18	15	20	31	12	88
V16	13	17	19	13	16	15	24	13	68
V17	16	26	18	22	22	16	27	19	76
V18	14	14	17	12	15	7	19	11	52
V19	15	16	16	13	16	8	19	10	55
V20	7	12	15	11	16	10	21	7	52
V>20	251	311	337	245	279	291	406	183	1501

En una distribución de frecuencias se establecen las siguientes relaciones:

Podemos apreciar, en primer lugar, que los vocablos de frecuencia 1 son los más numerosos[297]: el efectivo de vocablos empleados una vez es superior al de empleados dos veces, que este es superior al de tres veces, y así sucesivamente (Müller, Ch., 1965a: 35). Pero las clases de frecuencias medias presentan mayores irregularidades, pues puede haber alguna clase con un efectivo menor que la clase inmediatamente superior[298]. En el cuadro que

---

[297] En el *Trésor de la Langue française*, los hápax ascienden a 21.181, de un total de 71.415 vocablos (Imbs, P., 1971, Vol. II: VII).

[298] Müller, Ch. (1973a: 274) toma como referencia la frecuencia 20: "cuando los efectivos llegan a ser débiles, prácticamente a partir de los alrededores de 20, se producen irregularidades; se encuentra una frecuencia que tiene un efectivo más elevado que la frecuencia inferior; pero es visible que estas irregularidades realzan las variaciones aleatorias, y no acusan la ley general. Se puede por consiguiente afirmar con seguridad que si en un texto o un corpus hay 100 vocablos de frecuencia f, hay allí más de 100

ofrecemos, se observa, en el estrato total (ET), que las irregularidades se producen alrededor de V15 (vocablos de frecuencia 15): sus efectivos (88) superan a los de frecuencia 14 (84); lo mismo ocurre con los de frecuencia 17 y 19, en mayor número que los de 16 y 18 respectivamente. Algo parecido observamos entre los efectivos de distinta frecuencia que integran los demás estratos (E1, E2,...). Los hápax son algo más del doble de las palabras de frecuencia 2.

Cuadro 3: Distribución de las clases de frecuencia

RANGO	CLASE FREC.	FRECUENCIA	PORC. (%)	FREC. MEDIA
1	1 - 1000	364534	88'38	364'53
2	1001 - 2000	21632	5'24	21'63
3	2001 - 3000	9784	2'37	9'78
4	3001 - 4000	5529	1'34	5'52
5	4001 - 5000	3467	0'84	3'46
6	5001 - 6000	2343	0'56	2'34
7	6001 - 7000	1790	0'43	1'79
8	7001 - 8000	1000	0'24	1'00
9	8001 - 9000	1000	0'24	1'00
10	9001 - 10000	1000	0'24	1'00
11	10001 - 10353	353	0'08	1'00
TOTAL	10353	412433	100	39'82

Como decíamos en otro lugar (Cfr. 1.3.4.1.), la ley de Zipf establece que FxR = C. Si agrupáramos todas las clases de frecuencia y le asignásemos a

---

de frecuencia f-1, y menos de 100 de frecuencia f + 1". Para Dugast, D. (1981: 101), las irregularidades de las clases de frecuencias medias "se compensent dans un sens et dans l'autre, si bien qu'un schéma quasi régulier apparaît auquel les écarts relevés semblent bien tenir du hasard -hasard des distributions, des choix de l'ecrivain, de la norme de définition du mot et du vocable, et des erreur éventuelles de comptage-".

cada una de ellas un rango, se podría observar que esta ley se cumple efectivamente con algunos matices.

Pero para facilitar la representación, y siguiendo a F. Justica (1995: 154-156), hemos hecho un uso más genérico de este concepto, dividiendo así los 10353 vocablos del Vocabulario Usual (listado en orden decreciente de frecuencia) en grupos -o clase de frecuencias- de 1000 vocablos cada uno , a excepción del último que cuenta con 353. Obtenemos así 11 grupos, para los que hemos calculado el número de ocurrencias (frecuencia total), el porcentaje y la frecuencia media.

Con estos datos de nuestro trabajo podemos establecer comparaciones con otros que establecen las leyes estadísticas del vocabulario, las que han podido constatar una serie de hechos importantes.

Por ejemplo, la ley de Zipf, entre otros aspectos, deja constancia de que normalmente los 50 primeros vocablos por orden de frecuencia suelen ocupar el 50% de las ocurrencias. Otros datos cuantitativos que se manejan en relación a los trabajos de frecuencias son: los 100 vocablos[299] en orden decreciente suponen el 60%, los primeros 1000, el 85%, y los primeros 4000, el 97'5%, lo cual demuestra la intuición de que gran parte de las palabras de una lengua tienen un uso muy raro[300]. Esto, para Guiraud (1960: 94), no quiere decir que se identifique utilidad con frecuencia, pues aplicando la noción de

---

[299] En nuestro trabajo, los 50 vocablos de mayor frecuencia suponen exactamente el 55'57% de las ocurrencias, y los 100 el 65'18%.

[300] Véase Marcus, Nicolau y Stati (1978: 235); Guiraud, P. (1960: 17 y 93). No obstante, los datos no son "exactamente" coincidentes, pero sí muy cercanos, con recuentos posteriores, como *El léxico básico del español de Puerto rico*, (Morales, A., 1986: 32-33), en donde las cien palabras primeras tanto de uso como de frecuencia ocupan el 53%, y las primeras 500 el 68'31%.

información y su fórmula para calcularla, se apreciaría que las 100 primeras palabras comprenden el 30% de información, las 1000 primeras el 50% y las 4000 mil el 70%, lo cual conduce a una justa proporción de utilidad de un vocabulario de base, porque las palabras menos comunes, llevan mucha información, más "unités binaires d'information" que las frecuentes. Pero éstas, en mayor número, comportan más información en el conjunto que la palabra inusual.

Asimismo, se aprecia que las palabras de mayor frecuencia son las más cortas[301], las más antiguas, las más simples morfológicamente y las más extensas semánticamente, es decir, que poseen mayor número de acepciones[302] (según Zipf, las distintas acepciones de una palabra tienden a ser proporcionales a la raíz cuadrada de su frecuencia relativa[303]). Interpretaciones del hecho de que las palabras más cortas sean las más frecuentes nos las proporcionan H. Hörmann (1973: 127): "A quien no acepte la idea de que la longitud de las palabras empleadas está determinada de un modo puramente

---

[301] Guiraud, P. (1960: 30) se refiere al porcentaje de palabras según el número de sílabas, extraído del recuento de Kaeding. En este porcentaje, las palabras de una y dos sílabas ocupan más de ocho millones y medio de ocurrencias. Evidentemente, esto es posible porque las palabras más usuales son las llamadas palabras de relación. Casi todos los trabajos de frecuencias nos aseguran este aserto. Por ejemplo, Rodríguez Bou suprime las palabras más usuales de su lista, que son preposiciones, conjunciones, pronombres, etc. Otros testimonios los aportan Engwall, G. (1978: 86); Pastora Herrero, J. F. (1990: 57). En el caso de nuestro vocabulario, puede consultarse el cuadro 1 del presente apartado.

[302] Véase Marcus, Nicolau y Stati (1978: 65). Tal aserto lo ha rastreado Bustos Tovar (1966-7: 183-184), en el *Frecuency Dictionary...*, diccionario en el que se encuentran sustantivos polisémicos como *tiempo*, *cosa*, *mundo*, *casa*, *parte*, *pueblo*, *obra*, con un mayor índice de frecuencia, frente a otros sustantivos que aparecen en último lugar, de significación más restringida, como *aeroplano*, *aguardiente*, *almirante*, *bahía*, *balanza*, *boceto*...

[303] Los datos los tomamos de Marcus, Nicolau y Stati (1978: 49, 66).

casual, puede proponérsele otra reflexión: como el logaritmo de la probabilidad de aparición de una palabra expresa el contenido informativo de la misma, el número de fonemas de una palabra será proporcional a dicho contenido informativo (las palabras con muchos fonemas son raras, y tienen, por tanto, un mayor contenido informativo). Si en un texto se utilizaran las palabras frecuentes aún con más frecuencia (y las raras aún más raramente), el texto resultaría más corto, porque las palabras frecuentes son las cortas. Un texto más corto exige menos esfuerzo, pero su contenido informativo es menor en general. A la inversa, si se emplean sobre todo palabras raras y largas, aumenta la longitud y, sin duda, también el contenido informativo, pero el esfuerzo es mayor. La fórmula hallada por Zipf describe, por tanto, un equilibrio *observado también con arreglo a los principios de la teoría de la información:* aquel estado en que la transmisión de noticias muestra un óptimo económico". J. Lyons (1975:91) asegura que "las lenguas, cuando evolucionan a lo largo del tiempo y se <<transforman>> para satisfacer los cambiantes requerimientos de las sociedades que las emplean para comunicarse, pueden considerarse sistemas *homeostáticos* (o <<auto-reguladores>>), de modo que cada uno de sus estados en un momento cualquiera se <<regula>> por dos principios opuestos": el llamado principio del mínimo esfuerzo, según el cual se trata de conseguir la máxima eficacia comunicativa con el mínimo esfuerzo, y el segundo principio, el afán de ser comprendido. Zipf llamaba a estos dos fenómenos tendencia a la unificación y tendencia a la diversificación; la curva de Zipf sería la expresión del equilibrio de ambas tendencias.

Sobre la antigüedad de las palabras hallamos algún testimonio como el de J. M. Blecua[304], quien afirma que "el conjunto de las palabras de la lengua tiene tres fuentes básicas: a) *Voces patrimoniales* (las palabras procedentes del latín y de las lenguas primitivas); b) *los préstamos* de otras lenguas, y c) las

---

[304] La cita de Blecua la recoge Pastora Herrero, J. F. (1990: 57).

palabras de nueva *creación* que la lengua crea con sus medios estructurales propios. Las tres clases conviven armoniosamente en la proporción que puede estudiarse en la figura 1, aunque varía sensiblemente la frecuencia de uso que los hablantes hacen de ella (figura 2). Los datos proceden de los trabajos de William T. Patterson".

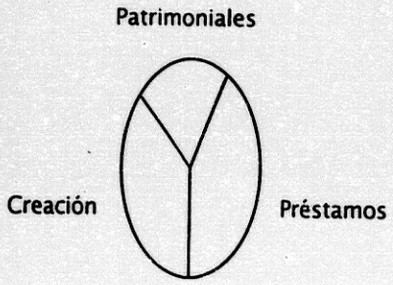


Figura 1

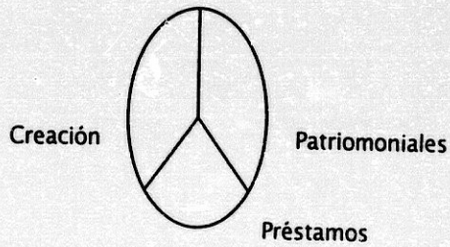


Figura 2

3.2.1. Conclusión

Nuestro recuento refleja las notas esenciales establecidas para cualquier distribución de frecuencias:

a) Los vocablos de mayor frecuencia están integrados por un solo efectivo.

b) A menor frecuencia, mayor número de efectivos, o lo que es lo mismo, la importancia de los efectivos varía en sentido inverso a la frecuencia.

c) Entre los primeros vocablos hay frecuencias que no están representadas, lo cual no ocurre entre los vocablos de frecuencia 1, 2, 3, etc.

d) Los vocablos de frecuencia 1 son los más numerosos.

e) Los efectivos de frecuencia 2 superan a los de frecuencia 3; a su vez, los de frecuencia 3 superan a los de frecuencia 4..., y así sucesivamente. Esta tendencia no es siempre uniforme. Por ejemplo, en nuestro trabajo los vocablos de frecuencia 15 (88) superan a los de frecuencia 14 (84), o los de frecuencia 19 (15) a los de frecuencia 18 (52).

3.3. Otros datos cuantitativos y cualitativos. Clases gramaticales

A continuación ofreceremos varios cuadros que aluden a criterios gráficos (las letras y sus datos cuantitativos) y criterios gramaticales.

1. Cuadro con el número de vocablos y palabras por letra inicial

LETRA	VOCABLOS	PALABRAS
A	1131	31911
B	334	6596
C	1376	31952
D	720	35006
E	808	62935
F	349	5831
G	236	5831
H	242	11946
I	480	7047
J	92	1576
K	15	128
L	281	7284
M	644	21977
N	186	14230
Ñ	0	0
O	195	5512

P	1024	34017
Q	38	16287
R	610	6015
S	611	31234
T	520	17595
U	55	13477
V	291	9273
W	3	4
X	2	4
Y	15	24010
Z	37	227

2. Cuadro con las clases gramaticales y sus frecuencias

CLASE GRAM.	V	N	% V	% N
ADJETIVOS	2780	3232	26'84	15'33
ADVERBIOS	338	25837	3'26	6'26
ARTÍCULOS	1	33744	0'009	8'18
CONJUNCIONES	16	35562	0'15	8'62
INTERJECCIONES	20	150	0'19	0'03
PREPOSICIONES	18	52929	0'17	12'83
PRONOMBRES	139	43757	1'34	10'60
SUSTANTIVOS	5413	80006	62'28	19'39
VERBOS	1628	77216	15'72	18'89

J. F. Pastora Herrero (1990: 58-59) recoge las cifras que aportan Peytard y Genouvrier con respecto al léxico individual medio, distribuidas según la clase gramatical, en primer lugar, y de acuerdo con la repetición de vocablos, en segundo lugar: 1. Sustantivos, el 62'5%; adjetivos, el 19%; verbos, el 15%;

formación de palabras, otros de naturaleza etimológica y, en fin, daremos cuenta de datos significativos de naturaleza lexicográfica (vocablos que no se incluyen en los diccionarios) y de nivel de uso, el cual nos servirá, entre otras cosas, para sacar conclusiones relativas acerca de si realmente existen diferencias importantes entre el léxico que emplean los jóvenes granadinos -en su expresión escrita- con respecto al empleado por la población adulta.

Algunos de estos apartados se relacionan con aspectos significativos de los diccionarios de frecuencias. Por ejemplo, serán útiles para confirmar postulados de la ley de Zipf, como aquel que asevera que las palabras más frecuentes son las más antiguas, afirmación que corroboran otros estudiosos, como por ejemplo J. M. Blecua (Cfr. 3.2.) cuando expone la teoría sobre el porcentaje de palabras usuales según su origen o su formación. Igualmente, otras observaciones también nos servirán para justificar la inclusión de algunos vocablos en el listado final de nuestra investigación, a pesar de que no figuran en el DRAE o en diccionarios escolares de categoría lexicográfica probada; vocablos que nos ofrecen una visión interesante de, por ejemplo, los recursos de creación espontánea de los escolares. Y fundamentalmente, todos estos matices proporcionan un conocimiento, siquiera mínimo, de aspectos cualitativos importantes del vocabulario usual de los alumnos; conocimiento que redundará en beneficio de la enseñanza sistemática del léxico.

Para poder establecer estos apartados, tomamos como punto de referencia el *Diccionario de la Lengua Española* de la RAE, en su edición de 1992. Y cuando existen dudas -vocablos de los escolares no incluidos en el DRAE, el origen de algunas voces, su formación...-, consultamos con el *Diccionario de uso del español*, de María Moliner, el *Diccionario crítico etimológico castellano e hispánico*, de J. Corominas, así como diversos diccionarios escolares: *Diccionario Anaya de la lengua* (1991), *Diccionario didáctico de español. Intermedio*, SM (3ª ed., 1994), *Diccionario esencial Santillana de la lengua española* (1ª ed., 5ª reimpresión, 1994) y el *Diccionario*

adverbios, el 3%; palabras útiles, el 0'5%; 2. en cambio, teniendo en cuenta las ocurrencias, los porcentajes cambian, y las palabras útiles ocupan el 50%; los sustantivos el 20%; los verbos el 17%; los adjetivos el 7'5% y los adverbios el 3%. Un trabajo de frecuencias como el de F. Justicia (1995: 45) nos proporciona los datos siguientes: 1. Calculamos[305], en primer lugar, las clases gramaticales tomando como referencia los vocablos: sustantivos, el 58'30%; adjetivos, el 20'55%; verbos, el 17'48%; adverbios, el 2'32% y palabras útiles (pronombres, preposiciones, conjunciones, interjecciones, artículos), el 1'32%; 2. basándonos en el número de ocurrencias, los porcentajes son estos: 34'25% de las ocurrencias para los sustantivos; 9'48% para el adjetivo; 4'73% para los adverbios; 16'71%, para los verbos y para las palabras útiles (pronombres, 7'19%, artículos, 10'62%, preposiciones, 9'28%, conjunciones, 7'65%, interjecciones, 00'43%), el 34'80%.

De cualquier modo, los datos que ofrecen los distintos trabajos dependen mucho de la norma léxica que se haya seguido, lo cual puede inclinar la balanza a favor de una determinada clase gramatical.

3.4. Breve análisis del vocabulario de los alumnos.

Hacer un análisis exhaustivo, desde todos los puntos de vista lingüísticos, del vocabulario que manejan los escolares de Enseñanza Media sería entrar de lleno en otra investigación, lo cual alargaría el trabajo en exceso. De ahí que no sea este el momento ni el lugar apropiado para emprender tan laboriosa empresa. No obstante, para que la consulta de nuestro vocabulario resulte más ilustrativa y se aborde con mayor entusiasmo, nosotros destacaremos algunos aspectos significativos: unos léxicos, de índole formal, como la

---

[305] Justicia, F. (1995: 45) ofrece el número de types y tokens para las distintas clases gramaticales, y con esos datos, nosotros hemos calculado los porcentajes.

Salamanca de la lengua española (1996) y otros diccionarios de idiomas, como el *Collins Concise Español-Inglés English-Spanish* (2ª ed., 1993)..

De acuerdo con estas indicaciones, pasamos a ofrecer los datos relativos al origen de las palabras y a su formación. Hemos seleccionado estos dos apartados porque podremos ofrecer resultados exhaustivos, tanto en types como en tokens.

3.4.1. Origen

	V	N	% V	% N
GRIEGO	680	6342	6'56	1'53
GRIEGO y PRÉSTAMO	8	30	0'07	0'007
GRIEGO Y LATÍN	36	316	0'34	0'07
LATÍN	7722	374054	74'57	90'69
LATÍN y PRÉSTAMO	5	7	0'04	0'001
ORIGEN EXPRESIVO	169	2023	1'63%	0'49
ORIGEN INCIERTO	306	2918	2'95%	0'70
O. PRERROMANO	47	1388	0'45%	0'33
PRÉSTAMOS	1382	25355	13'34	6'14

En los datos ofrecidos de la tabla anterior se tienen en cuenta sólo los lexemas. Es decir, algunas palabras se han formado con el lexema de origen latino y un sufijo o prefijo de origen griego, por ejemplo, pero las contabilizamos dentro del apartado LATÍN, etc. Los compuestos de GRIEGO y PRÉSTAMO o de LATÍN y PRÉSTAMO pueden responder también al orden inverso (préstamo y griego o préstamo y latín). Dentro de los PRÉSTAMOS, incluimos tanto las palabras de origen germano, árabe, arameo... como los indoamericanismos y los términos procedentes de lenguas modernas, incluidas las lenguas románicas de nuestro país: gallego y catalán . El mayor volumen dentro de este apartado lo constituyen, con cierta diferencia, los

galicismos. Le siguen arabismos, anglicismos, germanismos -entre los que incluimos tanto los términos de los antiguos pueblos germanos como del alto alemán y alemán-, italianismos, palabras procedentes de las otras lenguas románicas de nuestra península (gallego, catalán, portugués) y de otros múltiples idiomas (caló, coreano, checo, chino, escandinavo, esquimal, finlandés, hebreo, hindú, japonés, magiar, malayo, neerlandés, polinesio, ruso, tártaro, turco...). Dentro de las palabras de ORIGEN EXPRESIVO incluimos las que así son calificadas por el DRAE y las onomatopeyas. El calificativo ORIGEN INCIERTO responde a esta consideración del DRAE y a la de "etimología discutida". Estas suelen ser palabras muy arraigadas y conocidas en nuestra lengua. Para la mayoría de ellas, Corominas presupone un origen prerromano o latino, común a algunas lenguas románicas.

En cuanto a los datos cuantitativos, destacan las cifras relativas a las palabras pertenecientes al acervo latino: la mayoría de los vocablos derivan de la lengua madre. Los resultados son mucho más espectaculares si nos fijamos en el total de palabras: más del 90% de las ocurrencias tienen este origen. Todo esto confirma efectivamente que las palabras usuales son las más antiguas.

3.4.2. Formación de palabras: Derivación, composición, parasíntesis

Tampoco hay acuerdo absoluto entre los lingüistas a la hora de establecer los elementos derivativos y los de composición, sobre todo en el apartado de los prefijos. Como no es nuestra intención recordar de nuevo las opiniones existentes, y puesto que también ofreceremos algunos datos cuantitativos, baste recordar que seguimos el DRAE (1992).

Antes de nada, debemos hacer las siguientes aclaraciones: no consideramos derivadas las palabras que ya aparecen como tales en el latín o en el griego, según el DRAE, como *abertura*, *acumulación...*", de *apertura*,

accumulatio, *-onis*. Sí incluimos algunas que tienen la preposición latina *ad* (*acertar*, *alarma*, *aviso*) y *sin* (*simbiosis*). Recordemos por último que algunas palabras se componen de dos lexemas y algún prefijo o sufijo, y las incluimos dentro de la composición. Dentro de este mismo apartado hemos introducido también las formadas con elementos, no considerados ni sufijos ni prefijos, que han constituido palabras gramaticales (*aquel*, *cualquier*, *diecinueve*, *donde*, *vosotros*...). Aquellas que constan sólo de uno o más prefijos o sólo de uno o más sufijos (contabilizamos el sufijo *-ado*) las incluimos dentro de la derivación, como las que proceden de verbos, normalmente, y que no tienen una marca especial y que recogemos en el subapartado de *derivadas*. En el apartado de la parasíntesis englobamos aquellas voces formadas de prefijo (o prefijos) y sufijo (o sufijos) a la vez.

En primer lugar ofrecemos los datos cuantitativos:

1) COMPOSICIÓN

	V	N
- Lexemas y elementos compositivos	329	2214
- Gramaticales	90	13679
- Terminadas en *-mente*	243	1080
Total	662	16973

2) DERIVACIÓN

	V	N
- Prefijos	304	2343
- Sufijos	2733	20015
- Derivadas	271	3166
Total	3308	25524

3) PARASÍNTESIS

	N	V
	607	4732

4) TOTAL

	N	V
	4577	47229

3.4.2.1. Derivación

Ofrecemos a continuación una lista de los prefijos y sufijos usados en las palabras que componen el vocabulario de los escolares. El conocimiento de este y otros datos creemos que es interesante para la enseñanza sistemática del léxico, cuyos principios y postulados desarrollamos en otro lugar (Cfr. 4.9).

3.4.2.1.1. Prefijos

PREFIJO	SIGNIFICADO	EJEMPLOS
a-	-	*abatir, abono, abrazo, aflorar*
anti-	propiedad contr.	*antiaéreo, antidoping, antimisil*
co-, com-, con-	compañía, asociac.	*cooperación, coautor, cofrade,*

de-	refuerzo de sigdo.	*debajo, demacrado, deprisa*
des-	negación, privación	*desacuerdo, desaliento*
	inversión	*desandar, desatar*
	exceso, fuera de	*desmadre, desorden*
dis-	negac., contraried.	*discontinuo, disculpa*
em-, en-	dentro de, sobre	*embalse, encerrar, entorno*
	adquisic. de cualid.	*embrujar, encargar, encoger*
entre-	atenuación	*entrecortar, entretener*
	situación interm.	*entreabierto, entrecerrado*
es-	separación	*escoger*
	intensificación	*esforzar, estirar, estremecer*
ex-	privación, cese	*ex-amiga, ex-colegio, ex-marido*
extra-	fuera de	*extraescolar, extrovertido*
i-, im-, in-	negación	*ilegal, imperceptible, inadecuado*
inter-	entre, en medio	*intercambio,interestelar*
per-	refuerzo de sigdo.	*perdonar, perfilar, perfumar*
pos-, post-	después, detrás de	*posguerra*
pre-	anterioridad	*preescolar, prehistoria*
	prioridad	*predominar*

	encarecimiento	*predilecto*
re-	repetición	*rebote, recaer, recargar*
	intensificación	*recalar, reforzar, revuelo*
sub-, sus-	bajo, debajo de	*subacuático, subsuelo*
	inferioridad	*subcampeón, subdesarrollo, sustraer*
trans-, tras-	al otro lado	*transatlántico, traspasar*
	a través de	*transparente*

3.4..2.1.2. Sufijos

SUFIJO	CLASE GRAMATICAL RESULTANTE	SIGNIFICADO	EJEMPLOS
-áceo	adj.	sejemanza	*grisáceo*
-acho,a	adj., s.	despectivo	*picacho, ricach(ón)*
		-	*muchacho, muchacha*
-aco,a	adj., s.	despectivo	*maníaco, niñaco*
		gentilicio	*austríaco*
		relación	*policíaco*
-(a)da, -(i)da	adj., s.	relación	*camarada*
		acción, efecto	*acampada, cabida*
		conjunto	*bandada, manada*
		período	*madrugada*
		contenido	*bebida*

		golpe	*cachetada, pedrada*
		abundancia	*oleada*
		acc. peyorativa	*borregada, novatada*
-ado, -ido	adj., s.	dignidad	*marquesado*
		conjunto	*alumnado, calzado*
		cualidad	*atrevido, avispado*
		sonidos	*aullido, chillido*
		semejanza	*almendrado, azulado*
		acción	*bocado*
-aje	s.	acción, efecto	*aprendizaje, rodaje*
		lugar	*paraje, pasaje*
		conjunto	*equipaje, plumaje*
		-	*personaje, reportaje*
-ajo,a	adj., s.	despectivo	*hierbajo, salivajo*
		-	*horcajo, rodaja*
-al	adj., s.	relación, perten.	*ambiental, angelical*
		lugar	*arenal, bancal, portal*
		abundancia	*barrizal, manantial*
		objeto	*cigüeñal*
-ano	adj., s.	origen, relación	*aldeano, ciudadano*
		gentilicio	*accitano, africano*
-(a)nte, -(e)nte	adj., s.	agente	*alucinante, existente*
		empleo, cargo	*delineante, estudiante*
		objeto	*colgante*
-anza	s.	acción, efecto	*adivinanza, enseñanza*

-año,a	s.	relación	*ermitaño, hazaña*
-ar	v.	acción	*abusar, activar, actuar*
-ar	adj., s.	relación, perten.	*caballar, espectacular*
		colectivo	*castañar, olivar, pinar*
-ario,a	adj., s.	relación, perten.	*bancario, legendario*
		lugar	*campanario, herbolario*
		empleo, cargo	*bibliotecario, boticaria*
		agente, capacidad	*complementario*
-(a)rro, -(o)rro,	adj., s.	despectivo	*ceporro, chinarro, cotarrro*
-ato,a	s.	expresivo	*cubata, fogata*
		dignidad, instituc.	*califato*
		cualidad	*bachillerato, novato*
		acción, efecto	*cabalgata, campeonato*
-azgo	s.	estado	*noviazgo*
		acción, efecto	*hallazgo*
-azo,a	s.	aumentativo	*barcaza, pinchazo*
		expresivo	*esquinazo, puntazo*
		despectivo	*latazo, panzazo*
		golpe	*cabezazo, escobazo*
-ble	adj.	-	*apetecible, bailable*
-ción	s.	acción, efecto	*actuación, adaptación*
		-	*aviación*

-dad	s.	cualidad	*actualidad, fealdad*
-dor, -dora	adj., s.	agente	*acogedor, atracador*
		objeto	*carburador, cosechadora*
		lugar	*comedor, recibidor*
		empleo, cargo	*entrenador*
-dura	s.	acción, efecto	*andadura, magulladura*
		objeto	*cerradura*
		conjunto	*verdura*
-ear	v.	acción	*aletear, bromear, cabrear*
-eca	adj., s.	relación	*karateca*
-ecer	v.	acción	*apetecer, favorecer*
-eda	s.	colectivo	*arboleda*
		acción, efecto	*búsqueda, humareda*
-ejo,a	s.	-	*azulejo, moraleja*
-ense	adj., s.	gentilicio	*almeriense, jiennense*
-ento	adj.	semejanza	*amarillento, ceniciento*
		estado físico	*hambriento, harapiento*
-eño,a	adj., s.	semejanza	*aguileño*
		relación, cualidad	*pedigüeño, risueño*
		gentilicio	*alhameño, alpujarreño*

-eo,a	adj., s.	relac., perten.	*europeo*
		acción, efecto	*balanceo, bombardeo*
-ería	s.	cualidad peyorat.	*brujería, cabezonería*
		conjunto	*infantería, lencería*
		lugar	*cafetería, panadería*
		acción, efecto	*cacería*
-erío	s.	acción, efecto	*chillerío, griterío, poderío*
		lugar	*graderío*
-ero,a	adj., s.	relación	*aventurero, cortijero*
		empleo, cargo	*aceitunero, costalera*
		objeto	*azucarero, bañera*
		lugar	*bolera, merendero*
		árbol frutal	*morera, melocotonero*
		conjunto	*cabellera, cristalera*
		carácter	*casero*
-és, -esa	adj., s.	relación	*montés*
		gentilicio	*cordobés, escocés, siamés*
-ete,a	s.	diminutivo	*arete, avioneta, careta*
-ez	s.	cualidad	*brillantez, lucidez, madurez*
-eza	s.	cualidad	*alteza, bajeza, belleza*
-í	adj., s.	gentilicio	*marroquí, iraquí, israelí*
-ía	s.	cualidad, actitud	*maestría, alegría, cobardía*
		situación	*lejanía*

		acción	*travesía*
		estado	*sequía*
		lugar	*umbría*
-icio	adj., s.	relación	*alimenticio, desperdicio*
		intensidad	*bullicio*
-ico,a	adj., s.	relación	*afónico, apático, bíblico*
-iego	adj., s.	relación	*mujeriego*
-ificar	v.	acción	*planificar, plastificar*
-ijo,a	adj., s.	despectivo	*escondrijo*
		diminutivo	*lagartija, cortijo, vasija*
		-	*regocijo*
-il	adj.	relación	*estudiantil, varonil*
		lugar	*carril*
		objeto	*proyectil*
-illo,a	adj., s.	diminutivo	*armadillo, barbilla, bocadillo, bolsillo, bordillo*
-ín, -ino,a	adj., s.	diminutivo	*batín, fajín, futbolín*
		gentilicio	*argelino, mallorquín*
		agente	*andarín, cansino*
		profesión	*bailarín, bailarina*
		relación	*campesino, angorina*
		semejanza	*albino, blanquecino*
		-	*pegatina, neblina*

-ío	adj., s.	relac., perten.	*cabrío, regadío*
-ismo	s.	doctrina, movim.	*budismo, romanticismo*
		actitud	*fanatismo, fetichismo*
		deporte	*atletismo, motociclismo*
		abuso	*alcoholismo*
-ista	adj., s.	partidario de	*budista, madridista*
		inclinado a	*bromista, excursionista*
		ocupación	*analista, dentista*
-ístico,a	adj., s.	relac., perten.	*característico, futbolístico*
-ita	adj., s.	relac., perten.	*carmelita*
-ito,a	adj., s.	diminutivo	*caballito, cabezadita*
-ivo,a	adj., s.	capacidad	*altivo, caritativo, lectivo*
		empleo, cargo	*directivo*
		acción, efecto	*alternativa*
-izar	v.	acción	*colonizar, mentalizar*
-izo,a	adj., s.	semejanza	*blanquizo, cobrizo, rojizo*
		propensión	*enfermizo, pegadizo*
		relación	*levadizo, paliza, hortaliza*
		lugar	*pasadizo*
-mento	s.	acción y efecto	*cargamento*
		materia	*pegamento*

-miento	s.	acción y efecto	*agotamiento, pensamiento*
-ncia	s.	cualidad, acción objeto	*convicencia, importancia* *ambulancia*
- ol,a	adj., s.	gentilicio relación	*español, española* *farol, farola, perola*
-on,a	adj., s.	aument., intens. expresivo cambio de género y cambio de sigdo.	*bonachón, burlón, chillón* *chapuzón* *arcón, asperón, balón* *boquerón, cajón, camisón*
-or	s.	cualidad agente	*espesor, frescor, grosor* *tractor*
-orio	s.	relación, perten.	*jolgorio*
-oso	adj, s.	cualidad abundancia sigdo. activo intensificación	*amistoso, amoroso* *caluroso, caudaloso* *beneficioso, chirrioso* *boscoso, aparatoso*
-ote,a	adj., s.	aumentativo: sentido diminut.	*cabezota, pasota* *pegote, gaviota* *camarote, islote*
-rro	s.	despectivo	*cacharro, ceporro, chinarro*
-sco	adj., s.	propio de	*marisco, novelesco*

		acción, efecto	*mordisco*
		semejanza	*pardusco*
-sis	s.	enfer., estad. irreg.	*escoliosis, virosis*
-toria	s.	acción, efecto	*dedicatoria, eliminatoria*
-torio	adj.	relación	*giratorio, migratorio*
		lugar	*laboratorio*
-tud	s.	cualidad	*actitud, exactitud*
-ucho,a	adj., s.	despectivo	*aguiluc:ho*
-udo,a	adj., s.	gran tamaño	*barbudo, melenudo*
		intensidad	*huesudo, testarudo*
-uelo,a	adj., s.	diminutivo	*callejuela, gordezuelo*
-uno,a	adj.	relación, perten.	*vacuno*
-uño	s.	acción, efecto	*rasguño*
-ura	s.	cualidad	*altura, amargura, anchura*
-uzco,a	adj.	cualidad	*negruzco, paliduzco*

3.4.2.2. Composición

3.4.2.2.1. Raíces prefijas

R.PREFIJA	SIGNIFICADO	EJEMPLOS
acua-	agua	*acuífero*
aero-	aire	*aeroplano, aeropuerto, aerosol*
antropo-	hombre	*antropomórfico*
auto-	por uno mismo	*autoabastecer, autocar, autodefinir automarginar*
bi-, bis-, biz-	dos, dos veces	*bicicleta, bimotor, bisabuelo, bizcocho*
bio-	vida	*biografía, biología, biológico*
centi-	centísima parte	*centígrado, centímetro*
crono-	tiempo	*cronografía, cronometro*
di-	dos	*dióxido*
eco-	ámbito vital, casa	*ecología, ecologista, ecosistema*
foto-	luz	*fotografía, fotosíntesis*
geo-	tierra	*geológico*

hecto-	cien	*hectárea*
hidro-	agua	*hidroeléctrico*
hiper-	superiorid., exceso	*hipermercado, hiperpequeño*
homo-	semejante, igual	*homosexual, homosexualidad*
infra-	inferior., debajo	*infraestructura*
kilo-	mil	*kilogramo, kilómetro*
micro-	pequeño	*microbús, micrométrico, microondas*
mili-	milésima parte	*milímetro*
mini-	pequeño, corto	*minicadena, minicar, minifalda*
mono-	único, uno solo	*monopatín, monoplaza*
multi-	muchos	*multimillonario, multinacional*
neo-	reciente, nuevo	*neoclásico, neofascismo, neonazi*
peni-	casi	*penibético, penunbra*
penta-	cinco	*pentagrama*
poli-	muchos	*polideportivo, polivalencia*

psico-	mente	*psicología, psicólogo, psiquiatra*
semi-	medio	*semiabierto, semiartificial, seminatural*
sobre-	superposición	*sobrellevar*
	adición	*sobrenombre, sobrevivir*
	intensificación	*sobredosis, sobrehumano, sobrepasar*
	acción repentina	*sobrecoger, sobresaltar, sobresalto*
super-	superiorid., exceso	*superdotado, supermercado, supervisor*
tele-	lejos	*telediario, telesilla, televisión, telesquí*
tri-	tres	*triciclo, tridimensional*
ultra-	exceso	*ultrasur, ultratumba, ultravioleta*
uni-	uno	*unicelular*
vice-	"en vez de"	*vicepresidente*
xeno-	extranjero	*xenofobia*
zoo-	animal	*zoológico*

3.4.2.2.2. Raíces sufijas

R. SUFIJA	SIGNIFICADO	EJEMPLOS
-cidio	acc. de matar	*suicidio*
-fero	produce, tiene	*acuífero, plumífero*
-filia	afición, amor	*halterofilia*
-fobia	aversión, enemist.	*xenofobia*
-fonía, -fono	sonido	*telefonear, teléfono*
-fugo,a	que ahuyenta	*ignífugo*
-grafía, -grafo	representación gráfica, escritura	*fotografía, radiografía* *autógrafo, bolígrafo*
-logía, -logo	estudio, ciencia, especialista en	*tecnología, meneralogía* *futurólogo, psicólogo*
-mano,a	afic. en exceso	*pirómano, toxicómano*
-metro	medida	*cronómetro, kilometro*
-morfo	forma	*antropomórfico*
-nauta	navegante	*astronauta*
-ptero,a	ala	*helicóptero*

3.4.2.2.2. Raíces sufijas

R. SUFIJA	SIGNIFICADO	EJEMPLOS
-cidio	acc. de matar	*suicidio*
-fero	produce, tiene	*acuífero, plumífero*
-filia	afición, amor	*halterofilia*
-fobia	aversión, enemist.	*xenofobia*
-fonía, -fono	sonido	*telefonear, teléfono*
-fugo,a	que ahuyenta	*ignífugo*
-grafía, -grafo	representación gráfica, escritura	*fotografía, radiografía* *autógrafo, bolígrafo*
-logía, -logo	estudio, ciencia, especialista en	*tecnología, meneralogía* *futurólogo, psicólogo*
-mano,a	afic. en exceso	*pirómano, toxicómano*
-metro	medida	*cronómetro, kilometro*
-morfo	forma	*antropomórfico*
-nauta	navegante	*astronauta*
-ptero,a	ala	*helicóptero*

-teca	caja (lugar)	*discoteca*
-tomía	división, corte	*traqueotomía*

3.4.2.2.3. Lexemas

FORMANTES	EJEMPLOS
adj. + adj.	*altibajo*
adj. + sust.	*altavoz*
sust. + adj.	*aguardiente, boquiabierto, caradura, paticorto, pelirrojo*
sust. + sust.	*balompié, baloncesto, balonmano, balonvolea, bocacalle, coliflor*
adv. + adj.	*maleducado, malherido, maltratado, maloliente*
adv. + sust.	*bienvenida*
adv. + v.	*maltratar*
v. + sust.	*cazabombardero, cortacésped, cortafuegos, cumpleaños, girasol, lanzallamas*

3.4.2.3. Parasíntesis: prefijos y sufijos

Como veíamos al principio, algunos vocablos (607) de los usados por los escolares se habían formado por parasíntesis, entendida como la unión de prefijo, lexema y sufijo. Por tanto, y por poner algunos ejemplos, incluimos palabras como *compenetración*, *extraterrestre*, *inmadurez*, *irresponsabilidad*, *prehistórico*..., aun existiendo en el idioma una palabra derivada con sufijo o prefijo (por ejemplo, existe *compenetración*, *terrestre*, *madurez*...), como otras en las que no ocurre esa circunstancia: *adelgazar*, *agolpar*, *agradar*, *apilar*, *aprovechamiento*, *desbocar*, *descapotable*, *embarcar*, *embellecer*, *enriquecimiento*, *intoxicación*, *revolotear*...

3.4.3. El uso

Con este apartado queremos señalar o sugerir los distintos usos y registros de los escolares, según se comprueba en su vocabulario; es decir, introducir un breve análisis que atienda a las variedades lingüísticas que coexisten en una comunidad lingüística determinada -diatópicas, diastráticas, diafásicas- y cuyo reflejo se ve en el vocabulario escolar de nuestra investigación. Desde el punto de vista léxico, no es fácil determinar sobre todo cuándo una palabra pertenece a las variedades diastráticas[306] (que

---

[306] M. Seco (1973: 231-232) se refiere de la siguiente manera al problema de la delimitación de las variedades sociales: "Así como es perfectamente posible el encasillamiento geográfico de un hablante o un grupo de hablantes, es imposible su encasillamiento -en cuanto a la lengua- en niveles sociales precisos. Es evidente que éstos existen: no cabe duda de que la frase *Nos ha amolao* pertenece a un nivel particular, muy distante de aquella que pertenece la frase *Ha sido un enojoso contratiempo*; pero el paso de uno a otro nivel no se hace por escalones, sino por una rampa donde no es posible señalar <<hasta aquí es muy popular>>, <<desde aquí es menos popular>>, etc. Se trata de una escala donde, como en el arco iris, el paso de uno a otro color se produce en una forma totalmente continua y gradual entre dos extremos, de manera

atiende al nivel sociocultural: lengua culta, popular, lenguajes especiales[307]) y diafásicas (que tienen en cuenta la situación comunicativa, relación entre interlocutores, tono del escrito, modalidad oral o escrita, es decir, los distintos registros). Quizás por ello no existen demasiados estudios en los que se fijen referencias claras para poder considerar una palabra como perteneciente a uno u otro registro y sobre esta base se propongan inventarios concretos y extensos de vocablos, expresiones, construcciones... De ahí que nosotros hayamos recurrido a los trabajos lexicográficos existentes: los diccionarios. El DRAE a veces hace algunas matizaciones relativas al registro ("fam.") y a las variedades geográficas. Pero que contemple entre sus páginas indicaciones claras de uso y registro, hemos hallado uno que nos ha parecido interesante: *Diccionario Salamanca de la lengua española* (1996). En esta obra algunas palabras van acompañadas de marcas concretas que atienden a las variedades lingüísticas de las que venimos hablando; marcas como "restringido, rural, vulgar, jergal, coloquial, literario, elevado...". En el citado diccionario se hacen algunas aclaraciones previas que nos parecen interesantes: "Las marcas de uso pueden cruzarse con las de registro y llegar a ser extremadamente complicadas. Este diccionario pretende indicar una clasificación clara y relativamente útil. Para marcar el uso de las acepciones de cada palabra partimos de lo **normal o neutro**, que no se señala. Se considera neutro un amplio uso social, que abarca muchos registros. Lo normal es lo que se puede decir en público o lo que se puede escribir en medios públicos, que no destaca ni como especialmente elevado ni como vulgar, ni muestra especial confianza con el interlocutor. Es un amplio territorio en el que muchísimos términos aparecen. Lo que no está marcado se considera normal

---

que cualquier división que queramos hacer en la escala será -como en el propio arco iris- completamente convencional."

[307] Haensch (1982a: 138, 145-149) habla de obras lexicográficas de "marcación diatécnica" para referirse a aquellas que registran las terminologías específicas de una ciencia o profesión.

y se puede usar, por consiguiente, ante personas desconocidas sin peligro de ofenderlas o de hacer el ridículo"[308].

Hay que aclarar que en este diccionario no son demasiadas las palabras que aparecen con alguna marca determinada relativa al uso -decimos esto porque al cotejar nuestro vocabulario con el recogido en el vocabulario del que hablamos, hemos podido comprobar el aserto-, lo cual condiciona los datos que ofreceremos, sobre todo los cuantitativos.

3.4.3.1. Coloquial

Ofrecemos en este apartado aquellas palabras recogidas en nuestra investigación que están marcadas como coloquiales en el *Diccionario Salamanca de la lengua española* (1996)[309]. Aquí se alude muchas veces a usos coloquiales de determinadas palabras, es decir, no en todas sus acepciones ni empleos responden a esa denominación. Nosotros hemos recurrido en esos casos a la expresión de los escolares -sus composiciones- y sólo cuando había correspondencia, las hemos seleccionado. En cursiva ofrecemos algunas voces que se han formado con la base de la palabra considerada coloquial.

adefesio	almorrana	apañado
arrastre	asco	astronómico
baboso	barriga	barriobajero

---

[308] Estas indicaciones se recogen en la "Introducción" del Diccionario Salamanca de la lengua española (1996: IX), junto con otras en las que se especifican y se aclaran las distintas marcas.

[309] En el citado diccionario (p. X) se hace la siguiente referencia: "**Coloquial**: Perfectamente posible entre un grupo de amigos de confianza, en muchas ocasiones en público, pero hay que tener cuidado. Por escrito no suele ser aconsejable".

bestial	bocata	borregada
borrego	burro	cabezonada
cabezota	cabreado	cabrear
cabreo	caca	cachondear
cachondeo	cachondo	cagado
caña	caradura	cascar
cascarrabias	catear	cateto
ceporro	coco	colega
colocón	comilona	consumido
coña	cotarro	cotilla
cotillear	cotilleo	crío
cubata	culé	culo
currante	currar	curro
cursi	cutre	chao
chapucero	chapuza	chatarra
chaval	chavala	chirona
chivatazo	chorrada	chucho
chuchurrirse	chuleta	chuleta
chungo	chupa	chupado
chupi	chupón	desmadrado
desmadre	despachurrar	dominguero
empicado	empollado	empollar
empollón	encharcado	enchufado
enchufar	enrollado	enrollarse
enterado	escabullirse	escopeteado
esquinazo	estrafalario	facha
fanfarrón	fanfarronear	flechazo
flipar	fogata	follón
fulano	gandul	gandulear
ganso	garito	garrafal
golfo	gordinflón	gritón

gruñón
guantazo
guapamente
guaperas
guarro
guay
guiri
gusanillo
hartada
hincha
husmear
improperio
incordiar
incordio
ja
jaleo
jaleoso
jo
jolgorio
jolín
juerga
jugarreta
juguetaear
largar
latazo
ligón
ligotear
ligue
litrona
macarra
machacar
machaque
mandón
mangar
manta
marcha
marchoso
marica
mariquita
melenudo
mengano
mili
mirón
míster
mocedad
mogollón
molestoso
mollera
morado
moratón
mosqueado
mosquear
movida
narigudo
orejudo
pachanga
pachucho
paleta
pancho
panoli
panza
panzada
panzazo
parlanchín
pasada
pasota
patatán
patatín
patochada
patoso
pela
peliculero
pelota
pelotillero
pendón
piba
picado
picarón
Picio
pijo
pillo
pinito
pipa
pique
pirarse
pis
plantón
plasta
pompi
porquería
porrazo
porro
potingue

puntazo	puñeta	rabillo
rabioso	rabona	rechistar
rechoncho	regalón	regañina
regordete	remojón	resaca
respondón	resultón	rocambolesco
roñoso	rular	salido
súper	*supercolado*	*superenrollado*
superguay	tablón	tartajoso
tata	teta	tirado
tortolito	*tostoneo*	tragón
trasero	trastada	trastear
trasto	tristón	virguería
ye-yé		

3.4.3.2. Vulgar

Se registran aquí palabras que se denominan muchas veces "tabuizadas". En el *Diccionario Salamanca de la lengua española* (1996: X) se hace la siguiente observación: "**Vulgar**: Se ha marcado como vulgar cualquier acepción que no debe usarse en público, ni entre conocidos de escasa confianza. Es probable que entre amigos de cierta confianza se utilicen con frecuencia estos términos; sin embargo, conviene tener presente que es mejor, en principio, saber bien en qué ambiente, en público o entre amigos de poca confianza, puede resultar poco agradable".

cabrón	cagar	cojonudo
joder	jodido	maricón
mear	mierda	polla
puta	putada	putear
puto		

3.4.3.3. Jergal

El Diccionario *Salamanca de la lengua española* (1996: X) recoge las siguientes indicaciones para este apartado: "**Jergal**: Con este término marcamos las voces que no se consideran propias de la lengua general, sino de círculos de hablantes que quieren en este caso alejarse de ella (...). En el español peninsular, la marca se refiere siempre a la lengua juvenil o a ciertos usos relativamente marginales, aunque a veces se empleen también por razones de estilo en la lengua común".

canuto	colocado	costo
esnifar	papela	vacilar
vacilón		

3.4.3.4. Restringido

En este apartado, siguiendo el *Diccionario Salamanca...* (1996: IX), se recogen palabras muy heterogéneas. Pertenecen sobre todo al léxico regional o se emplean en determinadas circunstancias. Además del listado vamos a ofrecer a continuación, también podríamos considerar regionales o locales otras palabras como *alpaca*[310], *arruñar*[311], *carmen* (tipo de casa), *chino*[312],

---

[310] Tiene el sentido que recoge Rodríguez Titos (1996:61): "Paquete de paja o yerba prensada y atada. (En DRAE, paca)".

[311] Registrada también por Rodríguez Titos (1996: 157) con el sentido de "arañar".

[312] *Carmen* y *chino* aparecen en el DRAE, como localismo la primera, con el sentido de 'quinta con huerto o jardín', y como regionalismo la segunda , con el significado de 'piedrecita'.

clisado[313], *collejo*[314], *malafollá*[315], *manda*[316], *palomica*[317]..., según indicaciones del DRAE u otros trabajos de vocabulario local (García de Cabañas, Mª J., 1967; y Rodríguez Titos, 1996).

abasto	alcoba	almanaque
almuerzo	auto	balompié
botica	boticaria	cachetada
carrasca	chícharo	cuarta
defecación	desarreglado	doquier
engurruñido	fonda	mandoble
misto	mozo	orear

---

[313] Rodríguez Titos (1996: 102) recoge el lema *clisarse*, con el sentido siguiente: "Embobarse, abstraerse, quedarse mirando algo fijamente durante largo rato".

[314] Usada la palabra como adjetivo (*collejo/a*), y registrada también en el trabajo de Rodríguez Titos (1996: 102). Lo define como "pequeño y gracioso; se aplica especialmente a los niños que remedan en algo a los mayores, como dichos, vestimenta, etc.".

[315] Esta es una palabra típica en Granada, no sólo en la capital, sino también en la provincia, como lo demuestra el estudio de Rodríguez Titos (1996: 103), quien dedica varias líneas a su explicación: "Vocablo de amplio uso, granadino por antonomasia, no porque la gente de Granada sea así, sino por el empleo exhaustivo que de esta palabra se hace. Se refiere a personas sin gracia, o a sinvergüenzas, o, incluso, a mala gente. Sin. **malafondinga, malahostia, malasombra, malahora**".

[316] El DRAE, en una de las acepciones cuyo uso es propio de Andalucía, define esta palabra como " voto o promesa hecha a Dios o a un santo". Parecida es la definición de García de Cabañas (1967: 82): "Promesa, ofrecimiento hecho a Dios o a los santos de ejecutar una obra piadosa".

[317] Rodríguez Titos (1996: 85) define el vocablo como "cualquier mariposa muy pequeña (En **DRAE**, palomilla)".

orinar	roseta	talegazo
posada	veras	

3.4.3.5. Elevado

Son términos propios de la lengua escrita o de estilos muy formales y poco usadas coloquialmente.

albero	altanero	andadura
apocalipsis	aposento	arrebol
caos	crepúsculo	data
deambulante	deambular	diáfano
escuálido	esvástica	evento
exiguo	expedito	éxtasis
falacia	fastuoso	fisonomía
horda	huraño	ingente
ingerir	júbilo	límpido
magnánimo	malévolo	obeso
ocaso	oquedad	paupérrimo
quimera	recóndito	regazo
remoto	reverso	sarcasmo
solaz	supremacía	tórrido
yermo		

3.4.3.6. Tecnicismos

Registramos bajo este epígrafe las palabras señaladas como tales en el DRAE o en el *Diccionario Salamanca de la lengua española*. Pertenecen a terminologías muy variadas.

Disciplinas científicas

ácido	acné	afluente
alud	aminoácido	anabolizante
anfetamina	aorta	arcilla
arteria	asfixia	asteroide
astro	atmósfera	atmosférico
autopsia	bacteria	butano
calórico	carbono	cometa
conífera	cráter	crustáceo
cúbito	diámetro	dióxido
ecuación	encía	epilepsia
erosión	erosionar	escoliosis
esguince	esófago	estalactita
estalagmita	estómago	estrato
estratosfera	fémur	férula
fotosíntesis	gástrico	gen
gestación	glándula	glucosa
hemorragia	hidrógeno	hormona
intestino	jade	micrométrico
microorganismo	ornitológico	ortodoncia
ozono	paranoia	párkinson
proteína	radiación	radial
radioactividad	reflectar	roedor
sifosis	silicona	síndrome

tarso	terapia	testosterona
tibia	toxicólogo	traqueotomía
virosis	volcán	volcánico
zafiro	zinc	

Disciplinas humanísticas

Son términos que entran dentro del ámbito del arte, la música, la economía, el derecho, la mitología, la lingüística, la geografía, la comunicación....

arbotante	arlequín	bandurria
bóveda	capitel	cénit
clarinete	conservatorio	cornisa
crónica	cumbia	cúpula
deflación (economía)	elfo	fonema
golfo	hábitat	hemisferio
infracción	rajá	renacentista
republicano	retransmisión	rima
románico	saxofón	sílaba
single	solfeo	

Deportes, toros

antidoping	bate	centrocampista
chutar	colegiado	copero
cross	culturismo	enduro
escudería	halterofilia	líbero
lidia	linier	pichichi

penalti	regate	regatear
set		

Mundo del motor

biela	carburador	cigüeñal
pistón		

Navegación

borda	boya	eslora
foque	amarra	obenque
popa	proa	

Vida militar

búnker	contingente	corbeta
coronel	teniente	

Religiones

bautizar	budismo	cántico
corpus	cuaresma	párroco
persignar	presbítero	

3.4.4. Unidades fraseológicas

Con este apartado nos referimos, de manera general, a aquellos casos en que varias unidades morfosintácticas pueden representar una sola unidad de sentido. El problema de estas unidades ha sido tratado en otro lugar (Cfr. 2.11.3.). Baste recordar únicamente que las mayores dificultades radican en delimitar con exactitud el grado de lexicalización que presentan. De ahí que nosotros intentemos recoger no sólo las que se consideraran totalmente lexicalizadas, sino también algunas de aquellas que pueden encontrarse próximas a lexicalizarse. La idea de incluir aquí un listado con estas unidades no es otra sino dar cuenta también del uso que de ellas hacen los alumnos. Actuando de esta manera creemos que cumplimos con dos aspectos esenciales en un trabajo que recoge el vocabulario usual de una población determinada y que responde, al mismo tiempo, a las exigencias impuestas por la estadística léxica:

1) Atender al uso real que los escolares hacen de su idioma.

2) No desvirtuar o tergiversar los datos cuantitativos que exige una investigación de índole estadística.

En un trabajo como el nuestro que parte de un corpus tan amplio (412.433 palabras), analizar exhaustivamente todas estas unidades, como hemos dicho, es poco menos que imposible, ya que este mismo apartado puede ser objeto de una sola investigación. Pero ante la disyuntiva de no contemplarlas u ofrecer un listado, posiblemente incompleto, de ellas, elegimos la segunda opción. En nuestra lista tienen cabida, siguiendo a Gloria

Corpas (1997), las colocaciones[318], los distintos tipos de locuciones y los enunciados fraseológicos[319] (paremias y fórmulas rutinarias).

a altas horas de la noche	a base de
a cántaros	a causa de
a conciencia	a consecuencia de
a costa de	a cuestas
a diario	a duras penas
a escondidas	a excepción de
a falta de	a fin de cuentas
a finales de	a flor de piel
a fondo	a fuerza de
a gusto	a la francesa
a la mañana siguiente	a la vera de

---

[318] Corpas, G. (1997: 53) define así a estas unidades: "La primera esfera de nuestra clasificación comprende colocaciones, es decir, unidades fraseológicas que, desde el punto de vista del sistema de la lengua, son sintagmas completamente libres, generados a partir de reglas, pero que, al mismo tiempo, presentan cierto grado de restricción combinatoria determinada por el uso (cierta fijación interna)". La autora aporta numerosos ejemplos como *estallar una guerra, declararse un incendio, importancia capital, relación estrecha, una rebanada de pan, un diente de ajo, banco de peces, profundamente dormido...*

[319] Corpas, G. (1997: 132-133) expone: "Dentro de los enunciados fraseológicos distinguimos entre *paremias* y *fórmulas rutinarias*. Ambos tipos se diferencian fundamentalmente en dos aspectos. En primer lugar, las paremias poseen significado referencial (...), mientras que en las fórmulas rutinarias el significado es de tipo social, expresivo o discursivo fundamentalmente. Y en segundo lugar, las paremias gozan de autonomía textual, mientras que las fórmulas vienen determinadas por situaciones y circunstancias". La autora considera paremias expresiones del tipo *Los duelos con pan son menos, Si te he visto no me acuerdo, Contigo, pan y cebolla, Juntarse el hambre con las ganas de comer, Llover sobre mojado, A enemigo que huye, puente de plata, A falta de pan buenas son tortas...* Ejemplos de fórmulas rutinarias son, por ejemplo, *Que te sea leve, No es para tanto, Mira quien fue a hablar, ¡Válgame Dios!, ¡Lo que faltaba!, ¡No faltaría más!, ¿Cómo estás?, ¿Qué tal?, Muchas gracias, Feliz cumpleaños, Muchas felicidades...*

a la vez
a las afueras
a lo lejos
a lo mejor
a mediados de
a medida que
a menudo
a nivel de
a no poder más
a no ser que
a partir de
a pesar de
a pique
a punto de
a secas
a toda costa
a toda pastilla
a tope
a través de
a veces
abrirse todas las puertas (a alguien)
aburrido como una ostra
acabar de + infinitivo
acabar + gerundio
acabar por + infinitivo
accidente de tráfico
aceite de oliva
acerca de
agua medicinal
aire acondicionado
al cabo de
al día siguiente
al fin
al fin y al cabo
al final
al fondo
al lado
al lado de
al menos
al parecer
al principio
al respecto
al revés
alcanzar la vista
alimentarse del aire
altar mayor
alto cargo
andar + gerundio
andarse con rodeos
animal de compañía
aprovechar el día
apuntarse a un bombardeo
árbol de navidad
árbol frutal
arco iris
armario empotrado
arroz con leche
artes marciales
artillería pesada
asociación de padres
aspirar a ser
atracciones de feria

autobús de línea
banda de música
baúl de los recuerdos
blanco como la nieve
buen humor
buenas noches
buenas tardes
buenos días
caballo de batalla
cabeza de carrera
cabina telefónica
cadena musical
cadena privada
caer bien (a alguien)
caer bien (algo)
caer en la cuenta
caerse el mundo encima (a alguien)
cambiar de clase
cambiar de vida
campo de fútbol
canto del gallo
capa de ozono
casa de campo
casa flotante
castillo de arena
causar impresión
centro comercial
cerrar las puertas (a alguien)
ciencia experimental
ciencia natural
coche de choque
comenzar a + infinitivo
comer el coco
comerse una rosca
como caída del cielo
como en los viejos tiempos
como Pedro por su casa
como sardinas en lata
como si nada
como si no pasaran los años
complejo industrial
compuesto orgánico
compuesto y sin novia
con el fin de
con el paso de los años
con el paso del tiempo
con los nervios a flor de piel
con respecto a
con todos los gastos pagados
con una sonrisa de oreja a oreja
convento de clausura
corrida de cintas
cosa por el estilo
creerse con derecho a todo
cuarto de baño
cuarto de estar
cuarto de estudio
chapada a la antigua
dar a conocer
dar a entender
dar ánimos (a alguien)
dar caña

dar corte (a alguien)
dar el brazo a torcer
dar en el clavo
dar la casualidad
dar la impresión de
dar la nota
dar las gracias (a alguien)
dar lugar a
dar recuerdos (a alguien)
dar saltos de alegría
dar un repaso
dar una vuelta
darle vueltas a la cabeza
darse la vuelta
darse un chapuzón
de acuerdo
de boquilla
de higos a brevas
de mal en peor
de mírame y no me toques
de parte de
de repente
de tal manera que
de un lado para otro
de veras
deber + infinitivo
debido a
dejar como cosa perdida
dejar en paz (a alguien)
dejar la mente en blanco
dejar plantado (a alguien)

dar crédito (a algo)
dar el visto bueno
dar explicaciones
dar la espalda (a alguien)
dar la lata
dar la razón (a alguien)
dar las vacaciones
dar paso a
dar rienda suelta (a algo)
dar un paseo
dar una caminata
dar vergüenza (a alguien)
darse cuenta
darse prisa
darse una vuelta
de boca en boca
de buenas a primeras
de la noche a la mañana
de marcha
de nuevo
de pronto
de sobra
de todos modos
de veraneo
de vez en cuando
deber de + infinitivo
dejar + participio
dejar de + infinitivo
dejar hecho puré
dejar los libros a un lado
dejarse la vida en

dejarse llevar
día festivo
día tras día
echar a + infinitivo
echar Dios al mundo
echarse atrás
echarse una cabezadita
educación física
en broma
en busca de
en concreto
en cuanto a
en definitiva
en el fondo
en frente
en lo alto de
en principio
en resumen
en un abrir y cerrar de ojos
en vez de
es decir
estar a salvo
estar como una moto
estar cortados con la misma tijera
estar en lo más alto
estar entre rejas
estar loco por (alguien)
estar pisando los talones (a alguien)
estrella de mar
fin de semana
forma de andar
del dicho al hecho hay mucho trecho
día laboral
dicho y hecho
echar de menos
echar la bronca (a alguien)
echarse la culpa
echarte a la cara
empezar a + infinitivo
en bruto
en cambio
en contra de
en cuestión de segundos
en detrimento de
en fin
en general
en medio
en realidad
en torno a
en un santiamén
entrar en coma
estar + gerundio
estar al alcance (de alguien)
estar con la cara larga
estar en la flor de la vida
estar entre la espada y la pared
estar listo para el arrastre
estar muy visto
estirar la pata
fin de curso
forma de actuar
forma de pensar

forma de ser

formación profesional

ganarse la vida

gastar una broma de mal gusto

grandes almacenes

guardia civil

haber que + infinitivo

hacer acto de presencia

hacer el panoli

hacer la pelota (a alguien)

hacer los deberes

hacer un llamamiento

hacerse esperar

hacerse la puñeta

hoy en día

hundirse el mundo (a alguien)

ir a + infinitivo

ir dando tumbos

ir de bares

ir de excursión

ir de marcha

ir la marcha (a alguien)

ir para rato

juegos de azar

jugar una mala pasada

las apariencias engañan

liarse a + infinitivo

lomo embuchado

luna de miel

llegar a + infinitivo

llevar + gerundio

forma de vestir

galería de arte

gastar una broma

golpe de pedal

grupo social

haber de + infinitivo

hablando se entiende la gente

hacer caso

hacer falta

hacer la vida imposible (a alguien)

hacer pinitos

hacer una mala pasada

hacerse la boca agua

hincar los codos

hoy por hoy

ir + gerundio

ir al grano

ir de acampada

ir de compras

ir de mal en peor

ir de un lado para otro

ir para largo

irse la olla (a alguien)

juegos de mesa

lanzamiento de jabalina

levantar el ánimo (a alguien)

lobo de mar

lucha libre

llamar la atención

llenar de emoción

llevar + participio

llevar a cabo
llevarse una sorpresa
manos a la obra
mar de dudas
más allá
más o menos
medio ambiente
medio de comunicación
medio de transporte
medio de vida
merecer la pena
mesa camilla
mesa de estudio
mesita de dormir
meter boca
meter follón
meterse algo en la cabeza (a alguien)
meterse en los asuntos de alguien
montaña rusa
nariz respingona
necesitar de una mano amiga
negro como el carbón
nivel económico
no hay mal que por bien no venga
no obstante
no parar de hablar ni debajo de agua
o sea
ojo de buey
ojos saltones
oso de peluche
pájaro carpintero
palito de merluza
pantalón vaquero
para colmo
para que
parada del autobús
parar de + infinitivo
parecer mentira
parque de atracciones
partirse de risa
pasar a + infinitivo
pasarlo bien
pasarlo bomba
pasarse (a alguien) (algo) por la cabeza
pasarse de la raya
pasárselo de miedo
patatas fritas
pedazo de
pedir la mano
pedir perdón
perder la cabeza por
perder la cuenta
perder la noción del tiempo
picar la curiosidad
plaza de abastos
poco a poco
poder + infinitivo
poner a prueba
poner cada uno de su parte
poner de nuestra parte
poner manos a la obra
poner reparos

poner un granito de arena
ponerse a + infinitivo
ponerse de todos los colores
ponerse el vello de punta
ponerse en marcha
ponerse la cara muy larga (a alguien)
ponerse los pelos de punta
por arte de magia
por completo
por ejemplo
por fin
por nada en el mundo
por otro lado
por supuesto
por último
prestar atención
pros y contras
puesta de sol
punto de vista
que sea lo que dios quiera
quedar (algo) en agua de borrajas
quedarse a dos velas
quedarse hecho polvo
raíz cuadrada
recibir con los brazos abiertos
respirar aire puro
rueda de prensa
sacar a relucir
saco de dormir
salir de asombro
salita de estar
poner verde (a alguien)
ponerse de acuerdo
ponerse el sol
ponerse en contacto
ponerse hecho una fiera
ponerse la carne de gallina (a alguien)
ponerse manos a la obra
por cierto
por consiguiente
por favor
por las buenas
por otra parte
por parte de
por tanto
preparar la c
programa concurso
puerto marítimo
puesto que
que no cunda el pánico
que si patatín que patatán
quedar hecho polvo
quedarse frito
quitarse (algo, alguien) de la cabeza
rayo de sol
relación comercial
romperse en mil pedazos
sacar a la luz
sacar de quicio
sala de estar
salir de marcha
salón de actos

saltar loco de contento
salto de longitud
sano y salvo
semana cultural
sembrar el pánico
ser del montón
ser más feo que Picio
ser un cacho de pan
silla de ruedas
sin fin
soler + infinitivo
sudar la gota gorda
tal vez
tarde o temprano
televisión en color
tener cara de pocos amigos
tener en cuenta
tener la cabeza llena de pájaros
tener mucho morro
tener que + infinitivo
tener sus pros y sus contras
terminar de + infinitivo
tiempo libre
tirarlo todo por la borda
tomar el pelo (a alguien)
tomarse (algo, alguien) a mal
trabajar como un burro
traer de cabeza
traje de baño
valer la pena
venir a la cabeza
salto de altura
salvarse por los pelos
seguir + gerundio
semana santa
ser capaz de
ser más basto que un arado
ser pasto de
ser un golpe muy duro
sin embargo
sobre todo
soltarse el pelo
tablón de anuncios
tan pronto como
televisión en blanco y negro
tener algo en común
tener clase
tener golpes
tener las ideas claras
tener por seguro
tener sus más y sus menos
tenérselo bien montado
terminar listo para el arrastre
tienda de campaña
tomar conciencia
tomar riendas en el asunto
tomarse la vida como un juego
trabajos manuales
traer por el camino de la amargura
traje de gitana
valer más que las pesetas
venirse (algo) a la cabeza

venirse (algo) a la mente
viaje de estudios
vino blanco
volver a + infinitivo
vestir a la moda
viaje de fin de curso
vino tinto
ya que

3.4.5. Otros datos

En este apartado señalamos aquellas palabras que han utilizado los escolares y que por tanto aparecen en nuestro recuento, pero que no se registran en el DRAE y en los diccionarios escolares que citábamos más arriba. Estas referencias lexicográficas son imprescindibles para poder extraer alguna conclusión acerca de los procedimientos de formación que utilizan ocasionalmente los alumnos, y otros datos que pueden resultar interesantes. Es decir, para poder afirmar que una palabra ha sido creada ocasionalmente por una alumno determinado, hemos de justificarlo apoyándonos en el punto precedente. Por tanto, no está en nuestro ánimo (porque no es esa nuestra pretensión ni el motivo de este apartado) censurar a ninguno de los diccionarios mencionados por no incluir tal o cual vocablo. Es más, estamos de acuerdo, por ejemplo, con la decisión del DRAE de desterrar de sus artículos los xenismos y también aprobamos muchas de sus propuestas para acomodar los préstamos a la fonética del español.

Las palabras que no hallamos en el diccionario de la Academia[320], además de los adverbios en *-mente* (*abundantemente*, *accidentalmente*,

---

[320] El problema de registrar todos los derivados con sufijos usuales que se prestan a creaciones espontáneas es aludido por la Academia en las advertencias para el uso del diccionario, donde reconoce, para los adverbios en *-mente*, por ejemplo, "tampoco se incluyen todos los adverbios en *-mente* (...) por ser de formación fácil y a menudo ocasional. Pero su ausencia en el DICCIONARIO no significa que no existan en el uso o que sean incorrectos" (DRAE, 1992: XXII).

adecuadamente, *amistosamente*, *básicamente*, *copiosamente*, *deportivamente*, *posiblemente*, *sexualmente*, *tremendamente*, etc.) y de los participios en función adjetiva (*aborrecido*, *abrillantado*, *acaramelado*, *achatado*, *afligido*, *agotado*, *agujereado*, etc.), son muchas derivadas en *-ante*, *-ente* (*emocionante*, *ondulante*, *humeante*, *perteneciente*, *predominante*, *preocupante*, *proveniente*, *reluciente*, *resistente*, *serpenteante*, *vacilante*...), o con otros sufijos, prefijos o elementos compositivos (*inusual*, *reemprender*, *reptación*, *subacuático*, *vacilón*, *chuletón*, *subcampeón*, *chatarrería*, *chupa-chups*, *cortacésped*, *limpiacristales*, *autoabastecerse*...). Pero ante todo no se recogen neologismos, bien creados con los mecanismos propios del idioma, tanto léxicos -los ya citados derivación, composición...- (*litrona*, *vitrocerámica*, *senderismo*, *videojuego*, *estadounidense*, *futbito*), como semánticos (nuevas acepciones de palabras existentes: *encimera*, *movida*), bien xenismos o préstamos más o menos recientes que conservan la fonética propia de su idioma (*maître*, *souvenir*, *hámster*, *tatami*, *jiu-jitsu*, *camping*, *gánster*, *gag*, *hooligan*, *ketchup*, *look*, *match*, *party*, *playback*, *play-off*, *rap*, *self-service*, *rugby*, *scout*, *showman*, *skateboard*, *skinhead*, *snack-bar*, *spot*, *surf*...). También están ausentes algunas onomatopeyas (*ja*, *pi*, *plaf*, *ring*, *bum*...) y algunas siglas (GEO, NODO, ATS...).

De todos estos vocablos, 144 (sin contabilizar los adverbios en *-mente*, los participios en *-ado*, *-ido*, ni algunos gentilicios *-accitano*, *alhameña*...-) no se registran tampoco en los cuatro diccionarios escolares mencionados al principio. En total, estas 144 palabras suponen 306 ocurrencias, lo cual indica que su usualidad no es muy grande. El listado lo ofrecemos a continuación:

abs	acarminado	acb
acogible	aguaparc	aikido
aladino	aliblanco	alicortado
alsina	amarronado	antidoping

apiño
apompado
asaltaclases
aterrorizante
atontinado
autodefinir
automarginar
balsón
barbarroja
bombeta
bonico
buf
calvera
cañizada
carrito
cawboy
collejo
corregional
chapuceo
chas
chicane
chicotá
chillerío
chirrioso
chuletero
deportivista
deportivos
descivilizado
destellado
destelleante
doloriente
embarracado
empedregada
enduro
enredante
enrollador
entubamiento
erosionador
escombroso
estresivo
ex-amiga
ex-colegio
ex-director
ex-marido
extraescolar
goticista
grunge
gusanito
hiperpequeño
idolatrismo
impercibible
infiltrante
inflable
inleído
irreglamentario
juncaril
kata
katana
kempo
laboriego
ligoteo
lunero
llovizquear
malafollá
malformado
matear
minicadena
minigira
minitrén
motosierra
mountain-bike
ninja
niñaco
onduloso
palabrotada
paliduzco
palomica
paraviento
patero
paticorto
Pedro
perturbante
petrolífico
pinchito
pisanieves
pop-rock
posatizas
prescindibilidad
presididor
recopa
remontante
repeluznante
resignamiento

respetuosidad	resumista	rockabilli
rockera	salita	semiartificial
separamiento	sombrillero	sub-bético
subplaneta	superagrio	superapuesto
superbién	supercaro	supercolado
supercontaminado	supercontento	supercopa
superdivertido	superempinado	superenrollado
superfuerte	supergrande	superguay
superinteresante	superpoblado	superpoderoso
surfear	toc	tortuorio
tostoneo	toto	trasnochimiento
ultradefensivo	ultraderechista	valentina
voluminosidad	ziun	

Los vocablos que constituyen la lista precedente responden, a nuestro juicio, a diversos factores lingüísticos y sociales que venimos desarrollando a lo largo de este u otros capítulos. Para empezar, destacamos que algunos de esos vocablos son **neologismos** recientes, formados unos mediante la acronimia (ABS, ACB) y otros a través de la derivación de palabras bien de nuestro idioma (*recopa*, *supercopa*) bien del inglés (*rockera*, *surfear*) o compuestas de ambos (*aguaparc*); no faltan los xenismos, y como tales conservan la fonética originaria (*aikido*, *cawboy*, *chicane*, *enduro*, *grunge*, *kata*, *katana*, *kempo*, *mountan-bike*, *ninja*, *pop-rock*). Los **nombres propios** que aparecen han sido usados dentro de expresiones fijas (*Picio*, *Pedro*) o pueden considerarse comunes (*aladino*, *alsina*, *valentina*) de acuerdo con las frases en que se hallan (Cfr. 2.11.4.3.3.). Muchos de los términos podríamos englobarlos dentro de los **localismos** o usos peculiares de determinadas zonas, algunos de los cuales forman parte de terminologías especificas (*aliblanco*, *alicortado*, *bombeta*, *bonico*, *collejo*, *chicotá*, *embarracado*,

malafollá, *motosierra*, *palomica*, *patero*[321], *pinchito*, *tostoneo*, *toto*). Otras palabras, ya sean onomatopeyas (*buf*, *chas*, *toc*), ya se utilicen en el ámbito escolar (*extraescolar*) o fuera de él (*minicadena*), gozan de cierta usualidad en la comunidad lingüística. En la lista se reflejan también algunos usos muy de moda en la sociedad actual, sobre todo aquellos que obedecen a una inclinación pedante, afectada, "amanerada por la palabra alargada, hinchada, hueca" (Gómez Torrego, 1995: 172), como algunos vocablos con el sufijo *-dad* (*respetuosidad*, *voluminosidad*), a los que podríamos añadir otros que responden igualmente al gusto por el alargamiento léxico (*deportivista*). A este mismo estilo "hinchado o pomposo pertenecen también muchos verbos con el prefijo *auto-* para marcar de forma redundante la reflexividad"[322], como *automarginar*, *autodefinir*. El prefijo *súper-* goza hoy de mucha popularidad[323] y ha sido acogido por los jóvenes con cierto entusiasmo, a tenor de la fecunda utilización que de él hacen acompañando tanto a adjetivos como a adverbios (*superagrio*, *superapuesto*, *superbién*, *supercaro*, *supercolado*, *supercontaminado*, *supercontento*, *superdivertido*, *superempinado*,

---

[321] Las palabras *chicotá* y *patero* pertenecen a la terminología específica que rodea al mundo de las procesiones y cofradías de Semana Santa: *chicotá* se llama al período comprendido entre dos paradas de un *paso*, y *patero* es el *costalero* que ocupa una determinada posición bajo la imagen.

[322] Gómez Torrego (1995: 175). El autor aclara que este prefijo es necesario con sustantivos y adjetivos, pero no con verbos, ya que la reflexividad la marca el pronombre reflexivo. A pesar de todo, "el bombardeo de verbos con *auto-* es hoy demasiado llamativo. Verbos como *autocontrolarse*, *autoconfiarse*, *autoestimarse*, *autoprotegerse*, *autodefenderse*, *autoconsolarse*... desprenden un cierto tufo de pedantería" (Gómez Torrego, 1995: 176).

[323] Gómez Torrego (1995: 177-178) da cuenta de la difusión y popularidad de este prefijo, cuyo uso resulta empalagoso, sobre todo con adjetivos y adverbios. E incluso con sustantivos se utiliza en la publicada con más frecuencia de lo deseado: las entidades bancarias anuncian los *supercréditos*, los *superfondos*, las *supercuentas*, las *superlibretas*...; en los comercios es frecuente ver el cartel o anuncio de *superrebajas*, *superofertas* y en el mundo del deporte se habla de *SuperBallesteros*...