TRABAJO FIN DE MÁSTER EN COMUNICACIÓN SOCIAL. ITINERARIO DE INVESTIGACIÓN

Adquisición del español dentro de la competencia comunicativa

Revisión bibliográfica

Alumma: Leonor Celedonia Cano García
Director: Javier Campos Daroca
Co-director: Francisco Joaquín García Marcos

01/07/2013

Revisión bibliográfica de la adquisición de la destreza escrita en edad escolar dentro de la enseñanza del español como lengua materna. Análisis teórico de la enseñanza de la producción escrita y los modelos de investigación llevados hasta la actualidad.

ÍNDICE

1. INTRODUCCIÓN 4
2. PARÁMETRO ORTOGRÁFICO 5
2.1. Fundamentos ortográficos 5
2.1.1. Conocimiento ortográfico y fonológico 9
2.1.2. Alteraciones neuropsicológicas 12
2.2. Enseñanza de la ortografía española 14
2.2.1. Lengua materna 14
2.2.2. Lengua no materna 16
2.3. Simplificación de las normas. Propuestas 18
2.4. Comunicación y ortografía del siglo XXI 22
3. PARÁMENTRO LÉXICO 24
3.1.1. Léxicos frecuentes y léxicos básicos 25
3.1.2. Lexicometría y factores sociales 27
3.2. Disponibilidad léxica 32
3.2.1. Investigaciones en Francia y Canadá 32
3.2.2. Investigaciones en Hispanoamérica 34
3.2.3. Investigaciones en España 36
3.2.3.1. Andalucía 37
3.2.4. Aplicaciones de los estudios sobre disponibilidad léxica 43
3.3. Riqueza léxica 46
3.3.1. Inicios teóricos 46
3.3.2. Estudios hispanoamericanos 48
3.3.3. Estudios en España 49
3.3.4. Aplicación en el aprendizaje de la lengua no materna 50
4. PARÁMENTRO GRAMATICAL 52
4.1. Fundamentos gramaticales 52
4.1.1. Antecedentes: La Brant (1933) 52
4.1.2. Índices de madurez sintáctica: Hunt (1965) 53
4.1.2.1. Índices primarios 55

4.1.2.2. Índices secundarios 56
4.1.3. Memoria inmediata 58
4.2. Estudios de madurez sintáctica 59
4.2.1. Mundo anglosajón 59
4.2.2. Mundo hispánico 61
4.3. Aplicaciones de la madurez sintáctica 68
4.3.1. Modo del discurso 68
4.3.2. Enseñanza lengua no materna 69
5. CONCLUSIÓN 70
6. BIBLIOGRAFÍA 74

1. INTRODUCCIÓN

La presente investigación atiende a la estrecha vinculación entre la comunicación y la enseñanza del español como lengua materna. Ambos componentes son indispensables e inseparables desde el comienzo de la etapa escolar.

El desarrollo de la comunicación en la escolaridad es papel fundamental y transversal para todas y cada una de las asignaturas. Es por ello la necesidad de una revisión actualizada acerca de los métodos de enseñanza del idioma, en este caso el español.

La complejidad del tema a tratar me lleva a presentar en primer lugar un planteamiento estrictamente teórico con vistas a una futura investigación donde se desarrollarán de manera extensa y detallada tanto el corpus como el análisis.

La comunicación es el eje central sobre el que se fundamenta dicha investigación. En ocasiones encontramos serias dificultades entre los escolares a la hora de expresar sus sentimientos o razonar. Los procedimientos sobre adquisición del español dejan de lado la expresión para dar protagonismo a la corrección gramatical. El uso de un correcto español es una herramienta básica no solo para superar las etapas escolares sino para la vida en general.

Con todo ello, se puede resumir que es la lengua la base de nuestra comunicación. Estudiar, conocer, investigar y observar cómo somos capaces de adquirir nuestra lengua es en suma conocer las bases de nuestra comunicación verbal y escrita a través de un sistema de signos establecido.

A lo largo de las investigaciones llevadas a cabo y aquí recogidas, se cuestionan diversos métodos destinados a conocer el desarrollo escolar en cada etapa de aprendizaje del español.

Con el objetivo de acotar el estudio, me veo obligada a diferenciar tres parámetros sobre los que organizar el aprendizaje del español:

- Parámetro ortográfico
- Parámetro léxico
- Parámetro gramatical

La investigación abordará una revisión bibliográfica a partir de estos tres aspectos del español. Se han dispuesto dichos parámetros siguiendo el orden natural establecido por la adquisición de la lengua materna.

En el primer parámetro se abordan estudios sobre el conocimiento ortográfico y fonológico, así como las teorías acerca de la ortografía. Dentro del parámetro léxico serán los estudios sobre disponibilidad léxica y riqueza léxica los que mayor ejemplifican la adquisición del vocabulario. Por último, dentro del parámetro sintáctico, las investigaciones sobre madurez sintáctica vienen a descifrar un gran apartado dentro de la adquisición de la sintaxis española.

La competencia comunicativa es la más integral de todas, puesto que abarca desde la competencia lingüística, que es desarrollada en este trabajo, hasta las diferentes competencias como paralingüística, quinésica, proxémica, pragmática, estilística y textual.

El enfoque de mi investigación viene a desarrollar el apartado de la adquisición de la competencia lingüística, es decir, la adquisición de la capacidad de producir e interpretar signos verbales. Cabe mencionar que el resto de competencias hacen junto con la lingüística que la comunicación sea factible a la hora de comunicar dentro de unos parámetros socio-culturales, puesto que además de manejar una lengua, el hablante debe situarse en un contexto comunicativo dado.

2. PARÁMETRO ORTOGRÁFICO

2.1. Fundamentos ortográficos

A lo largo de la historia, la ortografía ha tenido un gran hueco dentro de las gramáticas tradicionales y ha sido siempre hueso de escolares y no tan escolares. Desde el paso del latín hasta un primer castellano desgastado por el tiempo y el uso, la aparición de una serie de normas ortográficas, que tenían como finalidad la regulación de la buena escritura, no siempre cumplían con el precepto marcado por Alfonso X "El Sabio" de un sistema alfabético.

Esta tradición marcada fuertemente por la etimología nos llega hasta la actualidad no sin polémicas a lo largo de los años, como veremos más adelante. Escribir correctamente español es una premisa general marcada por la correcta aplicación de unas normas ortográficas, que, en ocasiones, lejos de seguir un patrón unívoco, se convierte en un gran pesar para muchos de los hablantes.

La comunicación verbal entre hablantes puede efectuarse básicamente sobre dos soportes: oral y escrito. Si los hablantes de una misma lengua se comunican siguiendo la oralidad, utilizarán una serie de unidades fonológicas dispuestas según unas estructuras para formar en primer lugar sílabas, palabras, oraciones y textos. La comunicación entre ambos hablantes se efectuará siempre y cuando se disponga el texto oral de acuerdo a una estructura y vocabulario correcto y conocible entre su interlocutor, además de una serie de condicionadores propuestos ya por Jakobson a mediados del siglo XX para que la comunicación sea factible entre interlocutores. Si nos trasladamos al papel, o según marca la actualidad, a la pantalla, en ocasiones la comunicación puede verse interferida aun dándose los mismos factores que en la oralidad, puesto que la escritura cuenta además con una técnica complementaria: la ortografía.

Las faltas de ortografía pueden dificultar la comunicación escrita si su insistencia incurre en el texto, pudiéndose considerar como "ruido" al no permitir la comunicación perfecta, pero ¿realmente no se produce comunicación si escribimos "burro" con "v"? este apartado viene a demostrar los estudios y teorías que se han presentado sobre ortografía, algunos a favor y otros en contra de una posible simplificación de las normas ortográficas.

A través de la ortografía también podemos conocer a nuestro interlocutor, o por lo menos tener idea acerca de su estatus social, puesto que como algunos autores señalan (Ávila, 2009; Sánchez, 2009) la generación de errores ortográficos está fuertemente arraigada a un prestigio social. Ávila (2009) incide en la estratificación social marcada por la ortografía del hablante, puesto que el uso de una correcta ortografía es sinónimo de conocimiento y cultura, mientras que cometer errores ya califica al hablante de inculto, no siempre siendo así.

Hoy en día existen muchos profesores universitarios, sobre todo de las ramas científico-técnicas a los que se les "olvida" poner tildes, o simplemente comenten algún error. Es más, cuando nosotros mismos observamos que hemos cometido un error

ortográfico nos invade una sensación de rubor, porque claro está, nosotros conocemos todas las normas ortográficas y no nos podemos permitir rebajar nuestro conocimiento.

Es de esta manera como la ortografía arbitraria ha sido para muchos un verdadero calentamiento de cabeza así como su aprendizaje ha sido demostrado ineficaz en la actualidad.

Muchos de estos problemas sobre catalogación social de los hablantes arrastrados por la evaluación de los textos han seguido casi exclusivamente el parámetro ortográfico. Un texto puede estar compuesto por un vocabulario culto, con un uso sintáctico impecable y una disposición de ideas adecuadas, lo cual implica que el escritor conoce su lengua y su redacción es perfecta; pero si en ese mismo texto encontramos dos o tres faltas de ortografía, la evaluación sería negativa, pues el hablante no conoce la ortografía del español. Esta evaluación explicada *grosso modo* es la que actualmente está imperando en nuestras aulas, dándole prioridad a la correcta utilización de las normas ortográficas ante la correcta escritura:

> Hay que reflexionar sobre la importancia que se le da a la ortografía frente a la necesidad de lograr una alfabetización más eficiente en niños y adultos, por una parte, y la importancia de lograr una mayor calidad en la redacción por otra. (Ávila, 1992: 650).

Aparentemente, aprender a escribir parece una tarea fácil en español, puesto que el niño habla la lengua y conoce los fonemas que de ella consta; así, siguiendo el principio alfabético, basado en la transcripción fonográfica, no habría ningún problema ya que cada fonema estaría asociado a un grafema (símbolo) o representación en la escritura, estableciendo una relación biunívoca entre fonema y grafema. En efecto, el español se considera una lengua fonográfica, a diferencia de otros sistemas de escritura como el semiográfico (inglés) o el ideográfico (chino mandarín), pero existen una serie de excepciones que complican la ortografía española.

No se establece relación absoluta entre *grafema-fonema* y *fonema-grafema* en algunos sonidos para los cuales encontramos dos o más grafías:

- El fonema /k/ puede estar representado mediante las grafías "c", "qu" y "k".
- El fonema /θ/ puede estar representado por las grafías "z" y "c".
- El fonema /G/ puede estar representado por las grafías "j" y "g".

- El fonema /b/ puede estar representado por "v" y "b".
- El fonema /y/ puede estar representado por "y" o "ll", debido a la expansión del yeísmo en casi todas las zonas de habla hispana.

Además, encontramos la grafía "h", la cual no representa a ningún fonema y es el resultado de la antigua "f" aspirada (farina>harina) en muchas ocasiones, y en otras, se utiliza para los casos homófonos (ola, hola). Junto con estas posibles confusiones entre *fonema-grafema* y *grafema-fonema* basta añadir otra serie de normas ortográficas tales como "*antes de p y b va m"* tan recogida en los primeros cuadernos sobre ortografía.

Gignorini, Borzone y Diuk (2003) han clasificado en tres tipos las correspondencias entre los fonemas y grafemas del español:

1. Simples: se establece una relación biunívoca entre fonema y grafema, por ejemplo /t/ para "t".

2. Condicionales: los grafemas "j" - "g", y "z" - c" se utilizan mediante reglas de condición ante vocales. Para el fonema /k/ se utiliza la grafía "c" si es ante las vocales "a", "o" y "u"; en cambio, ante "i" y "e", utilizamos el grafema "k" o "qu" con dependencia etimológica además.

3. Irregulares: por ejemplo la distinción entre "b" y "v", que atiende a razones puramente históricas o etimológicas.

Además, también se considera falta ortográfica el uso inapropiado de la tilde o acento ortográfico. El acento marcado por la fonología en cada palabra está representado en español por una marca diacrítica que atiende a tres reglas de acentuación, las cuales se basan en la distinción entre las palabras llanas, agudas y esdrújulas.

Son muchas las normas que el hablante debe tener en cuenta a la hora de escribir, y en el caso de una nueva palabra que contenga el fonema /b/, para éste será imposible conocer si escribe con "b" y "v" hasta su visualización tal y como señala Cuetos (1989: 80) "solo el aprendizaje directo (visual) permite conocer cuál es la ortográficamente válida y cuáles las desechables" por tanto, uno de los factores base a la

hora de reproducir la escritura va a ser la memoria léxica, como ya se ampliará más adelante.

2.1.1. Conocimiento ortográfico y fonológico

Los procesos de lectura y escritura están muy ligados a los estudios sobre errores ortográficos porque los tres se basan en la fonología del idioma, y en el último caso, en su transcripción grafémica. Es por ello que ha habido una tradición errónea al considerar el desarrollo de la lectura y la escritura como dos procesos similares (Bryant y Bradley, 1980).

Bryant y Bradley (1980) llegaron a la conclusión de que los procesos cognitivos sobre la escritura y lectura son diferentes gracias a un estudio aplicado a alumnos en etapas educativas iniciales, cuyo resultado mostraba que había más niños que podían leer más palabras que escribirlas. Como ya se ha señalado anteriormente, hay que traducir al contexto de la lengua española tales objeciones, puesto que en inglés apenas existe correspondencia *fonema-grafema, grafema-fonema.*

A partir de los años 80, estudios teóricos sobre la adquisición de la escritura como los presentados por Bub y Kertesz (1982), Ellis (1982), Morton (1980), Nolan y Caramazza (1983) y Shallice (1981) aportan la teoría de las dos rutas sobre la escritura (directa e indirecta), expuesta también en los trabajos de Cuetos (1989) y García (2002) entre otros en España.

En el estudio de Signorini *et al.* (2003) encontramos el postulado de Liberman (1988), el cual sostiene que el aprendizaje de la escritura se sitúa en un tercer nivel para el escolar, puesto que el primer contacto con la palabra es a través del sonido (fonología) para descubrir después el significado; con estos pasos, el niño podrá hablar y utilizar las palabras para formar oraciones. Es después cuando aparece la escritura y con ella la hora de pasar de fonemas a grafemas.

En este momento señalado por Liberman (1988) los niños van a utilizar la ruta indirecta o fonológica para transformar en grafemas los fonemas que están acostumbrados a escuchar. De esta manera, los escolares podrán escribir cualquier letra atendiendo al sonido de los fonemas que la componen, puesto que en español existe bastante correlación entre *fonema-grafema*. Siguiendo la ruta fonológica el niño

escribirá la palabra *ueso**, puesto que la grafía "h" no tiene representación fonológica, por tanto, la ruta indirecta llevará a cometer errores ortográficos.

Por otro lado se encuentra la ruta directa o léxica, la cual se basa en una representación ortográfica de la palabra, la cual es recuperada del almacén del lexicón mental. Es decir, se basa principalmente en la memoria y si nos encontrásemos ante una palabra nueva con el fonema /b/ tampoco sabríamos escribirlo correctamente, pues no disponemos de esa palabra grabada en nuestro lexicón mental.

Cuetos (1989) realiza un experimento siguiendo los pasos marcados por Bryant y Bradley (1980). Para ello realiza un corpus formado por 104 alumnos de cursos comprendidos entre párvulos, primero, segundo y tercero de EGB. El test a realizar por parte de los alumnos está compuesto por 40 pseudopalabras, entendidas estas como "serie de letras pronunciables que no tienen representación léxica" (Cuetos, 1989: 75). Las 40 pseudopalabras estaban programadas para poner en práctica las grafías "j", "g", "s", "c" y "t". El experimento tenía una doble aplicación: para la lectura y para la escritura, aquí solo señalaré la última. El método escogido para que los hablantes escribieran las pseudopalabras fue el dictado, método fuertemente criticado por parte de muchos autores (Sánchez, 2009; Ruiz de Francisco, 1994; Tena, 1999).

En contraposición a los resultados obtenidos por Bryant y Bradley (1980) Cuetos señala que los aciertos en las palabras escritas fueron de un total de 38,60 sobre 40. Además los errores ortográficos más frecuentes eran los grafemas "j" con un 48,80% y "g" con un 20%, dichos datos indican la dificultad de los escolares ante la distinción de dos grafemas para un solo fonema. Los errores ortográficos iban en disminución conforme los escolares aumentaban de ciclos escolares, lo cual muestra un avance en el repertorio del lexicón mental.

El escolar va a cometer errores ortográficos al inicio de la escritura, puesto que la ruta primera es la fonológica y los casos de homofonía heterográfica (Signorini *et al.*, 2003) le harán caer en el error, pero poco a poco a medida que el escolar avance en escolaridad, este va a ir utilizando un léxico frecuente que va a ir construyendo su lexicón mental, basado en la reproducción mental de la palabra tras la memorización, lo cual le indicará cómo debe escribirse una palabra.

> Las reglas *fonema-grafema* y *grafema-fonema* permiten al niño hacer uso de la ruta indirecta, mientras tanto se desarrolla la ruta directa que le proporcionará mayor rapidez en la lectura y mejor ortografía en la escritura, pero que necesita mucho mayor tiempo de aprendizaje (Cuetos, 1989: 81).

Un reciente estudio publicado por Jiménez *et al.* (2008: 786-787) señalan también el uso de las dos rutas, fonológica y léxica, por parte de los escolares. Siguiendo a los estudios ingleses sobre adquisición de la escritura, Jiménez *et al.* diferencian tres etapas comunicativas en el escolar:

1. Etapa pre-comunicativa, en la que los niños aun no relacionan sonidos y letas.

2. Etapa alfabética: los niños son capaces de utilizar la ruta fonológica, pues realizan la conversión *fonema-grafema* y *grafema-fonema*.

3. Etapa ortográfica: los niños aplican la ruta léxica.

Estas tres etapas son reconocibles en la adquisición de la escritura para cualquier sistema de escritura, pero debe atenderse a que si el sistema es el fonográfico, los escolares tardarán menos tiempo en alcanzar la ruta léxica. En Jiménez *et al.* (2008) encontramos a Defior y Serrano (2005), quienes sostienen que en las lenguas menos transparentes, como el inglés, los escolares pueden tardar de uno a dos años más en alcanzar el nivel de escritura en comparación con las lenguas transparentes.

Jiménez *et al.* (2008: 787) lleva a cabo un estudio realizado sobre la premisa de que es a partir del segundo ciclo de primaria cuando los escolares ya tienen conocimiento pleno ortográfico y pueden resolver sin problema las dificultades presentes de la ortografía arbitraria, como las palabras homófonas y las excepciones ortrográficas. Partiendo de tal hipótesis, se realiza un análisis tripartito en el que se plantean diferentes búsquedas: conocer en qué momento se desarrolla el dominio de la ortografía arbitraria, analizar los errores ortográficos y comprobar si la producción de los errores está asociada a tareas de escritura tradicionales.

El corpus está formado por 1.045 alumnos con edades comprendidas entre los 7 y 12 años, es decir, alumnos desde segundo hasta sexto de primaria de EGB. El material utilizado ha sido el propuesto por Artiles y Jiménez (2007): *Proescri-Primaria*, Prueba de Evaluación de los Procesos Cognitivos de la Escritura, compuesta por un total de 15 pruebas.

Entre las conclusiones a las que llegan los investigadores, cabe destacar que sí se produce un "dominio de las representaciones ortográficas de las palabras que no están sujetas a reglas desde el segundo ciclo de Primaria" (2008: 792); pero es hasta cuarto de primaria cuando no se observa la aplicación de la ruta léxica, es decir, el pleno conocimiento ortográfico. Entre los errores ortográficos, los más numerosos son la confusión de los grafemas "b"-"v", "h", "c"-"s"-"z"-"x", recordemos que nos encontramos con el seseo canario, característica dialectal de la zona de estudio.

Según los resultados encontrados, se demuestra una tardía aplicación de la ruta léxica de manera justificada, puesto que esta se basa en la memoria y necesitamos dar tiempo a los jóvenes escolares para que vayan adquiriendo su propia "memoria léxica". Por ello, Jiménez *et al.* (2008) inciden en la importancia de una "adecuada instrucción en la ESO" sobre ortografía, puesto la instrucción basada en reglas de Primaria es realmente inapropiada.

2.1.2. Alteraciones neuropsicológicas

Dentro de la Lingüística Aplicada podemos encontrar estudios sobre alteraciones médicas que dan respuesta a según qué comportamientos con el lenguaje. Por ello, la adquisición de escribir y leer también ha tenido su estudio dentro de los parámetros de la neuropsicología o la psicología cognitiva.

Como ya se ha señalado, los procesos que interfieren en el aprendizaje de escribir se basan en dos rutas: fonológica y léxica, pero podemos encontrarnos con hablantes de una lengua que sufran alguna alteración en los procesos cognoscitivos que hacen posible tales tareas. Por ejemplo, podemos encontrar un hablante de una lengua que aparentemente no tenga ningún problema y encuentre dificultades para escribir pseudopalabras, lo cual señala un tipo de alteración neurológica.

La ciencia está cada vez más unida a la lingüística para llegar a conocer cómo son las relaciones cognoscitivas que entrañan cualquier hecho lingüístico para el ser humano. Si conocemos cómo es el funcionamiento de nuestros procesos, podremos también reconocer más a fondo la investigación sobre posibles casos de problemas neurológicos que afectan al lenguaje.

García Orza *et al.* (2002) presenta un estudio sobre alteraciones del procesamiento de la escritura basado en las dos rutas ya expuestas. En su estudio acerca a la investigación dos ámbitos complementarios, como son las alteraciones neurológicas (daños cerebrales) con la capacidad de escribir. Los investigadores también apoyan la teoría de que la ruta fonológica no tiene resultados en una ortografía basada en normas: "La ruta fonológica permite la escritura correcta de cualquier palabra y pseudopalabra, pero su empleo en palabras de ortografía arbitraria da lugar a errores ortográficos" (2002: 287).

A menudo creemos que una persona adulta con un buen nivel socioeconómico y estudios debe escribir sin faltas de ortografía, o puede reproducir gráficamente lo que oye, pero en ocasiones problemas como estos tienen su respuesta en algunos procesos alterados en la mente del hablante y que hasta ahora han sido casi desconocidos.

La investigación de García Orza *et al* (2002: 228-290) propone tres tipos de alteraciones que pueden provocar en el hablante serios daños en el lenguaje:

- Disgrafía fonológica: se trata de un tipo de alteración que se presenta en hablantes con dificultad a la hora de escribir pseudopalabras, lo cual señala que el proceso cognoscitivo de la ruta fonológica sufre una variación. Los hablantes con disgrafía fonológica no pueden representar mediante *grafemas* los *fonemas* que sí son capaces de interpretar.
- Disgrafía superficial: es el proceso contrario al anterior. El hablante sufre una alteración neurológica en la ruta léxica, es decir puede escribir perfectamente pseudopalabras porque no tiene ningún problema para transformar los *fonemas* en *grafemas* (y viceversa), pero sí tiene dificultad para escribir aquellas palabras con ortografía arbitraria, produciendo así constantes errores ortográficos.
- Disgrafía profunda: es la lesión más grave, puesto que ambas rutas, la fonológica y la léxica se encuentran alteradas. Los errores pueden llegar incluso hasta la confusión semántica.

Para ejemplificar un caso de disgrafía, García Orza *et al* realizan un experimento con un paciente dañado por un accidente vascular cerebral, este hecho dejará secuelas importantes en el cerebro que le han afectado al lenguaje. Antes del accidente, el

paciente era un hombre con estudios universitarios y cultura general, lo cual hace pensar que tenía conocimiento ortográfico.

A partir del experimento, los investigadores se dan cuenta que tras los análisis realizados el paciente comete más fallos en la escritura de palabras no-arbitrarias que en pseudopalabras, lo cual se traduce a una disgrafía superficial provocada por unas lesiones cerebrales. El paciente necesita ahora una rehabilitación para ir poniendo en marcha la ruta léxica; para ello, García Orza *et al* proponen ejercicios de asociación visual entre grafía y significado, pistas visuales, etc. pero un alejado aprendizaje basado en normas, pues como se ha demostrado, estas no infieren en la ruta léxica.

2.2. Enseñanza de la ortografía española

2.2.1. Lengua materna

Si nos adentramos en las aulas españolas, observaremos que la enseñanza de la ortografía está relegada a un par de normas dadas en cada unidad didáctica y a unos ejercicios sobre modelos repetitivos. Si aunamos la pésima enseñanza ortográfica con la cantidad de normas dan como resultado lo que observamos cada día, jóvenes con graves errores ortográficos.

El problema está en delimitar la raíz del "problema": ¿la enseñanza, los escolares, los profesores o las propias normas? Más adelante podremos en situación la polémica tradicional sobre la última hipótesis, en esta apartado analizaremos lo que ocurre con las tres primeras suposiciones.

A nadie le ha gustado tener que "memorizar" una serie de normas para luego "acordarse" de tener que ponerlas en práctica. La constante lucha con las reglas de acentuación o de la "b", por ejemplo, con normas como "Las palabras que comienzan con lab- rab- sab- tab- se escriben con "b" están plagadas de excepciones como "lavar, savia, etc." lo cual dejan al alumno en un mar de dudas igual que antes de conocer la regla. Además todas estas reglas se aplican en primaria y se extienden hasta el primer ciclo de Secundaria.

La enseñanza de la ortografía debería dejarse aconsejar por los resultados de los últimos estudios sobre la adquisición de la escritura y reforzar las técnicas ortográficas a

través de una renovación de actividades basadas en lo visual a partir de una edad aproximada de 10 años, cuando los alumnos están dispuestos a recoger y formar su lexicón mental. Los estudios señalan que es solo la ruta léxica la capaz de recuperar el léxico apropiado en palabras homófonas o con ortografía arbitraria, por tanto todas aquellas normas estudiadas desde muy pequeños apenas tienen aplicación, porque al inicio de la escritura, sólo utilizamos la ruta fonológica.

Según lo dispuesto, un sujeto de siete años tiene toda la capacidad para escribir pseudopalabras y cometerá errores ortográficos, por lo que la utilización de técnicas como el dictado, que se apoya en la visualización de errores, no ayudan para nada al fomento de la memoria léxica. Tendríamos que llevar a cabo una renovación total de las técnicas basándose en la visualización y sobre todo en métodos inductivos (Sánchez Jiménez, 2009) aunque debemos ajustarnos a la capacidad cognoscitiva de los alumnos según su nivel de escolaridad. En las primeras etapas los maestros deberían fomentar al alumnado hacia la escritura de manera que los errores ortográficos no sean siempre el centro de interés a la hora de evaluar la tarea, porque como señala Gómez Camacho (2005) uno de los principales alicientes para la didáctica ortográfica debe ser la actitud positiva del alumnado y el "deseo de escribir bien". Otras de las técnicas que ayudan a los escolares a ir acrecentando su lexicón mental serían la escritura y la visualización de un vocabulario frecuente para ellos, por lo cual escribir libremente o con temas amplios que determinen la redacción serán ejercicios que activarán la memoria léxica graduablemente. Estas son algunas de las prácticas que se deberían poner a prueba y sobre todo no confundir la evaluación escrita con la evaluación ortográfica por parte del profesor.

Un dato a tener en cuenta serían las variables dialectales, las cuales suponen a veces un elevado porcentaje de los errores ortográficos (Gómez Camacho, 2005). En ocasiones la ruta fonológica en los primeros años de escritura puede incurrir en errores a los alumnos por analogía con la pronunciación, si esta además es característica por ser "seseante" o "ceceante". En este caso, como hemos observado en el estudio de Jiménez *et al.* (2008) en Canarias, la confusión entre grafemas puede tener más aparición entre los resultados.

Con todo lo conocido hasta ahora, hay autores que señalan la posibilidad de una reforma de la enseñanza de la ortografía, y esta debería ser a partir de tercer o cuarto

curso de primaria, que es cuando empieza a activarse la ruta léxica (Jiménez *et al.*, 2008). Esta enseñanza se prolongaría hasta la ESO y Bachiller, pues los problemas con ortografía llegan hasta los niveles universitarios (Gómez Camacho, 2005).

Gómez Camacho (2005) señala que los universitarios andaluces comenten muchas faltas de ortografía y eso se debe en medida porque en la ESO y Bachiller no se refuerza la enseñanza de la ortografía y no se garantiza la "competencia comunicativa escrita". Este autor sostiene que "no hay un inventario exhaustivo de reglas que constituyan la ortografía del español, ni puede haberlo" (2005: 131) puesto que como señala en su investigación la *Ortografía* de la RAE de 1999 no habla de una serie de normas, sino de "notas orientadoras". Es cierto que cada vez encontramos recomendaciones y nuevas modificaciones, como las publicadas en por la RAE en 2010, que más que esclarecer, dejan al hablante libre a libre elección: "ye" o "i griega", "sólo" o "solo". Gómez Torrego (2011) realiza una síntesis de los nuevos cambios ortográficos estableciendo las diferencias entre las normas anteriores y las actuales.

La solución a todas estas recomendaciones podríamos recogerla en dos parámetros bien diferenciadores: por un lado una reforma sobre las normas creando un "conjunto de normas que regulen la escritura enunciadas de una forma clara y precisa, con pocas excepciones y que se aplicasen a muchas palabras de uso muy frecuente" (Gómez Camacho, 2005) o realizar una simplificación de las normas basándose exclusivamente en el criterio alfabético (Ávila, 1992, 2009).

2.2.2. Lengua no materna

Si el aprendizaje del uso correcto ortográfico ya es un reto para hablantes nativos, la tarea se complica sumamente cuando los alumnos son de lengua extranjera. Frente a otras lenguas como el inglés, la enseñanza de la escritura del español es más fácil por el sistema fonográfico, pero de nuevo encontramos complicaciones en palabras con ortografías arbitrarias y homófonas. Siguiendo el estudio sobre ortografía en clase de ELE de Sánchez Jiménez (2009) la enseñanza de ortografía con técnicas tradicionales, como la memorización de normas y dictados, tiene dos características: la baja motivación por parte del alumnado y la sobrecarga de información en forma de

normas. Además, tampoco se presta mucha atención al aprendizaje ortográfico salvo en los niveles iniciales, cuando la fonética es la protagonista de las primeras clases de ELE.

Este artículo recoge algunas recomendaciones didácticas para poner en práctica en las clases de ELE y muchas de ellas también sería conveniente introducirlas en Primaria y Secundaria. Lo más novedoso e imperante en una clase de ELE, y por experiencia propia, es la incursión de métodos inductivos por parte del alumnado. Para evitar la explicación de una norma y convertirla así en una regla con base memorística, lo interesante sería dejar que el alumnado advirtiese por sí solo a través de una reflexión la norma ortográfica a partir de la visualización de ejemplos de palabras bien escritas. De esta manera, señala Sánchez Jiménez (2009) que el alumno será activo en la clase y recordará la norma más fácilmente sin la necesidad de estudiar, solo con "adivinar".

En el lado opuesto se encuentran técnicas como la visualización de errores, puesto que estos también se pueden grabar en la memoria y tender a una confusión en reproducciones futuras con la forma correcta. Actividades como el dictado tradicional comenten este error. Lo correcto sería a partir de la visualización de palabras escritas correctamente llegar a una explicación del uso de la norma. Como se observa, la memoria léxica está cada vez más unida a la imagen visual y es ella la que se debe fomentar en el aula y fuera de ella. La lectura de siempre ha sido un método eficaz, así como la escritura, pues un estudio realizado por el mismo autor, Sánchez Jiménez (2006), que comenten más errores los alumnos que menor hábito de escritura tienen.

El parámetro léxico, como se verá más detallado en el siguiente capítulo, nos ofrece la posibilidad de conocer mediante léxicos básicos, frecuentes y disponibles, las palabras de mayor uso, por tanto, también se podría aplicar en la enseñanza de la ortografía. Barberá (1988) recoge un vocabulario básico, puesto que si procesamos las palabras de mayor uso el contacto con está hará que se incluyan en nuestra memoria léxica para la recuperación por a través de la ruta ortográfica. Esta aplicación puede utilizarse partiendo de un estudio previo del léxico básico de los alumnos, sobre todo para clases de ELE con fines específicos, por ejemplo *negocios*.

Entre otras de las actividades propuestas por Sánchez Jiménez (2009) cabe resaltar el carácter lúdico que se les debe otorgar a las actividades para potenciar así el aprendizaje de la ortografía ortográfica. Las clases pueden incluir juegos a través de imágenes visuales y representaciones audio-oral, para que los alumnos puedan

relacionar por sí solos a través de la reflexión. En Sánchez Jiménez (2009) se recogen algunas de las propuestas lúdicas de Camps *et al.* (2004) como adivinanzas, anagramas, etc.

2.3. Simplificación de las normas. Propuestas

Como se ha adelantado anteriormente, desde el siglo XIII con la escuela alfonsí el tratamiento del castellano ha ido variando según una serie de normas, hasta establecerse a partir de la Real Academia Española. El pensamiento de Alfonso X "El Sabio" era el de hacer una lengua basada en un sistema alfabético tal y como señaló años más tarde Nebrija en el siglo XV "tenemos de escribir como pronunciamos, y pronunciar como escrivimos; por que en otra manera en vano fueron halladas las letras" cita reproducida en la publicación de Ávila (2009).

Como se observa, la polémica acerca de una adecuación de las normas hacia la correspondencia biunívoca entre grafemas y fonemas se ha convertido en una tradición cuyas raíces remontan al principio de la lengua castellana. Han sido muchas las propuestas aportadas por autores que tenían como fin último la simplificación de unas normas que incurrían en errores ortográficos.

En el siglo XVII las aportaciones de Mateo Alemán, con la publicación en 1609 de *Ortografía castellana* y Gonzalo Korreas (Correas) con *Ortografía kastellana nueva i perfeta* se adscribían a la línea de autores fieles a la transcripción grafémica de los sonidos. Entre los gramáticos más defensores de tales ideas se encuentra Andrés Bello, quien defendió a ultranza en el siglo XIX la simplificación de la escritura española en boga de una menor diferenciación social. Su planteamiento es tan radical que incluso elimina las palabras en las que la "h" sirve para diferenciar homónimos, pues según su teórica, la pragmática es la encargada de diferenciar según el contexto la palabra referida. Otro de sus pensamientos que con mayor valor demostraba la arbitrariedad de la ortografía española es que si atendemos a razones etimológica también encontramos irregularidades tales como *boda* que proviene de la palabra latina *vota.*

Ya en el siglo XX encontramos a muchos escritores también adscritos a unos pensamientos que se alejan de la ortografía académica. Autores como Julio Casares (1941), Gabriel García Márquez (1997) con su "jubilemos la ortografía" o Julio Cortázar (1963), quien incluso introdujo un fragmento en su novela *Rayuela* con una

ortografía ya simplificada, van a proclamarse a favor de una simplificación de las normas.

De igual modo, en el lado opuesto se encuentran autores que no ven tanto problemas como señalan los anteriores, entre los defensores de las razones etimológicas e históricas se encuentran autores como Octavio Paz o Camilo José Cela, quienes opinan que las reglas ortográficas tienen su por qué en la ideología del hablante. Otra razón de ser aportada por autores a favor de las normas ortográficas actuales es idea imposible de realizar un alfabeto siguiendo únicamente los fonemas del español. Ferreiro y Teberosky (1979) o más recientemente Gómez Camacho (2005) señalan esa imposibilidad de transcribir el español debido a las grandes diferencias dialectales del idioma, por tanto, "tendríamos que aprender a escribir como hablan otros" señala Gómez Camacho (2005: 134).

Esta última idea es totalmente refutada por el actual lingüista Raúl Ávila, defensor a ultranza de la necesidad de una simplificación de las normas ortográficas. La conversión de fonema a grafema no debería realizarse siguiendo estrictamente la pronunciación de cada hispanohablante, lo cual daría lugar a diferentes escrituras sobre un mismo idioma, sino simplemente transformar nuestra escritura actual, común para todos, en una escritura basada en la relación *fonema-grafema.* De hecho, el propio Ávila ha creado un programa denominado ORTOFON que retranscribe de la ortografía actual a la fonológica.

Una de las razones por las cual muchos autores están convencidos de la necesidad de escribir tal y como ahora hacemos es la diferenciación de las palabras homófonas como señala Ávila (2009): "la ortografía busca evitar los homógrafos" (Zamudio Meza, 2007). Bello encontraba la respuesta aludiendo a la pragmática pero Ávila además señala el *principio de autorregulación* de la lengua: "la lengua se autorregula por la acción de los hablantes" (2009: 948) y pone como ejemplo en su publicación el hecho de que en América la mayoría de los hispanohablantes son seseantes, pues no distinguen los fonemas /s/ de /θ/, y aparecen homofonías propias como /Koser/, que puede hacerse referencia a la comida, "cocer", o a la ropa, "coser"; así, para poder diferenciar el sentido semántico de la palabra, la propia lengua se ha *autorregulado* utilizando para cada caso dos formar diferentes: /kosinar/ y /kosturear/.

Raúl Ávila (1992) lleva a cabo un estudio de campo para conocer más de cerca los errores ortográficos que comenten los alumnos, y en base a los resultados poder mostrar una simplificación ortográfica. Para el corpus se recogen las variables sociales: nivel de escolaridad (tercero, cuarto, quinto y sexto), sexo, zona (urbano/rural) y nivel socioeconómico (alto, medio y bajo). El número total de redacciones que compone el experimento es de 4.047 textos que después fueron transcritos manualmente señalando si existían errores ortográficos de sustitución (X=Y), adición (>X) y supresión (<X). Hay que añadir que Ávila introduce una nueva terminología: *lapsogramas* (errores ortográficos), *ortogramas* (aciertos ortográficos) y *alogramas* (lapsogramas más ortogramas).

De un total de 426.700 alogramas registrados, el 93,42% son ortogramas y el 6,57% son lapsogramas. El estudio llega a diferenciar dentro de ese 6,57% cuáles son los lapsogramas más comunes:

- letra "h": 22,38%
- fonema /b/: 11,90%
- fonema /j/: 5,87%
- fonema /y/: 4,72%
- fonema /s/: 3,55%

Los tres primeros puestos de lapsogramas se deben a una "inconsistencia de las normas académicas" (1992: 653), mientras que la explicación de los lapsogramas con los fonemas /y/ y /s/ es por el rasgo dialectal de los mexicanos, seseantes y yeístas.

En su estudio también se demuestra que conforme los escolares van avanzado en escolaridad se produce una disminución en los lapsogramas: desde un 8,57% obtenido en tercero hasta un 4,67% obtenido en sexto, siendo este último año el que mayor diferencia entre todos se encuentra, debido a la mayor edad de los escolares y a la práctica de la ruta léxica. En cuanto a la variable sexo, las niñas obtienen 5,66% lapsogramas frente a los niños, con 7,79%. Pero quizá sea la variable sociocultural la más significativa puesto que los niños con un nivel alto han obtenido tan solo 2,83%, los de nivel medio, 6,78% y los de nivel bajo 7,17%.

Ávila (1992) confirma con estos resultados las cuatro hipótesis que marcaba al inicio de su estudio: a partir de tercero los lapsogramas irán en disminución, las niñas producirán menos lapsogramas que los niños, se encontrarán menores resultados de

lapsogramas en los niños de las zonas urbanas, por el mayor contacto escrito existente, y el nivel socieconómico alto obtendrá menor porcentaje de lapsogramas.

Ya en esta publicación Ávila pretende dar un avance de la simplificación que más tarde detallará en 2009. Para llevar a cabo un proceso de escritura alfabético Ávila (2009: 958-959) se basa en tres normas a las que él las denomina: alfa (América), Beta (Andalucía y América) y Gama (Castilla). A partir de los rasgos fonéticos de las tres normas dialectales, Ávila señala que habría que coger 25 "inferfonemas" sobre los que determinar 25 grafías:

// a e i o u b p f d t s z y ch sh g k j m n ñ l ll r rr//

Aunque las grafías utilizadas sean las 25 anteriores, se seguiría manteniendo el alfabeto latino internacional para utilizarlo como ordenaciones de bibliotecas, etc. Las modificaciones que este sufriría sería la omisión de las grafías: c, h, q, v, w.

Las propuestas planteadas por Ávila (2009) se resumen en once postulados:

- Las reglas de acentuación se mantienen según la regla actual.
- Se suprimen las grafías: c, h, q, v, w.
- La z se mantiene.
- La grafía ll eliminada del alfabeto internacional se mantiene para el fonema líquido palatal.
- Para el fonema /k/ se utiliza la grafía "K" ante cualquier vocal: "ka, ke, ki, ko y ku". Lo mismo ocurre para el fonema /z/, el cual se utiliza para "za, ze, zi, zo y zu".
- El fonema velar sonoro /G/ no necesita de la "u" ante "a, e, i", por ejemplo: *gerra¸* esta simplificación lleva a la omisión de la diéresis en palabras como: *cigueña* o *linguística.*
- El fonema velar sordo /x/ se reproduce gráficamente ante "a, e, i, o, u" quedando palabras como: *jamás, jente, joven, jigante* o *juntos.*
- La grafía "h" solo se mantiene en casos de homonimia: *ola, hola,* en el resto de palabras, como *uevo* se omite.
- La grafía "ñ" se mantiene como en la actualidad, aunque no se encuentre en el alfabeto internacional.

- El grafema "sh" se incorporará para representar al fonema palatal sordo continuo de extranjerismos, como *show, shampoo.*
- Los nombres propios se pueden seguir escribiendo conforme a las reglas ortográficas actuales, lo cual no modificaría nombres, apellidos, países, ciudades, etc. Se podría seguir escribiendo *Cano* - /kano/, en lugar de *Kano* u optar por la última forma.

2.4. Comunicación y ortografía del siglo XXI

Una simplificación de las normas no resulta una modificación tan utópica tras comprobar que algunas lenguas han realizado ya modificaciones en sus estructuras ortográficas. Un ejemplo lo encontramos con la reforma del portugués realizada en 1990 en el postulado "Acordo ortográfico da lingua portuguesa" el cual se puso en práctica en 2008.

Un cambio más significativo fue el paso del alfabeto otomano al latino producido en 1928 en Turquía, dentro de una de las reformas de Attaturk: "Ley sobre la Adopción y aplicación del Alfabeto Turco", la cual cambió todas las grafías sobre la misma fonología del turco. Si cambios tan trascendentales han sido posibles, una simplificación de las normas ortográficas españolas tendría su cabida en la actualidad, cuya finalidad sería la realización por parte de todos de una correcta escritura, sin que esta desprestigiara a nadie por cometer faltas ortográficas.

De hecho, el desarrollo global en el que vivimos hace que las lenguas estén en contacto y los hablantes tengan un acercamiento sobre su lengua desde diferentes aspectos dialectales o plataformas de soporte.

Los avances tecnológicos han sido una puerta de acceso inmediato a la comunicación y el lenguaje se está viendo alterado por sus propios hablantes, quienes son capaces de construir su propio "sistema ortográfico" para cada una de las lenguas que componen el mundo.

La era de la telefonía móvil imponía un número determinado de caracteres al usuario para poder escribir un SMS (Short Message Service), lo cual determinó una serie de "simplificaciones" por parte de los hablantes para que en un mismo mensaje de

texto se llegara a comunicar con el menor número posible de letras. Este hecho, tan recurrido por millones de personas al día supuso para cada lengua un nuevo código gráfico *informal* entre los hablantes. El cometido final, comunicar, era posible gracias a un acortamiento de las palabras, y de nuevo la polémica sobre esta forma del lenguaje subía a la palestra. El miedo que infundía a los fieles del sistema actual ortográfico era que esta nueva "moda" de escribir llegar a la vida académica y se impusiera finalmente.

La comunicación del siglo XXI exige a la lengua que el principio de economía lingüística se ajuste a los nuevos medios de comunicación. A través de la llegada de internet y la creación de las redes sociales, como herramientas de comunicación, se solidifican aquellas modificaciones ya impuestas por los SMS, tales como *x* (por), *dsd* (desde) o *q*, *k*, (que). Pero con ello no significa que ahora todos los usuarios de las redes sociales no escriban "como deben", pero sobre todo el colectivo joven es que ha impuesto un sociolecto marcado por el factor edad.

Torrego González (2011) ha realizado un estudio reciente sobre la escritura de los jóvenes en una de las redes sociales más utilizadas por ellos en la actualidad: *Tuenti*. En este estudio se observa que la escritura de los jóvenes, entre edades comprendidas de 14 a 18 años, se basa más en la oralidad que en la forma correcta de escribir, alejándose en todo momento de las normas académicas impuestas.

Las conversaciones en las redes sociales son en realidad transcripciones de posibles conversaciones instantáneas, por tanto, muchos usuarios adoptan un rasgo de oralidad en su escritura. En el extremo opuesto encontramos el correo electrónico, el cual tiene un carácter más formal y ha sustituido a la carta ordinaria, en él la oralidad no ha permeabilizado tanto como en el resto de comunicaciones a través de internet.

En cuanto a la polémica sobre si trascenderá y perjudicará a los jóvenes este uso del lenguaje escrito tan radical entre sus chat virtuales, Torrego González señala: "al tratar el tema de la ortografía en las redes sociales, habría que plantearlo como una variedad diafásica del habla familiar, cuyo uso no se extenderá a otros contextos comunicativos" (2011: 39). Torrego González también incluye en su publicación opiniones totalmente contrarias a la defendida por ella, como la de Yus (2002): "la *antiortografía* sí ocasionará graves problemas en la competencia ortográfica". Ambas aportaciones demuestran que de nuevo nos encontramos un tema de debate lingüístico.

3. PARÁMENTRO LÉXICO

3.1. Fundamentos léxicos

En cuanto al parámetro léxico, también se han sucedido multitud de estudios destinados a conocer el vocabulario disponible de una comunidad de hablantes. Como veremos, los métodos llevados a cabo se han valido de la estadística lingüística, a través de fórmulas capaces de combinar sucesivas variables para hacer del estudio un fiel reflejo de la realidad.

El desarrollo de la lexicometría ha impulsado diferentes campos léxicos como la disponibilidad léxica y riqueza léxica, términos diferentes que tienen como denominador común conocer el léxico de sus hablantes. Los inicios de dichas investigaciones se remontan a mediados del siglo XX, cuando la proliferación de listas de vocabulario recogidas tras análisis de textos comenzó su andadura. La principal aplicación de los estudios sobre disponibilidad y riqueza léxica ha sido la enseñanza de la lengua materna en primer lugar, aunque como veremos adelante, esta disciplina ha llegado a incurrir en otros campos lingüísticos.

En el presente trabajo, se tratarán de forma separada disponibilidad léxica y riqueza léxica, puesto que sus estudios han sido muy delimitados y existe una gran diferencia entre ellos, la disponibilidad léxica ha sido centro de interés para muchos estudiosos desde la aparición de la léxico-estadística, mientras que el desarrollo de la riqueza léxica ha sido tardío.

El inicio del desarrollo de la disponibilidad arranca con unos primeros estudios sobre listas de frecuencias y vocabulario básico a partir de los años 30.

Tuvieron que pasar más de cincuenta años para que López Morales distinguiera de forma sencilla las líneas de investigación que definieron en los primeros años de la aparición de la léxico-estadística. El capítulo 6 de *Enseñanza de la lengua materna* (1986) pone de manifiesto los cuatro procedimientos que se han desarrollado para medir estadísticamente el vocabulario de una determinada comunidad. López Morales diferencia cuatro estudios:

- *Diccionarios normativos*: en ellos se intenta recoger la mayor cantidad del caudal léxico, por tanto se convierte en un "depósito" en el que tiene cabida todo tipo de palabras: arcaísmos, regionalismos, tecnicismos y cultismos, los cuales no forman parte del vocabulario utilizado en la cotidianeidad (1986: 54).

- *Listas de frecuencia*: han sido los estudios desarrollados por los pioneros en la lexicometría (Francia, años 50). Para López Morales las listas de frecuencia recuentan vocablos según su frecuencia de aparición, por tanto, se trata del vocabulario de uso cotidiano.

- *Léxico básico*: el léxico básico combina la frecuencia de aparición (método anterior) con la dispersión, por tanto, "nos ofrece un perfil muy adecuado del vocabulario usual de una comunidad lingüística dada, lo que sin duda constituye un valiosísimo instrumento para la planificación léxica de la escuela" (1986: 62).

- *Léxico disponible*: surge gracias a la revisión de los métodos anteriores, puesto que sus limitaciones eran acusadas. El léxico disponible está formado por el caudal léxico aplicado a una situación dada, por ello se desarrolló a partir de los centros de interés.

Los estudios del léxico básico y del léxico disponible han sido utilizados como métodos complementarios para conocer el léxico de una comunidad lingüística:

> El léxico básico trabaja a base de frecuencia ponderada con la dispersión para ofrecer índices de uso (...) El léxico disponible trabaja con frecuencia compensada con el grado para dar un índice real de disponibilidad (...) Las palabras usadas normalmente en cualquier contexto comunicativo se recogen en el léxico básico; aquellas que sólo se usan si hablamos de un tema determinado aparecen en el léxico disponible. Ambos exámenes nos dan el verdadero perfil léxico de la comunidad lingüística (López Morales, 1986: 66).

3.1.1. Léxicos frecuentes y léxicos básicos

Partiendo de la ordenación metódica propuesta por López Morales, los inicios de los estudios de vocabulario empezaron por las propuestas estadísticas de los léxicos frecuentes y los léxicos básicos. El desarrollo de la disponibilidad léxica se dará años

más tarde de la mano de Gougenheim *et al* (1967), por ello se dispone en el siguiente apartado.

Siguiendo a Mateo García (1998) y su repaso histórico sobre los antecedentes de la disponibilidad léxica llegamos al año 1921, en el que Thorndike publica *Teacher's Word Book*, en el que se recogen 10.000 palabras inglesas con mayor frecuencia dentro de un estudio de análisis de textos. Pocos años después, Buchanan (1927) publica *A Graded Spanish Word* con 1.200.000 entradas recogidas a partir de cuarenta textos, la recogida se dividió en siete áreas temáticas combinándose la frecuencia y el rango de aparición.

Verlée (1934) critica los estudios realizados hasta ahora sobre frecuencia léxica puesto que existen expresiones no recogidas que forman parte del vocabulario usual y frecuente de la comunidad de habla.

Las distintas publicaciones sobre análisis de vocabulario tenían en común la base literaria, pues los textos de los que se valían para recoger su léxico eran en su mayoría literarios y escritos en lengua francesa. Algunos de los autores que siguieron fueron Vander Beke (1935) o Aristizábal (1938), este último utilizó para su recuento palabras no procedentes de textos literarios, sino de escritos por niños y adultos.

Hasta los años 50 aparecen obras en lengua española, en 1952 el Consejo Superior de Enseñanza de Puerto Rico publica *Recuento del Vocabulario Español*, de Rodríguez Bou, iniciándose así los estudios léxico en español. Un año más tarde, García Hoz publica *Vocabulario usual, común y fundamental*, cuyo objetivo es la aplicación de un vocabulario básico español en la enseñanza de la lengua materna.

A finales de los años 50 ya contábamos con importante aportaciones investigadoras en el campo léxico de la lengua, por tanto, se fue desarrollando paulatinamente una serie de métodos estadísticos para definir mejor la realidad lingüística de los hablantes. Mateo García señala la importancia de la publicación en 1959 de Guiraud: *Problèmes et mèthodes de la stadistique linguistique,* obra que intenta poner orden a la estadística lingüística publicada hasta el momento, o siguiendo a García Marcos "primera gran obra de síntesis de la estadística lingüística" (1985:23). En ella, Guiraud aúna los términos: estadística y estilística literaria, dado el origen de la mayoría de las investigaciones presentadas hasta el momento. Por otra parte cabe

señalar la distinción entre vocablo y palabras, la cual años más tarde dará lugar al coeficiente de riqueza léxica. Para Guiraud, el vocablo son las "palabras diferentes" que contiene un texto, mientras que las palabras son el cómputo total de términos ("palabras totales") que aparecen (1959: 9).

Juilland publica una serie de trabajos sobre frecuencias bastante exhaustivos en 1964, entre ellos cabe destacar *Frecuency Dictionary of Spanish Words* con Chang-Rodríguez, estudio a partir de 5.024 unidades sacadas de un corpus de 500.000 unidades, dividas estas en cinco mundos. Lo destacable de esta publicación es el nuevo método utilizado, el cual combina la frecuencia con la dispersión. Juilland utiliza el "índice de uso" de 0 a 1, el cual se calcula multiplicando la frecuencia de aparición por la dispersión de cada palabra.

A partir de los años 60 la sucesión de publicaciones entorno al léxico básico y frecuente fue en auge gracias la evolución de la léxico-estadística y la determinación de la disponibilidad léxica. Como señala Mateo García, la distinción entre estudios hoy tan delimitados no estaba clara, pero sí se pone de manifiesto la complementación entre ellos, "la convergencia de ambos en el *léxico fundamental* de una determinada comunidad de habla sea considerada como uno de los mejores baluartes con los que puede contar una tarea de planificación lingüística" (Mateo García, 1998: 37-38), puesto que la aplicación de los estudios estadísticos del léxico ha estado ligada siempre a la enseñanza de lenguas.

3.1.2. Lexicometría y factores sociales

La aparición de la estadística léxica o lexicometría ha supuesto un adelanto a la sociolingüística (García Marcos, 1999: 67) puesto que desde el primer autor que dispuso el análisis del índice de disponibilidad léxica, Gougenheim *et al.* (1964), hasta nuestros días los factores sociales han sido indispensables para medir la realidad lingüística y ahora social de los hablantes.

Remitiéndonos al comienzo del presente estudio, López Morales (1986) señaló cuatro maneras diferentes de analizar el vocabulario, a partir de los estudios arrojados hacia la mitad del siglo XX. Entre esas cuatro maneras, los primeros estudios solo aportaban listados de frecuencias de aparición de vocablos, o siguiendo a Dimitrijévic

(1969) *unidades léxicas*, y fue en 1964 cuando Gougenheim *et al.* introdujeron la correlación social y su interpretación, formando así la base de la léxico-estadística.

La lexicometría tiene como unidad básica la *unidad léxica*, entendiendo esta como una serie de "unidades independientes categorizadas en signos lingüísticos" (Gómez, 2005: 44), puesto que la unidad léxica dispone de dos caras: por un lado la forma léxica, la representación por medio de la palabra, y por otro, el significado unitario, la semántica (Cerveró y Pichardo, 2000: 192).

Hasta 1964 las investigaciones sobre el léxico se basan en listas de frecuencias y recuentos de palabras para observar las incidencias más notorias, pero la metodología empleada dejaba atrás el factor social, introducido por Gougenherim *et al.*, lo cual supuso un cambio radical en los estudios léxicos. El "índice de estabilidad psicológica" es capaz de medir las diferencias de vocablos utilizados según la dependencia geográfica.

Un trabajo ya citado anteriormente dada su importancia dentro de la lexicometría es el presentado por Mackey *et al.* (1971) donde se estudian las diferencias lingüísticas entre el francés de Canadá y el de Francia, lo que parece apoyar drásticamente el índice de estabilidad psicológica.

Esta aplicación también ha sido llevada a cabo por Mateo (1998), en su estudio almeriense, donde por ejemplo había diferencias notorias en las palabras recogidas en el centro de interés "oficios y profesiones", donde "marmolista" se situaba entre los primeros rangos dentro de Macael, comarca del mármol.

Atendiendo a las palabras de García Marcos, el índice de estabilidad psicológica es un "instrumento analítico apto para calibrar las repercusiones y el grado de penetración real de la estandarización sociocultural y de sus consecuencias lingüístico-verbales" (1999: 70).

Poco a poco la léxico-estadística fue definiéndose y gracias a la aportación de Müller (1968) se fueron produciendo investigaciones en las que el rango de aparición de las unidades léxicas dentro de las palabras era un factor decisivo para su recuento. De esta manera, las simples listas de frecuencia se convirtieron en estudios sobre disponibilidad donde además de la frecuencia y dispersión, tenían cabida la posición.

La aparición de la dispersión en la escena léxico-estadística marcó un paso de avance en la disciplina: su función era determinar la estabilidad de la frecuencia a través de las categorías establecidas, gracias a lo cual los mundos léxicos, entidades homogéneas y diferenciadas entre sí, se convertían en las categorías representativas y básicas de cualquier ponderación (López Morales: 1983: 211).

A partir de los años 80, se llevó a cabo la búsqueda por una fórmula estadística capaz de medir diferentes frecuencias absolutas (por palabras y centros de interés) así como aunar la posición de cada palabra. Lorán y López Morales (1983) presentaron un "coeficiente con valor inferior a la unidad (valor lamda) que debería multiplicarse por la frecuencia alcanzada por la lexía en cuestión" (López Morales, 1996: 247). Para ello, lamda no actuaba en las apariciones del primer puesto, sino a partir del segundo, tercero, cuarto, etc. cuando el coeficiente aumentaba su elevación al cuadrado hasta la n potencia (Lorán, 1987).

A partir de la formulación de Lorán y López Morales, siguieron revisiones sobre la fórmula (Butrón 1987, 1991) al observar que el valor lamda quedaba neutralizado a partir del vigésimo tercer puesto de aparición de la lexía, con lo que las palabras registradas a partir de ahí, obtendrían valores similares.

Fue en 1987 cuando aparece la formulación más utilizada en los últimos estudios de disponibilidad léxica (Mateo García 1996, 1998; Muñoz Hernández, 2005, entre otros). Los autores mexicanos, López Chávez y Strassburger Frías presentaron una fórmula basada en un análisis de una matriz de vectores en la que participan:

- Frecuencia absoluta de cada palabra en cada posición.
- Frecuencia absoluta de la palabra: sumatoria de las diferentes frecuencias
- Número de sujetos participantes.
- Número de posiciones alcanzadas en el centro de interés.
- Posiciones en las que fue dicha la palabra.

$$D(Pj) = \sum_{i=1}^{n} e^{-2.3\left(\frac{1-1}{n-1}\right)} \times \frac{fji}{I_1}$$

en donde

n	=	máxima posición alcanzada en el centro de interés en esta encuesta.
i	=	número de posición en cuestión.
j	=	índice de la palabra tratada.
e	=	número natural (2.718181818459045).
fji	=	frecuencia absoluta de la palabra j en la posición i.
I_1	=	número de informantes que participaron en la encuesta.
D(Pj)	=	disponibilidad de la palabra j.

(López Morales, 1991)

El método llevado a cabo para la recogida de las encuestas de Osorio Mena (1986) ha sido el propuesto por el proyecto panhispánico y utilizado generalmente en los estudios del léxico disponible. El cuestionario estaba formado por los centros de interés a investigar y al tratarse de una prueba asociativa, las listas eran abiertas, es decir, no se imponía un número total de lexías, sino un tiempo de dos minutos para cada centro de interés. Este nuevo modelo ya consagrado, dista bastante con el utilizado en los primeros estudios sobre disponibilidad léxica, como Michéa (1953), en el que sí aparecía un número determinado de lexías a rellenar, en total veinte para cada centro de interés.

La acotación del tiempo ha sido estudiada en los estudios realizados sobre enseñanzas de lenguas extranjeras. Frey Pereyra (2007), desde la universidad de Alcalá de Madrid, ha estimado dejar sólo un minuto y medio de tiempo por centro de interés. Esto se debe al trabajar con alumnos de lengua extranjera, en los que ocasiones se había observado que a partir de dicho tiempo dejaban de escribir o realizaban breves descansos para pensar, lo cual no supone una actividad asociativa espontánea, sino reflexiva (2007: 369).

Las variables social se han ido definiendo cada vez más hasta llegar a reproducir lo más real posible los datos arrojados por las investigaciones. Se han utilizado muchas variables sociales con las que contrastar las variables lingüísticas, pero cabe reseñar

cinco de ellas como las más utilizadas y distinguidoras: sexo, edad, nivel sociocultural, geográfica y tipo de centro.

El factor sexo no se ha incluido desde los inicios de la disponibilidad léxica con Gougenheim *et al.* (1964), sin embargo sí se contempló en su estudio la distinción geográfica urbano/rural. Es a partir de los años 80 cuando las investigaciones comienzan a valorar dicho factor y contribuciones como las de Román (1985), Echevarría (1987) en Puerto Rico comenzaron su andadura. García Marcos y Mateo (1995) observaron que el factor sexo influye en la capital almeriense, puesto que las disponibilidades más elevadas en cada uno de los dos grupos eran muy distantes: los hombres obtenían mayor disponibilidad en el centro de interés "campo", mientras que para las mujeres lo eran "la alimentación" o "las partes de la casa".

En cuanto a la edad, esta no ha sido muy estudiada a lo largo de la historia sobre disponibilidad léxica, porque por lo general, se suele estudiar entre grupos homogéneos. López Chávez (1995) intenta estudiar la evolución de la disponibilidad léxica a partir de la escolaridad en alumnos mexicanos; por otro lado, Fuentes González (1994) demostró en su estudio sobre el léxico cañero-azucarero de Motril que sí aparece una distinción entre el vocabulario de diferentes estratos generacionales, donde los informantes de una media de edad de treinta años obtenían el mayor índice de lexías.

Otro factor estudiado recientemente ha sido el nivel sociocultural, para ello es necesario acudir a Samper y Hernández (1997) y su tabulación para recoger los datos de cada informante. La encuesta sobre datos de situación económica y nivel de vida puede parecer algo incómoda según su formulación, por ello, la tabulación presentada por Samper y Hernández se ha consagrado como el método más eficaz para recoger la realidad sociocultural que rodea al informante. La encuesta recoge datos acerca de la formación académica de los padres y su ocupación profesional, sobre un baremo de uno a seis. Se realiza dos encuestas, una para el padre y otra para la madre, por tanto, se puede recoger hasta un total de 24 puntos o el mínimo de 4. Según la puntuación obtenida, se procede a la distinción de los tres niveles socioculturales definidos:

- Alto: entre 20 y 24 puntos.
- Medio: entre 9 y 19 puntos.
- Bajo: entre 4 y 8 puntos.

Algunos investigadores han utilizado este método tal y como se propuso en 1997 (Mateo García, 1998) o bien han modificado algunos factores, como es el caso de Reyes Díaz (2008) y su estudio sobre riqueza léxica canaria.

La distinción geográfica clásica entre rural/urbano fue una de las estudiadas en los inicios de la disponibilidad, puesto que se presentaba como una de las variables con mayor capacidad de distinción entre los informantes. También Román (1985) la estudió en Puerto Rico entre escolares de áreas urbanas y rurales, así como Samper y Hernández (1997), los cuales aportan que dicha distinción no es significativa en una zona tan limitada como las Islas Canarias. Los datos de Mateo García (1998) también señalan la poca significación entre los resultados obtenidos, aunque sí se puede apuntar una mayor especialización del vocabulario en la zona del litoral este, puesto que uno de los índices con mayor disponibilidad sería "trabajos del campo y el jardín", sin duda influido por su contorno geográfico y su economía basada en la agricultura extensiva.

El tipo de centro escolar sí se ha definido como una de las variables más significativas a la hora de analizar el léxico de los jóvenes. Además, es un valor recogido por Humberto López Morales para el estudio panhispánico del léxico, aunque los estudios demuestran que es menos definitorio que la distinción sexo o nivel sociocultural.

3.2. Disponibilidad léxica

3.2.1. Investigaciones en Francia y Canadá

Los estudios de disponibilidad léxica empezaron en Francia a partir de la confección del *Français élémentaire*, publicado en 1958 por Gougenheim, Michéa, Savageot y Rivenc. El vocabulario recogido tiene un total de 1.475 palabras computadas mediante listas de frecuencias cuyo objetivo era la construcción de un vocabulario elemental del francés para la enseñanza a extranjeros. Lo importante de dicha investigación, fue que sus autores se dieron cuenta que el recuento de la frecuencia no recogía términos cotidianos como *métro, lettre, autobus* y *timbre*, debido a una insuficiencia metodológica (López Morales, 1996: 245).

En 1953 Michéa plantea la definición de palabras frecuentes y palabras disponibles, entendida por la tradición hasta nuestros días. Las palabras disponibles se entienden como el caudal léxico dado en una situación comunicativa dada y se desprende además una noción mental sobre el lexicón disponible:

> Existe en el lexicón mental una serie de términos que no se actualizan a menos que se necesiten para comunicar una información muy específica. Se trata de un léxico disponible, cuyo estudio no podría emprenderse manejando frecuencias, porque este factor es pertinente en el caso de las actualizaciones léxicas efectivas, no potenciales, no de disponibilidades que quedan sin manifestación en un corpus dado (López Morales, 1996: 245).

La metodología lleva consigo un cambio a partir del conocimiento del léxico disponible, puesto que las listas de frecuencias no daban respuesta a las necesidades de la lingüística. Las palabras recogidas se desarrollan en "sectores semánticos" (Michéa, 1953), mundos temáticos sobre los que encasillar el vocabulario. La reedición de 1964 de *L'Élaboration du Français Fondamental* ya utilizó los 16 centros de interés clásicos[1] de los estudios sobre disponibilidad. En la investigación se recoge una muestra de 904 encuestados con una edad comprendida entre 9 y 12 años de ambos sexos. Los requisitos a rellenar en los 16 campos de interés eran veinte palabras para cada uno de ellos en los que se mantenga una coherencia interna y sean universales. Con este tipo de metodología las pruebas asociativas fueron ganando terreno a las listas de frecuencia, dada su reacción espontánea con los campos temáticos. En la actualidad, las condiciones de reacción han cambiado para las pruebas, en lugar de imponer un número determinado de palabras, se deja un tiempo estimado de actuación, que suele estar alrededor de dos minutos para cada centro de interés.

Años más tarde, Müller (1968) desarrollará el inicio de la lexicometría o lingüística estadística al introducir el orden de aparición de las palabras en las listas, puesto que las palabras situadas en los primeros puestos indican además una mayor cercanía. Este apartado se desarrollará en el denominado "Lexicometría y factores sociales".

---

[1] *las partes del cuerpo humano; la ropa; las partes de la casa {sin muebles); los muebles de la casa; los alimentos y las bebidas de las diferentes comidas del día; los objetos colocados en la mesa para la comida; la cocina: sus muebles y utensilios; la escuela: muebles y material escolar; la calefacción y la iluminación; la ciudad, el campo; los medios de transporte; los trabajos del campo y el jardín; los animales; los juegos y las distracciones,* y *las profesiones y los oficios.*

Francia marcó los inicios de las pruebas de léxico disponible gracias al trabajo de Gougenheim *et al.* y también dejó constancia de un modelo de investigación que será llevado a cabo en Canadá para contrastar la lengua francesa en ambos países.

Mackey *et al*. (1971), del Círculo Lingüístico de Quebec, ajustó la metodología llevada a cabo por los franceses para poder comparar los resultados obtenidos, puesto que uno de los objetivos según Mateo García era conocer las diferencias culturales de ambos países a través del vocabulario obtenido mediante pruebas de disponibilidad léxica (1998: 40). La aportación de Mackey *et al.* es necesaria, puesto que Dulong presentó un estudio sobre disponibilidad realizado en la región de Manitoba con diferente metodología empleada, por lo que no se podrían comparar los resultados con los estudios anteriores venidos de Francia.

Las primeras investigaciones fueron fruto del Círculo Lingüístico de Montreal en 1959, en la región de Acadie (este de Canadá). Lo más destacable de este trabajo es que existe una diferencia sociocultural dada entre el francés de Francia y el de Canadá. Mateo García (1998: 40) resalta como dato curioso la diferencia entre los puestos recogidos para las palabras *cinéma* y *theâtre*, que reciben el puesto número uno y tres respectivamente en Francia, mientras que en Canadá se encuentran en el ciento treinta y ocho y doscientos cuarenta y dos.

3.2.2. Investigaciones en Hispanoamérica

El primer estudioso que llevó a cabo la disponibilidad léxica con fundamentos estadísticos fue Humberto López Morales (1973). Su investigación llevada a cabo en San Juan de Puerto Rico sigue las aportaciones de los autores anteriores. Uno de los puntos que llevaron a López Morales a investigar acerca del vocabulario de los puertorriqueños fue la teoría del déficit de Bernstein (1971/1989) basada en la inserción de los factores sociales, sobre todo el socioeconómico en el éxito escolar. Para este primer estudio, López Morales consideró tres factores sociales: sexo, nivel escolar (primero, tercero y quinto de escuela primaria) y nivel socioeconómico (medio y bajo). Los centros de interés utilizados fueron diez*: alimentos, animales, la casa, la cocina, el cuerpo humano, materiales de construcción, muebles, naturaleza, la ropa* y *transportes*. De las 6.841 palabras recogidas, sí se aprecia una diferenciación sociocultural, aunque López Morales (1979) utiliza los dieciséis centros de interés ya clásicos de los franceses

y advierte que teoría del déficit por sí sola no tiene cabida, sino que hay que tener en cuenta las diferencias sociodialectales (Mateo García, 1998: 42).

También en Puerto Rico se llevó a cabo el segundo estudio sobre disponibilidad léxica en lengua española, autores que seguirán la estela marcada por el profesor López Morales. Román (1987) analiza la comarca de Dorado siguiendo los postulados de López Morales. En 1987 aparece una obra con una metodología distinta, se trata de Butrón quien publica una investigación marcada por su fórmula metodológica. Cuatro años antes de su publicación, Lorán y López Morales (1983) realizan una fórmula distinta para medir la disponibilidad léxica[2]. A mediados de los años 80 surgen nuevas aportaciones matemáticas que harán diferente a la contribución de Butrón: López Chávez y Strassburger Frías (1986), Echeverría (1985) entre otros.

Por estos mismos años, en México aparecen publicados una serie de investigaciones que tienen como tema principal el estudio del léxico disponible, entre ellas, cabe destacar las aportaciones de Justo Hernández (1986), Canizal Arévalo (1987) y Ruiz Basto (1987), esta última destinada al análisis del léxico en enseñanzas superiores.

En 1988 López Chávez y Strassburger Frías van más allá del análisis cuantitativo del léxico y pretender relacionar el parámetro del léxico con las estructuras semánticas, eran los inicios de lo que hoy denominamos psicolingüística. Los postulados ofrecidos señalaban que existen unas "constelaciones" o "redes" léxicas en nuestras mentes, donde podemos organizar nuestro vocabulario a través de relaciones interdependientes, conocer cómo funcionan esas redes supondría para el hombre conocer el funcionamiento de los procesos cognoscitivos del ser humano.

En Chile, la Tesis Doctoral aportada por Mena Osorio en 1986 estudia el léxico en los primeros años escolares. Más tarde, Echeverría (1987, 1991) planteará el "índice de cohesión", con el cual se conoce el carácter abierto o cerrado de un centro de interés. Este índice se obtiene dividiendo el número total de palabras en un centro de interés entre el número total de palabras distintas dentro de ese centro de interés (vocablos).

[2] El apartado dedicado a la lexicografía ha sido ampliado anteriormente.

3.2.3. Investigaciones en España

Las investigaciones en España se suceden en diferentes focos espaciales: País Vasco, Madrid, Las Palmas, Castilla y León, Asturias, Valencia y Andalucía. Es importante resaltar la labor de muchos investigadores que apoyan el proyecto panhispánico coordinado por el profesor López Morales, Proyecto de Estudio de la Disponibilidad Léxica en Español, que pretende conocer la norma léxica a través de estudios sobre disponibilidad en estudiantes preuniversitarios.

En el País Vasco contamos con Azurmendi (1983), quien siguiendo las ideas aportadas por Mackey *et al*. (1971) en Canadá, llevó a cabo un estudio sobre la disponibilidad léxica en el plano del bilingüismo, por otro lado, Extebarría (1985) publica en Salamanca un estudio sobre la situación sociolingüística de Bilbao.

En Madrid, Benítez (1992) presenta una serie de investigaciones con las que colabora al proyecto panhispánico anteriormente mencionado.

Más tarde, en Las Palmas de Gran Canaria también suceden una serie de estudios sobre el léxico disponible, tales como los de Samper y Hernández (1995) que analizan una muestra formada por 539 alumnos de COU y un análisis de las variables sociales sexo, nivel sociocultural, tipo de enseñanza, ubicación del centro y situación del centro en el entorno ciudadano. Los cuatro primeros factores descritos se han convertido en factores clásicos para otros estudios posteriores, como los de Mateo García (1998) en Almería o Gómez Devís (2004) en Valencia. Siguiendo en Gran Canaria, la publicación de García Domínguez, Navarro y Pérez Martín (1994) utiliza los datos aportados por Samper y Hernández (1997) en función de los factores "tipo de educación" y "ubicación geográfica". En 1995, se inició un proyecto de investigación llevado a cabo por Benítez, Hernández y Samper sobre *Léxicos básicos de España* (*LEBAES) y de Canarias (LEBAICan)*, dada la escasa publicación de investigaciones sobre léxico básico. Dicho proyecto, sigue las bases teóricas de Morales (1986) y recoge cinco principios en su muestra: carácter sincrónico, pues los textos recogidos datan desde 1970 a 1994; carácter escrito; carácter homogéneo; carácter exhaustivo, solo contienen textos escritos en español en España o en Canarias, dependiendo del *LEBAES o LEBAICan* y clasificación en cinco mundos: drama, narrativa, ensayo, literatura

técnica y periodismo. Entre los factores lingüísticos aportados en los resultados se obtienen el índice de uso, frecuencia absoluta, dispersión y frecuencia absoluta en los cinco mundos ya descritos (Benítez *et al.* 1995: 16), válidas para la confección de dos diccionarios propuestos.

En Castilla y León cabe destacar la publicación de Galloso (1998, 2002) que estudia a los alumnos preuniversitarios salamanqueses y Bartol (2000) con el estudio del léxico soriano. En Asturias contamos con la aportación de Carcedo (2001).

En Valencia, Terrádez (1996, 2002) comenzó la andadura de los estudios sobre disponibilidad léxica del español hablado en Valencia, y su seguimiento con la autora Gómez Devís (2004) cuya Tesis Doctoral estudia a los alumnos preuniversitarios valencianos bajos los preceptos metodológicos del proyecto panhispánico. Esta investigación tuvo su origen en un estudio inédito de 1995 dentro del Grupo Val.Es.Co. en el que se encontraban además Briz y Gómez Molina entre otros.

3.2.3.1. Andalucía

En Andalucía los estudios sobre disponibilidad léxica surgieron en la Universidad de Granada a partir de la década de los 90 en el Departamento de Lingüística General y Teoría de la Literatura. A partir de Granada, Francisco Joaquín García Marcos, ya desde la Universidad de Almería fue quien impulsó los estudios de Lingüística Aplicada y Sociolingüística hacia otros autores como Fuentes González o Mateo García.

La primera obra la encontramos en 1994, año de publicación de una investigación realizada por Fuentes González, García Marcos y Mateo sobre el léxico cañero-azucarero de la localidad granadina de Motril.

El ya citado proyecto panhispánico coordinado por López Morales, también originó una versión andaluza en 1995, coordinada por el propio López Morales y García Marcos, dada la falta de planificación lingüística y el retraso de la comunidad andaluza con respecto al mundo hispánico en investigaciones de tipo léxico: "En Andalucía se carece prácticamente de una planificación lingüística entendida en sus justos términos; esto es, como una programación de la enseñanza de la lengua materna sistemática y científicamente formulada" (1995: 6).

Para dicho proyecto se siguen las pautas dadas años atrás para el estudio inter-hispánico, en el que los alumnos pertenecer al último año de educación secundaria. Se evalúan tres criterios: el proceso educativo, corregir y programar materiales para la enseñanza ELE y proporcionar índices de estratificación léxica. El corpus deberá está formado por 400 entrevistados por cada provincia andaluza, y los factores sociales estudiados serán sexo, nivel sociocultural y tipo de centro, puesto que la edad se neutralizaría. Estos factores sociales son los ya estudiados por Samper y Hernández (1997). Los centros de interés trabajados son los 16 clásicos más "los colores" y "la mar", aumentando la cifra a 18[3]. Estos centros de interés también serán estudiados por Mateo (1995, 1996 y 1998). Trabajos como los de Mateo para Almería o González (1997) en Cádiz forman parte de dicho proyecto.

En ese mismo año, 1995 aparece una publicación conjunta de García Marcos y Mateo García sobre la disponibilidad léxica en la zona oriental andaluza. En cuanto a los presupuestos teóricos, estos también han sido postulados en Almería de la mano de Fuentes González (1997) y García Marcos (1995, 1997).

Mateo García escribió su Tesis Doctoral (1995) centrándose en el centro de interés "la alimentación". Fruto de dicha investigación es su obra *Estratificación social del léxico de la alimentación en Motril (Granada),* (1996) en la que emplea una metodología rigurosa y estadística. El corpus está compuesto por 30 encuestas, que representan un total de 0.06% del total de la población a estudiar). Entre los factores sociales, como ya se ha adelantado, Mateo se centra en tres: edad, comprendida en tres parámetros: hasta 20 años, de 20 a 40 años y más de 40 años; sexo y nivel sociocultural: alto, medio y bajo. Entre las variables lingüísticas, se estudian el léxico frecuente de la alimentación de Motril (LFAM), léxico disponible de la alimentación de Motril (LDAM) y el léxico contextual de la alimentación de Motril (LCAM). Además, la autora aprovecha el corpus para realizar una revisión ortográfica de las encuestas.

Al centrarse en un centro de interés único, Mateo divide en cinco áreas el léxico recogido sobre la alimentación: "alimentos, elaboración, valores nutricionales, tareas domésticas y culinarias y útiles de cocina". De esta manera es posible estudiar la frecuencia (1996: 53).

[3] Benítez (1995) ya los propuso.

Los resultados del análisis se presentan por LFAM, LDAM y LCAM. Cabe destacar dentro del análisis de la frecuencia, que en los primeros rangos se sitúan productos autóctonos, como "aguacate, azúcar, chirimoya...". En general, son el grupo de los alimentos el más numeroso, puesto que un 74,31% lo ocupan léxicos alimenticios, siendo las verduras, hortalizas, frutas y legumbres las más destacadas (22,96%). En descenso encontramos el área de la "elaboración" con un 19,84%, seguido de "valores nutricionales" con un 2,58%, detrás encontramos "útiles de cocina" con un 2,41% y finalmente, "actividades domésticas y culinarias" con un 1,2%.

Entre las conclusiones a las que llega según los estudios de léxico disponible, no existe una dependencia entre los resultados de disponibilidad léxica y frecuencia, porque por ejemplo la palabra "chirimoya" se encuentra en el doceavo rango de frecuencia, mientras que su frecuencia relativa es de 33,33%. Entre los veinte primeros alimentos más disponibles encontramos a las hortalizas en cabeza, seguidas de las legumbres y frutas. Del estudio Mateo sostiene que el factor social más estratificado es la edad, puesto que los mayores han obtenido mayor disponibilidad en las hortalizas, caso contrario con los jóvenes, en el que el número de embutidos es más elevado. (1996: 99).

En cuanto al análisis contextual realizado a través de la publicidad, cartas de menús, carteles, etc. cabe destacar la elevada aparición de extranjerismos, la ausencia de frutas, pastas o guisos, relegadas estas últimas al ámbito doméstico. Es curioso observar una diferencia notoria entre el uso de la lengua de los hablantes y el de los restaurantes, por ejemplo: el LFAM sostiene la aparición de léxico como "jamón york", "aceituna", "paté", "mantequilla", mientras que el LCAM recoge vocablos diferentes que hacen alusión a las mismas realidades: "jamón cocido", "oliva", "foie-gras" o "margarina". (1996: 111).

Dos años más tarde, Mateo (1998) realiza un estudio con el que se inserta en el proyecto andaluz de disponibilidad léxica y es de mencionar la precisión descriptiva de la investigación que le otorga un claro ejemplo de conocimiento sobre la materia. Esta investigación ya recoge los dieciocho centros de interés de "los colores" a partir de Samper y Hernández (1997) y "la mar" propuesta teórica de García Marcos (1995).

Las variables sociales contempladas son cuatro: sexo, tipo de centro: privado o público, nivel socioeconómico y situación geográfica: capital, interior, litoral oeste,

litoral este. La edad queda neutralizada puesto el corpus lo forman alumnos de COU, con edades comprendidas entre los 16 y 18 años. Para la medición de factor sociocultural, sigue los preceptos de Samper y Hernández (1997: 229-239)[4]. La muestra está formada por 400 encuestados (López Morales y García Marcos 1995), y se dejó un tiempo total de dos minutos para cada centro de interés.

Siguiendo a García Marcos (1997), Mateo estudia dos factores: por un lado, lingüísticos: extranjerismos y grupos léxicos; y por otro lado, los sociales: disponibilidad léxica de vocablos, palabras y promedio de palabras por vocablos. La fórmula estadística utilizada es la propuesta por López Chávez y Strassburger Frías (1991) que mide la frecuencia, dispersión y rango de posición en la que aparece cada palabra. El tratamiento informático ha sido llevado a cabo con el programa ESDIS, diseñado por el Grupo de Sociolingüística Teórica y Aplicada de la Universidad de Almería y elaborado por J. L. Mateo y García Marcos (1997).

Entre los resultados obtenidos, Mateo los separa en los mismos dos factores analizados para llevar un orden:

- *Factores lingüísticos*: el general, el léxico recogido por los sujetos de COU viene a suponer una amplia representación del español estándar, puesto que sólo un 1,75% de los vocablos son extranjerismos, siglas, marcas y léxico específico. Dentro de ese 1.75%, el 64,38% son palabras del "léxico específico de la comunidad de habla", lo que supone que hay una cierta "disponibilidad local" en algunos centros de interés. Encontramos nueve centros de interés destacables por su léxico estándar: "la ciudad", "los alimentos", "las bebidas", "la cocina", "el cuerpo humano", "las partes de la casa", "los objetos colocados en la mesa para comer", "la escuela" y "los colores". El centro de interés denominado "profesiones y oficios" destaca por la incorporación de siglas; en "los muebles" aparece un número elevado de vocablos propios del habla almeriense, mientras que la influencia geográfica social del entorno hace mayor hincapié en "el campo", "los trabajos del campo", "el jardín" y "la mar". Por último, en este breve repaso cabe destacar la presencia de extranjerismos en "la ropa", "los medios de transporte" y "juegos y distracciones".

[4] La información sobre la tabulación del nivel socioeconómico de Samper y Hernández (1997) se reproduce en el apartado "Lexicometría y factores sociales".

- *Factores sociales*: existe una equiparación lingüística entre ambos sexos, por lo cual no es notoria su distinción. La variable más destacable es el nivel sociocultural de los sujetos, porque pese a la existencia de una gran base léxica, tienen lugar diferencias entre los tres parámetros socioculturales contemplados: alto, medio y bajo, aunque el nivel medio se no se encuentra muy alejado de ambos. Mateo propone matizar la "teoría del déficit" de Bernstein (1998: 248), puesto que las diferencias sociales en cuanto al léxico debían haber sido eliminadas por la escuela debido a que los sujetos se encuentran en la última etapa escolar. De manera resumida los resultados siguen la siguiente dirección:

  > El nivel de máxima producción léxica en Almería quedaría establecido en la correlación de los factores clase social alta, centros privados y capital. En el extremo opuesto se ubicarían los alumnos de nivel bajo, centros públicos, y de nuevo, la capital (1998: 249)

En cuanto a la metodología, Mateo sostiene la necesidad acuciante de introducir índices de ponderación puesto que la estadística ha sido un instrumento para conocer la realidad lingüística a través de los estudios de disponibilidad léxica (1998: 251).

Siguiendo la escuela iniciada en Almería, llegamos hasta la Tesis Doctoral presentada en 2005 por Muñoz Hernández sobre disponibilidad léxica en alumnos de enseñanza media almerienses. De dicha investigación se ha publicado recientemente un artículo firmado por la propia Muñoz Hernández y Rodríguez Muñoz sobre las palabras específicas utilizadas por los sujetos que participaron en el corpus de la tesis de Muñoz Hernández.

Dicho artículo sigue los preceptos marcados por Mateo García (1998) eligiendo los mismos dieciocho centros de interés, porque como se ha señalado arriba, sobre todo el centro de "el mar" es destacable por la influencia geográfica de Almería y su estudio amplía el campo léxico hacia una realidad más fiel.

El método escogido para analizar los datos proviene de Mena Osorio (1986), el cual destinó dos minutos para cada centro de interés y no limitó el número de palabras puesto que la prueba es asociativa, de esta manera los sujetos pueden poner todas las

palabras con las que asocien el campo de interés sin marcar un número determinado. La fórmula empleada, igual que Mateo (1998) fue la diseñada por López Chávez y Strassburger Frías (1991) en México. El programa informático utilizado ha sido LexiDisp, programa sobre disponibilidad léxica elaborado por José Enrique Moreno Fernández y Antonio García de las Heras.

El corpus recoge un total de 11.328 palabras a partir de la realización de 400 pruebas entre septiembre de 1998 y diciembre de 1999. Los factores sociales contemplados han sido: sexo, tipo de centro y un rango de edad desde los 15 hasta los 19. La situación geográfica quedó neutralizada puesto que las encuestas tuvieron lugar en Almería capital.

Los factores lingüísticos que se analizan son la descripción del uso de coloquialismos, extranjerismos, palabra patrimonio-dialectales, marcas comerciales, siglas y acrónimos entre las palaras formadas en los centros de interés.

En las conclusiones los autores realizan una exhaustiva interpretación de los datos obtenidos por cada factor lingüístico estudiado, cabe destacar aquí la acusada participación de marcas comerciales dentro de los centros de interés: "la ropa", "alimentos y bebidas", "la escuela: muebles y materiales", entre otros, signo de una influencia publicitaria. Estos centros de interés, junto con otros como "juegos y distracciones", "la mar" u "oficios y profesiones" también destacan por su alta aparición de extranjerismos. En cuanto al uso de un vocabulario propio de patrimonio dialectal almeriense, cabe destacar los campos de "alimentos y bebidas", y "la cuidad". Por último, cabe destacar un dato relevante, los datos arrojan una asimilación entre el campo "la mar" y la situación geográfica de la capital almeriense, dejando atrás "el campo"; esto se debe a las pocas referencias de palabras obtenidas en "trabajos del campo y jardín", es decir, a las palabras relacionadas con la agricultura, tan puntera en Almería, en cambio, sí que hay muchas referencias en el centro de "la mar". Quizás esta posible relación entre ciudad-mar podría estudiarse en un futuro sobre una investigación entre la capital y comarcas cuyo comercio se sustenta por la agricultura, para observar la incidencia del factor geográfico.

3.2.4. Aplicaciones de los estudios sobre disponibilidad léxica

Desde su aparición, asociada a la publicación de Gougenheim *et al.* (1984), se ha relacionado a los estudios de disponibilidad léxica la aplicación de la enseñanza de la lengua extranjera, como tal finalidad era la del *Français élémentaire*. Pero pronto se dio cuenta que a las conclusiones a las que se llegaba en dichas investigaciones arrojaban propuestas necesarias para poder guiar y planificar la enseñanza de la lengua materna; de hecho, López Morales y García Marcos (1995) ponen en marcha el proyecto de trabajo sobre disponibilidad léxica en Andalucía por la escasa planificación lingüística en la escuela y la posibilidad de realizar una programación de la enseñanza de la lengua materna acorde con los resultados sobre la realidad léxica, arrojados gracias a los diferentes estudios sobre disponibilidad léxica en cada una de las ocho provincias andaluzas.

Siguiendo la aplicación sobre enseñanza de lenguas extranjeras, cabe destacar las investigaciones sobre el vocabulario contenido en los manuales de español como lengua extranjera y su contraste con el léxico disponible como la que realizaron García Marcos y Mateo García (1994), Benítez (1994) o Norman (1994), este último especializado en el léxico encontrado entre los manuales suecos sobre ELE.

Entre las últimas publicaciones que relacionan a la enseñanza del español como LE a partir de un estudio sobre el léxico disponible, encontramos a Frey Pereyra (2007) quien llevó a cabo una investigación cuyo objetivo era aunar una investigación sobre disponibilidad léxica y textos escritos por alumnos de ELE. El corpus recogido consta de cuatro encuestas, todas ellas realizadas por alumnos extranjeros que están asistiendo a clases de español como lengua extranjera. En primer lugar, los alumnos rellenan un primer cuestionario sobre disponibilidad léxica; en diferentes sesiones, los alumnos escriben dos textos propios, uno personal y otro no personal, en los que se recoge la producción directa de su vocabulario activo; y finalmente, vuelven a realizar una encuesta sobre disponibilidad léxica. Esta metodología es novedosa y permite a la autora la comparación sobre los mismos sujetos entre su léxico pasivo y activo, dentro del contexto de la enseñanza de una lengua extranjera.

Los objetivos propuestos por Frey Pereyra son tres: llegar a conocer si existe semejanza entre la propia disponibilidad léxica y su aplicación en los textos escritos

(léxico pasivo/activo); conocer el vocabulario de inmersión que se encuentra en las encuestas; y, conocer si los errores léxicos producidos en las encuestas se producen también en los textos (Frey Pereyra, 2007: 368).

Entre las variables sociales recogidas, al tratarse de un corpus sobre estudiantes de ELE, también se han reunido además de sexo y edad, nacionalidad, nivel y método de enseñanza de ELE, lengua materna y conocimiento de otras lenguas.

Como ya se ha adelantado en el apartado sobre "Lexicometría y factores sociales", el tiempo para rellenar las listas de cada centro de interés ha sido un minuto y medio, en lugar de dos, que es el método clásico propuesto por Mena Osorio (1986). La autora también modificó la metodología de los dieciséis campos de interés clásicos, dejando catorce, pues omitió "trabajos del campo y jardín" e "iluminación, calefacción y ventilación".

Es significativo resaltar la base de datos de la que se sirvió la autora, puesto que recogía cuatro tablas: información sobre los alumnos (códigos y datos), cinco campos de vocabulario y las dos encuestas sobre disponibilidad (1, 2). Esta base de datos es capaz de relacionar las variables sociales de los informantes con sus listas sobre léxico disponible y el léxico recogido por sus escritos, además de las variables de palabras editadas u omitidas (Frey Pereyra, 2007: 370).

La muestra recogida está formada por quince alumnos de español avanzado procedentes de diferentes países, por lo tanto con diferentes lenguas maternas, aunque once de ellos tienen al inglés como lengua primera.

La cumplimentación de los tres objetivos iniciales pondrá de relieve si realmente los estudiantes de ELE utilizan de forma activa todo el vocabulario que conocen, gracias a la comparación entre las listas y los textos, ambos producidos por ellos mismos.

Otra de las nuevas aportaciones sobre la aplicación de la didáctica de la lengua extranjera ha sido la propuesta por Rodríguez Muñoz y Muñoz Hernández (2009), basada en el corpus recogido de Muñoz Hernández para su Tesis Doctoral *Disponibilidad léxica en alumnos de enseñanza media de Almería* (2005). Ambos autores creen que "la disponibilidad léxica proporciona datos útiles para la selección y nivelación del vocabulario en las clases de lenguas extranjeras" (2009: 10), aunque siempre se debe delimitar los rasgos propios delimitados por factores extralingüísticos.

Basándose en el mismo corpus que el ya comentado trabajo de 2011, llegan a la conclusión de que los centros de interés sobre "el campo" y "la mar" son los más asociativos, con un total de 680 palabras para el primero, y 667 para el segundo, lo cual se podría relacionar con el ámbito socioeconómico almeriense, lo cual no tendría mucha relación con la base para la planificación de ELE.

Entre las conclusiones cabe destacar que los métodos léxico-estadísticos aportan datos válidos para la didáctica, pero antes es necesario " efectuar un salto cualitativo al estudio del uso real de la lengua *hic et nunc*, en esta caso del léxico, concebida como proceso interactivo, cambiante y multiforme" (2009: 15).

Dejando atrás los numerosos estudios con aplicaciones didáctica, es de interés resaltar otras aplicaciones como las señaladas por López Morales (1996): dialectología, sociolingüística, psicolingüística y etnolingüística.

Los estudios con aportaciones sobre la dialectología han sido ampliamente investigados por López Chávez (1992) en Puerto Rico. A través de otros corpus ya realizados, como los de Benítez (1995) o Samper y Hernández (1994), estudió el léxico de Madrid y Las Palmas, en comparación con el de la República Dominicana (Alba, 1995) y Puerto Rico (López Morales, 1995). Existía mayor cercanía entre las localidades españolas, lo cual da pie a la corroboración de "la existencia de un español atlántico" (1996: 250).

La relación con los factores sociales ha hecho posible trazar una línea entre los estudios sobre el léxico y lo que rodea a la comunidad de hablantes, es decir, la proliferación de teorías sociolingüísticas arraigadas al estudio del léxico común. La *teoría del déficit* sostenida por Bernstein es una de las pioneras en resaltar la importancia de los factores sociales. Ha sido muy estudiada por López Morales, quien piensa que no existen léxicos "exclusivos" (1996: 251). Por otro lado, hay autores que reafirman la relación entre los factores socioculturales y los desniveles léxicos aportados por los estudios sobre disponibilidad léxica.

En cuanto a la psicolingüística, esta ha sido asociada a la aparición de los primero estudios sobre frecuencias por la base teórica sobre *teoría de la red* ya aludida anteriormente. Entre los últimos años, se han sucedido teorías psicológicas sobre la

distribución del léxico y la memoria semántica, especialmente encauzadas por López Chávez.

Por último, encontramos la relación del léxico disponible con la etnolingüística, apenas desarrollada entre los trabajos del español. La etnolingüística se sirve del "léxico disponible de cada sociedad como reflejo de su cultura" (1996: 253) lo cual va más allá de los factores sociales individuales, pues se estudia la sociedad en conjunto.

Gómez Devís (2004) publica su Tesis Doctoral en Valencia sobre disponibilidad léxica asociada a tres aplicaciones: sociolingüística, psicoligüística y didáctica. De especial relevancia es la relación marcada por la autora a la hora de intentar relacionar el grado de cohesión entre las palabras recogidas por cada centro de interés para poder llegar a conocer el lexicón mental defendido por las teorías psicolingüísticas.

El corpus recogido consta de un total de 465 informantes de COU entre centro valencianos. Entre las conclusiones a las que llega la autora basta señalar la necesidad de una "reflexión teórica-metodológica" acerca de los centros de interés, como la modificación de algunos de ellos que se encuentran alejados de la vida cotidiana y la incorporación de otros no contemplados y necesarios por los avances científico-tecnológicos de los últimos años. Entre las variables sociales, la más discriminatoria es el tipo de centro, siendo el público el que mayor disponibilidad recoge. Lo que realmente es nuevo en esta aportación, son las conclusiones acerca de la homogeneidad alcanzada por los léxicos recogidos (2004: 353): Gómez añade a los dieciséis centro de interés "los colores", contando con diecisiete, y la distribución entre el nivel de confluencia queda dividido en tres: mayor cohesión (6 centros de interés), cohesión media (6 centros de interés) y cohesión baja (5 centros de interés), lo cual se traduce en que "los treinta primeros vocablos han demostrado una fuerte estabilidad psicológica del conjunto de informantes", (2004: 354).

3.3. Riqueza léxica

3.3.1. Inicios teóricos

Si los estudios sobre disponibilidad léxica ponen de manifiesto el vocabulario pasivo utilizado por los informantes, debido a su mero contenido en listas de frecuencia, la riqueza léxica aportará nuevos indicios sobre el léxico activo recogido por los

hablantes, puesto que se basa en la distinción de palabra y vocablo, para definir así el número de vocablos diferentes recogidos por texto.

La distinción entre ambos términos fue definida por Müller (1973: 231), donde *vocablo* responde a "la actualización de un lexema en el discurso", y *palabra* es "toda ocurrencia de un vocablo cualquiera". Aunque las bases teóricas sobre el concepto de "Riqueza léxica" estaban vislumbradas en los años 70, los estudios sobre disponibilidad léxica fueron ampliamente investigados, dejando atrás otros índices como el que ahora nos atañe.

Años más tarde Menard define *Riqueza léxica* del siguiente modo: "Un lugar de comparación entre dos o varios textos en función de su extensión respectiva (número de palabras) y el número de vocablos que figura en cada uno de ellos" (1983: 16). Poco a poco las encuestas que tienen como objetivo final estudiar la riqueza léxica se van sucediendo paulatinamente, pero no es hasta 1984, cuando López Morales se refiere a *Densidad léxica* y señala la importancia de medir dicho índice para conocer de manera real la producción léxica de los alumnos.

Partiendo de los conceptos de *vocablo* y *palabras*, el índice de riqueza léxica se obtiene gracias al promedio del total de números de vocablos, entendiendo el número de vocablos como la cantidad total de palabras no repetidas dividido entre el número total de palabras.

Hay que tener en cuenta que el perfecto resultado de esta fórmula es 1; por tanto, todos aquellos valores que se aproximen a 1 significan que el hablante tiene un buen dominio de la lengua y es capaz de producir textos sin la acusante repetición. No obstante, se debe considerar que es científicamente imposible llegar a 1, ya que como palabras se contabilizan todas aquellas sin valor semántico, tales como los determinantes o preposiciones, que indudablemente el hablante necesita repetir (Cano García, 2012). Pero una gran riqueza léxica no es siempre sinónimo del saber hablar o escribir, puesto que comunicar haciendo uso del principio de economía es conocer y utilizar de manera apropiada la lengua: "una mayor riqueza da la posibilidad de comunicarse con eficiencia y economía o con redundancia y grandilocuencia; de utilizar no sólo un código extenso, sino también uno restringido, cuando la situación comunicativa así lo requiera" (Labov, 1972: 179).

Existen dos índices más para medir la riqueza léxica además del descrito arriba. Reyes Díaz (2008) utiliza de manera sistemática los tres en su estudio sobre textos escritos por alumnos de bachiller en Las Palmas, definiendo así los otros dos: el segundo índice de riqueza léxica se mide con "el intervalo de aparición de las palabras nocionales: N/PN". Es una fórmula válida para medir la riqueza léxica de nombres, adjetivos, verbos y adverbios, pues se contabilizan solo este tipo de palabras para dividirlas entre el total. Y el tercer método es el índice de hápax: "la variedad de hápax se obtiene con la aplicación de la fórmula que divide el número total de vocablos y la suma de las palabras de una ocurrencia. Conforme el índice va en aumento, va disminuyendo la variedad léxica (2008: 148).

3.3.2. Estudios hispanoamericanos

A partir de los años 90 aparecen las primeras investigaciones sobre riqueza léxica del español en el marco hispanoamericano. Será a finales de esta década cuando se empiece a estudiar la riqueza léxica dentro de España, gracias a la labor de autores que marcaron el camino como Raúl Ávila, López Morales o Echeverría, entre otros.

Raúl Ávila (1991) introduce en su estudio la densidad léxica además de los parámetros sobre frecuencia acumulada y proyección de vocablos en un corpus realizado a alumnos de tercero y sexto de primaria y textos orales recogidos a adultos. El corpus contó con un total de 205 entrevistas y 420.000 palabras sobre las que estudiar dichos parámetros arriba descritos. Se trata de una triple investigación, puesto que se estudia la riqueza léxica de dos grupos de hablantes distinguidos según el factor social generacional, y también se observan las diferencias entre textos escritos y orales. Entre las conclusiones a las que llega, Ávila destaca que aparece una mayor densidad léxica en los textos escritos, lo cual se corrobora la idea de oralidad y repetición de vocablos. Aunque Ávila solo se centra en el estudio generacional, también recoge otros factores sociales como sexo, nivel socioeconómico y situación geográfica de los hablantes (urbano/rural), aunque señala que dichos aspectos no se encuentran relacionados con la densidad léxica, pues los datos diferenciadores han sido poco significativos.

Uno de los factores a tener en cuenta para la recogida de las muestras es la extensión de los textos, Haché de Yunén (1991), siguiendo a López Morales, se centró

en un estudio para evaluar la riqueza léxica en escolares puertorriqueños. Los resultados obtenidos no pudieron ser comparados debido a un fallo en la metodología empleada, puesto que no se acotó a un número determinado de palabras por redacción, factor que incide en los resultados de riqueza léxica, pues resulta ventajoso para los textos más breves. Un año más tarde, también en Puerto Rico, Citrón Serrano (1992) llevo a cabo un estudio sobre riqueza léxica y promedios de categorías gramaticales, entre las conclusiones, basta destacar las diferencias encontradas entre las variables sociales, siendo el grupo femenino, nivel sociocultural algo y mayor escolaridad los que mayor riqueza léxica poseen.

Resultados parecidos obtuvo la investigación de Echeverría *et al.* (1992) en Chile, en el que la escuela privada, el nivel sociocultural alto, la residencia metropolitana y el tipo narrativo obtienen los valores más elevados. Por el contrario, Portela (1992) en la República Dominicana, estudia a estudiantes universitarios, y señala que el factor sexo no está asociado al índice de riqueza léxica.

3.3.3. Estudios en España

A partir de la proliferación de los estudios sobre riqueza léxica en Hispanoamérica, cada vez las universidades españolas comienzan a estudiar este parámetro ligado a la aplicación de la didáctica de la lengua materna.

Resultados parecidos sobre los arriba detallados son encontrados por Andrés Pérez (1997) en Madrid; de igual manera la investigadora Torres González (1999, 2003, 2003b) en Tenerife encuentra similitud entre los datos manejados, y se puede hablar de unas conclusiones generales en torno a la riqueza léxica del español: esta se da con mayor densidad en la variable femenina, nivel sociocultural alto, centros privados y en el aumento de la escolaridad. Las investigaciones llevadas a cabo por Torres González analizan de manera pormenorizada diferentes variables lingüísticas como la proporción de vocablos, el intervalo de palabras nocionales, índice de hápax y promedios de las diferentes categorías nocionales, lo cual data de un estudio rigurosamente estadístico sobre el léxico activo de un gran rango generacional tinerfeño, pues su corpus se base en alumnos de cuarto de primaria, segundo de secundaria y segundo de bachillerato.

Estudios posteriores y novedosos como el de Reyes Díaz (2008), también realizado en Las Palmas de Gran Canaria, estudia de forma comparativa las muestras

recogidas por alumnos de segundo de BUP (actual cuarto de E.S.O.) y COU (actual segundo de bachillerato), limitando las redacciones a 100 palabras. Las variables lingüísticas estudiadas son: la proporción de vocablos, el intervalo de palabras nocionales y el índice de hápax. Los resultados demuestran que existe un crecimiento en la proporción de los vocablos y el índice de hápax en tan solo dos años escolares, puesto que las redacciones de los alumnos de COU obtienen mayor riqueza léxica. Por otra parte, Reyes Díaz sostiene que "los comportamientos de los factores extralingüísticos no son similares en las dos muestras estudiadas" (2008: 162), lo cual explica que en la muestra recogida no aparece una relación directa que ligue factor social con riqueza léxica.

3.3.4. Aplicación en el aprendizaje de la lengua no materna

Las investigaciones sobre el léxico han sido ampliamente estudiadas con vistas a la planificación didáctica de las enseñanzas de segundas lenguas, y la riqueza léxica también ha sido objeto de investigación para conocer a fondo las producciones léxicas de alumnos de ELE.

La universidad de Alcalá de Henares, Madrid, ha sido pionera en la indagación del comportamiento léxico en estudiantes extranjeros. García Rosas (1996) realiza un estudio para conocer si la lengua materna y procedencia de los alumnos estudiantes de ELE tiene o no correlación con la riqueza léxica. Para ello, la muestra llevada a cabo se componía de 41 estudiantes de diferentes niveles de aprendizaje del español (medio, avanzado y superior), junto con la variable sexo, lengua materna y modo discursivo (descriptivo, narrativo y argumentativo). Las conclusiones a las que llega García Rosas demuestran que no existe una correlación directa entre la riqueza léxica con la edad o la lengua materna de los estudiantes; sin embargo, y como ya se había demostrado antes[5], sí existe una correlación entre el modo narrativo y mayor utilización de vocablos. A medida que los estudiantes avanzaban en nivel de enseñanza, se demostraba una mayor riqueza léxica, dato óptimo, pues resulta de una adecuada enseñanza.

La comparativa entre resultados obtenidos de alumnos inmigrados y nativos ha sido estudiada recientemente por Alberto García Marcos (2010), quien estudia la

[5] Echeverría *et al* (1992) ya estudió en Chile el comportamiento léxico en diferentes modos discursivos: narrativo, descriptivo, expositivo y argumentativo; siendo el modo narrativo el que mayor riqueza léxica obtuvo, en cambio, el que menor densidad léxica poseía era el argumentativo.

riqueza léxica además de la ortografía y la madurez sintáctica en dos grupos: escolares de Roquetas de Mar (Almería) y un grupo "experimental" formado por alumnos inmigrados de un curso de ELE del CEMyRI. Para el estudio sobre riqueza léxica los datos son tratados con un programa informático, *Lees*, bajo la fórmula: *número total de vocablos/número total de palabras*.

Con dicho estudio, se puso de manifiesto una mejora en el rendimiento lingüístico de los escolares inmigrados que habían seguido una instrucción específica en cursos de ELE, lo cual demuestra la necesidad de una incorporar métodos y personal cualificado en planificación lingüística en la escuela de hoy en día.

Siguiendo la estela creada por Alberto García, yo misma estudié en 2012 la riqueza léxica entre los alumnos de primero de E.S.O. en Antas, pueblo del levante almeriense. El estudio comparaba tres variables lingüísticas: lapsogramas, riqueza léxica y promedio de aparición de categorías nocionales en una muestra compuesta por 29 redacciones, 7 de alumnos inmigrados y 22 de nativos. La muestra no es representativa, pero es la realidad de un aula dentro de un instituto público actual.

El objetivo de dicho experimento era conocer las diferencias entre los alumnos nativos e inmigrados, los cuales comparten aula y lecciones, basándome en la importancia de la lengua como instrumento de comunicación. Junto con el factor nacionalidad, también se recogía el factor sexo, lo cual me ayudó a confrontar los datos en cuatro parámetros: *mujeres/hombres; extranjeros/nativos; mujeres extranjeras/mujeres nativas* y *hombres extranjeros/hombres nativos.*

En general, y de manera contradictoria a Citrón Serrano (1992), Andrés Pérez (1997) y Torres González (1999, 2003a, 2003b) mis resultaron arrojaron mejores datos sobre riqueza léxica entre los hombres, quienes obtuvieron 0.61 frente a 0.55 de las mujeres, lo cual indica una diferencia poco significativa. Otro resultado obtenido de mi corpus fue que tanto los hombres como las mujeres inmigradas obtenían promedios con mayor riqueza léxica, aunque realmente nos encontramos ante una muestra no representativa, lo cual incurre desfavorablemente en el grupo de los nativos, mucho más numeroso.

4. PARÁMENTRO GRAMATICAL

4.1. Fundamentos gramaticales

En el siglo XX tienen lugar una sucesión de teorías y estudios lingüísticos que desarrollarán un amplio camino dentro de gramática. El desarrollo de la sintaxis como ciencia proliferó que muchos autores propusiesen nuevos métodos de medición de la madurez y la complejidad sintáctica.

Los estudios aquí reseñados han llegado a demostrar que existe una correlación significativa entre el grado de madurez sintáctica y los años de escolaridad. A este punto cabe advertir al lector que en muchas ocasiones se han sucedido estudios en los que se han utilizado los términos "complejidad" y "madurez" como sinónimos alternativos. Por un lado, y basándonos en el padre de los índices de madurez sintáctica, Hunt (1970a: 178) plantea que " la capacidad de combinar más y más oraciones es señal de madurez", por lo contrario, Véliz señala que "la complejidad se juzga fundamentalmente por la mayor frecuencia con la que se van consolidando constituyentes oracionales y no oraciones dentro de oraciones" (Véliz, 1999).

4.1.1. Antecedentes: La Brant (1933)

Los estudios pioneros sobre el desarrollo sintáctico fueron diseñados por Lou L. La Brant en 1933 en Estados Unidos. La Brant propuso como unidad básica de medición la cláusula, puesto que esta es capaz de desarrollar un juicio completo y se compone de un predicado finito, marcado por un signo de puntuación de cierre: "since very finite predicate indicates the existence of a clause, approach to the investigation is made through the tabulation of predicates [dado que todo predicado finito indica la existencia de una cláusula, la aproximación a su investigación se hace mediante la tabulación de los predicados] (La Brant 1933:398).

Siguiendo la metodología de La Brant era posible entonces establecer un número de cláusulas en cualquier texto para comenzar su análisis. Los índices que marcarán el análisis en La Brant serán la longitud y el tipo de cláusulas (dependientes o independientes). El diseño de un criterio objetivo es la finalidad primera del estudio de La Brant y por ello, la delimitación de la dimensión a estudiar es muy compleja, porque

como resume Checa (2004:64-66) la labor de estudio se centra en delimitar los casos en los que se los problemas se alejan de la objetividad y casos como verbos elitidos ponen de manifiesto que la clasificación debe ser totalmente objetiva, sin valores semánticos añadidos.

La búsqueda de criterios objetivos le llevó a La Brant a diseñar el índice que verdaderamente destacamos hoy de su estudio, el IS, índice de subordinación: "The ratio of the number of dependet predicates used by that individual or group to the total number of predicates used by that individual or group, the ratio being expressed in percentage [La razón del número de predicados dependientes usados por cada individuo o grupo respecto del número total de predicados usador por ese individuo o grupo, expresada dicha razón en porcentaje]" (La Brant, 1933:415).

El índice de subordinación es la variable utilizada para medir la frecuencia utilizada en el uso de oraciones subordinadas (por tanto, más complejas), este índice fue el primero en plantearse en los estudios de madurez sintáctica. Los factores con los que contó La Bran en primer momento fueron sexo y edad. El IS se estableció como uno de los índices más fiables para medir la madurez sintáctica y pese a algunas críticas, como la de Anderson (1937), se consolidó como método cuantitativo.

Anderson (1937) estudió la fiabilidad del índice de subordinación y encontró diversos problemas, como la definición de cláusula como predicado o la inclusión de la clasificación de los infinitivos, por ejemplo, lo cual aleja al estudio de la objetividad, porque la dificultad de la semántica en ocasiones podía incurrir en la clasificación. De cualquier modo, el índice de subordinación se consolidó y se ha seguido utilizando hasta la actualidad.

4.1.2. Índices de madurez sintáctica: Hunt (1965)

Los estudios sobre madurez sintáctica se verán ampliamente desarrollados en los años 60 por el lingüista Hunt (1965), el cual llevó a cabo un estudio en el que se analizó una serie de redacciones de tema libre entre los escolares de cuarto, octavo y doceavo curso. Para ello recogió las primeras mil palabras de cada redacción y acotar así el contenido del corpus. Además, Hunt analizó un total de dieciocho artículos de revistas, nueve de *Harper's* y nueve de *Atlantic*, cogiendo de cada uno de ellos las mil

primeras palabras, para comparar los resultados obtenidos entre los escolares y la madurez sintáctica de los adultos profesionales.

El planteamiento de Hunt pone de manifiesto tiene como objetivo hallar una serie de medidas válidas para el desarrollo sintáctico tras la adquisición del lenguaje.

Existe una problemática a la hora de establecer la unidad mímima de estudio, ya que es difícil que todos los investigadores consideren como unidades básicas la oración, debido a su dificultad de acotación; por ello, se debían diseñar una serie de índices totalmente objetivos y significativos, entre los que destaca algunas aportaciones de La Brant (1933). La primera de las características permite al investigador tener como modelo un índice versátil y válido como modelo objetivo, dejando atrás el lastre de la dificultad de valores semánticos, tal y como señaló Anderson (1937). La significatividad debe servir para poder equiparar y comparar los datos entre los factores edad y sexo, los únicos que hasta ese momento se utilizaban.

Hunt resuelve estas dos características básicas a través de dos índices que desarrollará después (1965: 20-23 y 33-36). Entre los índices planteados por La Brant (1933), Checa señala: "De hecho, y aunque necesiten de cierta reformulación, hay dos índices que Hunt considera potencialmente objetivos y significativos: la longitud de la cláusula y la razón de subordinación". (Checa, 2004:81).

Hunt lleva a cabo una remodelación de los índices aportados hasta entonces y para el estudio de la longitud de la cláusula, esta se define como "any expression containing a subject or coordinated subjects and a finite predicate or coordinate predicates [cualquier expresión que contenga un sujeto o varios sujetos coordinados y un predicado finito o varios predicados coordinados] (Hunt, 1970a: 4).

La unidad básica de análisis se establece en la "Minimal Terminable Unit" [Unidad Terminal (u-T en adelante)], para con ello resolver el problema de la objetividad. La definición propuesta por Hunt es la siguiente: "one main clause plus any subordinate clause or nonclausal structure that is attached to or embebed in it [una oración principal más cualquier oración subordinada o estructura no oracional que es adjuntada o incrustada en ella] (Hunt, 1970a: 4).

La definición de cláusula y u-T permite establecer a Hunt una serie de tres unidades de segmentación siguiendo un orden de mayor a menor (Checa, 2004: 87).

Ante un fragmento, primero podemos separas las "oraciones puntuadas" (Hunt: 1065:8) las cuales se separan por los signos de puntuación (punto). El segundo nivel de análisis implica la localización de las u-T, para por último separar estas en cláusulas, unidades mínimas de análisis en las que solo se distingue sujeto y predicado.

4.1.2.1. Índices primarios

Llegados a este término, podemos conocer los índices primarios diseñados por Hunt y llevados a cabo en su estudio de 1965.

Entre los índices primarios se encuentran: tres longitudes y dos razones. Las tres longitudes son:

- Índice de longitud promedio de la oración, se trata del "promedio de palabras por oración en un determinado texto o conjunto de textos" (Checa, 2004: 88). Para poder hallar el índice de longitud se contabiliza el número total de palabras de las que consta el texto y se divide entre el número total de oraciones.
- Índice de longitud de u-T, el cual se obtiene dividiendo el número total de palabras por el número total de u-T. Este índice resuelve el promedio de palaras por u-T.
- Índice de longitud de cláusula, llegados a la unidad básica de análisis, el promedio de palabras por cláusula se calcula dividiendo el número total de palabras entre el número total de cláusulas dentro de un texto.

Expuestas las tres longitudes de Hunt llevadas a cabo en numerosos estudios[6], a continuación se deben definir las dos razones anteriormente señaladas:

- Promedio de u-T por oración: se halla dividiendo el número total de de u-T entre el número total de oraciones.
- Promedio de cláusulas por oración: se halla calculando el número total de cláusulas ente el número total de oraciones. Este promedio de cláusulas por oración se trata de una de las reelaboraciones aportadas por Hunt (1965) sobre las aportaciones de La Brant (1933) como el índice de subordinación (IS).

[6] El propio Hunt lo desarrolló en numerosos estudios: 1965, 1967, 1970 entre otros, Véliz (1986, 1988, 1999), Torres González (1994, 1996, 1977), Checa (2004) entre otros autores y estudios.

De las cinco novedosas aportaciones, Hunt (1965) desarrolla aun más dos nuevos índices válidos para medir longitudes que ayudarán al investigador para llegar a medir la madurez sintáctica:

- Longitud de la oración: podemos hallar la longitud de la oración partiendo del número de u-T y multiplicando dicha cuantía por la longitud promedio de u-T, descrita arriba.
- Longitud de la u-T: se calcula obteniendo la multiplicación del número total de cláusulas por la longitud promedio de cláusula.

Los nuevos índices hallados abren paso a una nueva etapa en la que el análisis estadístico es objetivo y significativo para la medición de la madurez sintáctica.

Por otro lado, Hunt estable otros criterios cuantitativos para el análisis sintáctico como son el índice de coordinación y el índice de subordinación[7]. Estos índices no han tenido tanta resonancia como los primarios, pese a ser clave para el análisis de la complejidad y válidos tanto para escolares como para escritores profesionales. Para contabilizar dichos índices basta con reconocer entre coordinación y subordinación y dividir entre el número de u-T en el texto. Como señala Checa (2004: 97), el recuento del número de oraciones coordinadas y/o subordinadas nos lleva de nuevo a una problemática a la hora de establecer el grado de dichas oraciones, puesto que Hunt (1967) no distinguía ente oraciones compuestas y complejas, sino simplemente u-T.

Según Checa (2004: 97) la denominación de "índices primarios" no fue del propio autor, sino Torres González (1992) será la encargada de etiquetar los índices entre "primarios" y "secundarios", de los cuales se diferencian los "secundarios clausales de los no clausales".

4.1.2.2. Índices secundarios

El índice de subordinación expuesto por La Brant (1933) y reformulado años más tarde por Hunt (1965) va a dar lugar a una clasificación del tipo de subordinación, con la cual se desarrollarán los "índices secundarios".

[7] Hunt le da el nombre de "razón de cláusulas subordinadas a unidades terminales" (*ratio of clauses to T-units)* (Hunt: 1965: 36). Se trata del índice de subordinación expuesto anteriormente por La Brant (1933), tomando como objeto de análisis la u-T.

Dentro del índice de longitud de cláusula, y estableciendo como criterio los tres tipos de subordinadas: adverbiales, sustantivas y adjetivas (Hunt, 1965), podremos llegar al índice secundario clausal, el cual mide el promedio de claúsulas subordinadas adverbiales, sustantivas y adjetivas por el número de u-T.

Estos nuevos índices han servido a varios investigadores[8] para conocer el grado de complejidad de las diferentes cláusulas subordinadas, lo cual puede aportar novedosas conclusiones que ayuden a la enseñanza de la lengua materna.

Por tanto, Hunt establece tres tipos de índices secundarios clausales:

- Promedio de cláusulas sustantivas: miden la proporción de cláusulas sustantivas frente al resto de cláusulas subordinadas. Se halla dividiendo el número total de cláusulas subordinadas sustantivas entre el número total de u-T.
- Promedio de cláusulas adjetivas: miden la proporción de cláusulas adjetivas frente al resto de cláusulas subordinadas. Es el resultado de la división del número total de cláusulas subordinadas adjetivas entre el número total de u-T.
- Promedio de cláusulas adverbiales: miden la proporción de cláusulas adverbiales frente al resto de cláusulas subordinadas. Es la proporción hallada del número total de cláusulas subordinadas adverbiales entre el número total de u-T.

Junto con los tres índices secundarios, Hunt plantea un nuevo índice, el cual no se desarrollará ni será tan investigado.[9] El índice de profundidad de subordinación (*depht of subordination,* 1965:25) se fundamenta en llegar a conocer el grado de profundidad de las cláusulas subordinadas. A la hora de ampliar una oración, la subordinación es un mecanismo que aporta información al verbo principal y se observa cómo pueden establecerse diferentes estratos de subordinación. Podemos encontrar una subordinada dentro de otra, y así sucesivamente, pues la medición de este índice tiene por objeto conocer el grado de profundidad de dichas cláusulas subordinadas. Lógicamente, este índice nos aportaría datos concluyentes acerca de la dificultad de creación de estructuras sintácticas subordinadas dentro de otras. Otro de los inconvenientes reseñados por Checa (2004) es la distinción de subordinadas y

[8] Se ha especificado en el estudios de los índices secundarios clausales para logar conocer cuáles son los tipos de oraciones subordinadas más complejas y su proyección en el estudio, como por ejemplo en los modos discursivos (Véliz, 1999) o también como experimento ante otros tipos de medición, como el de "memoria inmediata" (López Morales, 1994).

[9] Siguiendo a Irene Checa (2004) este nuevo índice no va a ser del todo definido por Hunt (1965) debido a su dificultad de concreción, pero sí lo utilizará Walter Loban (1976).

coordinadas: "Efectivamente, el hecho de que el índice no distinga las cláusulas coordinadas dentro de una *u-T* se debe a que no prima el grado de dependencia, puesto que de ser así, ambas coordinadas tendría el mismo grado de subordinación y no aumentarían la profundidad de subordinación de la unidad terminal" (Checa, 2004:99).

Una vez analizados los índices secundarios, Hunt propone de nuevo una clasificación, denominada "índices secundarios no clausales", que tiene por objetivo conocer cómo aumenta la longitud de la cláusula, no sólo a través de la subordinación, sino a través de sintagmas y modificadores que se van incluyendo y complementando información. Este índice es mucho más complejo y su utilización sería difícil y exhaustiva, puesto que el grado de concreción es muy elevado. Para hallarlo, basta con delimitar una serie de estructuras que alargan la cláusula, Hunt propone las siguientes:

> Sintagmas coordinados en una misma unidad terminal, modificadores del sintagma nominal (adjetivos, genitivos, sintagmas preposicionales), sintagmas nominales sin modificar, adverbios, formas no finitas del verbo, oraciones adjetivas (en tanto que modifican a un sustantivo), oraciones completivas (en tanto que provenientes de una oración subordinada), nombres adjuntos y aposiciones. (Checa, 2004: 101)

El promedio de frecuencia de dichas estructuras por u-T nos llevaría a conocer nuevos datos sobre la longitud de la cláusula. De hecho, Hunt señala además un nuevo índice de "profundidad de modificación" (*depht of modification*), el cual establecería los tipos de modificaciones que se establecen dentro de una cláusula, del tipo: modificadores de modificadores, pero no llegó a desarrollarlo.

4.1.3. Memoria inmediata

Entre las bases metodológicas para medir la madurez sintáctica, la propuesta por Hunt (1965) fue la más utilizada desde su aparición por autores como O'Donell, Griffin y Norris (1967), O'Donnell (1968, 1976), Crowhurst (1978), hasta estudios recientes como el de Checa (2004), pero también han surgido otras metodologías para llegar a alcanzar valores medibles en madurez sintáctica, aunque este método realmente llega a estudiar la complejidad sintáctica, puesto que no incurre en la creación de los sujetos, sino en la retención de estructuras sintácticas en la memoria.

López Morales (1986) introduce un nuevo método cuantitativo: la memoria inmediata. En el próximo apartado estudiaremos a fondo uno de sus experimentos (López Morales, 1994) en el que estudia la madurez sintáctica utilizando dicha metodología.

Las pruebas de memoria inmediata funcionan gracias a la retención de la memoria del sujeto con objeto de experimento. En primer lugar, se le leerán al sujeto una serie de oraciones según su grado de complejidad y una serie de palabras tras la oración. El estudio demuestra que el sujeto podrá: recordar la oración y las palabras, la oración solo o no recordar nada. Según el resultado de la memoria, así se demuestra el grado de complejidad sintáctica de las oraciones dictadas.

4.2. Estudios de madurez sintáctica

4.2.1. Mundo anglosajón

A partir de 1965 y la fundamentación de las bases teóricas de los índices de Hunt, se siguieron una multitud de estudios de madurez sintáctica en el mundo anglosajón, puesto que no será hasta los años 80 cuando este estudio no se inicie en el mundo hispánico (Véliz, 1986) (López Morales 1986).

Ya hemos destacado el estudio primero de Hunt (1965) en el que sienta las bases de una nueva metodología cuantitativa, es además en este primer trabajo donde se analizan con detalle los tres índices primarios en redacciones de tema libre en sujetos de cuarto, octavo y doceavo grado. Además, y como ya se dijo anteriormente, estos resultados serán contrastados con el análisis de dieciocho artículos publicados en dos revistas, para poder comparar así la madurez sintáctica en escolares con escritores profesionales. ´

Dos años más tarde, Hunt (1967) realiza de nuevo un experimento en el que escoge a 18 sujetos escolares, nueve femeninos y nueve masculinos con un coeficiente intelectual elevado, distribuido en: 90-110 y más de 130. Este estudio demostró que las estructuras sintácticas se van desarrollando con la escolaridad, al compararlo con los resultados obtenidos dos años antes (1965) de los escritores profesionales. Torres González (1993: 13) realiza una buena síntesis de las conclusiones obtenidas, destacando sobre todo el orden de frecuencias clausales, medidas con los índices

secundarios. Los resultados obtenidos fueron que la mayor frecuencias de cláusulas subordinadas se daba en las adjetivas, por lo que se deduce que son las primeras que se desarrollan en la tapa de escolaridad, y por tanto las menos complejas. En segundo lugar se encuentran las cláusulas adverbiales, las cuales avanzan poco a poco y siguen un desarrollo con los años de escolaridad. Por último, se encuentran las sustantivas, las cuales dependen en gran medida del modo del discurso.

En la década de los 60 no sólo aparecieron estudios de la mano de Hunt, sino que también O'Donnell (O'Donnell et al, 1967) (1968), aplicando los índices de Hunt, estudió la madurez sintáctica.

O'Donnell et al. estudiaron en 1967 la sintaxis en infantiles, en los cursos de tercero, quinto y séptimo, donde analizaron el promedio de u-T y el de transformaciones utilizadas por u-T. Entre los factores a destacar, la narración se medía en patrones orales y escritos, así como las variables destacadas eran las mismas que Hunt (1965, 1967, 1970), curso y edad. La variable nivel de la clase social se homogeneizaba al encontrarse en una escuela donde la clase predominante era media-alta. Los datos obtenidos les sirvieron para comparar este estudio con el de Hunt (1965) y ambos llegaron a un mismo denominador común: existía una correlación positiva entre el aumento de la madurez sintáctica y la escolaridad.

Un año más tarde, O'Donnell (1968) realizó un experimento con sujetos de enseñanza primaria y superior. De nuevo aplicó los índices de Hunt y observó una correlación positiva entre el aumento de longitud de cláusula y la escolaridad, aunque como señala Torres González: "No había pruebas que pudieran medir la complejidad de las estructuras de un individuo en diferentes situaciones" (Torres González, 1993: 14).

En la década de los 70 Hunt (1970) realiza un nuevo estudio en el que compara resultados de madurez sintáctica entre tres grandes grupos de sujetos: estudiantes, estudia a escolares de cuarto, sexto, octavo, décimo y duodécimo grado, adultos y profesionales. El texto estudiado es un texto expositivo, y no narrativo, como el utilizado en anteriores ocasiones. El texto fue realizado por O'Donnel (1965-66) el cual consistía en reelaborar una serie de treinta y dos oraciones. Entre las conclusiones destacadas por Torres González cabe distinguir la quinta de ellas: "La longitud de la cláusula es el índice más relacionado con la edad cronológica y mental" (Torres González 1993: 15).

En 1973 el investigador Endicott propuso una escala para medir la complejidad sintáctica, la cual será analizada en 1976 por O'Donnell. En 1974 Golub y Kidder introducen un nuevo índice "puntuación de densidad sintáctica"[10] que es una mera adaptación del ya presentado índice de profundidad de cláusulas subordinadas por Hunt (1965), en el que se busca un único índice que sea capaz de analizar el desarrollo sintáctico, para ello, Golub y Kidder presentan una escala con doce ítem lingüísticos. Stewart (1978) realiza un análisis en universitarios para observar si existe similitud con el desarrollo de la madurez sintáctica con escolares. Para ello, los sujetos de la muestra comparten la zona geográfica y la clase social, por tanto la variable que se quiere destacar es la edad. Los resultados demuestran que durante los tres primeros años de universidad, la longitud u-T se muestra estable y comparable con los últimos años de escolaridad aportados por Hunt (1970), mientras que en los años quinto y sexto, se produce un aumento considerable de dicho índice.

4.2.2. Mundo hispánico

Los estudios en lengua española se sucedieron a partir de los años 80 y como pionera encontramos a Mónica Véliz (1986)[11] en Chile. La madurez sintáctica del español de Hispanoamérica fue primeramente analizada en los países: Chile, México, República Dominicana y Puerto Rico, antes de su llegada a España. Todos los estudios publicados, hasta los más recientes, han aplicado los índices presentados por Hunt (1965), sobre todo haciendo hincapié en la correlación del aumento de los índices primarios con el aumento de la escolaridad.

Torres González (1993:17) realiza una buena síntesis de los estudios sobre madurez sintáctica en el mundo hispánico. En la presente revisión bibliográfica se van a desarrollar tan solo tres estudios hispánicos, dado su importancia para la futura investigación, pero cabe citar las investigaciones de Mónica Véliz y Gloria Muñoz (1983) en Chile, María Eugenia Herrera Lima (1991) en México, Lilliana de Olloqui de Montenegro (1991) en Santiago de Caballeros (República Dominicana), Leonilda Rodríguez Fonseca (1989) en Puerto Rico, entre otros. Estos no son los únicos estudios

[10] Siguiendo a Torres González (1993: 15) se demostró que O'Donnell (1976) estudió el índice de densidad sintáctica aportado por Gloub y Kidder (1974) y sus resultados son menos precisos que el índice promedio de longitud de la u-T.
[11] investigación inédita.

publicados, pero sí los precursores del análisis de la madurez sintáctica en cada uno de sus países y padres de los investigadores posteriores.

Habrá que esperar hasta el año 1993 para que la investigación de la madurez sintáctica llegue a España de la mano de Torres González (1993) y su aplicación en la isla de Santa Cruz de Tenerife. Es en 2004 cuando la investigadora Irene Checa realiza su tesis doctoral acerca del desarrollo sintáctico en escolares almerienses, coincidiendo entonces con la zona geográfica a estudiar en nuestra futura investigación.

Las tres investigaciones que a continuación se desarrollarán serán: Mónica Véliz (1986), por ser la primera autora en analizar la madurez sintáctica en español (Chile), Antonia N. Torres González (1993), la cual fue pionera en la presentación de una tesis doctoral en España sobre madurez sintáctica por la Universidad de la Laguna, Santa Cruz de Tenerife, y finalmente, Irene Checa (2004), precursora de los estudios sobre medición sintáctica en Almería, dentro del grupo de lingüistas de la Universidad de Almería.

Véliz define madurez sintáctica como "la capacidad para producir unidades sintácticas estructuralmente complejas" (Véliz, 1986: 41-42). También es conveniente recordar la traducción al español realizada por la propia autora sobre las unidades que en todo momento van a ser básicas no solo para su estudio sino para su posterior escuela. Siguiendo Hunt, Véliz define unidad terminal (u-T) como " una construcción constituida por una cláusula principal más las cláusulas subordinadas que puedan ir añadidas o incrustadas en ellas" (1986: 53) y cláusula como "una construcción libre o dependiente, articulada en sujeto y predicado, con el verbo en forma personal" (1986: 50).

El estudio llevado a cabo por Véliz en Tenerife tiene como muestra a cincuenta escolares de cuarto, sexto y octavo de primaria más segundo y cuarto curso de medio, es decir, el rango de edad establecido por la investigadora se sitúa entre los 9 y los 17 años. Además, y como contrastó en su día Hunt (1965) Véliz se vale de nueve textos de profesionales para poder comparar los resultados dados. Otro de los factores a tener en cuenta es el nivel es el socioeconómico, aunque este queda generalizado al tratarse de un estudio en un colegio privado, del cual la media socioeconómica es: medio-alto.

La metodología llevada a cabo es el análisis de los índices primarios y secundarios (Hunt, 1986), además de ocho índices secundarios no clausales: promedio de calificativos por unidad terminal, promedio de posesivos por u-T; promedio de frases preposicionales por u-T; promedio de aposiciones por u-T; promedio de infinitivos por u-T; promedio de gerundios por u-T; promedio de participios-adjetivos por u-T; promedio de participios-predicativos por u-T (Torres González 1993:18).

El experimento estaba formado por el análisis de dos composiciones por cada sujeto, cada una de ellas contaba con alrededor doscientas cincuenta palabras y se trataban de narraciones; la primera de ellas era sobre una película y la segunda era temática, pero de composición libre.

El resultado de los datos llevó a una serie de conclusiones bastante detalladas. Entre los escolares, aparecen diferencias en las estructuras sintácticas empleadas por estos, la longitud de las cláusulas va en aumento con el número de estructuras clausales y no clausales que se van interponiendo con mayor frecuencia según el aumento de escolaridad. Poco a poco, el incremento de estructuras subordinadas va desplazando el número de frecuencia de estructuras coordinadas y yuxtapuestas.

En el estudio de los nueve textos de profesionales, cabe destacar que la el índice secundario de promedio de cláusulas adjetivas y adverbiales es más elevado que en los escolares, lo cual demuestra su complejidad frente a las sustantivas; de forma general, Véliz apunta que los resultados han demostrado que la madurez sintáctica sigue avanzando tras los años de escolaridad.

También cabe destacar una serie de estructuras (más complejas) de dominio tardía, como son las adjetivas introducidas por: *que* en función de sujeto o complemento directo, *donde, lo que*, las cuales no se adquieren hasta pasado de cuarto básico. Los tipos de oraciones que más tardíamente se adquieren se tratan de las cláusulas adjetivas cuyos nexos introductorios son: *que* como complemento circunstancial, *cuyo* y *cual*, siendo estas las más complejas de producir. Por último, también hay que destacar la dificultad de producción de las cláusulas adverbiales temporales introducidas por el nexo *hasta que.*

Véliz (1986) realiza una comparativa entre los resultados arrojados por su estudio y los anteriores sobre inglés y llega a la conclusión de que en ambas lenguas, la

trayectoria de la madurez sintáctica es similiar, sobre todo en el índice de promedio de las cláusulas adjetivas, las cuales van aumentando considerablemente en las dos lenguas; por el contrario, serán las cláusulas adverbiales las que tengan trayectorias dispares, puesto que en español su aumento va en consonancia con la escolaridad, mientras que en la lengua inglesa se produce un estancamiento a una edad media. Otra conclusión bastante importante es que el promedio de clásulas sustantivas dependen más de los modos del discurso que se estudien que de la edad de los sujetos, lo cual se desarrollará posteriormente en estudios de madurez semántica según los modos discursivos, ya planteado por Crowhurst (1978).

Véliz también realiza una revisión sobre la metodología utilizada para llevar a cabo la medición de su investigación. Para ello, considera favorables los índices cuantitativos de Hunt, destacando los índices primarios, como mejores parámetros para medir la madurez sintáctica. El número total de índices secundarios fue once, un gran número de variables a estudiar, lo cual la lleva a destacar en las conclusiones solo a aquellos índices secundarios que realmente han aportado datos significativos a su investigación, siendo los siguientes: promedio de cláusulas adjetivas por u-T; promedio de infinitivos por u-T; promedio de adjetivos calificativos por u-T y promedio de frases preposicionales por u-T.

En España encontramos el primer estudio sobre madurez sintáctica elaborado por Torres González en Tenerife (1993). El objetivo que persiguió a la autora analizar las estructuras de los tinerfeños fue realizar un corpus en el área metropolitana de Tenerife a través de los índices de Hunt y poder comparar así los resultados obtenidos con las investigaciones realizadas recientemente en Hispanoamérica.

Entre los objetivos llevados a cabo, Torres González señala tres: conocer las estructuras sintácticas, observar su distinción y comparar datos. La finalidad máxima del estudio tiene por objeto la mejora en la programación lingüística de la enseñanza del español como lengua materna (Torres González, 1993: 39).

Son cuatro las variables dependientes utilizadas: nivel de escolaridad, siendo los cursos analizados cuarto y octavo de EGB y COU; el carácter de la enseñanza: privado o público; sexo y nivel sociocultural. Entre las variables independientes con objeto de estudio se encuentran los índices primarios de Hunt: L/uT; L/Cl y Cl/u-T. La definición

escogida por Torres González de *unidad terminal* es la traducida por Véliz (1986: 42) y ya reproducida anteriormente en el presente trabajo[12].

Además de los índices primarios, Torres González analiza los índices secundarios clausales (Cl adj/u-T; Cl sust/u-T y Cl adv/ u-T) además de los modificadores nominales, los cuales clasifica según los siguientes promedios: adjetivos calificativos/u-T; posesivos/u-T; frases preposicionales/u-T y aposición/u-T. Como observamos, los índices secundarios no clausales están bastante más resumidos en Torres González que Véliz, puesto que la propia Véliz concluyó que algunos de los once promedios secundarios utilizados en su estudio no arrojaron conclusiones pertinentes.

Una vez expuesto el planteamiento de las variables, queda resaltar que el corpus estaba compuesto por un total de ciento cuarenta redacciones de tema libre, con un mínimo de doscientas palabras por redacción. Los datos aportados por la autora ascienden a un total de veintiocho mil novecientos treinta y cuatro palabras (28.934), dos mil ochocientos once unidades terminales (2.811) y cinco mil trescientas noventa y un cáusulas (5.391) (Torres González: 1993: 79).

Son muchas las conclusiones a las que se llegan con el estudio de Torres González pero destaremos aquí las seis conclusiones de carácter general, siendo cuatro de ellas revalidadas con las cuatro variables dependientes utilizadas:

1. En referencia con la variable escolaridad, existe un aumento con orden ascendiente salvo en las cláusulas sustantivas y aposiciones en las que octavo de EGB supera a COU.

2. Los datos demuestran que la variable centro es significativa, puesto que el promedio L-uT aumenta en la enseñanza privada con notable diferencia; de igual modo aparece un aumento en el promedio de adjetivos calificativos/u-T; y también aumenta en la escolaridad privada el promedio de L-cláusula aunque es menos significativo que en los casos anteriores.

3. No existe una relación significativa en la variable sexo. De forma general, se puede destacar un aumento significativo de los promedio de cláusulas/u-T y cl sust/u-T en el

[12] Unidad mínima de análisis: unidad terminal: "capacidad de producir unidades sintácticas estructuralmente más complejas" (Véliz, 1986: 42).

sexo femenino; de menor importancia se muestra el aumento de cl adv/u-T en el sexo masculino.

4. En la variable sociocultural, no hay datos significativos importantes, salvo en el uso de adjetivos calificativos, que fue en aumento poco a poco. Se establece una significación estadística entre las medias de los cursos según los estratos socioculturales.

5. Torres González coincide con Véliz en dar validez fiable a los índices estudiados sobre madurez sintáctica.

6. En la comparación realizada entre los estudios del mundo anglosajón e hispánico, llega a conclusiones como que en ambas lenguas los índices primarios influyen por el nivel de escolaridad, no por la variable sexo, centro o sociocultural; también señala que los índices secundarios no clausales tienen una dependencia de la variable zona notoria, y finalmente, señala que el promedio de los modificadores nominales por u-T aumenta solo con la variable curso escolar.

Checa (2004) es la primera investigadora en centrarse en la capital de Almería, por ello, además de su proximidad cronológica expondremos aquí las principales aportadas en su tesis doctoral. El objetivo a analizar es la competencia sintáctica de los almerienses, y para ello, utiliza las siguientes variables: escolaridad: sexto de primaria, segundo de ESO, cuarto de ESO y segundo de Bachiller, haciéndolos coincidir con el fin de los ciclos escolares; carácter del centro: público o privado; ubicación del centro: divide Almería por zonas geográficas en: centro, barriadas, zonas marginales y anejos (esto supone para Checa una neutralización por cada centro según su zona del factor sociocultural, analizado anteriormente por Véliz, 1993) y sexo.

El corpus recogido muestra un total de 237 sujetos, por tanto se recogieron 237 redacciones de escritura libre, el tema se les indicaba a los escolares (las vacaciones) y no tenían límite (ni máximo no mínimo) para redactarlas, y 237 redacciones de reescritura, utilizando el texto de Hunt sobre la gallina (1977: 103), entre las instrucciones solo se les indicaban que debían "reescribirlo mejor", para no indicarles que las redacciones tenía por objeto último el unir oraciones. El tiempo estimado para la creación de cada texto era aproximadamente veinte minutos. El total de redacciones analizadas fueron cuatrocientas setenta y cuatro (474).

Las variables independientes utilizadas por Checa fueron las siguientes (2004: 397-98):

- LUT (Longitud de la u-T): palabras / total u-T
- IS (Índice de Subordinación): cláusulas / total u-T [valor mínimo de 1]
- LCL (Longitud de Cláusula): palabras / cláusulas
- PAL/TXT (extensión del texto en palabras): palabras en texto.
- UT/TXT (extensión del texto en unidades terminales): unidades terminales en texto.
- LUTLE (Longitud de la Unidad Terminal Libre de Error): palabras en unidades terminales libres de error / unidades terminales libres de error.
- PUTLE (Proporción de Unidades Terminales Libres de Error): unidades terminales libres de error / unidades terminales [valor entre 0 y 1]
- PLUTLE (Proporción de palabras en Unidades Terminales Libres de Error): palabras en unidades terminales libres de error / palabras (totales en el texto) [valor entre 0 y 1]
- LUTI (Longitud de la Unidad Terminal Incorrecta): palabras en unidades terminales incorrectas / unidades terminales incorrectas
- PUTI (Proporción de Unidades Terminales Incorrectas): unidades terminales incorrectas / unidades terminales [valor entre 0 y 1]
- PLUTI (Proporción de palabras en Unidades Terminales Incorrectas): palabras en unidades terminales incorrectas / palabras [valor entre 0 y 1]
- LNOM (Longitud de la expresión NOMinal): palabras en expresiones nominales / expresiones nominales
- PNOM (Proporción de expresiones NOMinales en el texto): expresiones nominales / unidades terminales [valor entre 0 y 1]

Las conclusiones a las que llega Checa son la que la clase social y la zona del centro son variables significativas, puesto que aparecen notables diferencias entre los resultados obtenidos. La zona del centro es la variable dependiente de la clase social, porque se puede considerar que los centros almerienses están distribuidos por zonas, las cuales coinciden con la clase social de las familias, así por ejemplo, encontramos los valores más altos en los centros del centro de Almería, donde la clase social media-alta está matriculada. Se predispone una distribución de igual forma entre los centros de los

barrios marginales, donde los sujetos pertenecen a una clase social baja y sus resultados quedan por debajo de los obtenidos en el centro.

Finalmente, resume Checa que la clase social es la variable dependiente con mayor influencia en el resto de variables, coincidiendo con la teoría sobre la variable etnia propuesta por Walter Loban (1976:87).

4.3. Aplicaciones de la madurez sintáctica

Entre las aplicaciones de los estudios de madurez sintáctica, hemos comprobado que la mejora de la programación sobre la adquisición de la lengua materna es una de ellas, incluso la primera finalidad de dichos estudios, tal y como señalaba Torres González (1993). Pero las aplicaciones de las conclusiones obtenidas por los investigadores también se abren a nuevos campos de interés lingüístico, aquí solo haremos mención a dos aplicaciones entre las muchas que se pueden desarrollar.

4.3.1. Modo del discurso

Ya era conocido por Crowhurst y Piché (1970) los resultados obtenidos al analizar una redacción dependiendo del modo del discurso. Más tarde, Crowhurst siguió con la idea e hizo de su investigación la especialización en el análisis de estructuras sintácticas en los diferentes discursos, otros autores, ya hispánicos empiezan a relacionar sus datos con las ideas aportadas por Crowhurst y llevan a cabo estudios en los que interfieren los modos del discurso como variables dependientes a tener en cuenta.

Entre los investigadores, debemos destacar la labor de Véliz et al (1991) que desarrollaron un corpus sobre estudiantes chilenos, en dicho corpus se demostró que en los textos narrativos descendían los valores de los índices de Hunt, mientras que en los textos argumentativos y expositivos no había descenso alguno en los parámetros obtenidos. De dicho análisis se extrae que los textos argumentativos y expositivos tienen mayor madurez sintáctica, debido al aumento del IS (Índice de Subordinación) y los tres índices primarios.

Será años más tarde cuando Véliz (1999) publique un artículo en el que se centre en el estudio de la incidencia de las estructuras sintáctica según los modos del discurso.

Para la investigación, la autora utiliza los índices de madurez sintáctica de Hunt. Utiliza un corpus formado por sesenta textos escritos por adultos, por lo cual la sintaxis escogida para la investigación queda exenta de posibles errores entre los escolares. Los textos escogidos son de tres tipos de discursos: argumentativos, descriptivos y narrativos.

El resultado del estudio demuestra que la tipología discursiva incide en la madurez sintáctica utilizada por los escritores, por tanto, se deduce que algunos tipos de discursos son más complejos de elaborar que otros gracias a las estructuras sintácticas utilizadas. El orden que arrojan los resultados de la investigación es el siguiente: entre los textos con mayor complejidad se encuentran los argumentativos, después tienen lugar los descriptivos, y por último, es decir, los menos complejos, se encuentran los narrativos.

4.3.2. Enseñanza lengua no materna

La enseñanza del español como lengua extranjera ha dado un giro en las últimas décadas. El español se encuentra en su mejor momento de extensión y la escasa bibliografía sobre ELE ha supuesto un hueco por rellenar para los investigadores. Esta aplicación de la sintaxis y su estudio sobre las programaciones ELE ha sido recientemente estudiada por Checa (2005).

Checa realiza su investigación basándose en una recogida de treinta fragmentos de lecturas para la enseñanza del español como lengua extranjera. La aplicación de la madurez sintáctica en la enseñanza de una lengua extranjera requiere además una metodología capaz de medir también los errores, tal y como ella utilizó en 2004 en su tesis doctoral con escolares y preuniversitarios, donde ya expusiese la necesidad de una "Unidad Terminal Libre de Errores"[13]. La metodología utilizada en todo momento fueron los índices primarios de Hunt.

Los treinta textos se clasificaban a su vez en tres niveles de dificultad: los de nivel inicial, tenían una extensión de 200 a 900 palabras; nivel medio, con una extensión de 900 a 1800 palabras, y por último, los de nivel avanzado, entre 1.800 y 3.000 palabras. El modo del discurso de dichos fragmentos era el mismo, narrativo, para

[13] Pese a su interés, no existe una definición universal sobre dicha unidad.

neutralizar así las posibles diferencias que se pueden extraer si los textos fuesen de diferentes modos discursivos. Entre las características de los textos, cabe señalar que se han elegido textos originales y adaptaciones, variable importante a tener en cuenta.

Las conclusiones a las que llega Checa son las siguientes: aparece un incremento de los índices de madurez conforme va avanzando el nivel de dificultad en los textos, aunque dicho incremento no es siempre continuado, la tendencia es ascendente. En aquellos casos en los que se suceden descensos de los niveles, es decir, rupturas en la tendencia de carácter ascendiente, son debido a textos adaptados, no originales, los cuales han sido manipulados para disminuir su complejidad. El crecimiento del promedio longitud de las unidades terminales es el índice más significativo entre las variables independientes estudiadas. Por último, también cabe destacar que el IS solo varía en los textos adaptados, con lo cual, la diferencia más notoria sería la encontrada entre los textos adaptados frente a los originales, puesto que el índice de longitud de cláusula y el IS es favorable a los textos originales.

5. CONCLUSIÓN

Tras la detallada revisión bibliográfica aquí recogida se puede decir que existe un marco conceptual sobre la adquisición del español como lengua materna. Las investigaciones llevadas a cabo y dispuestas en los tres parámetros estudiados: ortográfico, léxico y gramatical definen una metodología aplicable a las técnicas de enseñanza del idioma. La competencia comunicativa se nutre en primera instancia de la competencia lingüística, ampliamente desarrollada y potenciada a través del aparto escolar. Enseñar un idioma atiende a una serie de complejidades de tipo estructural que se pasan por cada una de los parámetros aquí descritos. La presente revisión pone de manifiesto un gran cauce sobre el que se puede trabajar en el futuro para no caer de nuevo en antiguas y obsoletas técnicas de enseñanza, que como ya se ha demostrado lejos se encuentran de cumplir su objetivo.

Por tanto, la apuesta por proyectos basados en los análisis experimentales con los que el español ya cuenta sería la mejor opción para conocer sus resultados en las

aulas. El giro de la enseñanza de una lengua debe estar basado en unos principios empíricos, sobre los cuales poder poner en marcha un proyecto a largo plazo.

Dentro del parámetro ortográfico, las polémicas sobre los sistemas de transcripción y correlación entre *grafema-fonema* y *fonema-grafema* en una lengua fonográfica han sido constantes a lo largo de la historia tal y como se ha podido observar. Las aportaciones sobre la *Teoría de las dos rutas* de Bryant y Bradlye (1980) han sido discutidas y desarrolladas en tanto que se consideran dos procesos cognitivos para el aprendizaje de la lectura y después escritura. A partir de las rutas fonológicas y ortográficas se ha podido conocer más sobre el por qué de los errores ortográficos arbitrarios. Una de las clasificaciones también recogidas en el presente estudio se trata de Gignorini, Borzone y Diuk (2003), quienes sostienen que las correspondencias entre *grafema-fonema* y *fonema-grafema* del español se pueden encontrar tres: simples, condicionantes e irregulares. Serán estas últimas quienes mayor necesidad tienen de adaptar a la enseñanza, puesto que hemos visto que las normas y técnicas como el dictado no realizan ningún bien en el alumnado. Las nuevas propuestas se apoyan en la ruta ortográfica, es decir, apuestan por una serie de actividades basadas en la visualización (nunca de errores) y en la deducción. Los experimentos de Cuetos (1998) y Jiménez *et al.* (2008) señalan que la tercera etapa de desarrollo en la escritura, la ortográfica coincide con las edades escolares intermedias, en las que el número de errores ortográficos arbitrarios disminuyen gracias al aumento del lexicón mental. De esta manera, se pone en tela de juicio la enseñanza tradicional de la ortografía. Hay autores que van más allá y están a favor de una simplificación de las normas ortográficas, como Raúl Ávila. Su propuesta es radical y señala solo 25 grafías para poder acomodar nuestra lengua a un sistema totalmente fonográfico, además, solventa el problema de la homonimia en el *principio de autorregulación* de la lengua. En cuanto a la relación del sistema ortográfico y la aparición de un nuevo sistema de comunicación a través de las redes sociales y el auge de las nuevas tecnologías, nos ha llevado a experimentar un nuevo lenguaje adoptado en su mayoría por el colectivo juvenil. Queda clara la necesidad de imponer una revisión metodológica guiada a través de una renovación surgida mediante los resultados de los estudios sobre adquisición del español. El favorecimiento del aumento del lexicón mental viene dado gracias a actividades visuales como la redacción libre, a través de la cual se favorecerá también la expresión del alumnado.

En el parámetro léxico, desarrollado a través de los principios de la lexicometría del siglo XX, aparecen con gran interés los resultados de las investigaciones sobre disponibilidad léxica y riqueza léxica. Partiendo de López Morales (1973) como principal investigador de la disponibilidad léxica del español, cabe resaltar los resultados obtenidos anteriormente por autores como Gougenheim *et al.* (1964) quienes desarrollaron índices tan utilizados como el *Índice de disponibilidad léxica*. Cada vez la estadística léxica fue aportando mayor eficacia a través de fórmulas, como la de López Chávez y Strassburguer Frías (1987). Los estudios de disponibilidad se centraron en el desarrollo de recogidas de materiales basados en metodología de reacción, en las que se dividían el vocabulario en centros de interés. Este método sigue aún practicándose entre los últimos estudios sobre disponibilidad, entre los que cabe destacar Mateo García (1998) en Motril. Los estudios sobre léxico son cada vez más necesarios a la hora de conocer no solo el vocabulario sino los condicionantes socio-culturales de los hablantes. Entre los proyectos más destacables, cabes señalar el guiado por López Morales, el Proyecto Panhispánico, del que forman parte estudios de toda España e Hispanoamérica. También encontramos una versión del proyecto andaluza, desarrollada en cada una de las provincias. Conocer el léxico disponible también es una ventaja a la hora de enseñar español, además, los estudios sobre densidad léxica han ofrecido un mapa sociolingüístico clave para conocer el tipo de actuación a seguir.

Finalmente, dentro del parámetro gramatical, el enfoque sintáctico es la clave de las investigaciones señaladas. La acuciante enseñanza de la sintaxis española, así como su análisis pormenorizado, deja a veces un vacío en la producción. La expresión de los alumnos preuniversitarios es a veces superficial, lo cual demuestra que su *índice de madurez sintáctica* no es en tanto elevado. A partir de los postulados de Hunt (1965), el desarrollo de la madurez sintáctica, tras las primeras nociones de La Brant (1933), acerca de la "cláusula" han servido hasta la actualidad como indicadores de la expresión y de la sintaxis entre los hablantes. Cabe diferenciar entre los índices primarios y secundarios, los cuales considero hoy importante de conocer y aplicar al estudio. Sin embargo, no se ha estudiado con tanto ahínco el *Índice de profundidad de subordinación* (Hunt, 1965) el cual creo que también debería ajustarse para su puesta en marcha. Conocer los grados de subordinación alcanzados por los escolares suponen en suma que ellos mismos conozcan su propia expresión, porque la construcción de oraciones es en suma una de las acciones a través de la cual se conoce mejor la lengua

del individuo y su capacidad de comunicar. En la metodología de proyecto que se suscita se debería realizar una renovación de las técnicas de análisis sintáctico en boga de la construcción sintáctica y el ejercicio. Estudios como el de Checa (2005) son claros ejemplos de investigaciones detalladas de un análisis sintáctico, otras aportaciones, como la de Torres González (1993), que también sigue los postulados de los índices de Hunt, acercan a la realidad el conocimiento de la escritura oracional en escolares.

Por último, cabe destacar que a través de los tres parámetros desarrollados queda cubierto cualquier ángulo de la escritura y su aprendizaje. Queda a un lado el análisis textual, el cual me llevaría un estudio mayor y se abordará en el futuro. Las aportaciones bibliográficas vienen a demostrar que hay un amplio contenido sobre el que trabajar, puesto que los resultados encontrados delimitan perfectamente una senda sobre la que basar la metodología a emplear. En general, se puede decir que las nuevas técnicas necesarias tienen su fundamento en técnicas inductivas para así hacer partícipe protagonista al alumnado. Esta revisión bibliográfica supone el primer eslabón de una investigación en la cual se llevará como estudio de campo los parámetros aquí descritos y sobre los que se fundamentará mi estudio de la adquisición del español.

6. BIBLIOGRAFÍA

Alemán, M. (1609) *Ortografía castellana*, (edición de 1950), México: Colegio de México.

Anderson, J. E. (1937). "An Evaluation of varios indices of linguistic development", *Child Development*, 8:62-68.

Andrés Pérez, B. (1997) *Riqueza léxica en textos escritos de tres niveles de EGB.* Memoria de licenciatura. Alcalá: Universidad de Alcalá.

Aristizábal, M. (1938) *Détermation Esxpérimentale du Vocabulaire Écrit pour servir de Base à l'Enseignement de l'Orthographie a l'École Primaire.* Paris: H: Champion.

Artiles, C. y Jiménez, J. E. (2007) *PROESCRI Primaria: prueba de evaluación de los procesos cognitivos en la escritura*. Las Palmas de Gran Canaria: Conserjería de Educación, Cultura y Deportes del Gobierno de Canarias y Universidad de Las Palmas de Gran Canaria.

Ávila, R. (1986) "Léxico infantil de México: palabras, tipos y vocablos". En *Actas del II Congreso Internacional sobre "El Español de América"*. Mexico: Universidad Nacional Autónoma de México. 510-517.

Ávila, R. (1991) "Densidad léxica y adquisición del vocabulario: niños y adultos". En *Actas del III Congreso Internacional de "El Español de América"*. Valladolid: Junta de Castilla y León. 612-630.

Ávila, R. (1992) "Ortografía española: estratificación social y alternativas" en *Nueva Revista de Filología Hispánica,* Colegio de México, XL, 2, 649-672.

Ávila, R. (2009) "Ortografía e ideología: alogramas y estratificación social" en *La lingüística aplicada actual: comprendiendo el lenguaje y la mente*, Carmen Bretones Callejas *et al.* (eds), Almería: Universidad de Almería, 934-981.

Azurmendi Ayerbe, M. J. (1983) *Elaboración de un modelo para la descripción sociolingüística del bilingüísmo y su aplicación parcial en la comarca de San Sebastián,* 3 vols. San Sebastián: Caja de Ahorros Provincial de Guipúzcoa.

Barberá, V. (1988) *Cómo enseñar la ortografía a partir del vocabulario básico*. Barcelona: CEAC.

Bartol, J. A. (2000) "Reflexiones sobre la disponibilidad léxica" en *Actas de las III Jornadas de Reflexión sobre la Lengua Española,* Salamanca: Universidad de Salamanca (en prensa). También en J. A. Bartol, S. Crespo, C. Fernández, C. Pensado, E. Prieto y N. Sánchez (eds.), *Nuevas aportaciones al estudio de la Lengua Española*. Salamanca, Luso Española de Ediciones, págs. 221-236.

Bello, A. (1847) *Gramática de la lengua castellana destinada al uso de los americanos,* (edición de 1984), Madrid: Edaf, S.L.

Benítez, P. (1992) "Disponibilidad léxica en la zona metropolitana de Madrid" en *B.A.P.L.E.*, I, 1 (2ª época): 71-83.

Benítez, P. (1994) "Léxico real, léxico irreal en los manuales de enseñanza de español para extranjeros". En *Actas del II Congreso Nacional de la Asociación para la Enseñanza del español como lengua extranjera*. Málaga: ASELE.

Benítez, P. (1995) "Disponibilidad léxica en Madrid: análisis cuantitativo". En Rodríguez Fonseca y Vázquez, eds., en prensa.

Benítez, P; Hernández, C. E. y Samper, J. A. (1995) "Léxicos básicos de España *(LEBAES)* y de Canarias *(LEBAICan)*. *Proyecto de investigación*, en *R.E.A.L.E.*, 3: 9-17.

Bernstein, B. (1971/1989). *Clases, códigos y control. Vol*. 1. Madrid: Akal.

Bryant, P. y Bradley, L. (1980) "Why children sometimes write words which they do not read". En U. Frith (Ed.): *Conginitive processes in Spelling.* Londres: Academic Press.

Bub, D. y Kertesz, A. (1982) "Deep agraphia" en *Brain and Language*, 17: 147-166.

Buchanan, M. (1927) *A graded Spanish word book*. Toronto: Toronto University Press..

Butrón, G. (1987) *El léxico disponible: índices de disponibilidad*. Puerto Rico: Universidad de Puerto Rico. Tesis doctoral inédita.

Butrón, G. (1987) *El léxico disponible: índices de disponibilidad*. Puerto Rico: Universidad de Puerto Rico (Tesis doctoral inédita).

Butrón, G. (1991) "Nuevos índices de disponibilidad léxica" en H. López Morales, ed. *La enseñanza del español como lengua materna*. Río Piedras: Universidad de Puerto Rico, 79-89.

Camps, A. *et al*. (2004) *La Enseñanza de la Ortografía.* Barcelona: Grao (3ª ed.)

Canizal Arévalo, A. V. (1987) *Disponibilidad léxica en escolares de primaria terminada*. *Análisis de seis centros de interés.* Memoria de Licenciatura inédita. México: U.N.A.M.

Cano García, L. C. (2012) *Estudio empírico del español entre el alumnado nativo e inmigrado,* Memoria de Máster inédita. Almería: Universidad de Almería.

Carcedo, A. (2001) *Léxico disponible de Asturias*. Asturias: Universidad de Turku, Publicaciones del Departamento de Lengua Española, Monografías, vol. 1.

Casares, J. (1941) "La reforma ortográfica" en *Nuevo concepto del Diccionario de la lengua y otros problemas de lexicografía y gramática*. Madrid: Espasa-Calpe.

Cerveró, M. J. y Pichardo, F. (2000) *Aprender y enseñar vocabulario*. Madrid: Edelsa.

Checa García, I. (2004). *Medidas del desarrollo sintáctico en escolares almerienses (12 a 18 años): hitos descriptivos y aportes metodológicos.* Tesis doctoral inédita. Almería: Universidad de Almería.

Checa García, I. (2005). "Medidas de madurez sintáctica aplicadas a lecturas ELE", *Interlingüística,* 16:273-285. Recuperado el 25 de marzo de 2013 en http://dialnet.unirioja.es/servlet/articulo?codigo=2514225

Citrón Serrano, F. (1992) *Índices de riqueza léxica en escolares de Barranquitas*. Tesis de maestría. Puerto Rico: Universidad de Puerto Rico.

Correas, G. (1630) *Ortografía kastellana nueva i perfecta,* (Ed. 1971), Madrid: Espasa-Calpe

Cortázar, J. (1963) *Rayuela*, (Ed. Andrés Amorós de 1989) Madrid: Cátedra.

Crowhurst, M (1978) *The effect of audence and mode of discourse on the syntactic complexity of the writing of sixth and tenth graders.* Tesis doctoral. Universtity of Minnesota (1977): [Dissertation Abstracts International, 38, 7300A - 7301A].

Crowhurst, M. y Piché, L.G. (1979) "Audience and mode of discourse affects on syntactic complexity in writing at two grade levels". *Research in the Teaching of English* 13: 101-110.

Cuentos, F. (1989) "Lectura y escritura de palabras a través de la ruta fonológica" en *Infancia y Aprendizaje*, 45: 71-48.

Defior, S. y Serrano, F. (2005) "The initial developmetn of spelling in Spanish: From globla to analytical" en *Reading and Writing*, 18: 81-98.

Dimitrijévic, N. (1969) *Lexical Availability*. Berlín: J. Groos.

Echeverría, M. S. *et al.* (1987) "Disponibilidad léxica en Educación Meda. Resultados cuantitativos". En *VII Seminario de Investigación y Enseñanza de la Lingüística*. Valparaíso: Ediciones Universitarias de Valparaíso, 102-116.

Echeverría, M. S. et alii (1987) "Disponibilidad léxica en Educación Media. Resultados cuantitativos", *Revista de Lingüística Teórica y Aplicada*, 25, 55-115.

Echeverría, M. S. et alii (1991) "Crecimiento de la disponibilidad en estudiantes chilenos de nivel básico y medio" en López Morales ed. (1991).

Echeverría, M. S. *et alii.* (1985) "Disponibilidad léxica en Educación Media" en *R.L.A.*, 15: 55-112.

Echeverría, M. y A. Valencia *el al*. (1992) "Evaluación de la riqueza léxica de estudiantes de último año de enseñanza media". En *Estudios Filológicos*, 27: 59-71.

Ellis, A. (1982) "Spelling and writing (and reading and speaking). En A. Ellis (Ed.): *Normality and pathology in cognitive functions.* Londres: Academic Press.

Endicott, A. I. (1973) "A proposed scale for syntaxtic density". *Research in the Teaching of English* 7: 5-12.

Etxebarría, M. (1985): *Sociolingüística urbana. El habla de Bilbao*. Salamanca: Universidad de Salamanca.

Ferreiro, E. y Teberosky, A. (1979) Los sistemas de escritura en el desarrollo del niño. México: Siglo XXI

Ferreiro, E. y Teverosky, A. (1979) *Los sistemas de escritura en el desarrollo del niño*, México: Siglo XXI.

Frey Pereyra, M. L. H. (2007) "Disponibilidad léxica y escritura del español como lengua extranjera: propuesta de comparación de dos corpus". En *Interlingüística*, 17: 366-373.

Fuentes González, A. D. (1994) *Actitudes y conciencia sociolingüísticas en la ciudad de Almería*. Tesis Doctoral. Granada: Universidad de Granada. Edc. microfichas.

Fuentes González, A. D. (1997) "Disponibilidad léxica y fines específicos" en García Marcos, ed., 1997: 127-146.

Fuentes, D; García Marcos, F. J. y Mateo García, M. V. (1994) "Índices de mortandad del léxico cañero en la Costa Granadina". En Escañuela, E. (ed.). *Lo dulce a la conquista de Europa*, Granada, Excma. Diputación Provincial de Granada, 219-255.

Galloso V. (2002) *El léxico de los estudiantes preuniversitarios en el distrito universitario de Salamanca*. Salamanca: Ediciones Universidad de Salamanca.

Galloso, V. (1998) "Hacia una definición de léxico disponible: posibles aplicaciones". *Actas III Congreso de Lingüística General. Lingüística para el siglo XXI,* Salamanca: Universidad de Salamanca, 711-717.

García Domínguez, M. J.; Marrero, V.; Pérez Martín, J. A. y G. Piñero (1994) "Estudio de la disponibilidad léxica en Gran Canaria. La variable geográfica y el tipo de eduación" en *R.E.A.L.E.*, 2: 65-72.

García Hoz, V. (1953) *Vocabulario Usual, Común y Fundamental*. Madrid: Instituto San José de Calasanz.

García Marcos, A. (2010) *Adquisición de español como segunda lengua. Evaluación empírica y alternativas en Lingüística Aplicada*. Memoria de Máster de Maestría inédito. Jaén: Universidad de Jaén.

García Marcos, F. J. (1985) *El segmento fónico **vocal** + S en ocho poblaciones de la Costa Granadina. Aportación sociolingüística, estadística e informática al reexamen de la cuestión.* Granada. Universidad de Granada. Memoria de Licenciatura inédita.

García Marcos, F. J. (1997) *Estudios de disponibilidad léxica*. Almería/Barcelona: GRUSTA/Sabir.

García Marcos, F. J. (1999) *Fundamentos críticos de sociolingüística*. Almería: Universidad de Almería.

García Marcos, F. J. y Mateo García, M. V. (1994) "Disponibilidad léxica y producción de materiales para la enseñanza de lenguas extranjeras". En *Actas de las II Jornadas sobre aspectos de la enseñanza de español a extranjeros*. Granada: Universidad de Granada, en prensa.

García Marcos, F. J. y Mateo García, M. V. (1995) "Disponibilidad léxica en Almería: análisis cuantitativo". En Rodríguez Fonseca y Vásquez, eds., en prensa.

García Márquez, G. (1997) Botella al mar para el dios de las palabras [en línea] Caracas: Biblioteca electrónica. Recuperado el 01 de febrero de 2013 en: http://www.analitica.com/bitblio/ggm/botella.asp

García Orza, J., Madrazo Lazcano, M., Viñals Álvarez, F. (2002) "Alteraciones del procesamiento de la escritura: la disgrafía superficial" en *Revista Española de Neuropsicología*, [en línea], 4: 283-300. Recuperado el 02 de marzo en: http://dialnet.unirioja.es/servlet/articulo?codigo=1007109

García Rosas, M. P. (1996) *Riqueza léxica en textos de estudiantes de español como lengua extranjera*. Memoria de máster. Alcalá: Universidad de Alcalá.

Golub, L. S. y C. Kidder (1974) "Syntactic Density and the Computer". *Elementary English* 51: 1128-1131.

Gómez Devís, M. B. (2004) *La disponibilidad léxica de los estudiantes preuniversitarios valencianos: reflexión metodológica, análisis sociolingüístico y aplicaciones*. Tesis Doctoral. Valencia: Universitat de Valencia. Recuperado el 02 de abril de 2013 en: http://www.tdx.cat/handle/10803/9814

Gómez Torrego, L. (2011) *Las normas académicas: últimos cambios*. Madrid: SM.

González Martínez, A. (1997) *Disponibilidad léxica de Cádiz*. Tesis Doctoral inédita. Cádiz: Universidad de Cádiz.

Gougenheim, G., Michéa, R., Rivenc, P. y Sauvageot (1956) *L'elaboration du Francais élémentaire. Etude sur l'establissement d'un vocabulaire et d'un grammaire de base*. Paris: Didier.

Guiraud, P. (1959) *Problèmes et mèthodes de la statistique linguistique.* Dordrecht: Reidel.

Haché de Yunén, A. M. (1991) "Aportes de las pruebas de riqueza léxica a la enseñanza de la lengua materna". En *La enseñanza del español como lengua materna*. En Humberto López Morales (ed.). Puerto Rico: Universidad de Puerto Rico: 49-90.

Hunt, K. W. (1965). *Grammatical structures written at three grade levels (Research Report No. 3).* Urbana Champaign, [I11]: National Council of Teachers of English.

Hunt, K. W. (1967). *Sentence structures used by superior students in grades four and twelve, and by superior adults.* Tallahassee, [F1]: Florida State University.

Hunt, K. W. (1970). "Recent measures of syntactic development" en M. Lester (Ed.) *Readings in Apllied Transformational Grammar*. New York: Holt, Rinehart y Winston. Pp, 179-192.

Hunt, K. W. (1970a). "Recent measures in syntactic development". *Reading in applied transformational grammar*. Ed. Mark Lester. Nueva York: Holt, Rinehart and Wiston. 179- 192.

Hunt, K. W. (1977) "Early blooming and late blooming syntactic structures", en Ch. R. Cooper y L. Odell (Eds.) *Evaluating Writing: Describing, Measuring, Judging.* Urbana Champaign. [I11]: National Council of Teachers of English. Pp. 91-104.

Jiménez, J. E., O'Shanahan, I., Tabraue, M. L., Artiles, C., Muñetón, M., Guzmán, R., Naranjo, F. y Rojas, E., (2008) "Evolución de la escritura de palabras de ortografía arbitraria en lengua española" en *Psicothema* [en línea], 20, 4: 786-794). Recuperado el 02 de marzo de 2013: http://www.psicothema.com/resumen.asp?id=3556

Juilland, A. y Chang-Rodríguez, E. (1964) *Frecuency Dictionary of Spanish Words.* La Haya: Mouton.

Justo Hernández, H. G. (1986) *Disponibilidad léxica en colores.* Memoria de Licenciatura inédita. México: U.N.A.M.

La Brant, L. L. (1933). "A study of certain language development of children in grades four to twuelve inclusive", *Genetic Psychology Monographs* 14: 387-491.

Labov, W. (1972) "The logic of non-standard English", en P. Giglioli (ed) *Language and social context*, Londres: Penguin, 179-215.

Liberman, A. (1988) "Reading is hard just because listening is easy" en *Haskins Laboratories, Status Report on Speech Research*, S-R-95/96: 145-150.

Loban, W. (1976). *Language Development: Kindergarten throught grade twuelve* (Research Report No. 18). Urbana [I11]: National Council of Teachers of English.

López Chávez, J. (1992) "Léxico fundamental panhispánico: realidad o utopía". En *Actas del IV Congreso Internacional sobre el español de América*, en prensa.

López Chávez, J. (1995) "El léxico en estudiantes mexicanos de primaria". En Rodríguez Fonseca y Vásquez, eds., en prensa.

López Chávez, J. y Strassburger Frías, C. (1987) *Otro cálculo del índice de disponibilidad léxica*. México: Universidad Nacional Autónoma de México.

López Chávez, J. y Strassburger Frías, C. (1991) "Un modelo para el cálculo de disponibilidad léxica individual". En López Morales, ed., 92-112.

López Morales, H. (1973), *Disponibilidad Léxica en los Escolares de San Juan.* Ms.

López Morales, H. (1983) "Lingüística estadística" en H. López Morales, ed. (1983). *Introducción a la lingüística actual.* Madrid: Playor, 209-225.

López Morales, H. (1986) *Enseñanza de la lengua materna*. 2ª ed. Madrid: Playor.

López Morales, H. (1991) *La enseñanza del español como lengua materna.* Actas del II Seminario Internacional sobre 'Aportes de la lingüística a la enseñanza de la lengua materna'. Río Piedras: Editorial de la Universidad de Puerto Rico.

López Morales, H. (1994) "Índices de complejidad sintáctica y memoria inmediata", *REALE* 1: 85-105.

López Morales, H. (1995) *Léxico disponible de Puerto Rico.* San Juan: Sociedad Lingüística del Caribe hispánico, en prensa.

López Morales, H. (1996) "Los estudios de disponibilidad léxica: pasado y presente" en *Homenaje a Rodolfo Oroz,* BFUCh XXXV: 245-259. Recuperado el 20 de abril de 2013: http://www.boletinfilologia.uchile.cl/index.php/BDF/article/viewFile/19231/20354

López Morales, H. y García Marcos, F. J. (1995) "Disponibilidad léxica en Andalucía proyecto de investigación" en *REALE*, 3: 65-76.

Lorán R. (1987) *La tasa de sustitución e índice de disponibilidad léxica*. Mayagüez: Dpto. de Matemáticas. Universidad de Puerto Rico (informe técnico).

Lorán, R. y López Morales, H. (1983) "Nouveau calcul de l'indice de disponibilité". Ms.

Lorán, R. y López Morales, H. (1983) "Nouveau calcul de l'indice de disponibilité". Ms.

Mackey, W. F. (1971) *Le sondage dans les enquêtes de disponibilité.* Québec: CIRB.

Mateo García, M. V. (1998) *Disponibilidad léxica en el COU almeriense*. *Estudio de estratificación social*, Almería: Universidad de Almería.

Mateo García, M.V. (1996) *Estratificación social del léxico de la alimentación en Motril* (*Granada*). Almería: GRUSTA.

Mena Osorio, M. (1986) *Disponibilidad léxica infantil en tres niveles de enseñanza básica* (tesina inédita). Concepción: Universidad de Concepción.

Mena, M. (1986) *Disponibilidad léxica infantil en tres niveles de enseñanza básica*. Tesis de *Magister*. Concepción (Chile): Universidad Nacional Autónoma de México.

Menard, N. (1983) *Mesure de la richesse lexicale. Théorie et vérifications expérimentales. Etudes stylométrique et sociolinguistiques.* Genève-Paris: Slatkine-Champion.

Michéa, R. (1953): "Mots fréquents et mots disponibles: un aspect nouveau de la statistique du langage", *Les Langues Modernes*, 47, 338-344.

Morton, J.(1980) *The logogen model and orthographic structure.* En U. Frith (Ed.): *Cognitive processes in Spelling*. Londres: Academic Press .

Müller, Ch. (1968-1973): *Estadística Lingüística*. Madrid, Gredos.

Muñoz Hernández, I. O. (2005) *Disponibilidad léxica en alumnos de enseñanza media de Almería*, (tesis inédita), Almería: Universidad de Almería.

Nebrija, A. (1942) *Gramática de la lengua castellana*, (Ed. A. Quilis, 1990), Madrid: Centro de Estudios Ramón Areces.

Nolan, K., y Caramazza, A. (1983) "An analysis of writing in a case of deep dyslexia" en *Brain and Language*, 20: 305-328.

Norman, J. (1994) "El vocabulario español en los manuales suecos. Comparación entre el estudio de disponibilidad léxica y seis manuales suecos", *Revista de Adquisición de la Lengua Española (REALE).*

O'Donnell (1968) "An objetive measure of structural complexity in children's writting". Unpublised paper delivered at the American Educational Research Association Meeting.

O'Donnell (1976) "A critique of some indices of syntactic maturity" *Research in the Teaching of English* 10: 31-38.

O'Donnell, R. C.; W. J. Griffin y R. C. Norris (1967). *Syntax of Kindergarten and Elementary School children: A transformational analysis*. Research Report No.8, Champaign [I11]: National Council of Teachers of English.

Portela, C. (1992) *Índices de riqueza léxica en estudiantes del primer año universitario*. Tesis de maestría en Lingüística. Santiago de los Caballeros: Pontificia Universidad Católica Madre y Maestra.

RAE, (1999) *Ortografía de la lengua española*, Madrid: Espasa-Calpe.

RAE, (2010) *Ortografía de la lengua española*, Madrid: Espasa-Calpe.

Reyes Díaz, M. J. (2008) "Riqueza léxica de textos redactados por alumnos de Bachillerato de Las Palmas de Gran Canaria". En *Anuario de Lingüística Hispánica*. XXIII-XXIV:147-163. Recuperado el 10 de marzo de 2013 en http://dialnet.unirioja.es/servlet/articulo?codigo=3209971

Rodríguez Bou, I. (1952) *Recuento del vocabulario español*, I y II. Río Piedras: Editorial Universitaria.

Rodríguez Muñoz, F. J. y Muñoz Hernández, I. O. (2009) "De la disponibilidad a la didáctica léxica", *Tejuelo*, 4: 8-18.

Rodríguez Muñoz, F. J. y Muñoz Hernández, I. O. (2011) "Disponibilidad léxica sobre palabras específicas en estudiantes de educación secundaria de Almería". En *Espiral. Cuadernos del Profesorado*, 4(8) 22-31. Recuperado el 08 de marzo de 2013 en http://www.cepcuevasolula.es/espiral

Román, B. (1985) *Disponibilidad léxica en escolares de Dorado*. *Puerto Rico* (tesina inédita). Río Piedras: Universidad de Río Piedras.

Román, B. (1987) *Disponibilidad léxica en escolares de Dorado*. Tesina inédita. Río Piedras: Universidad de Puerto Rico.

Ruiz de Francisco, I. (1994) *La enseñanza de la ortografía: Fundamentacion lingüística y didáctica*. Las Palmas de Gran Canaria: Departamento de Didácticas Especiales. Universidad de Las Palmas de Gran Canaria.

Ruiz, A. (1987) *Disponibilidad léxica de los alumnos de primer ingreso en el Colegio de Ciencias y Humanidades, Plantel de Naucalpan.* Memoria de Licenciatura inédita. México: U.N.A.M.

Samper, J. A. (1990) *Estudio sociolingüístico del español de Las Palmas de Gran Canaria.* Las Palmas: Cajacanarias.

Samper, J. A. y C. E. Hernández Cabrera (1995) "Disponibilidad léxica en Gran Canaria. Primeras estimaciones" en *Jornadas Internacionales de Lingüística Aplicada. Homenaje a Di Pietro.* Granada: Universidad de Granada. Ms.

Samper, J. A. y C. E. Hernández Cabrera (1997) "Estudio de la disponibilidad léxica en Gran Canaria. Datos iniciales y variación sociolingüística". *Homenaje a R. Trujillo Carreño, Revista de Filología.* Universidad de La Laguna, vol. II, 229-239.

Samper, J. A. y Hernández, C. E. (1994) "Aspectos de la disponibilidad léxica en Las Palmas. En *Homenaje a Ramón Trujillo*. La Laguna: Universidad de La Laguna, en prensa.

Sánchez Jiménez, D. (2006) *Análisis de errores ortográficos de estudiantes filipinos en el aprendizaje del español como lengua extranjera,* [en línea]. Ministerio de Educación y Ciencia. Biblioteca Virutal RedELE, 6. Recuperado el 02 de marzo de 2013 en: http://www.mecd.gob.es/dctm/redele/MaterialRedEle/Biblioteca/2006_BV_06/2006_BV_06_13Sanchez_Jimenez.pdf?documentId=0901e72b80e30c02

Sánchez, D. (2009) "Una aproximación a la didáctica de la ortografía en la clase de ELE" en *MarcoELE. Revista de didáctica ELE*, [en línea], 9. Recuperado el 22 de marzo de 2013 en http://marcoele.com/descargas/9/sanchez_ortografia.pdf

Shallice, T. (1981) "Phonological agraphia and the lexical route in writing" en *Brain*, 40: 413-429.

Signorini, A., Borzone, A. M., Diuk, B., (2001-2003) "Del conocimiento fonológico al conocimiento ortográfico. Alcance y limitaciones de la perspectiva "fonocéntrica" en el procesamiento de palabras escritas" en *Lenguas Modernas*, 28-29: 7-27.

Stewart, M. F. (1978) "Freshman Sentence Combining: A Canadian Project", *Research in the Teaching of English* 12 (3): 257-267.

Tena, P. (Junio, 1999) “Dictados para la enseñanza del español como LE”, en *Revista Espéculo, Revista de Estudios Literarios* [en línea], 11. Recuperado el 06 de marzo de 2013 en: http://www.ucm.es/info/especulo/numero11/dictados.html

Terrádez, M. (1996) *La disponibilidad léxica del español hablado de Valencia*. Tesis de Licenciatura (inédita). Valencia: Universitat de València.

Terrádez, M. (2002) *Frecuencias léxicas del español coloquial: análisis cuantitativo y cualitativo.* Cuadernos de Filología, Anejo XLI. Valencia: Universitat de València.

Thorndike, E. L. (1921) *The Teacher's Word Book.* NuevaYork: MacMillan.

Torrego González, A. (2011) "Eskriibo en el Tuenti komo pronunciioh. Apuntes sobre la ortografía en una red social" en *Tarbiya,* [en línea], 41: 33-52. Recuperado el 05 de marzo de 2013 en: http://dialnet.unirioja.es/servlet/articulo?codigo=3649718

Torres González, A. N. (1992). ""Índices secundarios de madurez sintáctica en escolares no-universitarios de Tenerife". Ponencia presentada en el I Congreso Internacional sobre "Enseñanza del Español". Centro Madrileño de Investigaciones Pedagógicas (CEMIP), Madrid del 27 al 31 de enero (en prensa).

Torres González, A. N. (1993). *Madurez sintáctica en estudiantes no universitarios de la zona metropolitana de Tenerife*. Tesis doctoral. Santa Cruz de Tenerife: Universidad de la Laguna. Recuperado el 04 de marzo de 2013 de http://exordio.qfb.umich.mx/archivos%20pdf%20de%20trabajo%20umsnh/aphilosofia/a%20tsa/cs3.pdf

Torres González, A. N. (1994). "Índices primarios de madurez sintáctica en escolares de Tenerife", *REALE* 2: 115-126.

Torres González, A. N. (1996). *Complejidad sintáctica en estudiantes de Tenerife*. Santa Cruz de Tenerife: Centro de la cultura popular canaria.

Torres González, A. N. (1997). "Complejidad sintáctica en escritores de las Islas Canarias", *REALE* 8:105-125.

Torres González, A. N. (1999) "Incidencia de las variables sociales en los índices de producción léxica de estudiantes del último curso de la enseñanza no universitaria". En *Contribuciones al estudio de la Lingüística Aplicada*. Edita Julián de las Cuevas y Dalila Fasla. Castellón: Asociación Española de Lingüística Aplicada. 393-401.

Torres González, A. N. (2003a) "Riqueza léxica en textos narrativos escritos por estudiantes de Tenerife". En F. Moreno Fernández, F. Gimeno Menéndez, J. A. Samper, M. L. Gutiérrez Araus, M. Vaquero y C. Hernández (eds.). En *Lengua, variación y contexto. Estudios dedicados a Humberto López Morales*. Madrid: Arco/Libros. 435-449.

Torres González, A. N. (2003b) "La producción léxica de estudiantes no universitarios de Tenerife". En *Estudios sobre el español de Canarias. Actas del I Congreso Internacional sobre Español de Canarias.* La Laguna y Las Palmas de Gran Canaria: Academia Canaria de la Lengua. 989-1009.

Vander Beke, G. (1935) *French Word Book*. Nueva York: MacMillan.

Véliz de Vos, M. (1986). *Evaluación de la madurez sintáctica en el discurso escrito.* Tesis doctoral inédita. Chile: Universidad de Concepción.

Véliz de Vos, M. (1988). "Evaluación de la madurez sintáctica en el discurso escrito", *RLA* 26:105-141.

Véliz de Vos, M. (1999). "Complejidad sintáctica y modo del discurso", *Estudios filológicos,* 34: 181-192. Recuperado el 08 de marzo de 2013 en http://www.scielo.cl/scielo.php?pid=S007117131999003400013&script=sci_arttext

Véliz de Vos, M; Muñoz, G.; Echeverría, Max S.; Valencia, A.; Ávila, E, y Núñez, N. (1991) "Evaluación de la madurez sintáctica en estudiantes chilenos de cuarto medio", *Estudios filológicos* 26: 71-81.

Verlée, L. (1934) *Basis-Woordenbock voor de Franse Taal*. Amsterdam: Neulenhoff.

Yus, F. (2001). *Ciberpragmática: entre la compensación y el desconcierto*. Congresos Internacionales de la Lengua Española: Congreso de Valladolid.

Zamudio, C. M., (2007) *Las consecuencias en la teoría lingüística*, [en prensa], México: El Colegio de México.