T E S I S D O C T O R A L

TITULO: "DETERMINACION Y ANALISIS DEL VOCABULARIO DEL NIÑO DE 6 A 10 AÑOS EN ANDALUCIA ORIENTAL."

DIRECTOR: Dr. D. JOSE BERNIA PARDO.
Catedrático de Psicología.

PONENTE: Dr. Dn. JAIME VILA CASTELAR.
Agregado de Psicología.

FERNANDO JUSTICIA JUSTICIA
DEPARTAMENTO DE PSICOLOGIA
FACULTAD DE FILOSOFIA Y LETRAS
UNIVERSIDAD DE GRANADA

Granada, Enero, 1983.

Esta investigación fue realizada mediante una beca patrocinada por la Comisión Asesora de Investigación Científica y Técnica del Ministerio de Universidades e Investigación.

Proyecto de Investigación 4107/79.

PROYECTO 4107/79:

DETERMINACION Y ANALISIS DEL VOCABULARIO DEL NIÑO DE 6 A 10 AÑOS EN ANDALUCIA ORIENTAL.

FERNANDO JUSTICIA JUSTICIA

INFORME FINAL.

COMISION ASESORA DE INVESTIGACION CIENTIFICO Y TECNICA.

AÑO 1.983

DETERMINACION Y ANALISIS DEL VOCABULARIO DEL NIÑO DE 6 A 10 AÑOS EN ANDALUCIA ORIENTAL.

INDICE GENERAL.

INTRODUCCION

INTRODUCCION.-

El vocabulario es probablemente el aspecto, dentro del proceso de adquisición y desarrollo del lenguaje, que ha sido objeto de mayor número de investigaciones. Investigaciones que fueron llevadas a cabo por lingüistas, psicólogos, pedagógos e incluso por muchos padres en quienes la palabra, su aparición y desarrollo, suscitó el interés y la curiosidad.

De los recuentos sumarios absolutos de palabras por edades se ha pasado recientemente a la realización de estimaciones más limitadas pero, al mismo tiempo, mejor ponderadas. Las últimas tendencias en los estudios del vocabulario infantil abordan el problema de su composición a partir de la realización de cálculos inferenciales basados en la frecuencia de uso que presentan las palabras en determinadas muestras del lenguaje oral o escrito del niño.

En casi todos los trabajos sobre el tema subyace un doble interés. De una parte, el interés científico por un mejor conocimiento del desarrollo de esta faceta del comportamiento humano. Y, en segundo lugar, un interés eminentemente práctico y pedagógico. En efecto, el lenguaje es uno de los pilares básicos que sustentan la enseñanza. La cantidad de exigencias, conocimientos y destrezas, que impone la sociedad al individuo impide que éste por sí mismo sea capaz de asumirlos satisfactoriamente. De aquí la necesidad de institucionalizar la escuela y exigir su entrada en ella a una edad cada vez más temprana. Por otra parte, el vehículo fundamental de transmisión de estos conocimientos es la

palabra, ya que resulta imposible que el sujeto pueda acceder a ellos a través de la acción o de la demostracción. Por consiguiente, es imprescindible el dominio del lenguaje y del vocabulario si queremos asegurar el éxito escolar de los alumnos.

La consecución de este objetivo exige, previamente, el conocimiento del vocabulario usual del que dispone el niño. Con ello no queremos decir que la enseñanza, en su totalidad, haya de girar en torno al vocabulario usual, ni que parcialicemos el vocabulario a los límites que marca el vocabulario usual. De actuar así, ciertamente estaríamos contribuyendo a un empobrecimiento del alumno cuando el objeto de la educación es precisamente lo contrario. Lo que queremos señalar es que el vocabulario usual debe servir de punto de partida no sólo para el aprendizaje de contenidos y tareas relacionados con el área del lenguaje sino, en general, con todas la materias de enseñanza. En este sentido, todo vocabulario básico sería un instrumento de trabajo,más que una meta o fin en sí mismo de enseñanza. Se sabe que el aprendizaje es más sólido y eficaz cuando enlaza con la experiencia previa del alumno, cuando la presentación de contenidos se adapta y conforma a su capacidad de organización cognitiva. A ello contribuye, indudablemente, el lenguaje y el vocabulario.

La vigencia e interés del tema es aún mayor al constatar el escaso número de trabajos que existen en nuestro país acerca de la frecuencia de uso de palabras empleadas por niños y, en particular, la inexistencia de un estudio que analice el vocabulario,desde esta perspectiva, en el período correspondiente a los primeros cursos de la E.G.B.

Nuestra investigación se propuso un doble objetivo:

1º) Determinar el vocabulario usual del niño, comprendido entre los seis y diez años de edad, dentro del ámbito geográfico de Andalucía Oriental. La usualidad la impone la frecuencia de aparición de la palabra en determinadas muestras del lenguaje. Para el análisis de la producción verbal infantil utilizamos textos escritos.

2º) Analizar la influencia que ejercen determinados factores en el desarrollo del vocabulario. Consideramos cuatro factores: la edad (nivel de escolarización), el sexo, la zona donde habitan los sujetos y la clase social.

El trabajo responde, en su presentación, al siguiente orden:

En primer lugar y como inicio de la revisión teórica del tema, ubicamos los estudios sobre la palabra y el vocabulario en psicolingüística (capítulo 1º). En segundo lugar, analizamos los planteamientos y presupuestos metodológicos implícitos en los recuentos de vocabulario que tienen por objeto el estudio de la producción verbal de sujetos adultos, para otras lenguas (capítulo 2º) y para el español (capítulo 3º). Acabamos la parte introductoria con la revisión de los estudios sobre el vocabulario infantil (capítulo 4º).

A continuación, procedemos a la exposición detallada de objetivos, asícomo de las principales hipótesis de nuestro trabajo (capítulo 5º). La muestra, los problemas inherentes a la

muestra, el tipo de muestreo, las distintas fases y etapas seguidas en la selección de los sujetos, se exponen en el capítulo 6º. La recogida de datos y las fuentes del vocabulario se analizan en el capítulo 7º.

Posteriormente, en el capítulo 8º, presentamos el diseño y los resultados correspondientes al análisis del incremento cuantitativo del vocabulario. La discusión de resultados y conclusiones aparecen en el último capítulo (capítulo 9º). En los Anexos se registra el Vocabulario Usual atendiendo a las dos formas de presentación al uso en los diccionarios de frecuencias, a saber, listado de vocablos en orden alfabético y en orden decreciente de frecuencia.

Este proyecto de investigación fue becado por la Comisión Asesora Científico Técnica del Ministerio de Universidades para su realización en el año 1980. Le agradecemos la ayuda y la asistencia prestada durante estos tres largos años, dedicados a la ímproba tarea de contar, clasificar y registrar palabras.

Igualmente debo extender mi agradecimiento por el apoyo y el interés mostrado por el trabajo al personal de la administracción, a los directores y maestros de los centros que colaboraron en el proyecto.

Mi gratitud también para D. Ramón Gutiérrez Jaimez, director del Departamento de Estadística de la Universidad de Granada (facultad de Ciencias) por sus acertadas orientaciones.

Asímismo, estoy agradecido a mis compañeros del Departamento de Psicología de la Universidad de Granada por su constante ánimo, su apoyo y colaboración, especialmente a D. Pío Tudela por

su inestimable ayuda en el análisis de los datos.

Por último, he de agradecer sobre todo, la comprensión y ayuda manifestada en todo momento por D. José Bernia Pardo, director del presente proyecto de investigación y a D. Jaime Vila Castelar, ponente del mismo por la Universidad de Granada.

CAPITULO I

I. EL VOCABULARIO.-

Uno de los aspectos en la adquisición y desarrollo del lenguaje del que poseemos mayor información es, sin duda, el vocabulario. Y, particularmente el vocabulario infantil.

Las palabras han sido repetidas veces unidades de observación y análisis con distintas finalidades: lingüísticas, psicológicas, sociólógicas... Nos interesa destacar aquí la investigación psicológica al respecto, si bien el carácter interdisciplinar que posee este tipo de estudios hace difícil el trazado de una línea divisoria entre unas disciplinas y las otras, particularmente entre la lingüística y la psicología.

No obstante, mientras las palabras y las sílabas han sido las unidades de la investigación psicológica, para los lingüistas, por el contrario, la unidad de observación y análisis es generalmente el morfema. Los morfemas o formas simples son las formas lingüísticas en las que no "existe semejanza parcial, fonológica o significativa, con otra forma lingüística" (Alcina y Blecua, 1975). Las palabras constan de uno o más morfemas, y éstos a su vez suelen estar compuestos por secuencias más cortas de fonemas (sonidos del habla) o de su representación gráfica, es decir, los grafemas. Los morfemas pueden ser fonemas o grafemas individuales o estar constituídos por más de un fonema. En este sentido, no necesariamente morfema y sílaba son términos equiparables o idénticos. Los detalles del análisis morfémico los proporciona la lin-

güística y más concretamente la morfología (Hockett, 1958; Alcina y Blecua, 1975).

Los estudios a los que aquí vamos a hacer referencia no abordan este carácter de interés particular para el lingüista. Consideran la palabra como unidad mínima a la que se puede reducir la producción verbal oral o escrita de un sujeto. Interesa la palabra sólo como secuencia de fonemas antes y después de los cuales es posible una pausa (en la producción oral), o bien son posibles espacios o signos de puntuación (en la producción escrita). En este sentido, no se establecen diferencias entre unas unidades y otras, entre unas palabras y otras. Todas tienen el mismo rango y se consideran cuantitativamente equivalentes.

Hacemos a continuación unas consideraciones generales sobre los estudios del vocabulario, para proceder después a un análisis más específico.

Las investigaciones sobre el vocabulario en psicología responden a dos orientaciones básicas (Goss, 1973):por un lado, tendríamos los estudios sobre la producción verbal en los que la palabra se produce en situaciones controladas de tal modo que su finalidad no reside en el análisis global de la producción lingüística de un sujeto sino de las palabras que se producen o seleccionan como respuesta bajo determinadas condiciones estímulo. En segundo lugar, los estudios cuyo objeto es el análisis de la producción verbal en sí misma.

1.1.- La Ocurrencia de las palabras en situación estímulo-respuesta.

Las palabras se derivan o seleccionan bajo una situación de estímulo determinada. Son una respuesta ante el estímulo. Los estímulos que se emplean suelen ser tanto de naturaleza verbal, es decir, otras palabras o sílabas, como no verbal, generalmente estímulos de naturaleza visual o auditiva. En este segundo caso, las palabras producidas corresponden a la denominación de las figuras, formas o colores que constituyen el material de estímulo utilizado.

Los resultados de estos trabajos no proporcionan datos directamente relacionados con el lenguaje de un sujeto o grupo de sujetos sino con otros procesos tales como percepción, discriminaciòn, asociación, reconocimiento de respuestas, adquisición del significado...

- Así, por ejemplo, las llamadas respuestas de reconocimiento implican reproducir, repetir o imitar sílabas y palabras presentadas bajo diversas condiciones estímulo. También pueden presentarse para su reconocimiento formas, colores o ambas cosas combinadas. Un caso particular de estos estudios ha dado origen a los tets de dibujos -vocabulario, cuya finalidad ha trascendido el establecimiento de unos determinados patrones de reconocimiento para convertirse en pruebas clásicas utilizadas para la medición y análisis del lenguaje espontáneo y particularmente del vocabulario.

El Peabody Picture Vocabulary Test, el Illinois

Test of Psycholinguistic Abilities y las pruebas de articulación y diferenciación constituyen una muestra representativa de los tets de vocabulario con dibujos más conocidos.

No son, ciertamente, las únicas pruebas existentes para la medición del lenguaje, ni nuestro objetivo aquí es hacer relación de las mismas. Sólo destacaremos algunos aspectos relacionados con su estructura y construcción por su vinculación con el estudio del vocabulario.

Por otra parte, la relación entre vocabulario e inteligencia (García Hoz, 1977) es manifiesta. Para los psicólogos que miden la inteligencia, la amplitud del vocabulario se cuenta como uno de los indicadores más fidedignos (García Hoz 1943, 1953, 1977, Therman - Merrill, 1944). La mayoría de los tests más conocidos de inteligencia contienen pruebas cuyos items prsuponen en la respuesta el dominio de un determinado vocabulario (pruebas de comprensión y fluidez verbal), requieren la comprensión de complejas instrucciones verbales o implican respuestas estrictamente verbales tales como definir palabras, detectar sinónimos, homónimos o antónimos, completar frases... Ahora bien, en estas pruebas también intervienen la memoria, el razonamiento y otros factores cognitivos que no deben confundirse con el desarrollo del lenguaje.

- El Peabody Picture Vocabulary Tets (PPVT, Dunn, 1965). Está compuesto por varias listas de palabras de dificultad variable. La presentación se realiza mediante la lectura de dichas palabras al sujeto. Su tarea consiste en identificarlas en uno de los cuatro dibujos que ilustran cada pala-

bra de la lista. Se trata, pues, de una prueba de vocabulario pasivo, el sujeto se limita a señalar la figura que correspon de a cada palabra que escucha. Es una prueba breve en su realización, de corrección fácil y normalizada para edades comprendidas entre los 2 y 16 años.

Se utiliza para evaluar el lenguaje si presuponemos que el incremento del vocabulario que manifiestan las edades comprendidas por la prueba es un reflejo directo del desarrollo len lenguaje. Ciertamente, no hay datos suficientes que justifiquen tal hipótesis. Con frecuencia se le utiliza también como prueba para medir la inteligencia debido a la correlación alta (sobre 0.70) que mantiene con pruebas clásicas en la medida de la misma, tales como el Stanford - Binet.

- Test de Illinois de capacidad psicolingüística (ITPA, Paraskevopoulos y Kirk, 1969). La versión actual está compuesta por doce subpruebas consecuentes a una revisión de la versión experimental de McCarthy y Kirk de 1961. Las modificaciones introducidas tienden a facilitar su aplicación, además se amplió el número de subpruebas de las nueve iniciales a las doce actuales.

Las pruebas de la versión original se elaboraron de acuerdo con la teoría del lenguaje de Osgood (1957). Para Osgood la conducta lingüística viene definida por tres dimensiones o procesos: receptivo, organizativo y expresivo. Tres subpruebas hay relacionadas con cada dimensión.

En relación con la primera, (modalidad del proceso

receptivo - perceptivo), las subpruebas miden la capacidad del sujeto para comprender símbolos auditivos o visuales. Sólo requieren respuestas afirmativas o negativas a respuestas cortas y directas (descodificación auditiva), o bien buscar las respuestas en ilustraciones estímulo (descoficiación visual). Las subpruebas relacionadas con el proceso de organización manifiestan la capacidad del sujeto para organizar, relacionar y manejar de manera lógica símbolos visuales y auditivos. Las tareas a resolver consisten en completar frases de enunciados análogos, o bien asociar conceptos presentados visualmente: una figura estímulo se ilustra con cuatro figuras entre las que hay que elegir la que manifieste una relación más estrecha con aquélla. Las subpruebas relacionadas con el proceso o dimensión expresiva suponen la puesta en juego de la capacidad para emplear símbolos verbales y gestos en la transmisión de ideas.

El ITPA se completa con tres subpruebas más cuya finalidad reside en la comprobación de la capacidad del niño para realizar tareas automáticas, no simbólicas, tales como completar frases y palabras (prueba de sintaxis y gramática) y recordar series de dígitos o secuencias de figuras que no representan nada concreto (prueba de memoria auditiva y visual a corto plazo).

- Otros tests de vocabulario con dibujos, aunque de menor uso, son el de Ammons y Ammons (1948) y el de Williamns y McFarland (1937). Ambos, el igual que los descritos anteriormente, constan de diversos dibujos que representan objetos, acciones o relaciones que actúan como posibles respuestas a

seleccionar por el sujeto ante las palabras estímulo que le propone el experimentador.

- Pruebas de articulación y diferenciación. En las pruebas de articulación, como en los tests de vocabulario, se muestran ilustraciones de objetos bien conocidos para el sujeto, y a continuación, se le pregunta por su nombre. Los objetos se seleccionan de acuerdo con los sonidos que se pretenden analizar. El objeto de la prueba consiste en evaluar de correcta o incorrecta la articulación verbal del sujeto (Templin y Darley, 1964; Fisher y Logeman, 1971).

Las pruebas de diferenciación emplean material verbal auditivo acompañado o no de distorsiones (ruidos) que dificultan su percepción (Eisenberg, Berlin, Dill y Frank, 1968). En otros casos, se muestran al sujeto pares de objetos bien conocidos cuyos nombres son semejantes pero difieren en algún sonido (pato-gato). El experimentador nombra uno de ellos y el sujeto lo identifica en su lámina correspondiente. Debido a la dificultad de encontrar un número suficiente de pares de palabras que cumplan estos requisitos, Eilers y Oller (1975) idearon un procedimiento más general constituyendo los pares con una palabra conocida y la otra similar a la anterior pero incluyendo algún fonema que hacía su pronunciación incorrecta para tratar de ver si el niño diferenciaba entre la articula- correcta o incorrecta de las palabras.

También son típicos los estudios de articulación y diferenciación que utilizan la lectura y el deletreo de palabras para constatar los errores de articulación (Bond y Dyks-

tra, 1967; Downing, 1967).

1.2.- El estudio de la palabra como respuesta: el análisis de la producción verbal.

El vocabulario, en cuanto producción verbal de un sujeto o grupo de sujetos, ha sido objeto de múltiples investigaciones desde principios de este siglo e incluso son conocidas algunas que se remontan al siglo pasado. Su finalidad estriba en el establecimiento de recuentos o cómputos de vocabulario. La mayor parte de estos trabajos han sido realizados por investigadores americanos, lo cual indica que la lengua base analizada es el inglés. Ello no supone la ausencia de tales recuentos en otras lenguas, pero su existencia es exigua y esporádica. Harkin (1957) recopila distintos recuentos realizados en 16 lenguas diferentes. Más adelante dedicaremos un apartado a los estudios más relevantes concernientes a nuestra lengua española.

Para una aproximación más precisa distinguiremos dos tipos de recuentos: a) Los Recuentos Generales confeccionados a partir de una producción verbal proporcionada por adultos, y b) Los Recuentos de Vocabulario Infantil. A los primeros haremos referencia en el próximo capítulo. Entre ellos diferenciaremos los recuentos realizados con otras lenguas de los que tienen por objeto el análisis de la lengua española.

CAPITULO II

II.- RECUENTOS GENERALES.-

2.1.- Orígenes de los recuentos de vocabulario.-

Los orígenes de los recuentos de vocabulario en las lenguas modernas se remontan a finales del siglo pasado y principios de este siglo. Responden a razones de muy diversa índole: lingüísticas, pedagógicas, sociológicas, e incluso políticas (Gougenheim, G.; Michéa, R., 1964).

Los países con grandes dominios territoriales fuera de sus fronteras, tras un largo periodo de dominación, se preocuparon de fomentar y facilitar la adquisición, el aprendizaje y la permanencia de sus lenguas respectivas con las propias de cada territorio, hasta el punto de que una vez independizados, muchos de ellos conservaron como idioma nacional el del país colonizador.

De cualquier forma, en todos los países con cierta preponderancia cultural, social, política y económica, se arbitraron procedimientos adecuados a las necesidades del mundo moderno para la difusión de sus lenguas respectivas. La mayoría de estos procedimientos radican sobre la idea de reducción y simplificación de los contenidos lingüísticos en general, y de la gramática y el vocabulario en particular. Por consiguiente, los estudios sobre vocabularios básicos o fundamentales aunque en su origen sean directamente dependientes de una necesidad de simplificación y adecuación de las distintas lenguas, indirectamente responden a la necesidad de man-

tener un predominio no sólo territorial sino también político, económico, comercial y cultural.

La mayoría de las lenguas modernas poseen un léxico extenso que se ve constantemente enriquecido con las nuevas aportaciones que hacen la tec-nología y la ciencia. No obstante, el vocabulario de uso cuotidiano dista mucho en su extensión del léxico que recogen los diccionarios. Gran parte de los vocablos que aquí se registran son letra muerta, apenas si los utilizamos, o cuando menos, ésto corresponde a una minoría de personas con un alto espíritu de refinamiento léxico.

La rápida difusión de las lenguas modernas trajo consigo, pues, una reducción de las mismas a sus elementos gramaticales y léxicos esenciales. Esto a su vez, supuso un cambio en el enfoque pedagógico tradicional de la enseñanza de las lenguas modernas. Este cambio pronto fué asumido por la administración de distintos países en los que se crearon comisiones especiales de estudio y seguimiento con el fin de adecuar a las nuevas necesidades la enseñanza de la lengua. Tal es el caso de Inglaterra, Estados Unidos o Francia.

Una de las primeras síntesis gramático-léxicas que se realizó es el "Basic English" (que responde -y no de forma fortuita-, a las iniciales de British American Scientific International Commercial). Es una de las simplificaciones de la lengua inglesa que más notablemente ha contribuído a la difusión del inglés, aunque la intención originaria de sus autores no fuera presentar un inglés simplificado y comprimido ni

dar una respuesta al problema de la expansión del idioma.

El BASIC responde más bien al objetivo de instrumentalizar una lengua universal con base en el inglés y estructurada sobre unos presupuestos lógicos y objetivos. La iniciativa, como señala Michéa (1964), responde a sugerencias filosóficas de siglos anteriores. Ya Bacon había apuntado el carácter universalista que poseían los sistemas de escritura - ideográfica empleados por los pueblos de Extremo Oriente, los cuales permitían el entendimiento a través de la escritura a pueblos que hablaban lenguas diferentes.

Más tarde, Leibniz en su obra la "Characteristica Universalis" vuelve sobre el tema al sugerir la posibilidad de sintetizar el pensamiento en expresiones simbólicas y hacer de la lengua hablada un instrumento universal, racional, lógico, objetivo y estable. Esta lengua por sí misma se compondría únicamente de elementos (las palabras) cuyo valor fuera claramente objetivo y permaneciera constante, de forma que reflejasen siempre la misma cosa para cualquiera que pudiera utilizarlos. Leibniz pretendía, de esta forma, aplicar el modelo del cálculo matemático a los conceptos y a sus relaciones; hacer de la lengua y las palabras una expresión algebraica.

La elaboración del BASIC, siguiendo estos supuestos, pretende la definición de todas las palabras con el uso de un restringido número de términos. Estos constituirían el vocabulario básico. Cuatro años dedicaron Ogden y Richards a la tarea de su determinación. En 1928 la lista definitiva de pala-

bras quedó fijada en 850, aunque posee algunas listas de palabras adicionales que elevan el total a 1993. En 1933 se iniciaron en Japón las primeras experiencias educativas de enseñanza del inglés a través de este método. Posteriormente, se extendieron a otros muchos paises del área anglófona.

El BASIC se presenta, pues, como una lengua universal susceptible de ser aprendida y hablada por cualquier sujeto sea cual sea su cultura y su lengua materna.

Nuevos vocabularios se confeccionaron a partir de los principios lógicos seguidos por el Basic (M. West, 1941). Todos ellos han sido de gran utilidad para la enseñanza del inglés, particularmente en los paises no angloparlantes. Sin embargo, a pesar del carácter eminentemente práctico y del rigor con que fueron confeccionados por sus respectivos autores no carecen de cierto subjetivismo y de gran número de limitaciones. (Para un examen crítico del Basic Englihs: M.H. Bongers, "The History and Principles of Vocabulary Control", Woerden, 1947).

Los métodos estadísticos aplicados al vocabulario presentan una nueva alternativa al método lógico. Una alternativa más objetiva cuya finalidad consiste en determinar el uso de las palabras a partir del análisis de frecuencias.

2.2.- Recuentos Generales.-

Con este epígrafe hacemos referencia a todos aque-

llos recuentos que utilizan los métodos estadísticos. Estos trabajos se enmarcan dentro de la denominada "filología dinàmica", una técnica de análisis del lenguaje fronteriza entre la psicología y la lingüística que considera "la producción del lenguaje como un fenómeno psicobiológico que debe investigarse con el espíritu de las ciencias exactas"(Zipf, 1953, pág. 3).

Son cómputos de frecuencia de palabras, aparecidas en diversas muestras del lenguaje oral o escrito. Su objetivo es cuantificar la ocurrencia de las palabras en una producción verbal determinada. Presentan distintas ordenaciones de las palabras por su frecuencia.

En los recuentos generales la recogida de datos se realiza sobre una gran diversidad de fuentes. En ellas se observa un claro predominio de las que utilizan material verbal escrito, impreso, frente a las que utilizan material tomado directamente de la expresión oral de unos sujetos dados constituyentes de la muestra. La literatura, los libros de texto, el lenguaje de los distintos medios de comunicación, las cartas, las conversaciones telefónicas, la producción verbal espontánea ante diversos temas estímulo, la prensa científico-técnica..., han sido algunas de las fuentes más frecuentemente utilizadas.

El número total de palabras registradas, los procedimientos empleados en la recogida de datos, el tamaño de las muestras en el caso de utilizar sujetos, los métodos de tratamiento y análisis estadístico, así como los planteamien

tos, hipótesis y objetivos son algunos de los aspectos que establecen diferencias sustanciales entre unas investigaciones y otras, hasta tal punto que muchas de ellas, como señala Francescato (1971), carecen de la suficiente justificación teórica o poseen un carácter ocasional y asistemático - que pone en duda el valor científico de sus conclusiones.

En la revisión de recuentos generales que presentamos a continuación prestamos una atención particular a aquellos que tienen una mayor relevancia o han sido el punto de partida de investigaciones posteriores sobre el vocabulario. En este capítulo el ámbito de la revisión comprende fundamentalmente los trabajos publicados en alemán, inglés y francés.

2.2.1.- F.W. Kaeding (1898).

En alemán, es conocido el recuento de F.W. Kaeding, uno de los pioneros en este tipo de trabajos. Su título: "Häifigkeitswörterbuch der deutschen Sprache", se realizó en 1898. Las muestras tomadas para este recuento estuvieron constituídas por debates parlamentarios, textos administrativos y comerciales, periódicos y revistas, escritos de teología y medicina, cartas personales y documentos militares. Su finalidad fue racionalizar las técnicas taquigráficas del alemán, de aquí que no posea solamente listas de palabras como los recuentos tradicionales, sino también de sufijos y letras con su frecuencia correspondiente. El diccionario de Kaeding no - es particularmente apropiado para la enseñanza del alemán. En la lista de los vocablos más frecuentes sobresalen los térmi-

nos relacionados con el comercio, la banca, la milicia, la administración y la política, lo cual evidencia un sesgo en la usualidad.

El recuento contiene una lista de 79.716 palabras (type) tomadas de 299 fuentes pertenecientes a once categorías o grandes temas, las cuales proporcionaron 10.910.777 palabras (token).

2.2.2.- E.L. Torndike y I. Lorge (1944).

Entre los recuentos generales, uno de los trabajos fundamentales fue el realizado por Thorndike E.L. y Lorge I. en 1944: "The Teacher's word book of 30.000 words". Se considera como la principal fuente de frecuencias de palabras para el inglés escrito. El recuento se realizó en fases sucesivas y complementarias. Se inició en 1921 con el "Teacher's word list of 10.000 words". Posteriormente se realizaron varias revisiones y ampliaciones. En la de 1931 se duplicó el número de palabras de la lista, 20.000, y en la definitiva de 1944 éste quedó establecido en 30.000.

El primer recuento (1921) se elaboró en base a diversas fuentes de material escrito con una distribución aproximadamente igual de los textos propios de niños frente a los de los adultos. En las ampliaciones y revisiones posteriores solamente se utilizó material escrito para adultos.

El procedimiento seguido por Thorndike consistió -

en ir registrando palabras sucesivamente hasta la obtención de las 30.000 diferentes que presenta el recuento final. A medida que avanzaba en el recuento el número de palabras que iban apareciendo era progresivamente menor, de tal suerte que para las 20.000 del recuento de 1931 precisó contar más de nueve millones y medio (9.565.000) de palabras tomadas de 279 fuentes diferentes; y esta cifra se vió prácticamente duplicada para la obtención de las 10.000 últimas que completan el corpus de 1944.

Thorndike no justifica la elección de las cantidades mencionadas. No obstante, dada su finalidad pedagógica: -ofrecer un conjunto de palabras cuyo aprendizaje por parte del niño le permita estar en condiciones de poderse comunicar con los demás-, cabe pensar que las palabras que primero ha de aprender han de ser aquellas que aparecen con mayor frecuencia. Y, difícilmente una palabra que no haya aparecido ninguna vez entre los dieciocho millones de palabras muestreadas se la puede considerar como palabra usual en la lengua inglesa y, por consiguiente nula la necesidad de que el niño la aprenda.

Desde la aparición del "The Teacher's word book of 10.000 words" de 1921, el desarrollo de las listas de vocabularios ha ocupado un volumen considerable en la investigación psicolingüística. En una revisión bibliográfica sobre la materia realizada por Dale y Razik (1963) aparecen referencias de más de 3.000 trabajos.

2.2.3.- Kucera y Francis (1967).

Para el inglés - americano actual, Kucera y Francis proporcionan un recuento de palabras, aún más amplio que el de Thorndike sin la necesidad de muestrear un número de palabras tan elevado. Ofrecen un ranking ordenado por frecuencias con un total de 50.406 vocablos extraídos de una muestra de 1.014.232 palabras del inglés escrito. Nos referimos al "Computational Analysis of Present-Day American English".

Para compilar estas palabras recurrieron a 500 muestras de otras tantas fuentes diferentes de unas 2000 palabras aproximadamente cada una. La homogeneidad de la muestra reside en el hecho de que todo el material recogido pertenece a un mismo periodo de tiempo: textos escritos durante 1961. La representatividad del vocabulario viene dada por las 500 muestras elegidas. Cada una de ellas abarcaba 15 categorías temáticas representativas de una amplia gama de contenidos y diversos estilos de prosa que van desde las páginas deportivas de los periódicos a la prensa científica, pasando por la novela romántica, la "prensa del corazón", la ciencia ficción y el ensayo filosófico.

Los listados finales de palabras se presentan, tanto en orden decreciente de frecuencias, como en orden alfabético. En la introducción se recoge abundante información en torno a la metodología seguida por este tipo de trabajos.

Un corpus semejante de palabras al de Kucera y Francis es el que presenta el "American Heritage Word Frequency

Boo-k" de Carroll, Davies y Richman (1971). Las diferencias con aquel estriban en el material utilizado. En este caso, los textos escritos elegidos estaban destinados básicamente a alumnos comprendidos entre el tercero y noveno grados de enseñanza. La lista se extrae de 1.045 muestras de 5.000 palabras cada una. El total de palabras muestreadas asciende, por tanto, a algo más de cinco millones de palabras. Contiene 86.741 vocablos (type) de los cuales 35.000 no aparecen más que una sola vez.

2.2.4.- Vander Beke (1929).

Para el francés existen también varios diccionarios de frecuencias. Los dos primeros fueron publicados en - Estados Unidos. Nos referimos al de Henmonn, "A French Word Book Based on a Count of 400.000 Running Words" de 1924, y al de Vander Beke, "French Word Book" de 1929.

- Henmonn intentó establecer científicamente cuáles eran las palabras más usuales del francés y determinar su grado de frecuencia. Se ayudó para ello de una lista de 5000 palabras tomadas de un diccionario. Posteriormente, detectó la frecuencia de aparición de éstas en textos de la literatura francesa. El total de la muestra de palabras elegido fué de 400.000.

- El diccionario de Vander Beke responde a una iniciativa del Committee on Investigation of the Modern Language Study, cuya finalidad era renovar los programas de las len-

guas modernas sobre presupuestos objetivos. Hemos de señalar que bajo esta misma iniciativa en el mismo periodo de tiempo se realizaron otros diccionarios, tales como el "A German Frequency Word Book" de Morgan (1928) para el alemán, y el "A Graded Spannish Word Book" de Buchanan (1927) para el español.

Vander Beke perfecciona el diccionario de Henmonn en una doble vertiente: por un lado, aumenta la muestra de palabras hasta 1.147.748, por otro, diversifica las fuentes del recuento, lo cual contribuyó a la obtención de un vocabulario más representativo. En cuanto a la distribución de las palabras, la novedad más importante que aporta Vander Beke al diccionario de Henmonn es la consideración del rango de las palabras junto al dato numérico de las frecuencias. Para Vander Beke una palabra que aparece una sola vez en cinco fuentes distintas (rango 5, frecuencia 5) es más importante en cuanto al uso se refiere, que una palabra que aparezca 10 en un solo autor (rango 1, frecuencia 10).

El diccionario proporciona una lista de 6.067 palabras tomadas de 88 fuentes pertenecientes a nueve géneros o categorías temáticas diferentes.

La crítica al diccionario de Vander Beke se centra en la caracterización de su vocabulario como elitista, literario, perteneciente a épocas pasadas y distante, por tanto,del vocabulario de la vida cuotidiana real y actual. No es apropiado para la enseñanza del vocabulario francés actual. No obstante, hasta la publicación del "Francés Fundamental" de Gougenheim (1959), fue objeto de numerosos trabajos y adapta-

ciones cuya finalidad era eminentemente didáctica: la enseñanza del francés para extranjeros (Maygood, M.J.B., 1936; Tharp, J.B., 1939; Schlyter, B., 1951; Verlée, L., 1954).

2.2.5.- Gougenheim (1959), (1964).

Más recientemente, y con un recuento total de palabras sensiblemente inferior, Gougenheim recopiló el "Francés Fundamental", publicado en 1959. El diccionario recoge 7.995 vocablos de una muestra de 312.135 palabras provenientes del registro de conversaciones mantenidas con 275 personas, todos adultos, pertenecientes a diversas regiones, niveles culturales, educacionales y socioeconómicos de Francia. Las conversaciones registradas presentaban una gran variedad en su temática: la familia y los amigos, los medios de comunicación, la sanidad, los transportes, profesiones, las artes, el deporte y los juegos, los viajes...

En un trabajo posterior (Gougenheim y Michéa, 1964) analizaron las relaciones entre la frecuencia de uso de las palabras -esta es la información que proporcionan los recuentos a los que venimos haciendo referencia-, y un nuevo criterio: la disponibilidad de las palabras. Si una palabra es usual por su alta frecuencia, por lo que se cree además que es bien conocida para el sujeto que la utiliza, debe disponer de ella con facilidad.

La mayoría de las investigaciones han demostrado, por el contrario, que las listas de frecuencia no contienen

en absoluto, o las recogen en muy escasa proporción, determinadas palabras concretas que corresponden a objetos, acciones o relaciones de uso cotidiano. Asi, por ejemplo, palabras como "silla", "puerta" o "cuchillo" que empleamos una o varias veces al día no dudaríamos en admitirlas como palabras frecuentes y, sin embargo, no aparecen como tales ya que raras veces las mencionamos. Son palabras disponibles en el uso cotidiano del hablante, pero aparecen con una frecuencia baja. Ello indica que el criterio de frecuencia como determinante del uso de las palabras ha de ser compaginado con otros criterios, tales como el de disponibilidad. Este es el fundamento del trabajo de Gougenheim, Michéa y colaboradores: confeccionar un nuevo vocabulario básico para el francés que compartiera los criterios de frecuencia y disponibilidad.

La investigación se llevó a cabo y consistió fundamentalmente en comparar dos listas de palabras. La primera, constituída por las palabras de mayor frecuencia del listado de frecuencias del Francés Fundamental. Un total de 1.063 palabras (todas las que aparecían con una frecuencia superior a 20). La segunda lista contrastada fué obtenida a partir de esta nueva investigación.

Para ello, interrogaron a 914 escolares de 9 a 12 años de distintos departamentos franceses (Dordogne, Marne, Eure y Vendée). Las instrucciones que se les dieron fueron las siguientes: "piensa en... (se le daba el título de uno de los temas elegidos como centro de interés) y escribe los 20 nombres más comunes que te sugiere y que crees son los más útiles que se han de saber en relación con dicho tema".

Los centros de interés que constituían el estímulo para la actividad lingüística eran, por consiguiente, sugeridos y no espontáneos. Estos fueron: las partes del cuerpo, los vestidos, la casa, el mobiliario, los alimentos, bebidas, animales, juegos, profesiones... Los sujetos disponían de 20 minutos para ejecutar la tarea.

Para cada uno de los 16 centros de interés sugeridos se ordenaron las palabras-respuesta por orden de frecuencia de mayor a menor. Esto permitió la comparación de ambas listas. De dicha comparación (la primera proporcionaba la frecuencia de uso, la segunda la disponibilidad), se obtuvo una nueva clasificación que dió lugar a una lista de 1465 palabras cuya distribución quedó de la siguiente manera: 692 nombres, 339 verbos, 98 adjetivos y 83 adverbios.

El "Francés Fundamental" de Gougenheim ha sido utilizado en numerosas investigaciones posteriores, las cuales han tenido casi todas como finalidad trascender la información estadística que proporciona el conocimiento de la frecuencia de uso de las palabras y estudiar su relación con otros fenómenos de naturaleza psicológica.

Es bien conocido que entre el sistema socialmente codificado de una lengua y el mundo real de los objetos, existe una relación funcional. Además, esta relación es particular para cada sujeto, lo cual indica a su vez una vinculación, una familiaridad entre el individuo concreto y las palabras que utiliza, (Fraisse, P., Noizet, G., Flament, C.; 1963; Noble, 1953). Se entiende, pues, por familiaridad, la estima-

ción de un sujeto particular hace de su propio lenguaje, suponiendo, claro está, que la disociación entre significante y significado, entre palabra y realidad designada, nunca se realiza por completo. Lo que equivale a afirmar que "es difícil que un sujeto pueda estimar la familiaridad de una palabra - por abstracción del objeto que designa, o de la realidad o situación a la que tal palabra remite directa o indirectamente" (Fraisse, P., y colab. pág. 189). La familiaridad es, pues un intermediario entre la frecuencia de uso de las palabras tal como aparecen en el lenguaje explícito y la frecuencia de las situaciones a las que hacen referencia. De lo que se deduce que la correlación entre familiaridad y umbrales perceptivos de las palabras debería de ser más fuerte que la correlación ya notable y a menudo subrayada entre los umbrales perceptivos y la frecuencia de uso.

Otros aspectos del Francés Fundamental han sido - puestos de manifiesto por Galison (1971) y Mackey (1971). Galison hace un análisis del francés fundamental de Gougenheim y clasifica el vocabulario en campos nocionales que permiten una mejor utilización del mismo con fines didácticos y pedagógicos. Mackey establece una comparación entre el "vocabulario disponible" de Gougenheim y Michéa y el vocabulario obtenido con un procedimiento similar tomando como muestra sujetos de la América francófona. Pone de manifiesto los aspectos comunes y las diferencias entre ambos vocabularios.

2.2.6.- Otros recuentos.

Entre los recuentos generales más recientes que tie

nen por objeto el establecimiento de frecuencias de uso de - palabras en otras lenguas hacemos mención del del Dabbs (1966) para el bengalí. Utiliza como material la prensa escrita. De igual manera, el holandés de los periódicos sirve de base para el recuento de Van Verckel (1965). Vakar (1966) proporciona un recuento del ruso hablado, y Josselson (1953) del ruso escrito. Landau (1959) lo obtiene de la prosa árabe moderna. Juilland y colbs. (1965) presentan el diccionario de frecuencias del rumano.

Todos los recuentos generales fundamentan sus estimaciones y conclusiones en la cuantificación de la aparición - de las palabras en diversos materiales del lenguaje oral o escrito. Los índices de frecuencia vienen dados por el total de ocurrencias de una palabra. En todos subyace el supuesto de - que las palabras más frecuentes son las más útiles, y éstas, por consiguiente, son las constituyentes del vocabulario usual de una lengua.

CAPITULO III

III.- RECUENTOS GENERALES DEL VOCABULARIO ESPAÑOL.-

La información que poseemos concerniente a la frecuencia absoluta y relativa de unidades lingúísticas de diferente orden (fonemas, morfemas, sílabas, vocablos, frases, categorías gramaticales...), en nuestro idioma, es suficientemente representativa si la comparamos con la existente en otras lenguas. La mayor parte de esta investigación, no obstante, ha sido realizada fuera de nuestras fronteras, particularmente la relacionada con el vocabulario y la determinación de listas y diccionarios de frecuencia de palabras. Hemos de exceptuar en este sentido, los trabajos de Victor García Hoz, pionero de este tipo de investigaciones en nuestro país.

Destacaremos en el presente capítulo los recuentos generales de vocabulario más significativos. Procederemos en nuestra revisión de forma cronológica. Los cuatro trabajos básicos a los que haremos referencia son los siguientes: Buchanan (1927), Rodríguez Bou (1952), García Hoz (1953) y Juilland y Chang-Rodríguez (1964).

Más adelante, revisaremos otros estudios cuyo objetivo es el análisis del vocabulario infantil. También tienen como base nuestra lengua y utilizan presupuestos teóricos y metodológicos semejantes a los que presentan los recuentos generales: Díaz Castañón (1975), Armayor y colbs. (1975, 1977, 1979).

3.1.- "A GRADED SPANISH WORD BOOK", Buchanan (1927).

Es uno de los recuentos más antiguos del español escrito. Su finalidad es proporcionar una lista de las palabras más usuales del español. Este propósito se refleja no sólo en la diversidad de fuentes utilizadas sino también en su pertenencia a diferentes periodos lingüísticos.

El universo de palabras que constituyen la base del recuento asciende a 1.200.000. Estas palabras se distribuyen por igual en cuarenta unidades de 30.000 palabras cada una. Dichas unidades pertenecen a siete grandes áreas temáticas que constituyen las fuentes del vocabulario. Estas son las siguientes:

- Literatura dramática: 7 unidades.
- Novela: 8 unidades.
- Textos en verso: 3 unidades.
- Folklore: 3 unidades.
- Prosa didáctica: 8 unidades.
- Vocabulario Técnico: 7 unidades.
- Periódicos: 4 unidades.

Las obras que integran cada unidad son ampliamente heterogéneas tanto espacial como temporalmente. Los textos, particularmente las novelas y las obras teatrales, pertenecen a autores representativos españoles e hispanoamericanos de distintas épocas comprendidas entre los siglos diecisiete y veinte.

El listado final de frecuencias registra como entradas a las palabras en cuanto vocablos o unidades léxicas tal como aparecen en los diccionarios clásicos. No hay referencias a las variantes de los vocablos debidas a los llamados "accidentes gramaticales".

3.2.- "RECUENTO DE VOCABULARIO ESPAÑOL". Rodríguez Bou (1952).

Es el recuento más extenso de los realizados en lengua española: 7.066.637 palabras. Su propósito es aún más ambicioso que el de Buchanan, a pesar de que los datos proporcionen una información más heterogénea dada la desigualdad entre los tamaños de las muestras recogidas.

Los materiales con los que trabaja contienen palabras provenientes del lenguaje oral y escrito. Las muestras recogidas pertenecen a tres grandes áreas divididas a su vez en un mayor número de sub-áreas:

1). Vocabulario de expresión:

a) Vocabulario Oral:	1.073.245 palabras.
b) Vocabulario de asociación libre:	1.730.026 palabras.
c) Vocabulario de redacciones escolares:	586.141 palabras.
TOTAL	3.389.412

2). Vocabulario de Reconocimiento:

a) Periódicos:	1.050.000	palabras.
b) Vocabulario de programas de radio:	465.000	palabras.
c) Vocabulario de literatura religiosa:	295.422	palabras.
d) Recuento de Buchanan:	1.200.000	palabras.
TOTAL	3.011.022	

3). Vocabulario de Autores:

a) Vocabulario de Libros de Texto:	255.569	palabras.
b) Vocabulario de Libros de Lectura:	410.634	palabras.
TOTAL	666.203	

La desproporción que se observa en los tamaños de las muestras no es necesariamente arbitraria o accidental como a primera vista parecen indicar los datos. La cantidad variable de palabras que muestra cada subárea ha sido determinada en función de una ponderación previa. Dicha ponderación pretende reflejar la importancia variable que el autor otorga a cada subárea en la realidad socio-cultural, y por consiguiente, en la conducta lingüística. Para Rodríguez Bou, el vocabulario de niños aparecido en redacciones o producido en situaciones o tareas de libre asociación es más apropiado y refleja de forma más clara la realidad lingüística que un gé-

nero literario concreto. De aquí, que su peso específico sea mayor y, por tanto, mayor también el número de palabras que aporta al total de las muestreadas.

Los criterios de ponderación adoptados por Rodríguez Bou no parecen excesivamente convincentes, como muy bien señala Juilland (1964), por la carga de subjetivismo que conllevan. Además, comprometen el valor comparativo de las palabras, la medida de su dispersión, uso y frecuencia de aparición.

3.3. "VOCABULARIO USUAL, VOCABULARIO COMUN Y VOCABULARIO FUNDAMENTAL". García Hoz (1953).-

"La finalidad de esta investigación no es puramente lingüística ni filológica... se concibe el vocabulario como un medio de conocimiento psicológico y como contenido de enseñanza, la primera tarea se reduce a determinar cuál es, de hecho, el vocabulario empleado por el hombre vulgar. Este vocabulario es el exigible en la educación general." (García Hoz, 1953. pp. 15).

El procedimiento utilizado por García Hoz para determinar las palabras que constituyen el vocabulario usual no fue el registro directo de la producción verbal del "hombre corriente" sino el análisis de fuentes escritas.

El criterio seguido para la selección del material escrito fue de carácter sociológico. Según García Hoz, "el vocabulario usual viene dado por las palabras utilizadas por los

adultos y puede, por tanto, ser considerado como la expresión lingüística normal de la sociedad". Las fuentes del vocabulario fueron elegidas teniendo en cuenta cuáles eran las manifestaciones fundamentales de las relaciones humanas, de las que el lenguaje es su máxima expresión.

Los cuatro aspectos que reflejan en mayor medida estas manifestaciones son, según García Hoz: la vida familiar, la vida social indiferenciada, la vida social regulada y la vida cultural. El origen del vocabulario usual radica en las cuatro fuentes mencionadas.

a) Vida Familiar:

100.000 palabras tomadas de 620 cartas privadas, pertenecientes a distintas regiones y provincias españolas. Este material intenta recoger y reflejar el léxico espontáneo e íntimo reservado a situaciones de vida privada.

b) Vida Social Indiferenciada:

100.000 palabras distribuídas en 25 bloques de 4000 palabras cada uno. El material escrito utilizado en cada bloque estuvo constituído por periódicos de diferente tirada y pertenecientes a diversas ciudades de la geografía española, editados entre 1943 y 1948. Cada bloque estuvo representado por un periódico.

Este área intenta reflejar el lenguaje del hombre de la calle. La prensa diaria enmarca los hechos y preocupaciones de la vida social, del vivir de cada día, los acontecimientos próximos al hombre y a la humanidad en un momento dado.

c) Vida Social Regulada:

100.000 palabras. Esta fuente de datos tiene su expresión en los documentos que regulan y representan la vida social organizada. Son un reflejo del lenguaje oficial, político, eclesiástico y sindical de un pais. Los medios de expresión de este lenguaje son los boletines oficiales del Estado, los boletines informativos de los distintos arzobispados y obispados y las publicaciones sindicales. De todos ellos se recogieron diversas muestras representativas.

d) Vida Cultural:

100.000 palabras. Los libros, en cuanto expresión del patrimonio cultural de un país, son el exponente más genuino de la vida cultural del mismo. Esta última fuente de material impreso estuvo constituída por 22 libros. Todos ellos se consideraban, tras una encuesta a diferentes editoriales, como los libros más vendidos desde 1943 hasta el momento de inicio de la investigación.

Del sucinto análisis de las fuentes presentado anteriormente, se deduce fácilmente que el total de palabras de la muestra del recuento de García Hoz asciende a 400.000, distribuídas por igual en las cuatro fuentes. Esta cifra, no obstante, fue fijada "a posteriori" una vez que se realizó el recuento. Los argumentos aducidos para su justificación no están exentos de cierto subjetivismo, que el propio autor, García Hoz, reconoce.

Los dos criterios básicos utilizados para fijar la muestra (número total de palabras), fueron, por un lado, las 12.911 invariantes (vocablos o palabras diferentes -"type"-) que aparecieron tras el recuento de las 400.000 palabras ("token") y, por otro, la frecuencia media de aparición observada en las palabras tras el recuento.

El volumen de invariantes o vocablos se aproxima a la cifra de 15.000 que Watts fija como promedio de palabras constituyentes del vocabulario de un adulto de habla normal de cualquier ciudad inglesa. García Hoz asume este criterio de Watts, a pesar de que la objetividad y los procedimientos que llevaron a Watts al establecimiento de esta cifra son poco precisos; y amén de otras salvedades no desdeñables como las diferencias de lengua que puedan existir entre el inglés y el español y las diferencias socioculturales y educativas entre los adultos de uno y otro pais.

Aún así, las cifras de ambos sólo son comparables de una forma meramente referencial ya que si tenemos en cuenta la diferencia real entre una cifra y la otra -más de dos

mil palabras-, ésta sería suficiente por sí misma para invalidar los cálculos establecidos. Para que las cifras fuesen realmente aproximativas entre sí, hubiese sido preciso continuar el recuento hasta la obtención de un total de palabras sensiblemente superior al empleado. De lo contrario, hemos de dar por supuesto que el hombre de habla normal, por razones de profesión, lugar, conocimientos específicos..., dispone de un vocabulario adicional específico de más de dos mil palabras sobre el vocabulario común usual. Este presupuesto parece ser decisivo, al ser asumido por García Hoz, para la detención del recuento en 400.000 palabras.

El segundo argumento lo creemos más convincente, aunque no se halla exento tampoco de subjetividad. A medida que se iba realizando el recuento se comprobaba que el número de palabras nuevas que aparecían era progresivamente menor, lo cual es obvio. Tras un análisis más explícito de esta constatación, García Hoz adoptó como criterio la cifra tope de 100.000 palabras para cada fuente dado que con un volumen de palabras semejante observó que éstas se repetían más de cien veces por término medio. Llegado este momento en que las palabras adquieren una frecuencia media de 100 difícilmente aquellas que no hayan aparecido ninguna vez se las puede considerar como palabras usuales.

A partir de estos cálculos, la lista de palabras registrada es la considerada por García Hoz como Vocabulario Usual. Esta se compone de 12.402 vocablos que sumados a los 253 considerados de significación restringida y a los 256 que no figuran en el diccionario de la Real Academia, dan el total

antes mencionado de 12.911.

Un nuevo concepto de vocabulario surge a partir del vocabulario usual, es el denominado Vocabulario Común. Está formado por 2.183 vocablos. La selección de palabras que lo integran obedece a un doble criterio: por un lado, todas las palabras (1971) que aparecieron simultáneamente en las cuatro fuentes utilizadas, independientemente de la frecuencia absoluta que estas presentaban. En segundo lugar, a las anteriores se adjuntó una relación adicional de 212 términos correspondientes a otros tantos vocablos que poseían en el ranking absoluto de frecuencias una frecuencia igual o superior a 40 con independencia del número de fuentes en que aparecieron.

Existe un tercer concepto de vocabulario más restringido: el Vocabulario Fundamental, compuesto por 208 vocablos solamente, extraídos del Vocabulario Común. Ciertamente, dentro del Vocabulario Común las palabras no tienen todas el mismo valor. Este viene dado por la distribución de frecuencias de una parte, y de otra, por la dispersión de frecuencias que cada palabra muestra en las cuatro fuentes. De tal suerte que, aunque todas las fuentes aporten el mismo número de palabras al vocabulario común, la participación de cada una de ellas en el mismo es desigual.

El análisis de correlación entre el vocabulario de unas fuentes y otras, sin tener en cuenta el grado de frecuencia de las palabras, fue alto (entre 0.7 y 0.9). El resultado contradice la afirmación que acabamos de hacer respecto a la heterogeneidad de los vocabularios de las distintas fuentes.

Sin embargo, un análisis más detallado llevó a la conclusión de que la alta correlación se debía sólo a la presencia de un pequeño número de palabras en las cuatro fuentes que presentaban una distribución homogénea. No es necesario precisar que estas palabras eran las de más alta frecuencia y, por tanto, las que absorbían gran parte de las ocurrencias totales del vocabulario.

Por consiguiente, se puede afirmar que sólo un pequeño núcleo de palabras comunes lo son realmente. El resto, aún estando presentes en las cuatro fuentes eran más propias de una de ellas, debido a la desigual distribución de frecuencias que presentaban. De aquí, que el vocabulario fundamental que presenta García Hoz sea tan reducido.

Esta conclusión, por otra parte, cuestiona la usualidad del propio vocabulario que se determina. Si existe una escasa correlación entre la usualidad de las palabras en unas y otras fuentes del vocabulario (análisis de correlaciones y análisis factorial), excepción hecha del pequeño grupo que forma el vocabulario fundamental, se demuestra que cada tipo de vida social tiene su propio vocabulario. Así, pues, las cuatro fuentes tomadas como origen y fundamento de un posible vocabulario usual tal vez no sean las más adecuadas para su determinación, dada la escasa coincidencia entre el vocabulario de unas y otras. El vocabulario usual,por tanto, que utiliza la gran masa social de hablantes, si existe, no es precisamente el que se recoge en el "Vocabulario Usual, Vocabulario Común y Vocabulario Fundamental".

Más recientemente, en 1977, el mismo García Hoz emprendió de nuevo la tarea de analizar el vocabulario. Esta nueva investigación tiene como objetivo la determinación de un Vocabulario General de Orientación Científica que sirviera de base para la programación de la enseñanza secundaria en España.

Las fuentes del recuento fueron los libros de texto de todas las materias científicas incluídas en los planes de estudio del Bachillerato, con excepción de los idiomas. La selección de textos de las diferentes editoriales se hizo mediante una encuesta entre los profesores que impartían dichas materias.

La muestra estuvo constituída por 2.302.360 palabras o frecuencias dentro de las cuales se registraron 25.402 palabras diferentes. Un corpus excesivamente amplio dada la finalidad didáctica del recuento.

Con objeto de facilitar su uso y manejo, la presentación de resultados no se limita a ofrecer una lista general de vocablos ordenados alfabéticamente o en orden decreciente de frecuencias sino que en función de diversos criterios se distinguen diferentes tipos de palabras dentro del Vocabulario General de Orientación Científica, que dan lugar a diversos estratos de vocabulario.

Los estratos del Vocabulario General de Orientación Científica:

- Vocabulario Común de Materias, constituído por las palabras que aparecen en todas y cada una de las materias exploradas.

- Vocabularios Específicos, de cada materia: 13 en total.

- Vocabulario Común de Areas, constituído por palabras que aparecen en algunas materias consideradas afines en los planes de estudio. Se distinguen cuatro áreas: Lingüística - Matemáticas - Sociales y Ciencias Naturales.

- Vocabularios Compartidos, constituídos por palabras que, sin aparecer en todas las materias, aparecen en un número determinado de ellas.

- Vocabulario General de mayor Frecuencia, constituído por las 6.100 palabras de mayor frecuencia.

Finalmente, en 1979, Lorenzo Delgado presenta su tesis doctoral con el título de "El Vocabulario televisivo y el Vocabulario de Orientación Didáctica". Es el último de los recuentos generales del español del que tenemos conocimiento.

Su finalidad, al igual que el Vocabulario General de Orientación Científica de García Hoz, es eminentemente didáctica: determinar los contenidos léxicos del curriculum escolar, particularmente los correspondientes a los primeros niveles de la E.G.B.

El Vocabulario procede de una muestra de 300.000 palabras. Utiliza como fuente única el mismo el lenguaje de la televisión. La metodología del recuento no difiere esencialmente de la establecida por García Hoz (1953, 1977).

Como aportación hemos de resaltar el análisis comparativo que hace entre el vocabulario de televisión y otros vocabularios, concretamente el Vocabulario Usual, Común y Fundamental y el Vocabulario General de Orientación Científica, ambos de García Hoz. El Vocabulario de intersección resultante de este contraste es el que se denomina Vocabulario de Orientación Didáctica. De él se ofrecen tres niveles:

a) Nivel Usual: 6.100 vocablos.
b) Nivel Común: 2.274 vocablos.
c) Nivel Fundamental: 608 vocablos.

3.4.- "FREQUENCY DICTIONARY OF SPANISH WORDS". Juilland, A. y Chang-Rodríguez, E., (1964).-

El diccionario de frecuencias de Juilland y Chang-Rodríguez se enmarca dentro de un conjunto de investigaciones destinadas al análisis de la estructura fonológica, gramatical y léxica de las cinco lenguas romances de mayor dominio lingüístico, a saber, el francés, el italiano, el portugués, el rumano y el español. Tiene, pues, el diccionario un carácter y una finalidad predominantemente lingüística que le diferencia de otros diccionarios de frecuencias cuyos objetivos son de índole más pedagógica, didáctica o psicológica. Estas investiga-

ciones se llevaron a cabo en las universidades de Stanford y Pensilvania en la década de los años sesenta: Como novedad, respecto a los trabajos anteriores, aportan el uso de computadores en la realización de los recuentos.

Metodológicamente, el diccionario de Juilland es uno de los recuentos mejor planteados. Con rigor y propiedad se tratan todas y cada una de las operaciones que conlleva la compilación de un diccionario de frecuencias, desde la selección del cuerpo de palabras que constituyen la muestra hasta el análisis de resultados, sin olvidar la problemática compleja ligada al procedimiento de la información, en este caso las palabras.

a) La Muestra.-

Viene dada por un conjunto de palabras cuya naturaleza, estructura y tamaño se determina en función de unos objetivos planteados por esta investigación. El propósito de Juilland, al respecto, es seleccionar un conjunto de palabras equilibrado, homogéneo y al mismo tiempo representativo de la diversidad del lenguaje estándard como un todo.

La homogeneidad se obtiene a partir de la selección de unos textos que poseen una unidad:

- de espacio: obras escritas por autores de la España peninsular. Se excluye, por consiguiente, cualquier referencia a obras de la literatura hispanoamericana.

- de tiempo: publicadas entre 1920 y 1940.
- sociocultural y de estilo: textos en prosa, originales, que reflejan un lenguaje maduro, de adulto.

La diversidad se manifiesta en las distintas fuentes tomadas como origen de la distribución del léxico español. La selección de las mismas se hizo según un criterio de autoridad proporcionado por los historiadores más relevantes de la literatura española de la época. Estas fuentes, cinco en total, remiten a otros tantos géneros literarios los cuales hacen referencia a un número amplio de facetas de la actividad del mundo sociocultural español. Asumen, por tanto, la representación del lenguaje usual escrito. Los cinco géneros literarios elegidos fueron los siguientes:

1. Teatro: 55 obras de 23 autores. 100.000 palabras.
2. Novela: 50 obras de 18 autores. 100.000 palabras.
3. Ensayo: 52 obras de 38 autores: 100.000 palabras.
4. Literatura técnico-científica: 32 obras de 29 autores. 100.000 palabras.
5. Literatura periodística: 100.000 palabras extraídas de las diferentes secciones de cinco diarios y cinco revistas. Se utilizaron varios números de cada uno de ellos, con distinta fecha de publicación a lo largo de varios años.

b) Tamaño de la Muestra:

De los datos anteriormente expresados se deduce que el número total de palabras del corpus del que se realiza el

recuento asciende a 500.000. Este conjunto de palabras se distribuye por igual en cinco subconjuntos de 100.000 palabras cada uno. Ahora bien, ¿por qué estas cifras y no otras?, ¿qué criterios justifican la elección de este corpus como representativo del lenguaje escrito?.

Estas cuestiones tienen respuesta en la concepción de Vocabulario Fundamental o Básico que mantienen los autores.

c) Vocabulario Básico.-

Cualquier tipo de recuento que se realice supone el análisis de un reducido número de palabras de cuantas constituyen el léxico de una lengua. De los objetivos planteados por el análisis dependerá su mayor o menor extensión. Estos dependen, a su vez, de la estructura del lenguaje que se analiza, de la finalidad que se persiga, de la frecuencia con que ocurren las palabras y del concepto de vocabulario básico que se pretenda establecer. No obstante, en la determinación de un diccionario de frecuencias el concepto de vocabulario básico es el elemento fundamental para delimitar la extensión del recuento.

Juilland y Chang-Rodríguez conciben el vocabulario básico como un todo discontínuo formado por clases de palabras con unas determinadas propiedades estructurales prosódicas, fonológicas, gramaticales, de frecuencia... Cada clase se compone de un número de elementos determinado que puede variar siempre que el cambio no afecte a las propiedades de la clase.

Un vocabulario básico es suficiente cuando las series de clases de palabras que lo integran adquieren estabilidad en el sentido de que cualquier aumento adicional del recuento puede modificar el número de elementos de las clases pero no las propiedades que las definen. Así, pués, el recuento se detiene cuando las clases de palabras se estabilizan.

Sobre la base de este concepto de vocabulario, un universo de 500.000 palabras (muestra utilizada por Juilland) es suficiente para la obtención de un vocabulario básico.

d) Técnica de Registro y Codificación.-

Es en este aspecto en el que se manifiestan las mayores discrepancias entre el recuento de Juilland y el resto de recuentos conocidos del español. En una crítica precisa Juilland cataloga de insuficientes las normas de registro adoptadas por la mayor parte de sus autores.

El registro de las palabras conlleva como tarea fundamental delimitar qué es una palabra y cuándo debe registrarse de manera diferenciada del resto. Cuestión difícil de precisar incluso para los propios lingüistas. No obstante, el despojo de palabras que se realiza en todo recuento precisa de una normativa de clasificación lexicográfica que solvente los problemas de registro que se plantean: problemas de identificación y reducción de las palabras, por un lado, y por otro, de diferenciación y distinción entre unas palabras y otras. En resumen, problemas relacionados con la polivalencia

de uso y significado que presentan muchas palabras en un contexto dado.

Juilland procede de forma rigurosa y exhaustiva. Clasifica y registra todas las palabras tal como aparecen en el contexto. Con posterioridad, se realiza la tarea de reducción de cuantas variantes presenta una misma palabra ("token") a la invariante correspondiente ("type"). Se presta particular atención al problema de la homonimia. Al respecto, el diccionario de frecuencias de Juilland es la obra que proporciona una información más precisa acerca de las distintas formas en que esta se manifiesta: homonimia léxica, sintáctica y morfológica.

e) Los resultados.-

Los resultados se presentan en una doble lista. En primer lugar, la lista alfabética de vocablos acompañada de todas las variantes que éstos presentan con su correspondiente frecuencia en cada una de las cinco fuentes origen del diccionario. La segunda lista registra solamente las entradas o palabras tipo (vocablos) clasificadas en orden decreciente según su coeficiente de uso, frecuencia y dispersión. La lista aparece dividida en 10 grupos de 500 palabras cada uno correspondientes a las 5.000 que componen el vocabulario básico.

CAPITULO IV

IV.- EL VOCABULARIO INFANTIL.-

Existe una segunda vertiente en los estudios del vocabulario que tiene por objeto el estudio del léxico desde un punto de vista genético y evolutivo. Esta perspectiva responde a dos objetivos primordiales. Por un lado, el análisis del origen del lenguaje y más concretamente de la palabra: ¿cuándo aparecen las palabras?, ¿A qué edad se inicia el lenguaje articulado?. Y, en segundo lugar, la determinación del perfil evolutivo del vocabulario tanto a nivel cualitativo como cuantitativo: es decir, cuáles son las primeras palabras, cuál es su significado y, según la edad, qué incremento sigue el vocabulario, cuál es el número medio de palabras, qué palabras son las más frecuentes y qué aspectos diferenciales del vocabulario se deben a la influencia de otras variables además de la edad, como son el sexo, nivel de instrucción, procedencia sociocultural, cociente intelectual o determinadas pautas de respuesta psicofisiológica.

4. 1.- Etapa preverbal.-

La aparición del lenguaje, cómo y cuándo se produce, de qué factores depende o cuáles intervienen en su adquisición son algunas de las cuestiones previas que hemos de precisar antes de plantearnos el problema de la determinación del vocabulario infantil.

Uno de los factores determinantes en la aparición y consolidación del lenguaje del niño es el lenguaje de los adul

tos. Sin embargo su adquisición no es debida a un proceso de aprendizaje meramente receptivo u observacional, sino que precisa de una capacidad intrínseca al niño que, por otra parte, parece específica de la especie humana.

Siendo la palabra el medio de comunicación por excelencia, no es el único ni el primero que marca el inicio de las relaciones sociales. Antes de que el niño comience a hablar muestra atención por las palabras de los adultos, atiende a la entonación, a las flexiones de la voz, al ritmo, etc.. Los comienzos de la comunicación y la interrelación social son anteriores al conocimiento y uso del significado de las palabras y del lenguaje mismo.

Existe un período prelingüístico, preverbal en la adquisición del lenguaje por parte del niño, constituído por un sistema complejo de comunicación que encuentra su fundamento en la experiencia sensorial y motora del niño. Esta comunicación, esencialmente no verbal, se establece a partir del nacimiento y, a medida que crece el niño, aumenta su riqueza expresiva en gestos y contenido. Son numerosos los estudios que ponen de manifiesto la importancia, relación, e incluso dependencia de la comunicación verbal respecto a este otro tipo de comunicación prelingüística, preverbal o gestual que, por otra parte, tampoco desaparece con la llegada de la palabra sino que continúa en estrecha vinculación con ella.

La negación de esta fase previa en la adquisición del lenguaje no supondría sólo el rechazo de la evidencia experimental que atestigua su existencia, sino la misma independen

cia y autonomía de la comunicación del niño en las etapas iniciales de su desarrollo (Siguán, M., 1978; Schaffer, 1979). Existe una auténtica comunicación interpersonal antes de que tenga lugar la aparición de la palabra.

La interpretación del lenguaje en esta fase inicial es difícil. Comienza con sonidos o vocalizaciones cuyo papel en el desarrollo del lenguaje aún no está plenamente establecido. Días después del nacimiento, el niño responde con claridad al lenguaje hablado de los adultos, a la voz humana. El lenguaje hablado parece resultarle grato. A su vez, este dato nos indica un comportamiento selectivo en la relación del niño con otros seres humanos, preferencial frente a cualquier tipo de objetos o cosas de su entorno, para la que muestra una madurez cerebral que supone respuestas diferenciales ante distintos tipos de sonidos ya a una edad muy temprana (Molfese, D.L., 1973).

El período de las vocalizaciones corresponde a los seis primeros meses. En torno a esta edad se inicia el balbuceo. Los sonidos que ahora emite el niño se aproximan más al lenguaje que las vocalizaciones previas, cuyo valor lingüístico no parece definido. No obstante, las vocalizaciones cumplen otras funciones ya que son respuestas perfectamente coordinadas con situaciones biológicas de necesidad o satisfacción por las que atraviesa el niño. De aquí que la gama de vocalizaciones sea muy variada, desde el llanto a los arrullos pasando por otros sonidos menos definidos. La exploración de los mecanismos bucales articulatorios (Lewis, M., 1963), la ejercitación de los labios y la lengua, son otras de las hipótesis que

se han formulado respecto a la finalidad de estos primeros sonidos emitidos por el niño.

Tras la fase de vocalizaciones se inicia el periodo del balbuceo, entre los seis y doce meses. Los sonidos de arrullo y llanto se hacen más característicos y diferenciados y aparecen nuevos elementos vocálicos y consonánticos que se emiten asociados a modo de sílabas. El inicio del balbuceo parece ser una cuestión de maduración fisiológica ya que independientemente del tiempo de exposición al lenguaje a que haya estado sometido un niño, sucede aproximadamente en torno a la misma edad, incluso en los niños sordos de nacimiento. No obstante, la continuidad del balbuceo y su desarrollo con posterioridad a los 8 ó 9 meses depende de que el niño sea capaz de oirse a sí mismo y a los demás. Este hecho pone de manifiesto que la teoría que intenta explicar la adquisición del lenguaje en base a un proceso de aprendizaje condicionado por el refuerzo (gratificación de los padres de los sonidos emitidos por el niño) y por el aprendizaje observacional (el niño imita sus propios sonidos y los que escucha de los adultos) es insuficiente e inclompleta.

La pauta de desarrollo del balbuceo es muy similar en todos los niños, sea cual sea su lengua materna de origen. Todos los niños balbucean igual en este estadio (Nakazima,1962). Los primeros sonidos consonánticos son guturales (/g/, /k/). Más tarde, éstos tienden a desaparecer y son sustituídos por sonidos labiales y dentales (/b/, /p/. /d/). Unas y otras, en combinación con sonidos vocálicos, dan origen a las primeras secuencias de sonidos (sílabas) emitidas por el niño. Tampoco

estos cambios y combinaciones que se producen en el período de balbuceo pueden explicarse por un proceso de reforzamiento social o de imitación, aunque uno y otro contribuyen, y de forma particular en los momentos finales a aumentar la frecuencia del balbuceo. Será entonces cuando el aporte del lenguaje de los adultos comience a producir efectos diferenciales sobre el lenguaje infantil.

Antes de la denominada etapa holofrástica, la primera propiamente dicha que podemos considerar fuera de la fase preverbal, algunas autores consideran aún otro período intermedio hacia los 9 - 10 meses (Hernández Pina, 1979), caracterizado por la coincidencia de alguna de las vocalizaciones esporádicas del niño con vocablos propiamente dichos. Muchos niños a esta edad tienen en su repertorio emisiones con una pronunciación relativamente estable y de fácil identificación. Sin embargo, estas primeras "palabras" no deberían considerarse como el inicio de una lengua concreta, aunque sí manifiestan un mayor control sobre la articulación del lenguaje, pero el significado que encierran es aún muy difuso.

El paso del balbuceo a las primeras palabras, como señala Viliers (1980), representa un cambio de dirección en la articulación del lenguaje desde la práctica de un juego sin restricciones con los sonidos hasta el lenguaje planificado y controlado que suponen las palabras. Las secuencias de sonidos que ha de producir para hacerse entender con sus interlocutores precisan un orden que no aparece en el período de balbuceo. De hecho, los niños simplifican bastante la pronunciación de las primeras palabras de manera que su interpretación al

margen del contexto y los gestos de que van acompañadas resulta difícil pues su significado no se corresponde con el que generalmente tienen para el adulto.

Las primeras palabras se construyen, pues, sobre núcleos silábicos repetidos varias veces tales como "pa-pa", "ma-ma", "ta-ta" o "au-au". Aunque la estructura formal sea relativamente semejante en su articulación a la del adulto, su significado, como anteriormente apuntábamos es vago e indeterminado.

4. 2.- Las primeras palabras.-

El niño emite sus primeras expresiones verdaderamente lingüísticas entre los doce y dieciocho meses aproximadamente. En este punto coinciden casi todos los observadores (McCarthy, 1954). A este período se le denomina de la palabra-frase o etapa holofrástica. Se caracteriza porque la mayoría de las emisiones realizadas por el niño son frases de una palabra o un sólo elemento y, por tanto, carentes de las relaciones paradigmáticas y sintagmáticas pertinentes. Gramaticalmente se puede decir muy poco de ellas y semánticamente el análisis del significado no es posible si se extraen del contexto en el que se producen.

Desde el punto de vista fonético son bastante regulares y estables, normalmente constan de una o dos sílabas y las sílabas están constituídas casi siempre por la secuencia consonante-vocal.

En cuanto al significado hemos de decir que es muy genérico, amorfo e impreciso. El niño utiliza las palabras para referirse a una serie de objetos y acontecimientos libremente asociados y, normalmente, fuera del ámbito de referencia que tienen para los adultos. Amplían el significado de las palabras, no tienen un campo semántico definido. Identifican el significado de una palabra con alguna propiedad de un objeto: su forma, tamaño, color o sonido que produce, pasando a utilizar después la misma palabra para referirse a cuantos objetos comparten dicha propiedad. Estas generalizaciones del significado no están restringidas solo a los sustantivos sino también a los verbos. Lo que parece indicar que el niño para cada vocablo tiene un referente preferido, sea un objeto o acontecimiento o una acción y este referente suele coincidir con aquel al que más frecuentemente el niño está expuesto. Posteriormente, extiende el uso de la palabra (sustantivo o verbo) para denominar a cualquier otro objeto que comparta alguna o algunas de las propiedades o atributos de su referente preferido.

Por consiguiente, el vocabulario inicial empleado por el niño no hace referencia a la misma gama de objetos, acontecimientos o relaciones presentes en el vocabulario del adulto. La mayor o menor extensión semántica de una palabra para el niño viene determinada por el proceso de generalización, al que hemos aludido. No obstante, a pesar de que lo característico de la etapa de la palabra-frase sea la enorme latitud semántica de que gozan las palabras para el niño, no podemos restringir su vocabulario a las palabras que utiliza porque el significante de éstas no responde en realidad a un

significado concreto. La comprensión del lenguaje está por encima de lo que refleja su vocabulario. El niño comprende y responde adecuadamente a muchas más palabras de las que puede producir.

Las primeras palabras incluyen por lo general los nombres de los objetos o acontecimientos más destacados del mundo infantil, las personas de mayor importancia para el niño: padres, hermanos, abuelos, personas que le cuidan, sus alimentos favoritos, los juguetes, o los lugares y momentos más gratificantes de su vida cotidiana: el baño, el aseo, la cama... Los animales también ocupan tempranamente la atención del niño (Nelson, 1973). El vocabulario alcanza una extensión de unas cincuenta palabras en este período anterior a la aparición de la frase propiamente dicha o al menos de la etapa en que comienza a relacionar y asociar dos o más palabras.

Los niños usan sus primeras palabras de varias formas. Rara vez las emplean simplemente como nombres aunque esto ocurre cuando el objeto mencionado suscita fuertes emociones, por ejemplo, cuando el niño ve una pelota conocida y dice: "pelota!". Es más normal que vea las zapatillas de papá y diga: "papá" (Dale, 1980). Las primeras palabras con frecuencia cumplen una función explicativa y descriptiva de los objetos y acontecimientos de su medio ambiente. En ocasiones, a estas funciones de designación, descripción o clasificación de los objetos se añaden otras como el uso de las palabras para dar órdenes a los demás o dárselas a sí mismo o para manifestar actitudes de rechazo y negación.

De cualquier forma, lo cierto es que las primeras palabras parecen ser algo más que palabras aisladas, como afirmábamos más arriba. El niño tiene en su mente algo semejante a una oración pero no dispone de los términos adecuados para expresarla. Varias han sido las hipótesis propuestas ante el hecho de la palabra-frase: existe una limitación en la capacidad de atención y memoria del niño en esa edad que le impide el uso simultáneo de varias palabras (McNeill, 1970); se debe a un problema de aprendizaje: el niño carece aún de la destreza lingüística necesaria para poderse expresar a través de oraciones (Ingram, 1971). Algunos autores,no obstante, niegan la existencia de este período de la palabra-frase. Este es el caso de Bloom (1973). Para Bloom la insuficiencia del niño para construir frases se debe a un desconocimiento del significado de las palabras en un contexto, es decir, al significado que se deriva de la relación de unas palabras aisladas, le falta la capacidad para codificar lingüísticamente las relaciones entre las palabras, por tanto, no se les puede atribuir una función que no tienen. La aparición de una sintaxis rudimentaria encerrada en la palabra-frase no es sostenible lingüísticamente. La aparición de la sintaxis corresponde al período siguiente denominado del lenguaje telegráfico y que se caracteriza por la combinación de 2 ó más palabras.

En resumen, hasta finales del primer año no aparecen las primeras palabras, mientras tanto y aún después, la comunicación entre el niño y el adulto discurre a través de un complejo ritual de gestos. Más tarde, su vocabulario se irá ampliando y descubrirá la gramática, pero por importantes

que éstos puedan ser, no constituyen todo el conjunto de elementos que el niño aprende para comunicarse con sus semejantes. Durante la edad preescolar el lenguaje va ganando en flexibilidad y llega el momento en el que el niño manifiesta poseer una riqueza léxica y gramatical que utiliza para sus propios fines.

Sería complejo hacer una revisión en profundidad de todos los aspectos implicados en el desarrollo y adquisición del lenguaje del niño. No es esta nuestra intención. Si hemos hecho referencia sucinta a las primeras etapas no ha sido con otro interés que el de subrayar la aparición de la palabra como unidad léxica. En torno a ella gira nuestro trabajo.

4.3.- El Vocabulario Infantil: el incremento del léxico.

Los primeros trabajos sobre el lenguaje infantil se centran en la observación y descripción del mismo. Todos analizan la palabra como dato elemental. La tarea de los investigadores consistió en observar y registrar la producción verbal del niño. Existen numerosos estudios donde se recogen datos sobre el proceso de adquisición del lenguaje por niños de distintas lenguas.

Algunas de las recopilaciones del lenguaje infantil en los primeros años de desarrollo datan de finales del siglo pasado y principios de este siglo. Fueron hechas generalmente por Lingüístas que utilizaron como sujetos experimentales a sus propios hijos. En este sentido, son clásicos los estudios

de Antoine Gregoire (1937) para el francés; de Werner Leopold (1939-1949) para el alemán y los diarios acerca del aprendizaje del inglés de Chamberlain A.F. y Chamberlain J.J. (1904) y de H.V. Velten (1943). Recientemente, en español, Fuensanta Hernández Pina (1979) ha realizado un estudio sobre la adquisición y desarrollo del lenguaje del niño, utilizando a su hijo como sujeto único y del que hace un seguimiento del lenguaje desde las primeras emisiones hasta los tres años.

En cualquier caso, se trata de estudios descriptivos longitudinales con sujeto único. Este método posee sus ventajas: es de índole longitudinal de tal modo que ponen bien de manifiesto el proceso de aprendizaje de un sólo niño; por otra parte, es llevado a cabo por personas que conocen bien al niño, sus hábitos, sus costumbres, sus antecedentes.. Pero presenta, asimismo, ciertos inconvenientes no sólo de registro sino también de generalización de conclusiones de modo que no podemos estar seguros de que un diario sea representativo del lenguaje infantil en general y si los correspondientes datos son válidos para otros niños.

En los últimos años se han perfeccionado las técnicas de registro, de codificación y análisis. Las aportaciones de la tecnología moderna, el magnetófono, el cine, la televisión, el video..., han contribuído a un mejor conocimiento del contexto en el que el niño actúa y se comunica con las personas de su entorno. Esto ha supuesto un mayor acercamiento no sólo a la palabra como tal, sino también al contenido semántico, es decir, al significado.

El aumento del vocabulario a partir de la producción de las primeras palabras, tiene lugar a un ritmo acelerado a medida que aumenta la edad del niño. La experiencia y no pocos estudios así lo manifiestan.

Las palabras pierden pronto el carácter difuso, elemental, simpráxico (según la terminología de Luria, 1979), es decir, de relación con la acción; dejan de designar una situación y se convierten en autónomas, independientes del contexto simpráxico. Comienzan a tener un significado objetal, adquieren el carácter del sustantivo, es decir, la designación del objeto. Hasta entonces (18 - 20 meses) lo más propio son las palabras amorfas en cuanto al significado, ya que son utilizadas por el niño para designar cualquier cosa, pues su significado depende de la situación donde se emplean, de los gestos que le acompañan o de la entonación con que son pronunciadas. Esto explica, según Luria, que sus necesidades de comunicación se vean satisfechas con un reducido número de palabras. Sin embargo, a partir de esa edad, los significados se reducen y el vocabulario se amplia. El niño precisa enriquecer su vocabulario para poder reflejar más adecuadamente el mundo que le rodea; precisa nuevas palabras que le pongan en contacto con una realidad cada vez más compleja no sólo por el mayor número de objetos que percibe cuanto por la amplitud de las acciones, relaciones y cualidades con las que aquellos se relacionan.

La mayoría de las investigaciones coinciden en señalar la existencia de un reducido número de palabras (entre 12 y 20) en torno a los dieciocho meses que pronto asciende a las

100 (21 meses), 200 (22 meses), 300 (24 meses), para situarse en el millar pasados los tres años.

Este rápido aumento, según Luria, se explica por el paso de un habla simpráxica al habla sinsemática, es decir, por el distanciamiento progresivo entre acción y palabra. La palabra nace en un contexto simpráxico y se separa progresivamente de la práctica "se convierte en un signo autónomo, que designa un objeto, una acción, una cualidad (y más adelante una relación). En este momento ocurre el verdadero nacimiento de la palabra diferenciada como elemento del sistema complejo de códigos de una lengua" (Luria, A.R. Conciencia y Lenguaje, pp. 34).

Por esta razón, los cálculos directos y absolutos del número de palabras que componen el vocabulario infantil en cada edad difícilmente se pueden precisar más allá de los dos o tres primeros años. A partir de entonces, los resultados que reflejan las investigaciones se obtienen de cálculos inferenciales. Otros procedimientos de recopilación de datos absolutos por encima de los tres años resultan materialmente imposibles (McCarthy, 1954).

McCarthy (1954), tras un exhaustivo análisis de numerosos estudios sobre el desarrollo del vocabulario realizados por diversos autores con diferentes objetivos, sujetos y metodología, resume los distintos enfoques en las siguientes categorías:

1. Recuentos generales por edades.
2. Recuentos analíticos: clasificación de las palabras según las diferentes categorías gramaticales.
3. Recuentos de vocabulario agrupando los términos por temas o conceptos.
4. Recuentos de vocabulario del lenguaje oral (conversaciones).
5. Recuentos de vocabulario del lenguaje escrito.
6. Cómputos de vocabulario obtenidos mediante la técnica de la libre asociación.
7. Cómputos obtenidos mediante el uso de tests de vocabulario.
8. Recuentos de frecuencias.

Observamos que la clasificación responde a dos criterios. Por un lado, se tiene en cuenta la fuente de procedencia del vocabulario y el tipo de lenguaje que se analiza (oral o escrito), por otro, la finalidad del recuento -proporcionar un cómputo absoluto o relativo-, y la clasificación de las palabras (establecimiento de categorías gramaticales y cálculo de frecuencias). Por consiguiente, podemos reducir a cuatro las eategorías fundamentales:

1. Recuentos sumarios por edades.
2. Vocabularios de reconocimiento.
3. Vocabularios analíticos: aproximación al signifi

cado.

4. Cálculos de frecuencia.

4.3.1.- Recuentos sumarios por edades: el incremento cuantitativo del vocabulario.

La determinación del número de palabras que el niño adquiere, sirve para establecer una descripción cuantitativa de la evolución que sigue el desarrollo de su vocabulario. Las investigaciones clasificadas en este grupo son las más numerosas. Habitualmente, los recuentos de las primeras edades se llevaron a cabo con un sólo sujeto, como anteriormente mencionamos, o utilizando un reducido número para edades posteriores. De cualquier manera, tanto si se trata de establecer cómputos absolutos como relativos, siempre hay un dato que subrayan todos los observadores: el rápido aumento del vocabulario con la edad. Este aumento, como afirma McCarthy, es relativamente lento, al principio, sufre una aceleración en la época preescolar para disminuir nuevamente después.

Sin entrar ahora en la consideración de los presupuestos metodológicos, las muestras utilizadas, las técnicas de recogida de datos, la exhaustividad de los recuentos o su fundamentación teórica (que es compleja y variada, lo que hace que los resultados de unas investigaciones y otras difícilmente sean comparativos entre sí), hemos de precisar como muy bien señala Dale (1980), que el estudio del desarrollo infantil en cualquiera de sus aspectos cambia constantemente, y ésto no sólo se debe al perfeccionamiento de la metodología y a la aparición de nuevos modelos de estudio, sino también a los

cambios experimentados en el mismo objeto sujeto al análisis, es decir, el niño. Múltiples aspectos del desarrollo que pueden ir desde el inicio de la pubertad (Tanner, 1973) hasta el rendimiento en las pruebas de inteligencia, se han incrementado en los últimos cincuenta años. El lenguaje, por ser una realidad viva constantemente sujeta al cambio, es particularmente uno de ellos.

Con la cautela, pues, con que han de ser valorados los datos absolutos que proporcionan los recuentos sumarios, ofrecemos los resultados de algunas investigaciones. Hemos de añadir que el objetivo de todas ellas es fijar el vocabulario de cada edad y analizar el incremento mediante un muestreo de la producción verbal del niño.

a) Descoudres, A. (1921), datos relativos a cinco sujetos franceses (se trata de cifras medias: la diferencia entre los distintos sujetos aumenta con la edad).

Edad	Número de palabras	Incremento
1,9	174	174
2,9	639	465
3,1	1394	755
5,0	1954	560
7,0	2903	949

b) Stern, C. y Stern W. (1907), para niños alemanes:

Edad	Número de palabras	Incremento
1,6	100	100
2,0	300-400	200-300
3,0	1000-1100	700
4,0	1600	500-600
5,0	2200	600
6,0	2500-3000	300-800

c) Averril, M. (1956):

Edad	Número de Palabras	Incremento
2,0	300	300
4,0	1500	1200
6,0	2500	1000
8,0	3500	1000
10,0	5500	2000
12,0	7500	2000
14,0	9500	2000

Las cifras, como se puede observar, difieren de unos autores a otros. Los números tienen un valor relativo si admitimos que proporcionan únicamente indicaciones acerca del progreso que el vocabulario experimenta con la edad, sin que nos lleven a pensar que las cifras establecidas para una edad concreta expresan en realidad el vocabulario que entonces posee el niño. De cualquier modo, se trata siempre de promedios y

las diferencias individuales, principalmente en las primeras edades, son considerables como asimismo lo son las que provienen de la influencia de otros factores como la inteligencia o la condición sociocultural de los sujetos, la situación experimental o la metodología con que se investigue. Estas mismas diferencias se aprecian también en la evaluación cuantitativa del vocabulario de los adultos y en los recuentos generales revisados en los capítulos anteriores.

4. 3. 2.- Los Vocabularios de Reconocimiento.-

Con objeto de subsanar las deficiencias que presentan los recuentos absolutos, sobre todo a partir de una edad en que el vocabulario del niño es ya muy extenso, aparecieron otros métodos para evaluar el incremento. Este es el caso de los vocabularios de reconocimiento. La evaluación del incremento se realiza por medio de tests basados en el vocabulario adulto. Los cómputos absolutos de palabras se obtienen del promedio de palabras reconocidas por la muestra de sujetos de cada edad sometidos a prueba.

Presentamos a continuación, a título ilustrativo, los datos obtenidos por M. E. Smith en 1926 con este método. El procedimiento seguido consistió en confeccionar un test con un número determinado de palabras pertenecientes al vocabulario adulto. Exactamente, se eligieron 203 palabras de entre las más frecuentes de las 10.000 que componen la lista de Thorndike (1921). Las palabras se sometieron al reconocimiento de varias muestras de niños de diferentes edades. El núme-

ro de items que contenían las pruebas para cada edad oscilaba entre 5 y 52. La presentación de las palabras se hizo a través de dibujos. La tarea de experimentador consistía en constatar si el niño evocaba la palabra cuando se le presentaba el dibujo. En caso negativo, se le formulaba una pregunta utilizando la palabra correspondiente y se valoraba la comprensión que el niño tenía de ella en la respuesta que daba.

La proporción de palabras conocidas se correlaciona con el vocabulario total por una simple multiplicación.

Hemos de señalar que los resultados obtenidos por Smith no expresan el vocabulario que posee el niño sino la proporción de palabras que reconoce en una lista parcial.

Edad	Número de Palabras	Incremento
0,8	0	0
0,10	1	1
1,0	3	2
1,3	19	16
1,6	22	3
1,9	118	96
2,0	272	154
2,6	446	174
3,0	896	450
3,6	1222	326
4,0	1540	318
4,6	1870	330
5,0	2072	202
5,6	2289	217
6,0	2562	273

Incremento del Vocabulario y Edad (Smith, 1926).

Los puntos discutibles de este procedimiento residen en la selección de las palabras base, en el número de items de la prueba, en la amplitud de la muestra de donde proceden y en los criterios de selección de las mismas. Basta alterar cualquiera de ellos para obtener unos resultados completamente diferentes.

Los resultados obtenidos por Smith tiene, pues, un valor muy relativo y no revelan, como muy bien señala Miller (1969), la amplitud del vocabulario del niño medio. Los datos tienen un valor meramente indicativo acerca del desarrollo

verbal en diferentes edades. Por otra parte, H. Williams ya subrayó en 1932 otra incongruencia de estos trabajos que dió origen a nuevos tests de vocabulario y a nuevos procedimientos de selección de palabras más adecuados. Nos referimos a las diferencias existentes entre el vocabulario infantil y el vocabulario de los adultos. Es obvio, que a pedida que crece el niño su vocabulario se va acercando progresivamente al vocabulario de los adultos y en último término al léxico que contienen los diccionarios. En consecuencia, los datos obtenidos por este método son tanto más válidos cuanto mayor es la edad de los sujetos muestreados, pero presentan claras deficiencias en los niveles de edad más inferiores.

Ciertamente, la mayor parte de estas deficiencias se han ido subsanando en investigaciones posteriores (Gansl, 1939; Seashore y Eckerson, 1963; García Hoz, 1946, 1977; Ehrlich y colb., 1978), sin embargo, la objeción principal sigue residiendo en el carácter más o menos fortuíto de la selección de palabras de base (Francescato, 1971).

García Hoz (1977), basándose en los resultados de algunas de sus investigaciones sobre el número de palabras conocidas por alumnos españoles de diferentes grados de enseñanza, trató de fijar analíticamente las diferencias observadas. La representación de las mismas le llevó a la formulación de dos hipótesis sobre el incremento del vocabulario.

Las pruebas utilizadas para medir el incremento del vocabulario de reconocimiento fueron dos. Ambas se elaboraron siguiendo un procedimiento semejante al empleado por Smith. En

ambos casos también, el test de la prueba estuvo constituído por cien palabras. No obstante, las diferencias más notorias entre ambas eran relativas a la fuente de procedencia de las palabras del test y al procedimiento de selección.

La primera de ellas tomó como base el vocabulario general de la lengua, considerando como tal el conjunto de entradas que registra el diccionario de la Real Academia. Las cien palabras seleccionadas se extrajeron mediante un sistema de catas. Este primer test se pasó a una muestra de 2774 sujetos. En la Tabla A aparecen los resultados:

TABLA A

Edad de los sujetos	Porcentaje de palabras conocidas
9 años	29.54
10 "	37.48
11 "	40.83
12 "	43.34
13 "	51.22
14 "	54.56
15 "	59.20
16 "	62.87

La segunda prueba se elaboró a partir del Vocabulario Usual del propio García Hoz (1953). La investigación se realizó con 425 alumnos de ocho a doce años. Resultados en la tabla B:

TABLA B

Edad de los sujetos	Porcentaje de palabras conocidas
8 años	32.34
9 "	38.77
10 "	42.00
11 "	47.28
12 "	51.40

La observación inicial de los datos llevóle a formular la primera de las hipótesis: el aumento del vocabulario sigue una función lineal. Hipótesis que parecían confirmar otras investigaciones (Averril, 1956).

Ajustando las rectas correspondientes, observó que los resultados eran casi idénticos:

Experimento A: $y = 26.396 + 4.622\ x$

Experimento B: $y = 28.369 + 4.663\ x$

A pesar de la coincidencia de los resultados, la adecuación de las rectas ($s_r = 1.30$ para la serie A y $s_r = 0.97$ para la serie B) abrió la posibilidad de una segunda hipótesis.

La segunda hipótesis postula que el incremento del vocabulario sigue una función parabólica. Hechos los correspondientes ajustes y adaptaciones de ambas parábolas (Experimento A: $y = 25.04 + 5.467\ x - 0.0904\ x^2$ y Experimento B: $y = 27.080 + 5.767\ x - 0.8184\ x^2$), se llegó a la conclusión

de que la evolución cuantitativa del vocabulario se puede expresar por medio de una función parabólica. Sin embargo, los datos no permitían el establecimiento de la ecuación exacta de la parábola. Lo único que afirman es que el coeficiente de la incógnita de segundo grado es pequeño y que la aproximación por tanto, entre la representación lineal y parabólica es muy alta. Tal vez, la confirmación de una de las dos hipótesis hubiese sido posible si se hubiesen muestreado mayor número de grupos (edades) que mostrasen más claramente la tendencia del incremento para poder determinar con exactitud la ley de aumento del vocabulario.

4. 3. 3.- Recuentos analíticos: la aproximación al significado.

Desde los primeros estudios sobre el aumento cuantitativo del vocabulario, el aspecto morfológico ha sido también uno de los examinados más atentamente. Muchos autores han inventariado el vocabulario clasificando los términos según las categorías gramaticales: nombres, verbos, adjetivos, adverbios, preposiciones, conjunciones... En la cuantificación de las partes de la oración, en el incremento y cambios que en éstas se producen, consiguientes al desarrollo, han pretendido ver un índice no sólo del desarrollo cuantitativo sino también del incremento cualitativo del léxico.

La insuficiencia teórica de estas investigaciones es, sin embargo, más notoria en este campo que en otros. A pesar de ello, es abundante la información que al respecto proporcio

na la lingüística infantil clásica. A título informativo, reproducimos aquí la clasificación obtenida por McCarthy (1930).

	18 meses	2 años	3 años	4 años
Sustantivos	50	38	23	20
Pronombres	10	15	19	22
Adjetivos, artículos	10	10	16	15
Verbos	14	21	23	26
Adverb. Prep., Conj.	8	11	16	16
Interjecciones	8	2	1	1
Varios	0	2	1	1

CUADRO 4.2.- Distribución de las palabras (porcentajes) según las categorías gramaticales (McCarthy, 1930).

Se observa una evolución que sigue un sentido de aproximación a la distribución que se aprecia en el lenguaje adulto. El niño, al principio, sólo emplea palabras semánticamente "llenas" (nombres, adjetivos, verbos...), y omite las palabras "vacías" o carentes de significación propia, es decir, preposiciones, artículos, conjunciones, interjecciones... (Guillaume, 1927). El predominio de las palabras con función léxica sobre aquellas que desempeñan una función meramente gramatical comienza a desaparecer gradualmente a partir de los dos años. Este cambio está relacionado con el progresivo desarrollo del uso de frases que se ajustan cada vez más a la es

trutura de las empleadas por el adulto.

En la actualidad, se tiende a dar una menor importancia a este tipo de clasificaciones, no sólo por el hecho de que la información que proporcionan es escasa sino también por la dificultad que entraña el establecerlas. Dificultad que se deriva de la polivalencia semántica (polisemia) de un gran número de palabras y que se acentúa, como hemos visto, en las primeras fases del desarrollo correspondientes al período holofrástico, debido a que el niño no dispone aún de un lenguaje suficiente y adecuado para la comunicación. Además, desde la perspectiva de las modernas teorías lingüísticas apenas si tiene sentido la distinción de categorías gramaticales con objeto de extrapolar de ellas un determinado contenido semántico, es decir, si se pretende ver reflejado, por ejemplo, en la distribución e incremento que se produce en sustantivos y verbos un índice del dominio del sujeto sobre sustancias y acciones.

La aparición de ciertas palabras en el vocabulario del niño, la descripción del crecimiento progresivo del léxico o la clasificación gramatical de los vocablos, no pueden considerarse en sí mismos una respuesta adecuada y científica al problema de la adquisición del significado. Una aproximación correcta al problema del aprendizaje gramatical supone afrontar y resolver previamente otra cuestión, la del conocimiento y utilización que el niño hace del significado.

Este hecho explica en cierta medida la decadencia y la falta de interés por los estudios del vocabulario, mostrada por la mayoría de los investigadores del lenguaje infantil después de la década de los años cuarenta, una vez acumulados nu-

merosos informes con gran cantidad de datos sobre el crecimiento cuantitativo del léxico. A ellos hemos hecho referencia en los epígrafes anteriores.

Desde entonces, los modelos propuestos para el estudio del vocabulario han girado en torno al análisis del significado. Las diferencias entre ellos se derivan básicamente de la perspectiva lingüística, filosófica o psicológica en la que se sitúan sus autores. Excede nuestros planteamientos y sería prolijo un tratamiento del tema en profundidad. Una revisión de estos trabajos puede consultarse en Francescato, 1971; Geckeler, 1976; Wotjak, 1979; Oleron, 1976, 1981; Dale, 1980... Aquí, sólo hemos querido resaltar la relación que presentan algunos de los primeros trabajos del vocabulario entre el incremento cuantitativo y la adquisición del significado. Relación fundamentada en la teoría lingüística del referente vigente en los comienzos de siglo.

Esta teoría sostiene que el significado de una palabra es un referente. Las palabras son símbolos que representan las cosas del mundo, es decir, sus referentes. Ciertamente, para explicar el significado de no pocas palabras, basta con señalar algo de lo que nos rodea. Las palabras no tienen significación alguna sino en la medida en que nos remiten a algún objeto o cosa. Aquello a lo que se refiere una palabra se denomina referente. El referente, pues, implica la existencia de una relación funcional entre los objetos reales y los signos que son las palabras. En esta teoría, el significado se construye a partir de los objetos, a partir de la realidad. Se supone la existencia de una correspondencia entre las uni-

dades verbales y las unidades reales. El lenguaje se reduce a una nomenclatura de la realidad. El aprendizaje de la lengua no supondría más que el aprendizaje de las denominaciones. Si cada objeto tuviera un sólo nombre y cada nombre un único referente no existiría ningún problema. El aprendizaje de la lengua sería sencillo y, en consecuencia, el perfil del aumento cuantitativo podría considerarse a su vez un índice adecuado del desarrollo del pensamiento y del significado.

En la actualidad esta teoría es insostenible. Según los principios que la definen el desarrollo o evolución en el significado de las palabras, como señala Vygotsky (1973), es inexplicable e imposible si se postula la existencia de una relación inexorable entre sonido (imagen acústica) y contenido (significado).

> "Las antiguas escuelas de psicología consideraban que el enlace entre palabra y significado era un vínculo de asociación que se establecía a través de percepciones simultáneas y repetidas de determinados sonidos y objetos. Una palabra sugiere a la mente su contenido como el sobretodo de un amigo nos hace pensar en él o una casa en sus moradores. La asociación entre palabra y significado puede tornarse más fuerte o más débil, enriquecerse por conexiones con otros objetos de la misma especie, extenderse sobre un campo más amplio o restringirse a otro más limitado, puede sufrir cambios cuantitativos y externos, pero no puede cambiar su naturaleza psicológica. Para que esto sucediera tendría que dejar

de ser una asociación... El desarrollo (del significado) se reducía a cambios en las conexiones asociativas entre determinadas palabras y objetos. Un vocablo podría denominar primero un objeto y luego asociarse con otro, de la misma forma en que un sobretodo, habiendo cambiado de dueño, puede hacernos recordar primero a una persona y luego a otra. La lingüística no comprendió que con la evolución histórica del lenguaje también cambian la estructura del significado y su naturaleza psicológica. El pensamiento verbal se eleva de las generalizaciones primitivas a los conceptos más abstractos. No cambia sólo el contenido de la palabra, sino el modo en que se generaliza la realidad y se refleja a través de la palabra". (Pensamiento y Palabra, Vygostky, 1973).

Al identificar los objetos con las palabras y éstas con sus significados, toda posibilidad de desarrollo conceptual y, por tanto, del pensamiento queda reducida al establecimiento de nuevas asociaciones. Las modificaciones en el significado, su naturaleza psicológica, su ampliación o reducción, la evolución cualitativa que sufre, la formación de conceptos quedan excluídos en esta teoría. En una palabra, la posibilidad de un desarrollo del pensamiento y el lenguaje son inviables desde el momento en que se considera adquirido el significado con la aparición de la palabra. Según la teoría del referente, la palabra, el significado y los objetos son una misma cosa.

La consideración del lenguaje como nomenclatura cuya finalidad se reduce a la denominación de los objetos no tiene sentido frente a la consideración del lenguaje como sistema tal como lo definiera F. Saussure. La relación entre realidad y lenguaje no se explica por la simple asociación entre objetos y palabras. Las palabras son algo más que unidades independientes. Para Saussure, el signo lingüístico (la palabra) es una entidad psíquica con dos caras: el significado o concepto y el significante o imagen acústica. El signo lingüístico no relaciona un objeto con un nombre, sino un concepto con una imagen acústica.

4. 3. 4.- Cálculo de Frecuencias.-

Los recuentos de palabras que tienen por objeto el cálculo de frecuencias son relativamente recientes y escasos todavía, por lo que al vocabulario infantil se refiere. Los presupuestos metodológicos en los que se apoyan no difieren, sin embargo, de los utilizados por los recuentos generales de vocabulario expuestos en los capítulos segundo y tercero.

El estudio de la frecuencia de uso de las palabras se acentuó en el período subsiguiente a la aparición de la teoría de la información (Oleron, 1981). Con anterioridad, son muy pocos los trabajos que podemos reseñar sobre frecuencias del vocabulario infantil. En efecto, la teoría de la información concedía una gran importancia a la frecuencia con que ocurren cuantos elementos integran un mensaje. La frecuencia es considerada el criterio base en función del cual se puede

determinar la probabilidad de aparición de una palabra en un mensaje.

Son numerosas las experiencias que han demostrado el papel desempeñado por la frecuencia en tareas de percepción, memorización y evocación de palabras. La frecuencia forma parte de las características que definen el estímulo. Es un factor importante en la asociación entre un estímulo y una respuesta. La frecuencia en el marco de los modelos conductistas fue invocada como explicación del aprendizaje bastante antes de que hiciera su aparición en los estudios consagrados al lenguaje.

No existen evaluaciones precisas sobre la frecuencia de uso de palabras producidas por niños bien sea con muestras del lenguaje oral, bien sea con muestras del lenguaje escrito. Los estudios cuantitativos sobre el desarrollo del vocabulario se han centrado más en el análisis de datos relativos al número de palabras emitidas y al incremento que éste sufre con la edad, que al análisis de la frecuencia con que ocurren las palabras mismas. Por otra parte, los escasos trabajos que analizan este aspecto presentan serias deficiencias metodológicas, particularmente los primeros que se realizaron. En su mayor parte, las frecuencias que comunican se refieren a palabras emitidas por un sólo sujeto. Así, lo señalan diversas revisiones sobre la literatura del tema (Doran, 1907; Whipple, 1909, 1915; McCarthy, 1930, 1946, 1954; Uhrbock, 1936; Weir, 1962).

Con mayor número de sujetos, Horn (1925) compiló una lista de las palabras más frecuentes emitidas por niños de habla inglesa entre uno y seis años. Sin embargo, las frecuencias establecidas por Horn, a pesar de que povenían de una muestra considerable (270.000 palabras emitidas), no son específicas para cada palabra ni para las edades separadamente. Se limita a presentar una lista con las palabras más frecuentes y su correspondiente frecuencia, en el total de las emitidas.

A partir de entonces, los datos disponibles sobre frecuencias de palabras emitidas por niños de habla inglesa han aumentado considerablemente. Buckingham y Dolch (1936) proporcionan una lista con palabras obtenidas en asociaciones libres de niños y extraídas de otros recuentos de la producción verbal infantil así como de palabras aparecidas en silabarios y libros de lectura para niños. Rinsland (1946) compiló otra lista sobre el vocabulario escrito. Murphy (1957) analizó el vocabulario de un grupo de niños en situaciones espontáneas de habla, pertenecientes a los primeros grados de enseñanza. Los resultados registran 6.318 entradas diferentes de un recuento total de 1.195.098 palabras. Templin (1957) presenta el vocabulario usual expresado por niños en 50 emisiones.

Beier y colbs. (1965) registraron conversaciones de niños entre 12 y 16 años hasta completar una muestra de 200.000 palabras (40.000 por grupo). Determinan la frecuencia de uso y analizan el incremento que se produce con la edad. Un estudio semejante fue realizado por Wepman y Hass (1969)

con niños de 5, 6 y 7 años. En este caso, el material verbal utilizado estuvo constituìdo por historias narradas por niños ante las las làminas del T.A.T.

En 1971, Newman y Baily confeccionaron una lista con las 1000 palabras más frecuentes usadas por niños de preescolar. Se basaron para elaborar la lista en ocho recuentos de frecuencias aparecidos entre 1928 y 1971, que reunían en total una muestra de dos millones de palabras aproximadamente. Reich y Reich (1977), en Toronto (Canadá), llevaron a cabo una investigación cuyo objetivo fue contrastar el vocabulario de niños de primero, segundo y tercer grado de enseñanza primaria, con listas previas de vocabulario recomendadas para la enseñanza de la lectura en estos mismos niveles. Los resultados confirmaron la existencia de grandes diferencias entre uno y otro.

En español, los recuentos de frecuencias de palabras emitidas por niños son aún escasos. Sólo disponemos de dos recuentos realizados recientemente bajo el patrocinio del Instituto de Ciencias de la Educación. Ambos se llevaron a cabo dentro de un ámbito geográfico reducido y responden a unos objetivos eminentemente pedagógicos: "poner al alcance de los docentes una información sobre el léxico real del alumno, el léxico de su lengua materna" ("Vocabulario básico del español y sus aplicaciones a la enseñanza".D.Castañón, 1975, pp. 4). Nos referimos a los trabajos de Armayor y Colbs. (1975, 1977, 1979) y de Díaz Castañón y colbs. (1975).

Los recuentos de Armayor (1975, 1977) determinan el vocabulario de uso del niño cordobés de 6 y 14 años. Se eligieron estas edades que marcan el comienzo y el final de los estudios correspondientes a la E.G.B., con el fin de observar la evolución que seguía el vocabulario durante este periodo educativo.

Los resultados proporcionan una lista ordenada por frecuencias de 2574 vocablos para los niños de 6 años y de 4286 para los niños de 14 años. En ambos casos la muestra de palabras pertenecía al lenguaje oral. Las conversaciones registradas se obtuvieron en diversas situaciones de habla espontáneas y sugeridas.

Estos dos trabajos sobre el léxico activo del alumno se complementan con el "Vocabulario Técnico escolar para alumnos de 8º de E.G.B." extraído de textos escolares (Armayor, 1979). Este recuento de textos sigue en su elaboración las pautas marcadas por García Hoz en el Vocabulario de Orientación Científica (1977).

El trabajo de Díaz Castañón (1975), realizado en la provincia de Oviedo, contiene una información sobre el vocabulario más rica y extensa. Utilizó ~~una~~ muestra de 3150 alumnos (1890 de E.G.B. y 1260 de B.U.P.), agrupados en cinco bloques según la edad (tres correspondientes a E.G.B.: 8 - 9; 10 - 11 y 12 - 13 - 14 años; y dos correspondientes a B.U.P.: 15 - 16 y 17 años).

El vocabulario se recogió de la producción verbal escrita de los sujetos mediante una encuesta con dos modalidades:

a) Encuesta libre: consistente en escribir durante cinco minutos todos los términos que libre y espontáneamente se le ocurran al sujeto.

b) Encuesta controlada: consistente en escribir los veinte primeros términos que acudan a la mente tomando como tema ("centro de interés") un campo familiar al alumno. 11 fueron los temas de interés sugeridos.

De esta forma se obtuvieron 170.000 (encuesta libre) y 680.000 (encuesta controlada) palabras respectivamente que dieron lugar a un corpus total de 850.000 palabras. Una vez hecho el recuento y calculada la frecuencia de los vocablos los resultados se presentan en una triple lista: la primera, ordenada alfabéticamente; la segunda, en orden decreciente de frecuencia y la tercera, según los temas o centros de interés sugeridos.

Las hipótesis en que se fundamentan los trabajos mencionados permiten probar y establecer que las palabras más frecuentes que proporcionan los cálculos son las que se aprenden más temprano, están más disponibles y son más fácilmente movilizables por el sujeto para explicar y aprehender la realidad que le rodea.

En relación con la adquisición del significado, no existen aún estudios suficientemente consistentes y sistemáticos que permitan hacer una valoración acerca de la información que al respecto proporcionan los datos sobre la frecuencia. En ciertos casos, ha sido posible relacionar la frecuencia y el orden con que se adquieren ciertos significados. La teoría de los rasgos semánticos (una de las alternativas más recientes en el estudio del significado), así parece confirmarlo.

En efecto, E. Clark (1972) en la discusión de resultados sobre el orden de adquisición de los antónimos espaciales y espacio-temporales reconoce la existencia de una correlación significativa entre aquel (el orden) y la frecuencia de uso de las parejas de antónimos estudiadas (por ejemplo, big / small" es cinco veces más frecuente que "tall / short" y de 4 a 10 veces más frecuente que "wide / narrow", "thick / thin" y "deep / shallow"). Ciertamente Clark, a pesar de la correlación, objeta la tesis que atribuye a la frecuencia un papel determinante en la adquisición del significado. Sin embargo, como sugiere Oleron (1981), no podemos caer en el error de considerar la frecuencia como un dato meramente estadístico carente de cualquier otro valor o significación. Cuando se recurre a la frecuencia para interpretar comportamientos relativos a palabras significativas, no se puede excluir la significación.

Los cálculos de frecuencias y la misma experiencia sugieren, respecto a las parejas de antónimos, que los términos positivos se utilizan con más frecuencia que los negativos. Los términos positivos son prioritarios, se aprenden antes, co

rresponden a lo que se presenta en primer lugar. Los negativos están constituídos por su negación. La negación presupone de algún modo la afirmación y, por tanto, un elemento suplementario a aquella, con lo cual se introduce una complejidad en el término que no se haya implicada en los términos que denotan rasgos positivos. (En el Anexo nº 1, Vocabulario Usual, podemos observar la mayor frecuencia de términos que denotan rasgos positivos, como ALTO, MAS, GRANDE, BUENO, LARGO, BONITO, GRUESO, ANCHO... respecto a sus correspondientes antónimos. Junto a la mayor frecuencia, en algunos casos se observa también su aparición más tardía en el tiempo). Lo mismo sucede con los términos no marcados respecto a los términos marcados. (En la teoría de los rasgos semánticos los términos marcados y no marcados se utilizan para describir y distinguir ciertos antónimos dimensionales, como "largo - corto", "mucho - poco", "más - menos",... Los términos no marcados tienen una significación más amplia, incluyen más rasgos semánticos y, por consiguiente, se utilizan con mayor frecuencia que los términos marcados a los que se reconoce como única dimensión significativa la del contraste, es decir, servir de referencia, remitirnos al polo opuesto expresado por los términos no marcados).

La correlación entre la frecuencia de las palabras y su complejidad significativa (el carácter positivo o negativo, marcado o no marcado, el uso amplio o restrictivo) permiten comprender por qué razón esas características están enlazadas a la facilidad o dificultad de la adquisición y, en último término, al carácter precoz o tardío con que aparecen en el lenguaje del niño. Podemos afirmar que la presencia de pa-

labras de alta frecuencia en el vocabulario de uso va unida a la facilidad o dificultad, a la complejidad o simplicidad semántica que esas mismas palabras entrañan.

El papel desempeñado por la frecuencia se puede extender también a los aspectos cognitivos. Lo que es más difícil cognoscitivamente (caso de lo complejo, de lo negativo respecto a lo positivo, de lo abstracto frente a lo concreto...) es de uso menos frecuente en las formulaciones verbales cotidianas. Así parece confirmarlo la observación que se ve asistida, de otra parte, por razones de otra índole como son el principio de economía y del menor esfuerzo (Zipf), la mayor éficacia de la comunicación cuando el mensaje es fácil de captar, la rapidez en la elaboración del mensaje por parte del emisor cuando este es simple...

La frecuencia, en el plano del contacto del niño con la realidad lingüística, sigue siendo un elemento clave. Las palabras que incrementan progresivamente el vocabulario del niño no podemos aislarlas de las condiciones en que se ejercen, es decir, de las ocasiones de encuentro con los objetos. La frecuencia en el contacto con los objetos y su asociación con las palabras propicia la adquisición del vocabulario, la profundización en el significado y, al mismo tiempo, marca la rapidez y dirección del desarrollo del léxico.

Para concluir diremos que los recuentos de vocabulario basados en el cálculo de frecuencias lejos de constituir un fin en sí mismos, de concluir con la aportación de datos meramente estadísticos, tienen una importancia considerable

en el estudio no sólo cuantitativo sino también cualitativo del vocabulario infantil. Por otra parte, desde un punto de vista práctico y aplicado, de un adecuado conocimiento de la formación del léxico y de su crecimiento con la edad se podrían derivar notables consecuencias para una ordenación y planificación más racional y objetiva de la enseñanza del vocabulario en la escuela. La elaboración de textos escolares, la determinación de vocabularios básicos para la enseñanza de la lectura, la consideración del vocabulario de los distintos niveles o edades como punto de partida para la ampliación y enseñanza del nuevo vocabulario, la determinación del incremento por niveles de enseñanza en función de la dirección que éste sigue con la edad..., serían algunas de las aplicaciones de inmediato.

CAPITULO V

V.- METODOLOGIA Y OBJETIVOS DE INVESTIGACION.-

Como ya se ha señalado en la revisión del capítulo anterior, las investigaciones sobre el vocabulario infantil han han respondido a diversos planteamientos y objetivos, desde que se iniciaran con carácter sistemático a finales del siglo pasado.

Las distintas alternativas de análisis han sufrido numerosas modificaciones y correcciones con el paso del tiempo. Modificaciones que han surgido de las objecciones y la crítica a que han estado sometidas por parte de lingüistas y psicólogos. La aparición de nuevas teorías lingüísticas, de nuevos modelos de estudio del comportamiento infantil y, en último término, el progresivo perfeccionamiento introducido en las técnicas de análisis y la metodología permiten que los estudios sobre el vocabulario tengan aún vigencia.

Tal como ocurriera en otros sectores de la psicología infantil, el estudio del desarrollo del vocabulario tuvo su punto de partida en la observación casual de hechos aislados que se registraron en documentos biográficos. Después, durante las dos primeras décadas del siglo actual se recopilaron gran cantidad de datos relativos a la adquisición del vocabulario desde que el niño articula la primera palabra hasta el cuarto o quinto año de vida. Sin embargo, gran parte de este material de observación directa no posee las suficientes garantías de validez científica, dada la metodología tan dispar utilizada en unas y otras investigaciones, lo limitado de las observaciones (en su mayor parte con sujetos aislados) y la subjetividad de los

informes, compilados generalmente por los padres de los sujetos estudiados.

En torno a 1930, se inician nuevos trabajos con el método biográfico. Esta vez con técnicas de registro más precisas y mayores precauciones en el tratamiento de las cuestiones de carácter lingüístico implícitas en todo recuento. No obstante siguen siendo poco representativas en cuanto al establecimiento de una normativa general acerca de los patrones de desarrollo seguidos por el vocabulario dado el escaso número de sujetos utilizados en las muestras (Low, A.A., 1936; Gregoire, A., 1937; Leopold, W.F., 1939...).

De forma paralela, algunos años antes (Smith, M. K., 1926)los trabajos sobre el vocabulario habían adoptado ya un nuevo modelo para afrontar el problema del desarrollo cuantitativo. Los recuentos absolutos de palabras, por edades, resultan materialmente imposibles en muestras amplias de sujetos pasada una determinada edad debido al fuerte incremento que se produce en el léxico. Esta dificultad planteó la necesidad de buscar modelos alternativos,distintos al registro y la observación directa. A partir de aquí, se abren dos nuevas perspectivas: los vocabularios de reconocimiento y los diccionarios infantiles de frecuencias.

En los primeros, se adopta como modelo de referencia para la explicación del desarrollo e incremento que se produce en el vocabulario, el léxico de la lengua, el vocabulario de los adultos o un extracto de ambos. Mediante la aplicación de pruebas que evalúan el reconocimiento de una lista exigua de palabras pretenden definir las dimensiones del vocabulario total disponible por el niño en cada edad.

Los diccionarios de frecuencias, por el contrario, no pretenden establecer una evaluación global acerca de las dimensiones del léxico infantil. Son cálculos relativos basados en la relación tipo/figuración ("type/token"), es decir, en la cantidad de veces que aparece una misma palabra en proporción al número total de palabras emitidas o analizadas. El vocabulario de uso propio de una edad determinada se establece a partir de la frecuencia con que ocurren las palabras.

Esta segunda perspectiva, relativamente reciente en el estudio del vocabulario infantil, entronca metodológicamente con los recuentos generales de palabras. Nuestra investigación se encuadra dentro de este tipo de trabajos. Pretende dar, simultáneamente, una respuesta al problema del desarrollo del vocabulario y definir los límites del vocabulario mismo, correspondiente a los niños comprendidos entre los seis y diez años. Participa, por ello, en la doble problemática, tanto teórica como metodológica que atañe, por una parte, a los estudios sobre el vocabulario infantil y, por otra, a los recuentos generales de vocabulario.

5.1.- Presupuestos Metodológicos.-

Sin prejuzgar el tratamiento que daremos, posteriormente, de manera puntual y en profundidad, a todos y cada uno de los aspectos que atañen a la tarea que conlleva nuestro trabajo, haremos una referencia previa, con carácter general, a las cuestiones más importantes que, desde el punto de vista metodológico, concurren en la realización de un recuento y en la determinación de un vocabulario básico.

a) Léxico y Vocabulario.-

Léxico y vocabulario son términos estrechamente vinculados entre sí. En la lingüística, el léxico hace referencia al sistema de palabras que componen una lengua. Las unidades que integran el léxico se denominan lexemas. Ahora bien, cuando estas unidades virtuales se actualizan en el discurso, reciben el nombre de vocablos.

El Vocabulario se define como el conjunto de vocablos que aparecen en una o varias muestras del lenguaje oral o escrito bien sea de un autor, de un hablante, de una comunidad lingüística o de una ciencia. Representa a una parte del léxico, ya que el léxico comprende no sólo los vocablos recogidos en un muestreo parcial del habla (en nuestro caso, de un grupo de sujetos) sino también aquellos vocablos virtuales o pasivos cuya apreciación excede las posibilidades de una observación controlada. Por consiguiente, sólo es posible cuantificar, con cierto rigor experimental el vocabulario. Nuestra ignorancia acerca de la dimensión que posee el vocabulario de unos sujetos dados nos impide efectuar una estimación real del léxico que los caracteriza.

b) Las unidades de análisis: Vocablos y Palabras.

En el lenguaje corriente, vocablo y palabra resultan ser términos sinónimos. En estadística lexical se reserva el tèrmino de vocablo ("type") para designar a cada una de las unidades que componen un vocabulario. Se consideran unidades léxicas. Los vocablos son las diferentes entradas que registra un diccionario de frecuencias. El término palabra ("token") (figuración) se refiere a todas y cada una de las unidades elementa-

les, perfectamente distinguidas por la tipografía y la escritura, que constituyen un texto. Palabra es toda ocurrencia de un vocablo cualquiera.

c) Proceso de Cuantificación.-

La relación palabra-vocablo (razón P/V), también denominada relación tipo-figuración ("type-token") se establece tras la cuantificación del vocabulario. Cuantificar un vocabulario implica realizar sucesiva o simultáneamente dos operaciones distintas:

1ª) Contar las palbras que componen la muestra. El total (P) nos dará la medida de la extensión del recuento.

2ª) Reducción de las palabras a vocablos. De esta operación resulta el total de vocablos (V), equivalente a la extensión del vocabulario.

Ambas operaciones precisan de una normativa clara que permita diferenciar las unidades de análisis (palabras y vocablos) y, al mismo tiempo, hagan operativo el recuento. No obstante, el lenguaje que en sí mismo es un todo continuo, se resiente más a nivel lexical que a cualquier otro nivel cuando intentamos atomizarlo o fragmentarlo. Es por esta razón por la que la normativa de clasificación ha dado lugar a serias controversias entre los estudiosos de la estadística léxica (Juilland, 1966). Controversias de índole esencialmente lingüística. Y esto, a sabiendas de que el español no es una lengua que ofrezce un alto porcentaje de casos dudosos en la delimi-

tación entre palabra y vocablo. En el anexo primero relatamos las normas de clasificación lexicológica utilizadas en nuestro recuento.

Si la diferenciación de la palabra plantea relativamente pocos problemas en español es porque la tradición ortográfica y tipográfica nos da una solución que, la mayor parte de las veces, es aceptable también para el lingüista. Consideramos como palabra, toda "unidad gráfica" o conjunto de signos separado de los grupos vecinos por espacios en blanco o algún signo de puntuación. Como excepciones a esta norma tendríamos aquellos casos en que una misma unidad gráfica comprende varias palabras, por ejemplo, las contracciones "al" y "del", o el caso inverso, en el que varias unidades gráficas constituyen una misma palabra, por ejemplo, las formas compuestas de los verbos y las locuciones verbales.

Una vez establecidos los criterios se inicia el recuento y registro de formas (palabras). La fase siguiente consiste en reunir bajo una misma rúbrica (vocablo)todas las palabras que representan a un mismo vocablo. Esta segunda operación es más compleja y dificultosa que la primera, particularmente, en una lengua como la española donde son frecuentes las formas homográfas debido a la polivalencia semántica y sintáctica que presentan no pocas palabras. En estos casos, unas veces será el contexto (polivalecia sintáctica) y otras, el diccionario de la lengua (polivalencia semántica) quienes ayuden a establecer el criterio que facilite la reducción a uno o más vocablos.

d) Codificación.-

Una vez realizado el recuento y obtenido un índice

alfabético de formas, se procede a la codificación y categorización de los vocablos con objeto de diferenciar los homógrafos. Esta operación consiste en clasificar los vocablos dentro de su categoría gramatical correspondiente.

e) La extensión del Vocabulario.-

Estadísticamente, el problema fundamental que plantea todo recuento es el de su representatividad. Todo vocabulario como subconjunto de un léxico, es siempre una muestra finita de éste, sin que esta afirmación lleve implícito, en un sentido estríctamente matemático, que el léxico sea una realidad infinita. Sin embargo, es imposible saber el número exacto de unidades que constituyen el léxico de un idioma y aún resulta más complejo todavía delimitar el léxico propio de una comunidad o de un determinado grupo de individuos. Sabemos que el léxico varía de un estado de lengua a otro, de un locutor a otro e incluso para un mismo locutor es diferente según el momento o las circunstancias que determinan la comunicación. La probabilidad de uso de los vocablos es, pues, muy variable. Para confirmar este extremo basta una rápida ojeada a la distribución de vocablos contenida en los recuentos ya existentes. También es cierto que las frecuencias que presentan los vocablos entre sí son muy desiguales y se requiere el recuento de una gran cantidad de palabras para conseguir una cierta estabilidad en la ordenación de las frecuencias. A medida que pregresamos en la realización de un recuento es cada vez menor el número de unidades nuevas que aparecen, pero nunca podremos estar seguros de haber agotado el léxico de un sujeto, de una comunidad o de un idioma. Ahora bien, si el vocabulario no cesa de crecer, ¿existe algún criterio que nos permita considerarlo significativamente representativo en un momento dado?

La respuesta a esta cuestión ha preocupado a diversos investigadores que intentaron profundizar en el conocimiento de la estructura (distribución de frecuencias) del vocabulario y de las leyes empíricas que la regulan (ley de Zipf, índice de Yule-Herdan, fórmula de Waring-Herdan). Sin embargo, en la actualidad, aún no tenemos a disposición (a pesar de los intentos) la fórmula que permita construir un modelo teórico consistente que dé respuesta a los problemas planteados en este punto, es decir, que permita establecer una relación precisa entre palabras (P) y vocablos (V) y, en último término, que resuelva el problema de la extensión de un vocabulario.

De hecho, hemos podido constatar en los recuentos generales las diferencias de muestra tan acusadas que sirven de base para el recuento en unas investigaciones y otras. Ciertamente, esto nos induce a pensar que los autores sostienen, cada cual de acuerdo con sus objetivos de investigación, un concepto diferente acerca de lo que es un vocabulario básico. Más adelante, volveremos de nuevo sobre esta cuestión cuando planteemos los límites que definen nuestro vocabulario. En este momento, sólo queremos resaltar la preocupación de los diversos autores por la realización de un muestreo científico, adecuado a unos objetivos, que confiera solidez al recuento, aunque a ciencia cierta no sepamos todavía cuál es el número ideal de datos ni cuál sea el mejor método para obtenerlos. Las distintas opciones quedan supeditadas en cada investigación a los objetivos que se pretendan conseguir.

5.2.- Objetivos de Investigación.-

Nuestro trabajo intenta responder a dos objetivos

complementarios entre sí:

a) Analizar la producción verbal espontánea del niño, a partir de muestras escritas, con objeto de fijar el vocabulario básico, usual, que le caracteriza. El período de edad elegido comprende, tanto niñas como niños entre los seis y los diez años. Esta edad se corresponde con los cinco primeros cursos de la E. G. B.. El estudio se llevó a cabo en el ámbito geográfico de Andalucía Oriental, es decir, en Almería, Granada, Jaén y Málaga.

Es el objetivo que ocupará la parte central de nuestra investigación dada la ausencia de evaluaciones precisas que fijen y determinen el vocabulario del niño español en estas edades.

b) En el desarrollo normal del lenguaje se ha detectado en numerosas investigaciones la presencia de importantes diferencias individuales. En concreto y en relación con el aspecto que nosotros estudiamos, el vocabulario, diversas evaluaciones han puesto de manifiesto las diferencias que se producen en el incremento debidas a la edad, el sexo de los sujetos, la zona donde habitan y la clase social a la que pertenecen.

Con objeto de replicar los resultados de estos estudios sometemos a verificación las siguientes hipótesis, de acuerdo con el diseño que más adelante presentaremos:

1ª. Entre los cinco grupos de edad señalados deberán aparecer diferencias en la producción verbal, siendo este factor (edad) el que mayor peso específico tenga en la explicación del incremento observado en el desarrollo del vocabulario.

2ª. Los resultados deberán confirmar la mayor precocidad de las niñas. El rendimiento en tareas verbales que suponen el manejo del vocabulario será superior en las niñas en relación con los niños.

3ª. Los sujetos procedentes de zonas rurales deberán manifestar una producción verbal inferior a la observada en los sujetos de zona urbana.

4ª. En relación con la clase social, el análisis de la producción verbal de los sujetos permitirá establecer un orden jerárquico entre los índices de rendimiento. Las puntuaciones promedio de la clase alta serán superiores a las de la clase media y éstas se impondrán, a su vez, sobre las de la clase baja.

CAPITULO VI

VI- LA MUESTRA.-

6.1.- Clase de muestreo.-

Las características de una muestra, la clase de muestreo y el tamaño de la misma han de ser fijados en función de los objetivos a conseguir por toda investigación. En nuestro caso, teniendo en cuenta las variables, los niveles o valores adoptados por cada una de ellas y el ámbito geográfico en que se realiza el trabajo, consideramos como procedimiento de muestreo más adecuado la realización de un muestreo estratificado con afijación óptima. Se entiende por tal el procedimiento que se sigue en la elección de una muestra procedente de una población dividida en diversas subpoblaciones o estratos (Azorín Poch, 1976).

La principal ventaja que proporciona la estratificación estriba en el aumento de precisión que se logra en la estimación de los tamaños de la muestra en sus distintos estratos. Por el contrario, este tipo de muestreo exige como requisito un conocimiento de la población más exhaustivo que el supuesto para otros muestreos aleatorios más simples.

Las variables de estratificación que hemos considerado son las siguientes: Edad - Zona y Provincia. Estas tres variables a su vez poseen distintos grados o niveles que dan lugar a los diferentes estratos, a saber:

Edad (6, 7, 8, 9 y 10 años): 5 estratos.
Zona (Urbana - Rural) : 2 estratos.
Provincia (Almería, Granada
Jaén y Málaga) : 4 estratos.
Total: 5 x 2 x 4 = 40 estratos.

Una vez determinado el número de estratos, el tamaño de cada estrato se fija a partir del N total de la muestra, es decir, el número total de sujetos que se muestrean. Para ello se precisa la adopción de un tipo de afijación o reparto que distribuya el tamaño muestral en los 40 estratos. Por consiguiente, una cuestión preliminar a la distribución de N en los 40 estratos es precisar el tamaño de N.

Ahora bien, esta cuestión está relacionada con otra que consideramos crucial en cualquier recuento de vocabulario. Se trata del número total de palabras que constituyen la base del recuento.

6.2.- Universo Léxico: Número total de palabras.-

Para la mayoría de los recuentos conocidos, el universo léxico utilizado no plantea excesivos problemas de muestreo. Al menos, desde la perspectiva de tener que elegir un determinado número de sujetos. La casi totalidad de los recuentos se limitan a muestrear palabras que aparecen en diferentes fuentes escritas, generalmente textos o cualquier otro material impreso. Pero no suelen utilizar la producción verbal de unos sujetos dados en una situación de habla, bien sea espontánea, bien sea sugerida, tanto si emplean el lenguaje

oral como el escrito.

Los criterios para la elección de fuentes son generalmente subjetivos o basados en el principio de autoridad.El elenco de autores y los textos base del recuento son elegidos por expertos entre aquellos que se consideran más relevantes para el tipo de lenguaje que se pretende estudiar: literario, científico, periodístico, de televisión...

De aquí que la problemática que presenta nuestro recuento respecto a la muestra difiera de la que normalmente plantean los recuentos tradicionales. Para éstos, la preocupación fundamental reside en la elección y justificación de un determinado número de palabras y, al mismo tiempo, la justificación de ciertos textos y autores considerados los más cualificados para alcanzar unos objetivos y de donde se extrae la muestra de palabras. Por lo común, se analizan muestras extensas (libros) del lenguaje escrito de pocos sujetos. De estas muestras se obtienen diccionarios de vocablos cuya frecuencia de uso se generaliza después para toda una lengua.

En nuestro caso, las palabras soporte del recuento no son independientes de los sujetos que las producen. Nuestros objetivos no se centran en la compilación y análisis de un vocabulario en abstracto. Pretenden la determinación de un vocabulario concreto, propio de los sujetos de una edad, y el análisis de las diferencias que puedan existir en las fuentes de producción del mismo, es decir, los grupos de sujetos que se muestrean.

En los recuentos tradicionales, por consiguiente, las fuentes del vocabulario y el número total de palabras a muestrear, no guardan una relación muy estrecha con los sujetos de donde unas y otras proceden. La representatividad de los sujetos es secundaria frente al lenguaje, a la palabra, como tal. De hecho, las razones que llevan a la conclusión de un recuento están relacionadas,no tanto con el número de sujetos muestreados, cuanto con el contenido y la estructura que presenta el recuento.

Siguiendo la nomenclatura de Ch. Muller (1973), entendemos por contenido la naturaleza de las unidades que integran un recuento. La estructura viene dada por la distribución de frecuencias. La estructura de un vocabulario comprende elementos cuantitativos simples que son el número de los vocablos y la frecuencia de cada uno de ellos. Cuando reunimos los vocablos de una misma frecuencia se forman clases de frecuencia. La estructura de un vocabulario incluye, además, las relaciones cuantitativas y cualitativas que se establecen entre los efectivos (vocablos de una misma frecuencia) de clases sucesivas de frecuencia. El conjunto de estos datos numéricos se conoce con el nombre de "distribución de frecuencias". Hemos de clarificar que los elementos cualitativos a los que se hace referencia en la estructura son la naturaleza gramatical de los vocablos y las relaciones de asociación (gramatical o semántica, paradigmática o sintagmática, sinonímica o antonímica) que existen entre estos vocablos.

Pues bien, un recuento finaliza o un determinado número de palabras-muestra se considera suficiente cuando la es

tructura o el contenido que representan los vocablos posee ciertas características de índole lingüística y/o estadístico-matemática.

Entre las características que delimitan la extensión de un recuento o hacen que éste se detenga, podemos considerar las siguientes:

a) Número de vocablos o palabras tipo:

El recuento continúa hasta que aparece cierto número de vocablos: 1000, 2000, 3000, etc. El número de palabras de la muestra no se fija hasta tanto no se han conseguido los vocablos deseados.

b) Estructura estable: el recuento se detiene cuando se consigue una distribución de frecuencias tal que la continuidad en el muestreo de palabras no modifica esencialmente el número de efectivos de las distintas clases de frecuencia.

c) Ordenación según un coeficiente decreciente de frecuencia:

Esta característica supone la concepción del vocabulario como un continuo jerarquizado compuesto por unidades ordenadas. El orden viene dado por el coeficiente de frecuencia decreciente. Cuando de un recuento surge un vocabulario con un coeficiente de frecuencia similar al que proporcionan otros recuentos y las palabras ocupan aproximadamente el mismo orden éste se da por concluído.

d) Proporción en clases de palabras:

El recuento se detiene cuando se obtienen distribuciones de vocablos con cierta proporción de nombres, adjetivos, verbos, palabras autosemánticas o léxicas frente a palabras relacionales o funcionales....

e) Razón P/V (Cociente entre número total de palabras utilizadas como muestra y número de vocablos obtenidos). Es un índice que disminuye a medida que aumenta el recuento por lo que no lo podemos considerar una medida fiable que establezca diferencias entre unos recuentos y otros, sobre todo si se han obtenido a partir de tamaños de muestra (número total de palabras) muy desiguales. Sin embargo, a pesar de esta limitación, la razón P/V refleja, por un lado el grado de riqueza o diversidad léxica que contiene un vocabulario: Por otro, la podemos considerar como un indicador de la usualidad de las palabras ya que la razón P/V lo que nos indica en realidad es la frecuencia media, en términos absolutos, que tienen los vocablos en un vocabulario dado.

Ciertamente, esta diversidad de criterios en la delimitación de la extensión de los recuentos, lo que postula es la existencia de distintas concepciones de un vocabulario básico o fundamental. Cada autor define su concepto en función de la prescripción de una o varias de las características antes mencionadas, a la estructura y contenido de su recuento. Por esta razón hay tanta variedad entre los tamaños de muestra de unos recuentos y otros.

De cualquier forma, sean cuales sean los criterios que se adopten, hemos de tener en cuenta que todo vocabulario

es una muestra limitada del léxico de una lengua o bien del léxico de un grupo de sujetos representativo de la misma en determinada situación de habla. Esta reducción implica una extensión finita, más o menos variable, de vocablos, dada la imposibilidad material de utilizar en un recuento todas las palabras que los hablantes pueden emitir. Incluso, si esto fuera posible, tales palabras no serían sino la manifestación de una parte del léxico: el léxico activo.

Es sabido que la competencia lingüística de un sujeto viene dada tanto por el léxico activo como por el léxico pasivo. Cualquiera que sea la extensión de las observaciones (muestras) tomadas de un sujeto, la lista de vocablos empleados por él no será nunca más que un subconjunto de su léxico. Sabemos, además, que muchos vocablos conocidos y comprendidos por el sujeto tienen escasa o nula probabilidad de ser utilizados normalmente; otros, el sujeto los utiliza en cuanto oyente, los comprende, pero no los emplea en cuanto locutor o escritor. Unos y otros forman parte de su léxico pasivo.

Si tenemos en cuenta todo lo expuesto con anterioridad, el número de palabras que constituyen la muestra debe ser consecuente a ciertas restricciones que definan los límites del vocabulario que se pretende conseguir.

En nuestro caso, los criterios que definen la estructura y contenido de nuestro recuento, y a partir de aquí el número de palabras necesario, son:

1º) La consecución de un número de vocablos que se

aproxime al considerado como vocabulario usual promedio de los sujetos de las edades que estudiamos. Este quedó fijado entre 5500 y 6000 vocablos.

2º) Una frecuencia media en los vocablos suficientemente alta como para que podamos considerarlos usuales.

Estamos convencidos de que tales límites encierran cierta subjetividad. De ella no carecen el resto de recuentos conocidos. La subjetividad se deriva del desconocimiento existente en estadística lingüística acerca de una ley exacta que precise la relación entre P y V. Ello hace que todas las argumentaciones que se ofrecen para justificar el tamaño de muestra (P, número de palabras) en los recuentos sean "a posteriori", una vez conocida la estructura del recuento.

"A priori", no existen unos límites estrictamente objetivos que impongan la necesidad de un determinado número de palabras para la obtención de un vocabulario básico. En otros tér-minos, no hay una ley precisa que fije la relación cuantitativa entre vocablo y palabra: cualquier cifra de vocablos que se elija impide fijar con exactitud el número de palabras necesario para conseguirla. Y no porque entre ambos (P y V) no exista relación, sino porque esta es aún mal conocida.

Está claro que V es función de P: V crece cuando lo hace P. También es cierto que V aumenta más lentamente que P, puesto que cada palabra que representa a un vocablo que ya ha aparecido infiere un retraso al crecimiento de V con relación

a P. Es evidente, que a medida que se realiza un recuento cada vez se necesita mayor número de palabras para que surja un vocablo nuevo.

Por otra parte, V nunca cesa en su crecimiento, y menos si consideramos el caso de la población infantil: Los niños aprenden cada vez mayor número de palabras nuevas.

- Número de Vocablos:

Los estudios sobre la amplitud del vocabulario infantil respecto al promedio de vocablos por edad son poco precisos una vez superados los primeros años de infancia. G. Miller (1969), por ejemplo, recogiendo datos de la literatura sobre el tema, proporciona para niños de 10 años unas cifras que oscilan entre 5000 y más de 50000 palabras.

¿A qué se deben estas diferencias? ¿Por qué hemos considerado adecuado para nuestro recuento una cifra aproximada de 6000 vocablos más cercana al límite inferior que al superior de los datos aportados por Miller?.

Para una respuesta a estas cuestiones precisamos hacer unas consideraciones sobre los distintos procedimientos y objetivos que caracterizan a los estudios sobre amplitud del vocabulario infantil.

Dos son los procedimientos básicos utilizados para determinar la amplitud del vocabulario de niños. Según adoptemos uno u otro los resultados promedio por edad diferirán. Por un lado, existen los llamados "vocabularios de reconoci-

miento"; por otro, los que podríamos denominar "vocabularios muestra".

Los "vocabularios de reconocimiento" tienen por objetivo la estimación cuantitativa global del vocabulario infantil, o del vocabulario propio de una edad concreta. Responden a la pregunta: cuántas palabras componen el vocabulario medio. Para ello, el procedimiento más frecuentemente utilizado consiste en la realización de un sistema de catas de palabras de un diccionario dado. Con la lista-test de palabras elegidas se muestrea una determinada población infantil. La tarea consiste en constatar si las palabras de la lista son reconocidas o no por los niños. Existen distintos índices para la medida del conocimiento. Con los resultados se elaboran unos puntajes cuya amplitud es proporcional a la dimensión (número de palabras) del diccionario utilizado para extraer la lista-test (Smith, 1926; Seashore-Eckerson 1940; García Hoz, 1977; Ehrlich y colb., 1978).

Los "vocabularios muestra", por el contrario presentan una perspectiva diferente. Su finalidad es ofrecer una muestra del lenguaje del niño. En estos vocabularios las cifras globales tienen un valor relativo frente a las características que presentan las palabras por sí mismas. Son muestras del léxico infantil. Ofrecen una menor precisión respecto al número total de palabras conocidas (amplitud del vocabulario), sin embargo, proporcionan una mayor información respecto a cuáles son las palabras más utilizadas, con qué frecuencia aparecen y el contexto en que se dan. De aquí que se les denomine vocabularios fundamentales, básicos o usuales. El promedio de

palabras que presentan por edad es sensiblemente inferior al que proporcionan los vocabularios de reconocimiento. Las diferencias estriban en la dificultad para llevar a cabo grandes recuentos de palabras y al coste del muestreo de un elevado número de sujetos.

Por otra parte, las diferencias también resultan obvias si consideramos el tipo de competencia lingüística que unos y otros analizan. Los vocabularios muestra utilizan como base del recuento palabras producidas por el niño. Muestrean, por tanto, sólo una parte de su léxico activo. Los vocabularios de reconocimiento muestrean un léxico mucho más amplio: palabras no sólo producidas por el sujeto sino también otras que el niño puede reconocer y comprender aunque no emplee normalmente y que forman parte, por tanto de su léxico pasivo.

Es en el contexto de los vocabularios muestra en el que inscribimos el presente trabajo. De aquí que la extensión de nuestro vocabulario (5750 vocablos) la consideremos adecuada. L.A. Averril (1956) en su obra "La vida psíquica del escolar", sitúa la amplitud media del vocabulario de niños de 6 a 10 años entre 2500 y 5500 palabras. El número de palabras que tuvimos que muestrear para la obtención de los 5750 vocablos antes mencionados fue de 258983.

- <u>La Frecuencia media ($\bar{f}$ = P/V).</u>

$$\bar{f} = \frac{258983}{5750} = 45.04$$

La frecuencia media de un vocablo en el total del recuento viene dada, como ya hemos señalado, por el cociente P/V. Esta nos indica en términos absolutos, la frecuencia media de aparición de los vocablos que integran un vocabulario. Si esta es alta, podemos asegurar que la repetición media de las palabras es alta y, por tanto, que aquellas que no hayan hecho su aparición cuando la repetición media es de 45 difícilmente podrían considerarse como usuales.

En función de estos criterios, creemos que el número de palabras elegido: 25o.ooo, es suficiente para la obtención de un vocabulario básico para niños de 6 a 10 años.

En verdad, la cifra no coincide con las 258.983 que resultaron al final del recuento. Esta es muy superior. Sin embargo, las diferencias son mínimas si tenemos en cuenta que éstas no se debieron a errores de cálculo sino a problemas de registro y recuento. Exactamente, el número de palabras previsto para los 2166 sujetos fue de 250.978, como más adelante expondremos. Ahora bien, entre estas no se consideraron como palabras diferentes ninguna de las formas enclíticas (muy abundantes entre las formas verbales), ni tampoco se desdoblaron las formas contractas tales como "al" y "del". Piensese, por ejemplo, que sólo estas dos últimas registraron más de 2.500 frecuencias que se sumaron simultáneamente a la preposición "a" y el artículo "el", por un lado, y a la preposición "de" y el artículo "el", por otro. Esto hizo que el número total de frecuencias y, por consiguiente, de palabras muestra en el registro definitivo fuese superior.

6. 3.- Tamaño de la muestra: número de sujetos.-

El número total de sujetos utilizado en la muestra lo constituyen 2166 niños, entre 6 y 10 años, pertenecientes a las zonas urbana y rural de las cuatro provincias de Andalucía Oriental: Almería, Granada, Jaén y Málaga. Dicho número corresponde a la suma de los sujetos que integran los 40 estratos de la muestra.

El criterio adoptado para la fijación del número de sujetos ha sido el número de palabras necesarias para el recuento. Una distribución proporcional de las mismas en función de la variabilidad observada en la producción verbal de los sujetos objeto del trabajo sirvió de punto de referencia para fijar el tamaño de la muestra.

En nuestro caso, la inexistencia de datos respecto a la producción verbal de los sujetos en cuestión hizo necesario el estudio de una muestra piloto, como etapa previa a la fijación de la muestra.

Este procedimiento de muestreo estratificado bietápico con afijación óptima será el que detalladamente expondremos a continuación.

6. 4.- Etapas del Muestreo.-

6.4.1.- 1ª Etapa: Muestra Piloto.

La muestra piloto la formaron 1640 sujetos de 7, 8, 9 y 10 años. De ellos 832 eran niños y 808 niñas. 1008 pertenecían a zona urbana y 632 a zona rural.

ZONA	CURSO	ALMERIA		GRANADA		JAEN		MALAGA		TOTAL
		-os	-as	-os	-as	-os	-as	-os	-as	
URB.	2º	18	9	26	20	25	37	39	34	208
	3º	31	29	39	28	31	31	52	29	270
	4º	29	33	35	30	30	33	37	28	255
	5º	37	22	36	30	30	50	33	37	275
TOTAL		115	93	146	108	116	151	161	128	1008
RUR.	2º	21	24	26	17	10	17	3	20	138
	3º	24	26	20	20	15	18	10	16	149
	4º	19	27	25	25	17	15	26	21	175
	5º	20	16	25	34	18	15	25	17	160
TOTAL		84	93	96	96	60	65	64	74	632

CUADRO 6. 1.- Distribución de la muestra piloto por Zonas, Curso, Provincias y Sexo.

La presencia de una muestra piloto en el trabajo se justifica por la necesidad de conocer determinadas características de una población dada para extraer una muestra más precisa de ella.

Nuestro interés al respecto reside en el conocimiento de la varianza que presentan los sujetos en cada estrato en cuanto a producción verbal. Esto nos va a permitir hacer una distribución proporcional más precisa. Así, aunque un estrato tenga menos sujetos que otro en la población de origen no por ello va a requerir menor representación en la muestra si la producción verbal que tales sujetos presentan es muy variable. Por el contrario, estratos de mayor tamaño pueden requerir menor representación si la dispersión que presentan es menor.

La recogida de datos de la muestra piloto se llevó a cabo durante el primer trimestre del curso 1979-80 (de Octubre a Diciembre). A continuación, se procedió al recuento del total de palabras que cada sujeto había escrito en los protocolos de prueba: un test de vocabulario de asociación libre con tiempo limitado y un texto libre o redacción. Se calcularon para cada estrato la media, la varianza y el total de palabras escritas. Se observará en los resultados que no hacemos diferenciación entre niños y niñas, ni tampoco entre las dos pruebas (test y texto libre). Los datos son globales para cada estrato.

Estos fueron los resultados:

PROV.	ZONA	CURSO	N	MEDIA	VARIANZA	PALABRAS
ALM.	URB.	2º	27	109.333	1972.159	2852
		3º	60	112.850	931.226	6771
		4º	62	130.452	1833.054	8088
		5º	59	179.932	2681.622	10616
	RUR.	2º	45	89.066	1418.276	4008
		3º	50	92.120	2118.865	4606
		4º	46	95.804	749.500	4407
		5º	36	198.220	2587.726	7136
GR.	URB.	2º	46	74.434	388.072	3424
		3º	67	110.209	1194.186	7384
		4º	65	156.661	2498.930	10183
		5º	66	188.366	3107.770	12432
	RUR.	2º	43	53.488	456.668	2300
		3º	40	95.925	676.005	3837
		4º	50	119.500	1160.129	5975
		5º	59	140.830	2108.140	8309

PROV.	ZONA	CURSO	N	MEDIA	VARIANZA	PALABRAS
JAEN	URB.	2º	62	75.080	1064.750	4655
		3º	62	108.140	754.113	6705
		4º	63	125.030	2025.808	7877
		5º	80	166.240	2540.180	13299
	RUR.	2º	27	75.037	495.516	2026
		3º	33	118.600	1369.147	3914
		4º	32	135.000	2874.500	4320
		5º	33	128.940	3190.480	4255
MA.	URB.	2º	73	84.740	1610.440	6186
		3º	81	104.457	1413.260	8461
		4º	65	139.877	2992.477	9092
		5º	70	146.070	1724.700	10225
	RUR.	2º	23	59.900	366.340	1378
		3º	26	115.500	2402.400	3003
		4º	47	186.766	4948.100	8778
		5º	42	188.190	3665.153	7904
TOTAL		32	1640			204506

CUADRO 6.2.- Media, Varianza y total de palabras por provincia, Zona y Curso en la muestra piloto.

6. 4. 2.- Tamaño de los Estratos.-

Los tamaños de los estratos se determinaron madiente el procedimiento de afijación, asignación o distribución óptima. Con este tipo de afijación, también llamada afijación de Chuprow-Neyman (Azorín Poch, 1976), el tamaño de un estrato se fija proporcionalmente al tamaño de dicho estrato en la población de origen por la variabilidad que presenta en el carácter que se estudia. La variabilidad viene dada a su vez por la varianza o dispersión que presentan los sujetos en un carácter determinado. En nuestro caso, el carácter estudiado es el vocabulario y, más concretamente, el número de palabras escritas por niños bajo determinadas condiciones de prueba.

La elección de este criterio (varianza en la producción de palabras) contribuye a aumentar la precisión en el muestreo, ya que al extraer la muestra de una población estratificada no sólo tenemos en cuenta el tamaño de los estratos en dicha población, sino también la variabilidad que éstos presentan respecto a aquello que estudiamos, es decir, la producción verbal.

Una afijación o reparto de esta naturaleza supone, pues, dos condiciones: el conocimiento del tamaño de los estratos en la población y su variabilidad en producción verbal.

La fórmula utilizada para calcular el reparto proporcional de la muestra en los diferentes estratos fué la siguiente:

$$n_h = \frac{N_h \cdot S_h}{\sum N_h S_h}$$

Siendo,

n_h = tamaño de muestra de cada estrato. Proporción de la muestra total correspondiente a cada estrato.

N_h = tamaño del estrato.

S_h = estimador de la variabilidad. $S_h = \sqrt{\sigma^2}$

La variabilidad viene dada normalmente por la cuasi-desviación estándar. Sin embargo, en ocasiones, se utiliza como estimador de la variabilidad la desviación estándar o raiz cuadrada de la varianza verdadera observada en el estrato. Prácticamente, no hay diferencias entre una y otra. No obstante, se recomienda el uso de la desviación estándar como estimador de la variabilidad en aquellos casos en los que se presentan grandes diferencias en la variabilidad de los estratos (Azorín Poch, 1976). Uno de estos casos es el que nos ocupa.

Por consiguiente para hallar n_h precisamos conocer N_h y S_h.

N_h = la población de cada estrato.

Los datos necesarios para el cálculo de N_h los tomamos de la "Estadística de la Enseñanza" para el curso 1978-79. Se trata de un resumen que confecciona anualmente la Delegación del Instituto Nacional de Estadística en el Ministerio de

Educación y Ciencia sobre los niveles de escolarización de la población española. En él se ofrecen datos generales y específicos por provincias relativos a alumnado, centros, profesores, tipo de enseñanza...

$S_h = \sqrt{\sigma^2}$, siendo σ^2 la varianza observada en la producción de palabras en los estratos de la muestra piloto. La raiz cuadrada de la varianza es la desviación estándar S_n.

En el cuadro siguiente aparece la proporción de muestra y, por tanto, el porcentaje de ésta correspondiente a cada estrato, una vez aplicada la fórmula de la afijación en cada uno de los 32 estratos.

PROV.	ZONA	CURSO	N_h	S_h	$N_h S_h$	%
AL.	URB.	2º	4596	44.410	204108.360	2.056
		3º	4727	30.516	144249.132	1.453
		4º	5001	42.814	214223.814	2.156
		5º	4780	51.744	247335.364	2.491
	RUR.	2º	4008	37.660	150941.280	1.520
		3º	3643	46.031	167690.933	1.689
		4º	3854	27.377	105510.958	1.062
		5º	3684	50.870	187405.080	1.887
GR.	URB.	2º	7909	19.699	155799.391	1.569
		3º	8001	34.557	276490.557	2.785
		4º	8007	49.990	400264.325	4.031
		5º	8153	55.747	454505.291	4.578
	RUR.	2º	7834	21.370	167412.580	1.686
		3º	7925	26.005	206089.625	2.075
		4º	7999	34.060	272445.940	2.744
		5º	8075	45.915	370763.625	3.734

PROV.	ZONA	CURSO	N_h	S_h	$N_h S_h$	%
T.	URB.	2º	6713	32.630	219045.190	2.206
		3º	6867	27.461	188574.687	1.899
		4º	6861	32.028	219744.108	2.213
		5º	7433	50.400	374623.200	3.773
	RUR.	2º	6387	22.260	142419.480	1.434
		3º	6546	37.002	242215.092	2.439
		4º	6540	53.614	350637.522	3.531
		5º	7084	56.484	400134.781	4.030
MA.	URB.	2º	16980	40.130	681407.400	6.863
		3º	16745	37.593	629494.785	6.340
		4º	16890	54.703	923933.670	9.306
		5º	16966	41.529	704587.800	7.097
	RUR.	2º	5687	19.140	108849.180	1.096
		3º	5608	49.010	274848.080	2.768
		4º	5657	70.343	397928.654	4.008
		5º	5682	60.540	343988.280	3.464
TOTAL		32			9927557.163	100.

CUADRO 6.3.- Distribución de la muestra por estratos. Proporción de muestra para cada estrato.

6.4.3.- Análisis de la muestra piloto.-

Los sujetos componentes de cada estrato en la muestra piloto fueron elegidos para calcular la variabilidad en el número de palabras escritas. Hubiese sido suficiente con una muestra piloto inferior a la elegida: se podría haber calculado la varianza de un grupo de sujetos de los cuatro niveles de edad, en zona rural y urbana, de una o dos provincias y tomarla como criterio para el resto. Este fue el procedimiento seguido con los estratos del primer nivel.

No obstante, tanto por razones del coste del muestreo, como por razones de precisión elegimos una muestra tan amplia para los primeros 32 estratos. Téngase en cuenta que los sujetos de la muestra piloto pasaron a formar parte, en su mayoría, de la muestra definitiva, y la diferencia de la muestra final para los 32 estratos (1764 niños) con relación a la muestra piloto (1640), salvo algunas modificaciones en la proporción de los estratos, fue sólo de 124 sujetos. De no haber aprovechado la muestra piloto hubiésemos tenido que duplicar el presupuesto asignado al muestreo en el proyecto.

Sin embargo, no fue esta la razón principal que nos movió a la elección de una muestra piloto tan amplia. Por el contrario, la razón fue lograr una mayor precisión y homogeneidad en la muestra. Precisión, por cuanto cada estrato tuvo su representación y calculamos su media y varianza propias. Homogeneidad, en el sentido de que la mayor parte de los sujetos fueron muestreados en un mismo periodo de tiempo (octubre diciembre), salvo aquellos cuya recogida de datos se hizo en

la segunda etapa, para completar los tamaños de los estratos que por su variabilidad requirieron mayor número de sujetos que los utilizados en la fase de muestra piloto.

Una vez hecho el recuento de las palabras escritas por los 1640 sujetos se obtuvieron 204506. Hemos de advertir que el criterio seguido para realizar este cómputo fue muy simple: se consideró como palabra a toda unidad verbal precedida o seguida en el texto de un espacio en blanco o de signos de puntuación.

La distribución de palabras por estratos aparece en el CUADRO 6.2. en su última columna. A continuación ofrecemos un resumen del mismo por niveles:

CURSO	ZONA URBANA		ZONA RURAL		TOTAL	
	Niños	Palabras	Niños	Palabras	Niños	Palabras
2º	208	17217	138	9712	346	26929
3º	270	29321	149	15360	419	44681
4º	255	35240	175	23480	430	58720
5º	275	46572	170	27604	445	74176
TOTAL	1008	128350	632	76156	1640	204506

CUADRO 6. 4.- Total de palabras en la muestra piloto.

Se observará, que en la muestra piloto los tamaños de los estratos no guardan ningún tipo de proporcionalidad entre sí. Esta se estableció una vez conocida la variabilidad que éstos presentaban en producción verbal. Esto nos lleva a

la conclusión de que el número total de palabras que arroja la muestra piloto no debe coincidir con el número de palabras que esta misma muestra arrojaría si los sujetos se hubiesen distribuído de acuerdo con las proporciones fijadas.

Las diferencias son obvias si tenemos en cuenta que los niños de los niveles más bajos (y, al mismo tiempo de menor producción verbal) muestran una mayor homogeneidad (menor varianza) y, por tanto, precisan menor representación en la muestra que aquellos niveles, -los superiores-, donde la variabilidad observada fue mayor.

Por consiguiente, una distribución proporcional de los 1640 sujetos de la muestra piloto, de acuerdo con las proporciones fijadas en el CUADRO 6.3., nos daría aproximadamente 214000 palabras.

CURSO	ZONA URBANA		ZONA RURAL		TOTAL	
	Niños	Palabras	Niños	Palabras	Niños	Palabras
2º	208	17846	94	6529	302	24375
3º	205	21993	147	15781	352	37774
4º	290	40808	187	27257	477	68065
5º	294	48755	215	33974	509	82729
TOTAL	997	129402	643	83541	1640	212943

CUADRO 6.5.- Distribución proporcional de palabras en la muestra piloto.

Como anteriormente expusimos en el punto 6.2., del

presente capítulo, el número de palabras que constituirían el universo léxico base del recuento se fijó en 250000. De las 250000 palabras, 229000 corresponden a los 32 estratos de la muestra piloto, el resto a los ocho estratos del primer nivel.

Fijado el número de palabras y la proporción de los tamaños de los estratos, para hacer una distribución apropiada sólo nos resta conocer el tamaño de la muestra total en lo que a número de sujetos se refiere.

El cálculo es fácil tomando como referencia el CUADRO 6.5.: si con 1640 sujetos conseguimos un universo léxico de 212943 palabras, para obtener 229000 precisaríamos 1764 sujetos.

Una vez conocido el tamaño de la muestra (1764 niños), los tamaños de cada estrato se determinan mediante un reparto proporcional de ésta de acuerdo con los porcentajes fijados (CUADRO 6.3.):

ESTRATO	PROV.	ZONA	CURSO	%	N
1	AL.	URB.	2º	2.056	36
2			3º	1.453	26
3			4º	2.156	38
4			5º	2.491	44
5		RUR.	2º	1.520	26
6			3º	1.689	30
7			4º	1.062	18
8			5º	1.887	34
9	GR.	URB.	2º	1.569	28
10			3º	2.785	50
11			4º	4.031	70
12			5º	4.578	80
13		RUR.	2º	1.686	30
14			3º	2.075	36
15			4º	2.744	48
16			5º	3.734	66

ESTRATO	PROV.	ZONA	CURSO	%	N
17	J.	URB.	2º	2.206	38
18			3º	1.899	34
19			4º	2.213	40
20			5º	3.773	66
21		RUR.	2º	1.434	26
22			3º	2.439	44
23			4º	3.531	62
24			5º	4.030	70
25	MA.	URB.	2º	6.863	122
26			3º	6.340	112
27			4º	9.306	164
28			5º	7.097	126
29		RUR.	2º	1.096	20
30			3º	2.768	48
31			4º	4.008	70
32			5º	3.464	62
TOTAL				100.	1764

CUADRO 6.6.- Distribución de la muestra por estratos.

6. 4. 4.- Los estratos del primer nivel.-

Se habrá observado que el número total de estratos o grupos representados en la muestra piloto es de 32 y no de 40 como habíamos previsto. Los estratos que restan corresponden al nivel de primer curso (seis años) en las dos zonas de las cuatro provincias (1 x 2 x 4 = 8).

De los sujetos de este nivel no se registraron datos en la primera etapa del muestreo (octubre - diciembre, 1979). Estos se tomaron durante la segunda etapa (enero - marzo, 1980).

La razón que nos llevó a tomar esta decisión con los sujetos de este nivel fue la constatación del escaso o nulo dominio de la escritura que muestran los niños de seis años en el primer trimestre del curso. La posibilidad de respuesta a los protocolos de prueba era muy limitada por parte de estos niños salvo en un reducido número de alumnos que se iniciaron ya en la escritura en preescolar. Hemos de recordar que la enseñanza obligatoria, la E.G.B., se inicia precisamente con seis años, y la Preescolar no está lo suficientemente extendida aún como para atender en su totalidad los niveles de 4 y 5 años. Esta deficiencia es mayor en las zonas rurales. Pensamos, pues, que durante la segunda fase del muestreo estarían en mejor disposición para responder de una forma más adecuada. Y así sucedió.

No obstante, hemos de señalar por lo que se refiere al muestreo, que el procedimiento seguido con estos ocho es-

tratos del primer nivel es el mismo que hemos realizado con los 32 estratos restantes. Sin embargo, el desfase producido motivó que el muestreo de estos 8 estratos apareciera diferenciado del resto.

a) La Muestra Piloto:

Estuvo constituída por 92 niños de Almería y Granada para la zona urbana, y 71 de Almería y Jaén para la zona rural.

ZONA	PROV.	NIÑOS	NIÑAS	TOTAL
URB.	AL.	26	17	43
	GR.	31	18	49
RUR.	AL.	14	19	33
	J.	22	16	38
TOTAL	4	93	70	163

CUADRO 6. 7.- Muestra piloto del primer nivel (1er curso: 6 años).

Se hizo el recuento de palabras, siendo el total de 8885. Las medias y varianzas por estrato fueron las siguientes:

ZONA	PROV.	N	$\bar{X}$		PALABRAS
URB.	AL.	43	63.35	43.31	2724
	GR.	49	68.22	28.64	3343
RUR.	AL.	33	36.24	18.09	1196
	J.	38	42.68	15.90	1622
TOTAL	4	163			8885

CUADRO 6. 8.- Media, varianza y total de palabras en los estratos del primer nivel.

b). Tamaño de los estratos:

El tamaño se fija con el conocimiento previo de N_h y S_h.

Según podemos observar en el CUADRO 6.7. la muestra piloto del primer nivel proporciona datos relativos a cuatro de los ocho estratos que lo integran. Esto indica que no todos los estratos estuvieron representados en la muestra piloto tal como sucedió con los 32 estratos analizados anteriormente. Ahora bien, si el objeto de la muestra piloto es calcular la variabilidad que presenta una determinada población en un carácter dado, la muestra elegida es suficiente para ello. Las varianzas de los estratos no representados las podemos calcular a partir de las varianzas de los estratos representados; siendo aquellas unas varianzas ponderadas obtenidas a partir de las varianzas observadas.

Así pues, las varianzas para los estratos del primer nivel son:

- Para Almería y Granada, Zona Urbana, las obtenidas en la muestra piloto con sujetos de primer curso (6 años) en dichas provincias.

- Para Almería y Jaén, Zona Rural, las obtenidas en la muestra piloto con sujetos de dichas provincias.

- Para Jaén y Málaga, Zona Urbana, calculamos las varianzas a partir de las varianzas observadas en los sujetos de Zona Urbana de las provincias de Almería y Granada, según la siguiente expresión:

$$6^2 = \frac{\sum (6_i^2 + \bar{x}_i^2)\, n}{\sum n_i} - \left(\frac{\sum n_i \bar{x}_i}{\sum n_i} \right)^2$$

Tomando los datos del CUADRO 6. 8. y haciendo las correspondientes sustituciones, tenemos que:

$$6^2 = \frac{49\ (820.09 + 68.22^2) + (1875.85 + 63.35^2)\ 43}{49 + 43} -$$

$$- \left(\frac{49 \times 68.22 + 43 \times 63.35}{49 + 43} \right)^2 = 1319.464$$

$$S_h = \sqrt{6^2} = 36.324$$

- Para Granada y Málaga, Zona Rural, calculamos las varianzas de igual forma, es decir, a partir de las varianzas observadas en los sujetos de zona rural de las provincias de Almería y Jaén:

$$6^2 = \frac{33\ (327.212 + 36.24^2) + (252.85 + 42.68^2)\ 38}{33 + 38} -$$

$$- \left(\frac{33 \times 36.24 + 38 \times 42.68}{33 + 38} \right)^2 = 297.64$$

$$S_h = \sqrt{6^2} = 17.25$$

Una vez obtenidas las varianzas y sacados los N_h de la Estadística de la Enseñanza, procedimos al cálculo de la proporción de muestra correspondiente a cada estrato mediante la afijación de Chupow-Neyman: En la columna (%) aparecen los resultados.

ZONA	PROV.	N_h	S_h	$N_h S_h$	%
URB.	AL.	5331	43.31	230885.61	16.41
	GR.	8367	28.64	239630.28	17.04
	J.	7144	36.32	259470.08	18.45
	MA.	18623	36.32	676387.36	48.10
TOTAL	4			1406373.33	100.
RUR.	AL	4093	18.09	74042.37	17.10
	GR.	8288	17.25	142968.00	33.03
	J.	6809	15.90	108263.10	25.01
	MA.	6237	17.25	107588.25	24.85
TOTAL	4			432861.72	100.

CUADRO 6.9.- Proporción de muestra para los 8 estratos del primer nivel (6 años), según la zona.

c). Tamaño del universo léxico y número de sujetos:

El número de palabras que se fijó para estos estratos fue de 22000. El cálculo de sujetos necesarios para obtener tal número de palabras se realizó de la misma forma que en los 32 estratos iniciales.

Así, procedimos a la distribución de los 163 sujetos-muestra en los ocho estratos según la proporción observada (CUADRO 6.11) obtuvimos 8894 palabras.

En función de este dato, calculamos proporcionalmente el número de sujetos necesarios para conseguir las 22000 palabras fijadas. El resultado fue de 402. (En el cálculo redondeamos la cifra de 8894 en 9000).

PROV.	ZONA URBANA			ZONA RURAL			TOTAL	
	N	$\bar{x}$	Palabras	N	$\bar{x}$	Palabras	N	Palabras
AL.	15	63.35	950	12	36.24	435	27	1385
GR.	16	68.24	1092	23	39.69	913	39	2005
J.	17	65.94	1121	18	42.68	768	35	1889
MA.	44	65.94	2901	18	39.69	714	62	3615
TOTAL	92		6064	71		2830	163	8894

CUADRO 6. 10.- Distribución proporcional de palabras en la muestra del primer nivel.

Una vez conocido el tamaño de la muestra, llevamos a cabo la distribución de acuerdo con la proporción observada. De esta forma, precisamos el tamaño de muestra para cada estrato:

ESTRATO	ZONA	PROV.	%	N
1	URB.	AL.	16.41	40
2		GR.	17.04	42
3		J.	18.45	44
4		MA.	48.10	116
5	RUR.	AL.	17.10	28
6		GR.	33.03	52
7		J.	25.01	40
8		MA.	24.85	40
TOTAL				402

CUADRO 6. 11.- Tamaño de muestra en los estratos del primer nivel.

6.4.5.- 2ª Etapa: La Muestra.

La segunda etapa del muestreo se llevó a cabo entre los meses de Enero y Marzo de 1980. El objetivo de esta segunda fase fue reajustar los estratos a los tamaños previamente fijados. El trabajo fue semejante al que realizamos en la primera etapa: se continuó la recogida de datos de niños pertenecientes a estratos que por su tamaño requerían una representatividad mayor que la que tuvieron en la muestra piloto.

De esta forma se completó la muestra que quedó definitivamente, así:

ESTRATO	PROV.	ZONA	CURSO	NIÑOS	NIÑAS	TOTAL
1	AL.	URB.	1º	20	20	40
2			2º	18	18	36
3			3º	13	13	26
4			4º	19	19	38
5			5º	22	22	44
6		RUR.	1º	14	14	28
7			2º	13	13	26
8			3º	15	15	30
9			4º	9	9	18
10			5º	17	17	34
11	GR.	URB.	1º	21	21	42
12			2º	14	14	28
13			3º	25	25	50
14			4º	35	35	70
15			5º	40	40	80
16		RUR.	1º	26	26	52
17			2º	15	15	30
18			3º	18	18	36
19			4º	24	24	48
20			5º	33	33	66

ESTRATO	PROV.	ZONA	CURSO	NIÑOS	NIÑAS	TOTAL
21	J.	URB.	1º	22	22	44
22			2º	19	19	38
23			3º	17	17	34
24			4º	20	20	40
25			5º	33	33	66
26		RUR.	1º	20	20	40
27			2º	13	13	26
28			3º	22	22	44
29			4º	31	31	62
30			5º	35	35	70
31	M.A.	URB.	1º	58	58	116
32			2º	61	61	122
33			3º	56	56	112
34			4º	82	82	164
35			5º	63	63	126
36		RUR.	1º	20	20	40
37			2º	10	10	20
38			3º	24	24	48
39			4º	35	35	70
40			5º	31	31	62
TOTAL			40	1083	1083	2166

CUADRO 6. 12.- Distribución de los sujetos de la Muestra por Estratos, Provincia, Zona, Curso y Sexo.

En las dos etapas del Muestreo participaron niños y niñas de 120 unidades (clases) de los cinco primeros niveles de E.G.B. en las cuatro provincias.

Los Centros Educativos que colaboraron en la presente investigación fueron los siguientes:

ALMERIA

ZONA URBANA:

C.N. de Prácticas Masculino de Almería
C.N. de Prácticas Femenino de Almería
Colegio de la "Compañía de María" de Almería

ZONA RURAL:

C.N. Mixto de Fiñana (Almería)

GRANADA

ZONA URBANA:

C.N. de Prácticas Masculino de Granada
C.N. de Prácticas Femenino de Granada
Colegio de la "Compañía de María" de Granada
C.N. "Reina Fabiola" de Motril (Granada)

ZONA RURAL:

C.N. Mixto de Moraleda de Zafayona (Gr.)
Escuela Unitaria Mixta de Moclín (Granada)
C.N. Mixto "Antonio Machado" Puerto Lope (Gr.)

JAEN

ZONA URBANA:

C.N. de Prácticas Masculino de Jaén.

C.N. "Virgen de la Capilla" de Jaén.

C.N. "Virgen de Guadalupe" de Ubeda (Jaén)

C.N. "Sagrada Familia" de Ubeda (Jaén)

ZONA RURAL:

C.N. Mixto "Virgen de la Fuensanta" de Huelma (Jaén)

C.N. "Víctor García Hoz" de Beas de Segura (Jaén).

MALAGA

ZONA URBANA:

C.N. de Prácticas Masculino de Málaga.

C.N. de Prácticas Femenino de Málaga.

C.N. "Isaac Albéniz" de Málaga.

Academia de "Santa Teresa" de Málaga.

C.N. Mixto "Miguel de Cervantes" de Ronda (Málaga)

ZONA RURAL:

Escuela Hogar "Sagrado Corazón" de Antequera (Málaga)

Escuela Hogar "El Castillo" de Ronda (Má.)

C.N. Mixto "S. Sebastián de Archidona (Ma.)

C.N. Mixto "S.Fernando" de Archidona (Ma.)

CAPITULO VII

VII.- LA RECOGIDA DE DATOS.-

7.1.- Lenguaje oral versus-lenguaje escrito.-

Los protocolos de prueba utilizados para la recogida de datos del vocabulario fueron dos: un test de asociación libre de palabras y una redacción o texto libre. Además, se pasó un pequeño cuestionario para una mejor identificación y control de las variables independientes, particularmente de la clase social de los sujetos.

El vocabulario se extrajo, por consiguiente, a partir de muestras del lenguaje escrito. ¿Por qué elegimos este lenguaje y no el oral? ¿Existen diferencias entre uno y otro?.

En principio, las ventajas del lenguaje escrito frente al lenguaje oral son obvias si nos atenemos al sistema de registro. La recogida de muestras orales (conversaciones), particularmente en niños, se ve dificultada en primer lugar, por la falta de medios técnicos y condiciones adecuadas para llevarla a cabo. En segundo lugar, por la pérdida de tiempo que supone la observación de situaciones espontáneas de habla. Y en el caso de utilizar estímulos o temas-centro de interés que controlen la situación y provoquen el habla, corremos un doble riesgo: o bien el niño (sobre todo si es pequeño) se inhibe y no habla ante el experimentador, o bien parcializamos el lenguaje de tal manera que condicionamos en exceso la es-

tructura de frecuencias del vocabulario.

Por otra parte, la elección del lenguaje escrito no se hizo sólo por las ventajas que respecto a la recogida de datos proporciona en tiempo y economía. Influyeron otras características que son propias de este lenguaje y consonantes con los objetivos que perseguimos. Nuestro objetivo, ya lo hemos expuesto, consiste en ofrecer una muestra del lenguaje del niño.

Ciertamente, la forma ideal de conseguirlo hubiese consistido en el registro de conversaciones espontáneas mantenidas por los niños en diversas situaciones. No obstante, ya hemos visto que esto resulta difícil, tanto técnica como económicamente, con una muestra amplia de sujetos. De aquí, nuestra decisión por el lenguaje escrito, a sabiendas de las diferencias que hay entre un tipo de lenguaje y el otro. Diferencias que tienen una clara fundamentación psicológica (Luria, 1980), pero que no impiden a cualquiera de ellos ser el soporte válido del vocabulario que pretendemos.

El lenguaje escrito es un lenguaje sin interlocutor, proyectado completamente por el propio sujeto que escribe. No contiene casi ningún otro medio de expresión complementario, no verbal. En cuanto a su origen, aparece como resultado de un aprendizaje especial que comienza con el dominio consciente de la escritura. El lenguaje oral, por el contrario, se forma en el proceso natural de comunicación que, desde el nacimiento, se entabla entre el niño y el adulto. Posee un carácter simpráxico -de vinculaciòn con la situación

práctica- que facilita la transmisión y comprensión de las palabras. En este sentido, podríamos afirmar que el lenguaje oral es más vivo. pero, al mismo tiempo, "menos" voluntario y consciente que el lenguaje escrito, al menos hasta que éste se automatiza.

El sujeto que escribe realiza un mayor esfuerzo para poderse comunicar y hacerse comprender. Este esfuerzo se traduce en una búsqueda voluntaria y consciente de los términos más adecuados, y precisa un mejor conocimiento de las normas que relacionan unas palabras con otras, hasta tal punto que deben ser completamente suficientes para expresar lo que se quiere transmitir. Así pues, el uso del lenguaje escrito supone una construcción consciente de las frases y de sus elementos mínimos que son las palabras. Este hecho, evidentemente, está mediatizado no sólo por el dominio de las técnicas de la escritura sino también del léxico y la sintaxis.

El léxico y, consecuentemente el vocabulario, en el lenguaje escrito es más consistente que en el lenguaje oral. El lenguaje escrito implica una selección de palabras más consciente y cuidada que la que conlleva normalmente el lenguaje oral. El lenguaje oral es más automático, más fluído y se suele apoyar en otros medios de expresión no verbal tales como los gestos, la entonación, las reacciones del interlocutor..., que facilitan la comunicación.

Como podemos observar, la estructura de uno y otro lenguaje son diferentes. Tanto uno como otro los consideramos válidos para conseguir nuestros objetivos. Con la elec-

ción del lenguaje escrito lo que perdemos en naturalidad, espontaneidad y viveza lo compensamos con la mayor consistencia de la palabra escrita.

7.2. Producción espontánea versus producción sugerida.-

Tanto el test como la redacción, tiene como objetivo obtener palabras producidas por el niño de una forma libre y espontánea. De ahí que desistiéramos de otros procedimientos para determinar el vocabulario.

En otras investigaciones (Armayor y colab., 1975; 1976; Díaz Castañón y colb., 1975) además de las pruebas de producción libre se utilizan otras en las que se proporcionan al sujeto estímulos que desencadenan su producción verbal. Dichos estímulos suelen girar en torno a temas o centros de interés familiares al niño: el campo, los animales, la ciudad, la casa, la escuela, el juego, los medios de transporte, el vestido... El objeto de estos temas estímulo es obtener un vocabulario más rico y diversificado. Los temas se reparten proporcionalmente entre los sujetos de la muestra de tal manera que todos los temas son tratados por un mismo número de sujetos.

No obstante, creemos que con una muestra tan amplia como la que hemos elegido, en primer lugar, la variedad en temas de la producción espontánea iguala y supera la de aquellos otros procedimientos donde se sugieren los centros de interés ya que estos necesariamente han de ser limitados

por razones obvicas de muestreo. En segundo lugar, y es más importante, a esto hemos de añadir las diferencias que se introducen en la estructura del recuento. Nos referimos a diferencias en la frecuencia de las palabras. El utilizar determinados centros de interés tratados por un mismo número de niños implica imponer una estructura al vocabulario que en realidad no tiene, o cuando menos que manipulamos artificialmente al proporcionar una mayor homogeneidad a la frecuencia de las palabras relativas a dichos temas.

Por el contrario, si analizamos la producción verbal libre observamos una mayor desigualdad en el tratamiento de los temas, pero esta desigualdad la imponen las preferencias del niño. Esto nos indica que tanto el contenido como la estructura del vocabulario obtenido responde más adecuadamente al vocabulario real.

Este argumento nos hizo desistir de la sugerencia de temas de interés. Por estas razones desistimos de utilizar en nuestro trabajo temas de interés o producción verbal sugerida.

7. 3.- Los Protocolos de Prueba.-

Ya hemos hecho referencia, con anterioridad, a las pruebas utilizadas en la recogida de datos: el test y la redacción. Ambas, son pruebas de fácil y rápida ejecución, al alcance de los niños de los cinco niveles de edad. Las dos pruebas se pasaron por el mismo experimentador a todos los su

jetos de la muestra.

7.3.1.- El test de asociación libre.-

La prueba consistió en escribir palabras durante un tiempo determinado. Previamente, con una pequeña muestra de sujetos se hicieron distintas pasaciones para fijar el tiempo de duración y el formato de la hoja de respuestas.

INSTRUCCIONES: (Se dieron verbalmente).

"Rellena los datos que aparecen en la parte superior de la hoja: Nombre y Apellidos, Curso y Ciudad. (Pausa).

A continuación, cuando se indique, vas a escribir todas las palabras que quieras. Escribirás una en cada renglón, procurando no repetirlas: todas deben ser diferentes.

Pueden ser: nombres, adjetivos, verbos, adverbios, pronombres, artículos, preposiciones y conjunciones. Los únicos que no se pueden escribir son los nombres propios. Por ejemplo, vuestro nombre, el de vuestros amigos, los nombres de las ciudades, de los ríos... Estos no se escriben. Podeis escribir todas las demás palabras que se os ocurran. ¿Comprendido?.

Seguidamente, para cerciorarnos de que han comprendido (particularmente con los niños de 6 y 7 años), se hace un ejemplo.

Fijaos en el primer renglón, el que tiene delante un 1; pensad una palabra, cada una la que quiera, y escribidla en ese renglón. (Pausa). ¿Está? Pues bien, cuando se indique vais a continuar escribiendo palabras una debajo de otra. No olvideis que pueden ser nombres (los propios, no), adjetivos, verbos, adverbios, pronombres, artículos, preposiciones y conjunciones, y no se pueden repetir. Todas las palabras han de ser distintas."

TIEMPO: 4 minutos.

Este tiempo se mantuvo con los sujetos de todos los niveles, excepto con los del primer nivel, los niños de 6 años. En este caso, dada la mayor dificultad que presentaban para escribir el tiempo se prolongó 2 minutos más, siendo, pues, 6 los minutos de duración de esta prueba en este nivel.

7.3.2.- Redacción (texto libre).

Se proporcionó a cada niño una hoja con un ecabezamiento en que se pedían los mismos datos de identificación que en la prueba anterior.

INSTRUCCIONES:

"En primer lugar, vais a rellenar los datos que arriba se indican. Cada uno deberá escribir su NOmbre y Apellidos, Curso y Ciudad. (Pausa).

A continuación, vais a escribir una redacción, un texto libre sobre lo que vosotros querais. Lo primero que teneis que hacer es pensar el título, el tema sobre el que vais a escribir. Una vez hecho esto, donde pone TITULO escribir el tema que hayais elegido y comenzais. Es importante que cada uno trabaje en lo suyo y no se fije en lo que hacen los demás.

Ahora no teneis un tiempo fijo, como en la prueba anterior, sino que podeis escribir todo cuanto querais utilizando para ello la hoja por las dos caras."

TIEMPO:

Como ya hemos indicado esta prueba no fue cronometrada. Sin embargo, el tiempo de realización osciló en torno a los 20 minutos. Durante este tiempo, la tarea del experimentador se limitó a controlar el trabajo del alumno de modo que este resultara personal, sin ningún tipo de recursos: copia, textos... En este sentido, en las instrucciones se evitó dar ningún tipo de pistas o ideas que pudie-

sen sugerir un tema concreto, con la finalidad de recoger el vocabulario presente en el mayor número de temas de interés para el niño.

7.3.3.- Cuestionario o Ficha-control.

Más que de un cuestionario se trata de una ficha-control que se adjuntaba a las dos pruebas. Normalmente era cumplimentada por los mismos y revisada posteriormente por el maestro tutor que acababa de rellenarla o rectificaba datos incorrectos. En los cursos inferiores (6 y 7 años), esta tarea la realizó el tutor.

La finalidad de la ficha fue recabar datos fidedignos en función de los cuales poder establecer un mejor control sobre las variables y, en particular, sobre la variable clase social.

7.4.- Fuentes del Vocabulario.-

Finalizada la recogida de datos y antes de iniciar el recuento de palabras, hicimos un análisis de los temas que presentaban las redacciones o textos libres. El objeto de este análisis fue precisar, al menos de una forma descriptiva, cuáles podrían ser las fuentes del vocabulario, a qué temática respondía y en qué contexto habíamos de inscribirlo.

Esto mismo no se pudo hacer con las palabras del

test de asociación libre ya que éstas carecían de relaciones sintagmáticas propias, es decir, no aparecían en contexto alguno puesto que presentaban una alineación sin posibilidad de encadenamiento claro entre ellas.

Por otra parte, la aportación que hizo el test de asociación libre al total de palabras registrado fue muy inferior a la que prestaron las redacciones: aproximadamente de una cuarta parte. En la muestra piloto de los 32 primeros estratos, por ejemplo, de las 204.506 palabras que se contabilizaron, sólo 50.819 (menos del 25%) correspondían al test de asociación libre.

Este análisis nos permitió, además, constatar ciertas diferencias en los temas de interés más frecuentemente elegidos por edades, sexo, zona y clase social. Diferencias que ponen de manifiesto, al menos superficialmente, tendencias de distinto signo entre los niños y las niñas, la zona urbana y la rural, y la clase social alta, respecto a la media y la baja.

Estas diferencias podrían estar relacionadas con la calidad y riqueza del vocabulario. Sin embargo, nuestras conclusiones no nos permiten más que reflejar una tendencia cuya confirmación hubiese requerido de un análisis más profundo.

La clasificación de los temas se hizo agrupando bajo un mismo epígrafe aquellos textos que presentaban un mismo título o un mismo contenido. Tuvimos dos excepciones a este

criterio general de clasificación. En primer lugar, la de aquellos sujetos que en la prueba de redacción escribieron palabras sueltas sin llegar a formar una sola frase; y, en segundo lugar, la de aquellos cuyo contenido no presentaba la suficiente homogeneidad que permitiera clasificarlas dentro de un tema: escribieron frases sueltas sin relación con un tema concreto.

Los primeros, -sólo algunos niños de 6 años-, ponen en evidencia la dificultad que aún tienen muchos niños para escribir correctamente en este nivel, y el desconocimiento de unas normas mínimas para relacionar unas palabras con otras en la escritura. Estos niños convirtieron el texto libre en una prolongación de la prueba de asociación libre.

Los segundos se limitaron a construir frases con sentido pero sin apenas relación unas con otras. Generalmente, el contenido hacía relación a distintas situaciones vividas en su medio ambiente familiar y escolar. En la clasificación las denominamos frases sobre medio ambiente. También observamos aquí, aunque en menor grado, la dificultad del niño para hacer consciente en la escritura unas normas sintácticas complejas que le sirvan de soporte para el desarrollo de una idea, de un tema o de una pequeña narración. Al menor con el mismo nivel de complejidad con que es capaz de exponerlas oralmente. En este nivel se encontraban casi el 50% de los niños y niñas de 6 años, 35 de 7 años y 6 de 8 años.

En el CUADRO 7.1., aparece el elenco de temas y la frecuencia con que fueron tratados por Cursos, Zona y Sexo.

Nº	TEMAS	1º				2º				3º				4º				5º					
		U.		R.		U.		R.		U.		R.		U.		R.		U.		R.			
		-os	-as	-os	-as	-os	-as	-os	-as	-os	-as	-os	-as	-os	-as	-os	-as	-os	-as	-os	-as	TOTAL	%
1	Cuentos	7	20	6	4	62	45	18	19	21	17	18	18	45	24	24	20	40	28	16	18	470	21.70
2	Frases sobre Medio - Ambiente	51	61	26	33	8	8	12	7	6	-	-	-	-	-	-	-	-	-	-	-	212	9.79
3	Los Animales	12	5	2	2	6	10	5	2	23	15	8	1	20	16	11	8	14	8	16	5	189	8.73
4	La escuela	7	2	5	4	2	5	-	4	4	6	5	12	5	20	7	14	9	10	8	13	142	6.55
5	La ciudad - el pueblo	-	-	-	2	2	5	1	4	6	7	2	8	17	10	8	3	24	16	7	11	133	6.14
6	El Campo	1	1	1	2	6	8	2	-	5	6	14	5	4	6	7	8	11	6	18	15	126	5.82
7	La Naturaleza	4	7	2	-	4	2	-	1	7	6	7	4	14	11	2	5	12	6	9	7	110	5.08
8	Otros temas originales	-	-	-	-	1	-	-	-	5	11	1	4	14	13	10	4	13	7	11	13	107	4.94
9	Vacaciones	12	9	9	12	5	6	4	2	2	1	1	-	4	6	5	9	7	3	3	3	103	4.75
10	Las flores y las plantas	-	2	1	2	5	6	-	1	-	14	5	10	6	9	8	14	-	11	1	6	101	4.66
11	La familia y la casa	14	2	2	2	2	4	-	2	4	8	-	4	4	8	3	4	3	18	2	1	87	4.02
12	El Mar y la playa	1	-	1	-	1	3	2	3	8	4	3	-	8	11	3	-	4	5	2	3	62	2.86
13	La Primavera	-	1	1	4	-	1	1	1	-	5	1	3	1	6	-	4	1	10	6	7	53	2.45
14	El Verano	2	-	1	2	4	-	2	3	-	1	2	4	-	1	3	2	2	9	5	5	48	2.22
15	Navidad - Reyes	-	-	1	3	-	4	-	-	-	2	-	1	2	8	-	2	7	9	2	5	46	2.12
16	Medios de Locomoción	-	2	3	1	2	1	-	-	11	2	-	1	5	2	3	-	6	-	2	-	41	1.89
17	Palabras sueltas	5	5	19	6	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	35	1.62
18	Deportes - Juegos - Juguetes	-	-	-	1	1	1	3	1	6	3	-	-	6	2	2	-	2	2	4	-	34	1.57
19	Profesiones y Oficios	-	-	-	-	1	2	-	1	3	-	11	2	1	-	1	1	1	3	3	-	30	1.39
20	El Otoño	8	-	-	-	-	1	-	-	-	2	-	-	-	1	-	-	1	5	1	1	20	0.92
21	El Invierno	-	1	-	-	-	-	1	-	-	1	1	2	-	2	2	1	1	2	-	3	17	0.78
	TOTAL	124	118	80	80	112	112	51	51	111	111	79	79	156	156	99	99	158	158	116	116	2166	100.

CUADRO 7 . 1 . Fuentes del Vocabulario: tema de las redacciones por Curso, Zona y Sexo.

Un análisis detallado de los resultados pone de manifiesto que en las preferencias de unos temas y otros se establecen diferencias notorias según el sexo, la zona y la clase social de los sujetos.

CUADRO 7.2.- Frecuencia de los temas y rango que ocupan según el Sexo, la Zona y la Clase Social.

En el primer lugar de la clasificación aparecen los Cuentos. Bajo este epígrafe recopilamos todos los textos que tenían un contenido fantástico, de ciencia ficción, o historias populares de tran tradición infantil. En esta categoría no hay grandes diferencias si bien el porcentaje es mayor en niños que en niñas, en la zona urbana respecto a la zona rural y en la clase alta sobre la media y la baja. No obstante en todos ocupa el primer rango. Otro tanto sucede en el tema "Animales", que aparece en segundo lugar.

A partir de aquí las diferencias se diversifican. Vemos como las preferencias del niño se inclinan más por "el campo", "los temas originales", "la naturaleza", "el mar y la playa", "los medios de locomoción" y "las profesiones y oficios". Mientras que las niñas manifiestan un mayor interés por "la escuela", "las flores y las plantas", "la familia y la casa","la primavera" y las fiestas hogareñas como son las de "navidad y reyes". Es evidente que los datos no permiten más que una constatación superficial de los intereses, pero ésta no dista mucho de las preferencias que popularmente se asignan al niño y a la niña.

Nº	TEMAS	SEXO				ZONA				CLASE SOCIAL					
		Niños		Niñas		Urb.		Rur.		Alta		Media		Baja	
		N	R	N	R	N	R	N	R	N	R	N	R	N	R
1	Cuentos	257	1º	213	1º	309	1º	161	1º	46	1º	86	1º	338	1º
2	Frases sobre medio ambiente	103	3º	109	2º	134	2º	78	2º	18	2º	46	2º	148	2º
3	Los Animales	117	2º	72	5º	129	3º	60	5º	17	3º	37	3º	135	3º
4	La escuela	52	8º	90	3º	70	6º	72	3º	9	6º	25	7º	108	4º
5	La Ciudad - el Pueblo	67	5º	66	6º	87	4º	46	8º	5	9º	27	4º	101	6º
6	El Campo	69	4º	57	7º	54	10º	72	4º	5	9º	17	11º	104	5º
7	La Naturaleza	61	6º	49	11º	73	5º	37	10º	3	16º	27	5º	80	7º
8	Otros temas originales	55	7º	52	9º	64	8º	43	9º	14	4º	23	8º	70	1o º
9	Vacaciones	52	9º	51	10º	55	9º	48	6º	4	12º	20	9º	79	8º
10	Las flores y las plantas	26	13º	75	4º	53	11º	48	6º	8	7º	18	10º	75	9º
11	La Familia y la Casa	34	10º	53	8º	67	7º	20	14º	11	5º	26	6º	50	11º
12	El Mar y la Playa	33	11º	29	14º	45	12º	17	16º	4	12º	8	14º	50	12º
13	La Primavera	11	19º	42	12º	25	15º	28	12º	4	12º	6	16º	43	13º
14	El Verano	21	16º	27	15º	19	17º	29	11º	3	16º	6	17º	39	14º
15	Navidad - Reyes	12	18º	34	13º	32	13º	14	17º	6	8º	9	13º	31	15º
16	Medios de Locomoción	32	12º	9	20º	31	14º	10	19º	3	16º	11	12º	27	17º
17	Palabras sueltas	24	14º	11	17º	10	20º	25	13º	1	20º	5	19º	29	16º
18	Deportes - juegos - juguetes	24	14º	10	18º	23	16º	11	18º	4	12º	7	15º	23	18º
19	Profesiones y Oficios	21	16º	9	21º	11	19º	19	15º	3	16º	4	20º	23	19º
20	El Otoño	10	20º	10	18º	18	18º	2	21º	5	9º	6	18º	9	21º
21	El Invierno	5	21º	12	16º	7	21º	10	19º	1	20º	4	21º	12	20º
	TOTAL	1086		1080		1316		850		174		418		1574	

CUADRO 7. 2.- Frecuencia de los temas y rango que ocupan según el Sexo, la Zona y la Clase Social.

Lo mismo podríamos decir de las diferencias entre zonas. Los niños urbanos centran sus preferencias en torno a su propio ambiente: "la ciudad", "la familia y la casa", "el mar y la playa", "los medios de locomoción". Mientras que en la zona rural todo parece girar en torno a límites más estrechos y reducidos: "la escuela", "el campo", "la naturaleza", "las plantas" y "las vacaciones".

La clase social también establece sus diferencias, particularmente, si contrastamos la clase social alta con la clase social baja. Aquella presenta una mayor proporción de temas originales y un mayor interés por "la familia y la casa", "las flores y las plantas", las fiestas y "los deportes, juegos y juguetes". En cambio en la clase social baja predomina el interés por "la escuela", "el campo","la naturaleza", "la ciudad".

CAPITULO VIII

VIII.- EVOLUCION CUANTITATIVA DEL VOCABULARIO. ANALISIS DIFERENCIAL (I). DISEÑO Y RESULTADOS.-

Otra de las investigaciones abiertas en el estudio del desarrollo cuantitativo del vocabulario se refiere al análisis diferencial que sufre el incremento en función de diversas variables. Este es el segundo objetivo que pretendemos estudiar en esta parte de nuestro trabajo.

Nos proponemos analizar las diferencias que presenta la producción verbal de los sujetos que constituyen la muestra. Tomamos como variables, la edad, el sexo, la zona de procedencia y la clase social.

8. 1.- DISEÑO.-

Utilizamos un Diseño Factorial Fijo de 5 x 2 x 2 x 3, con grupos de n-desiguales. Tenemos, por tanto, cuatro factores como variables independientes (V1 a V4) y la producción verbal (cantidad de palabras escritas) como variable dependiente (V5 a V12).

8. 1 . 1.- Los Factores o variables independientes: V1 a V4.-

- V 1. FACTOR 1: EDAD.-

Este factor presenta cinco niveles, correspondientes

a las cinco edades de los sujetos que integran la muestra, es decir, 6, 7, 8, 9 y 10 años.

En relación con la definición de V1 (Edad) hemos de hacer las siguientes puntualizaciones. A saber: los sujetos que componen cada uno de los cinco niveles de esta variable no coinciden exactamente en edad con la que les corresponde por el nivel en el que se adscriben. En la asignación de los sujetos a los grupos se siguió el criterio del nivel de escolarización en lugar de la edad cronológica. La correspondencia se establece de la siguiente forma:

- 6 años - 1º de E.G.B.
- 7 años - 2º de E.G.B.
- 8 años - 3º de E.G.B.
- 9 años - 4º de E.G.B.
- 10 años - 5º de E.G.B.

La razón de esta diferencia es fundamentalmente pedagógica. La identificación de la edad con el nivel de escolarización que presentaban los sujetos se hizo con objeto de determinar el vocabulario real de los distintos niveles escolares, más que el vocabulario de unas edades en concreto. La edad, en sí misma, no es una variable de gran significación o relevancia, a menos que sea considerada como manifestación de una compleja suma de efectos hereditarios, madurativos y ambientales concurrentes en un momento dado. Si la edad establece diferencias significativas es debido, en gran parte, a las

diferencias en experiencia y aprendizaje que reflejan los sujetos pertenecientes a los distintos grupos de edad. Y, no podemos olvidar, que la experiencia y el aprendizaje verbal, particularmente en las edades que estudiamos, están mediatizadas por la influencia de la escuela. De aquí, que considerásemos oportuno identificar los diversos grupos de edad con sus correspondientes niveles de escolarización. Las estadísticas, por otra parte, de que dispusimos para establecer la muestra fueron las relativas a escolarización y no las relativas al número de sujetos por edad.

A pesar de la puntualización comentada, observamos, no obstante, que las diferencias entre edad y nivel de escolarización en primera etapa de E.G.B. son de escasa importancia. En primer lugar, porque el criterio edad (tener seis años cumplidos) es determinante para el comienzo del primer curso de E.G.B., y prácticamente la totalidad de los niños se encuentran escolarizados debido a la obligatoriedad y gratuidad de la enseñanza básica. Y en segundo lugar, porque los desfases que se producen entre edad y curso escolar durante la primera etapa de E.G.B. son mínimos si tenemos en cuenta la ausencia de criterios selectivos que facilitan la promoción del alumno de un nivel a otro.

En el CUADRO 8.1. aparece la distribución de este primer factor:

V 1 - E D A D

NIVEL	CODIGO	FRECUENCIA ABSOLUTA	FRECUENCIA RELATIVA (%)
SEIS años	1.	402	18.6
SIETE años	2.	326	15.1
OCHO años	3.	380	17.5
NUEVE años	4.	510	23.5
DIEZ años	5.	548	25.3
TOTAL		2166	100.

- V 2. Factor 2: SEXO.-

Este factor presenta dos niveles: Niños y Niñas. La Distribución en la muestra está tomada al 50%.

V 2 - S E X O

NIVEL	CODIGO	FRECUENCIA ABSOLUTA	FRECUENCIA RELATIVA (%)
NIÑOS	1.	1086	50.1
NIÑAS	2.	1080	49.90

CUADRO 8. 2.- Distribución del Factor V2 - SEXO.

Se observará que en la distribución aparece una

pequeña diferencia a favor de los niños. Se debió a un problema de muestreo. Se contabilizaron tres niños más en lugar de tres niñas, por error en la trascripción de datos al ordenador. Concretamente, ésto sucedió en el nivel de seis años, zona urbana, de la provincia de Almería. No creemos que el error sea de importancia, dadas las dimensiones de la muestra.

El criterio del 50% en la distribución de la variable Sexo se adoptó una vez observado el número de niños y niñas escolarizados en las cuatro provincias. Aunque en todas ellas se observó que el número de niños escolarizados por nivel era superior al de niñas, sin embargo, la diferencia no la creímos significativa para tener que ajustar la muestra a la proporción real observada (la diferencia total, en porcentaje, no era superior al 1.2% a favor de los niños).

- V 3 - Factor 3. ZONA.-

Consideramos dos niveles en este factor: Zona Urbana y Zona Rural.

La definición de esta variable se hizo tomando como índice de referencia el número de habitantes. Situamos el criterio de diferenciación de zonas en 15.000 habitantes. Así pues, las dos zonas comprendían:

a) Zona URBANA: Poblaciones de más de 15.000 habs.
b) Zona RURAL: Poblaciones de menos de 15.000 habs.

Para la distinción de zonas nos basamos en los datos comunicados por el Instituto Nacional de Estadística sobre la población de derecho de los municipios de las cuatro pro-

vincias según la Rectificación del Padrón Municipal de Habitantes en 31 de Diciembre de 1979 (Para Almería, utilizamos los datos de 31 de diciembre de 1975).

Según estos datos clasificamos los municipios dentro de cada provincia en dos grupos: rurales y urbanos. A continuación, computamos el número de alumnos correspondientes a cada grupo (Estadística de la Enseñanza del M.E.C. para el curso 1978-79) y establecimos los porcentajes de representación para cada zona (CUADRO 8.3.).

PROVINCIA	HABITANTES	%	ALUMNOS	%
ALMERIA				
Z. URBANA	176.397	45.40	37.187	56.48
Z. RURAL	212.095	54.60	28.654	43.52
GRANADA				
Z. URBANA	357.596	45.28	62.015	50.24
Z. RURAL	432.096	54.72	61.434	49.76
JAEN				
Z. URBANA	314.756	46.04	52.739	51.20
Z. RURAL	368.910	53.96	50.257	48.80
MALAGA				
Z. URBANA	760.374	72.93	129.235	74.91
Z. RURAL	282.240	27.07	43.282	25.09
TOTALES				
Z. URBANA	1.609.123	55.40	281.176	60.49
Z. RURAL	1.295.341	44.60	183.627	39.51

CUADRO 8.3.- Distribución de la muestra por Zonas.

Las diferencias que se observan entre estos porcentajes y los que en la práctica se utilizaron (Cfr. CUADRO 6.6. y CUADRO 6.12.) se deben al efecto del redondeo y aproximación realizados en la fijación de los tamaños de los estratos al repartir la muestra definitiva de los 2166 sujetos. En estas aproximaciones salió favorecida la zona urbana respecto a la zona rural. Esta es la diferencia que se observa al comparar los porcentajes teóricos totales fijados (CUADRO 8.3.) y los porcentajes que arroja esta variable en la muestra definitiva (CUADRO 8.4.):

V 3 - Z O N A

NIVEL	CODIGO	FRECUENCIA ABSOLUTA	FRECUENCIA RELATIVA (%)
Z. URBANA	1.	1316	60.8
Z. RURAL	2.	850	39.2

CUADRO 8.4.- Distribución del Factor V 3 - ZONA.

- V 4. Factor 4. CLASE SOCIAL.-

La Clase Social no fue considerada entre las variables de estratificación de la muestra. La clasificación social de los sujetos se hizo "a posteriori", una vez seleccionada la muestra, tomando como referencia los datos aportados por el Cuestionario o Ficha Control de cada alumno (Cfr. Anexo 5.3.).

Los estudios de investigación sobre el fenómeno de las clases sociales han hecho hincapié en la existencia de di-

versos índices que marcan la pertenencia a una determinada clase social. También han puesto de manifiesto las implicaciones que de ello se derivan. Implicaiones de importancia notable en el ámbito de la educación y de las que se ocupa la sociología de la educación (Kahl, 1957; Warner, W.L.; Meeker y Eels, K., 1960; Havighurst, R.J., 1961; Havighurst y Neugarten, 1967; Informe Foessa, 1970; S. del Campo, 1974; Bozal V. y Paramio L., 1975; Lerena, 1976).

Los efectos de la clase social sobre el desarrollo del lenguaje, en general, también han sido ampliamente analizados. Son conocidas las tesis de Bernstein, al respecto. Bernstein (1961), afirma que el lenguaje hablado es la base principal sobre la cual una estructura social determinada llega a formar parte de la experiencia individual de un sujeto. El lenguaje hablado, conforme se aprende, "provoca, refuerza y universaliza distintos tipos de relaciones con el ambiente y crea así dimensiones particulares importantes". Llega a la conclusión de que "ciertas formas lingüísticas suponen para el que habla una pérdida o adquisición de aptitudes, tanto cognoscitivas como sociales, que son estratégicas para el éxito educativo y profesional, y estas formas del uso del lenguaje estám determinadas culturalemnte y no individualmente" (pp. 288-289).

Bernstein identifica dos estilos de comportamiento verbal o códigos de comunicación que son característicos de las diversas clases sociales. El código "elaborado" o el "lenguaje formal" es propio de la clase media. Este tipo de lenguaje posee una estructura gramatical y sintáctica peculiar. El orden y la estructura de las frases es correcta. Se eligen con cuidado los adjetivos y adverbios. La construcción de oraciones es

gramaticalmente compleja debido al uso de una amplia gama de conjunciones y cláusulas subordinadas. Un simbolismo expresivo discrimina entre los significados incluídos en las secuencias del habla. No se refuerzan las palabras o frases dominantes, ni se presentan las secuencias de manera difusa y generalizada. Los mensajes comunicados con este código se adaptan mejor a los requerimientos de cualquier situación particular. Un código, como éste, permite que la lengua funcione en una variedad más amplia de situaciones, que se establezcan diferenciaciones más sutiles en el significado, y, en último término, que la comunicación sea más precisa. La posibilidad de usar un código semejante, según Bernstein, queda limitada a la clase media, aunque necesariamente ésto no sucede siempre así.

Sugiere, por otra parte, que los individuos de la clase baja no aprende un código tan elaborado y se limitan al uso de unos códigos más "restringidos". El lenguaje de la clase baja se distingue por ser más corto, gramaticalmente más sencillo, a menudo aparecen oraciones inconclusas, pobres en su forma sintáctica. Apenas se emplean las cláusulas subordinadas para explicar el significado de lo que se quiere expresar o comunicar. Se utilizan en forma repetitiva y de manera rígida las conjunciones, adjetivos y adverbios. Muchas de las afirmaciones denotan una confusión entre las premisas y la conclusión, lo cual conduce a pronunciamientos simplistas y categóricos. Los códigos restrictivos proporcionan un lenguaje de significado implícito que fácilmente se entiende dentro de la comunidad familiar. No obstante, por tratarse de un lenguaje estereotipado y condensado, carece de los detalles y de la exactitud necesaria para la conceptualización y la diferenciación del contenido del mensaje que se pretende transmitir.

Bernstein (1964), tras la comprobación experimental de sus hipótesis, llega a la conclusión de que estos dos tipos de lenguaje, elaborado y restringido, tienen su origen en diversos patrones de interacción social y familiar.

Por lo que se refiere al desarrollo cuantitativo del vocabulario, aspecto que nos ocupa en esta investigación, también son numerosos los estudios que señalan la existencia de una relación entre el nivel sociocultural de procedencia y la riqueza de vocabulario que el niño manifiesta. Todos coinciden en afirmar que los niños pertenecientes a medios favorecidos muestran una superioridad en producciòn verbal sobre los niños que proceden de medios más modestos y pobres. Sólo a título de ejemplo para no hacer la referencia muy extensa, citamos algunos de los trabajos más recientes y significativos en este sentido (Templin, 1957; Stodolsky, 1965, 1967; Bourdieu y Passeron, 1966; Delis y Wittwer, 1971; Marc, 1972; Armayor y colbs., 1975, 1977; García Hoz, 1977; Brossard, 1978; Erhlich y colbs., 1978).

Los criterios adoptados en nuestro trabajo para la clasificación social de los sujetos fueron dos: la profesión de los padres y el nivel de estudios de los padres. De la combinación de estos dos criterios resultaron los tres niveles que componen la variable clase social:

1. Clase Social Alta (Profesiones relacionadas con los cuadros superiores de mando de la industria lā administración y los servicios; profesiones liberales; empresarios. Estudios con titulación de rango universitario).

2. Clase Social Media (Profesiones que requieren para su desempeño un cierto grado de especialización y cualificación. Estudios con titulación de de grado medio o diplomatura universitaria).

3. Clase Social Baja (Obreros, empleados con o sin estudios primarios).

En el CUADRO 8. 5. aparecen los resultados totales relativos a la clasificación social de la muestra:

V 4 - C L A S E S O C I A L

NIVEL	CODIGO	FRECUENCIA ABSOLUTA	FRECUENCIA RELATIVA (%)
ALTA	1.	174	8.00
MEDIA	2.	418	19.30
BAJA	3.	1.574	72.70

CUADRO 8.5.- Distribución del Factor V4 - CLASE SOCIAL.

8. 1. 2.- La Producción Verbal (Variable Dependiente): V 5 a V 12.-

Consideramos como variable dependiente la producción verbal, es decir, la cantidad de palabras escritas por cada sujeto en los protocolos de prueba. Ahora bien, con objeto de realizar un análisis diferencial exhaustivo acerca del desarrollo cuantitativo del vocabulario, presentamos la producción verbal diferenciada por categorías gramaticales.

Las 8 categorías gramaticales establecidas fueron:

V 5 - NOMBRES (No se incluyeron los nombres propios)

V 6 - ADJETIVOS

V 7 - VERBOS

V 8 - ADVERBIOS

V 9 - PRONOMBRES

V 10 - ARTICULOS

V 11 - PREPOSICIONES

V 12 - CONJUNCIONES

El análisis y el recuento de la producción verbal de cada sujeto dió lugar, por tanto, a ocho puntuaciones parciales, correspondientes a las ocho categorías establecidas. La categorización de las palabras se hizo en función de una normativa de clasificación lexicográfica adoptada previamente (Cfr. Anexo nº 1).

La adscripción de las palabras a las distintas categorías se hizo siempre en función del contexto en el que aparecían en las dos pruebas utilizadas para registrar la producción verbal.

En el test de asociación libre la clasificación apenas si planteó dificultades, dada la ausencia de relaciones asociativas entre las palabras a nivel paradigmático. En cambio, éste fue el mayor problema para la clasificación de las palabras en la prueba de redacción o texto libre.

8. 1. 3.- Descripción de la Población.-

MUESTRA: 2.166 sujetos.

SUBPOBLACIONES DE LA MUESTRA:

VARIABLE	CODIGO	NIVEL	N
V1	1.	SEIS	402
V2	1.	NIÑOS	204
V3	1.	URBANA	124
V4	1.	ALTA	23
V4	2.	MEDIA	43
V4	3.	BAJA	58
V3	2.	RURAL	80
V4	1.	ALTA	1
V4	2.	MEDIA	10
V4	3.	BAJA	69
V2	2.	NIÑAS	198
V3	1.	URBANA	118
V4	1.	ALTA	13
V4	2.	MEDIA	28
V4	3.	BAJA	77
V3	2.	RURAL	80
V4	1.	ALTA	3
V4	2.	MEDIA	7
V4	3.	BAJA	70

VARIABLE	CODIGO	NIVEL	N
V1	1.	SIETE	326
V2	1.	NIÑOS	163
V3	1.	URBANA	112
V4	1.	ALTA	8
V4	2.	MEDIA	29
V4	3.	BAJA	75
V3	2.	RURAL	51
V4	1.	ALTA	2
V4	2.	MEDIA	3
V4	3.	BAJA	46
V2	2.	NIÑAS	163
V3	1.	URBANA	112
V4	1.	ALTA	12
V4	2.	MEDIA	26
V4	3.	BAJA	74
V3	2.	RURAL	51
V4	1.	ALTA	1
V4	2.	MEDIA	7
V4	3.	BAJA	43

VARIABLE	CODIGO	NIVEL	N
V1	1.	OCHO	380
V2	1.	NIÑOS	190
V3	1.	URBANA	111
V4	1.	ALTA	13
V4	2.	MEDIA	36
V4	3.	BAJA	62
V3	2.	RURAL	79
V4	1.	ALTA	2
V4	2.	MEDIA	5
V4	3.	BAJA	72
V2	2.	NIÑAS	190
V3	1.	URBANA	111
V4	1.	ALTA	17
V4	2.	MEDIA	26
V4	3.	BAJA	68
V3	2.	RURAL	79
V4	1.	ALTA	3
V4	2.	MEDIA	6
V4	3.	BAJA	70

VARIABLE	CODIGO	NIVEL	N
V1	1.	NUEVE	510
V2	1.	NIÑOS	255
V3	1.	URBANA	156
V4	1.	ALTA	21
V4	2.	MEDIA	44
V4	3.	BAJA	91
V3	2.	RURAL	99
V4	1.	ALTA	2
V4	2.	MEDIA	5
V4	3.	BAJA	92
V2	2.	NIÑAS	255
V3	1.	URBANA	156
V4	1.	ALTA	28
V4	2.	MEDIA	39
V4	3.	BAJA	89
V3	2.	RURAL	99
V4	1.	ALTA	1
V4	2.	MEDIA	5
V4	3.	BAJA	93

VARIABLE	CODIGO	NIVEL	N
V1	1.	DIEZ	548
V2	1.	NIÑOS	274
V3	1.	URBANA	158
V4	1.	ALTA	14
V4	2.	MEDIA	53
V4	3.	BAJA	91
V3	2.	RURAL	116
V4	1.	ALTA	1
V4	2.	MEDIA	7
V4	3.	BAJA	108
V2	2.	NIÑAS	274
V3	1.	URBANA	158
V4	1.	ALTA	8
V4	2.	MEDIA	34
V4	3.	BAJA	116
V3	2.	RURAL	116
V4	1.	ALTA	1
V4	2.	MEDIA	5
V4	3.	BAJA	110

8. 2.- Análisis de Varianza.-

En el actual diseño hemos utilizado, pues, cuatro variables independientes y la producción verbal como variable dependiente. Por lo que se refiere a la muestra, se trata de un diseño factorial con grupos de n desiguales dentro de cada uno de los niveles que presentan los factores o variables independientes. Asímismo, se trata de un diseño factorial de efecto fijo, por cuanto el reparto de sujetos por grupos agota todos los niveles posibles dentro de cada factor.

Con el análisis de varianza podremos inferir los efectos independientes (efectos principales) e interactivos (de 2º, 3º y 4º orden) que ejercen la edad, el sexo, la zona y la clase social sobre la producción verbal.

Debido a las ocho subclasificaciones introducidas en la variable dependiente hubimos de realizar ocho análisis de varianza, uno para cada categoría gramatical.

Los pasos seguidos en el análisis de varianza para cada variable dependiente se estructuran en el CUADRO 8.7..

El análisis estadístico se realizó en el Centro de Cálculo de la Universidad de Málaga. Se utilizó el programa SPSS (Statistical Package for the Social Sciences - Realise 6.03, Noviembre, 4, 1976).

FUENTE DE VARIANZA	SUMA DE CUADRADOS (SC.)	GRADOS DE LIBERTAD (Gl.)	MEDIA DE CUADRADOS (MC.)	F.
A	SC_A	(a-1)	SC_A/ $Gl._A$	MC_A/ MC_{error}
B	SC_B	(b-1)	SC_B/ $Gl._B$	MC_B/ MC_{error}
C	SC_C	(c-1)	SC_C/ $Gl._C$	MC_C/ MC_{error}
D	SC_D	(d-1)	SC_D/ $Gl._D$	MC_D/ MC_{error}
AB	SC_{AB}	(a-1)(b-1)	SC_{AB}/ $Gl._{AB}$	MC_{AB}/ MC_{error}
AC	SC_{AC}	(a-1)(c-1)	SC_{AC}/ $Gl._{AC}$	MC_{AC}/ MC_{error}
AD	SC_{AD}	(a-1)(d-1)	SC_{AD}/ $Gl._{AD}$	MC_{AD}/ MC_{error}
BC	SC_{BC}	(b-1)(c-1)	SC_{BC}/ $Gl._{BC}$	MC_{BC}/ MC_{error}
BD	SC_{BD}	(b-1)(d-1)	SC_{BD}/ $Gl._{BD}$	MC_{BD}/ MC_{error}
CD	SC_{CD}	(c-1)(d-1)	SC_{CD}/ $Gl._{CD}$	MC_{CD}/ MC_{error}
ABC	SC_{ABC}	(a-1)(b-1)(c-1)	SC_{ABC}/ $Gl._{ABC}$	MC_{ABC}/ MC_{error}
ABD	SC_{ABD}	(a-1)(b-1)(d-1)	SC_{ABD}/ $Gl._{ABD}$	MC_{ABD}/ MC_{error}
ACD	SC_{ACD}	(a-1)(c-1)(d-1)	SC_{ACD}/ $Gl._{ACD}$	MC_{ACD}/ MC_{error}
BCD	SC_{BCD}	(b-1)(c-1)(d-1)	SC_{BCD}/ $Gl._{BCD}$	MC_{BCD}/ MC_{error}
ABCD	SC_{ABCD}	(a-1)(b-1)(c-1)(d-1)	SC_{ABCD}/ $Gl._{ABCD}$	MC_{ABCD}/ MC_{error}
Error	SC_{error}	N - abcd	SC_{error}/N - abcd	
Total	SC_{total}	N - 1	SC_{total}/ N - 1	

CUADRO 8. 7.- DISEÑO FACTORIAL FIJO. ANALISIS DE VARIANZA.
(5 x 2 x 2 x 3).

La interpretación de la información que proporciona el programa SPSS requiere unas puntualizaciones previas sobre la forma de proceder el programa en el análisis de los datos.

Los diseños factoriales con grupos de n-desiguales son considerados por el programa SPSS como no ortogonales, es decir, considera a los factores asociados entre sí y, por tanto no independientes. Sólo en el caso de que el número de sujetos por grupo, no siendo iguales, guardasen alguna proporcionalidad en la distribución, los efectos principales seguirían siendo ortogonales. Sin embargo, los efectos de interacción de cualquier orden en ningún caso serían independientes de los efectos principales.

En razón de esta consideración, el procedimiento seguido por el programa en la asignación o reparto de las sumas de cuadrados (SC) es el siguiente. Obsérvese que proponemos un ejemplo de diseño con dos factores con objeto de simplificar y exponer con mayor claridad la estrategia de división del programa. Las consideraciones son extensivas a un diseño más complejo como el que nosotros hemos introducido.

El método clásico de análisis de varianza descompone el total de la suma de cuadrados (SC_t) en un diseño de dos factores en los tres tipos siguientes:

$SC_{A,B}$ = Suma de Cuadrados debida a los efectos aditivos de A y B.

SC_{AB} = Suma de Cuadrados debida al efecto de interacción = $SC_{A,B,AB} - SC_{A,B}$

SC_{error} = Suma de Cuadrados de error = $SC_t - SC_{A,B,AB}$

En el reparto de las Sumas de Cuadrados los tres componentes se hacen ortogonales uno al otro al imponer o establecer entre ellos una determinada jerarquía. En la división de los efectos aditivos de A y B en efectos principales separados, la SC_A y la SC_B no se sumarán en $SC_{A,B}$ ya que A y B no son ortogonales. Por tanto, la suma de los efectos de cada factor por separado no coincide normalmente con la suma de cuadrados de los efectos aditivos de A y B ($SC_A + SC_B \neq SC_{A,B}$). La interacción se define como la diferencia entre las sumas de cuadrados explicadas por el efecto de unión total de A y B, y la suma de cuadrados de los efectos aditivos de A y B. Es decir, el componente de interacción viene dado por los componentes residuales de A y B. Y, por último, el componente del error viene dado igualmente por las sumas residuales de cuadrados.

En un modelo de efecto fijo, como el que nos ocupa en este trabajo, la razón apropiada de F para probar el significado de cada componente es la siguiente:

$$F = \frac{\text{MS para cada componente.}}{MS_{error.}}$$

La diferencia importante que establece el programa en el tratamiento de los datos en diseños de n-desiguales respecto a los diseños de n-iguales (ortogonales) estriba en el grado de interrelación que manifiesten los factores entre sí. De tal manera que si A y B están estrechamente asociados sería posible obtener unos datos en el análisis de varianza que indicaran que el efecto de los dos factores, asociados como un todo, fuese significativo sin que ninguno de los e-

fectos principales, por separado, lo fuera. La razón de ser de esta hipótesis estaría explicada si partimos del hecho de que cada factor adquiere relevancia sólo en la medida en que se ve incrementado por los efectos del otro factor.

También podría ocurrir el caso contrario: que el programa no considere ortogonales los factores al analizar los datos y dividir las sumas de cuadrados y, sin embargo, los resultados confirmen la existencia de una baja asociación entre ellos, en cuyo caso no habría grandes diferencias entre el diseño ortogonal y el diseño con grupos de n-desiguales, debido al grado de independencia que manifiestan los factores.

En el programa SPSS caben distintas opciones en el análisis de los datos para cada uno de los tipos de diseños según se consideren ortogonales o no los efectos de los factores. Esto depende, como ya hemos expuesto, de que los grupos tengan un número igual o desigual de sujetos. En nuestro caso, dada la naturaleza del problema que pretendemos resolver y los factores o variables independientes que consideramos entre las que existe evidentemente una interrelación o dependencia, optamos por el modelo clásico de análisis de varianza. Es el método apropiado para aquellas si-tuaciones, como la nuestra, en que los factores no manifiestan un orden causal conocido, están correlacionados entre sí y no poseen en los grupos igual número de sujetos.

Hechas estas consideraciones sobre el programa y antes de pasar a una exposición de resultados, haremos referencia a otra cuestión previa: la prueba de homogeneidad de varianzas.

Para cada uno de los ocho análisis de varianza probamos la homogeneidad de varianzas de la población utilizada. Aplicamos la prueba de Bartlett y estos fueron los resultados (chi-cuadrado) obtenidos:

NOMBRES:	26.63
ADJETIVOS:	53.59
VERBOS:	54.52
ADVERBIOS:	74.35
PRONOMBRES:	57.93
ARTICULOS:	42.83
PREPOSICIONES:	66.18
CONJUNCIONES:	62.94

Comparando los chi-cuadrado obtenidos con el valor crítico de chi-cuadrado con 59 Gl. y α = .05, observamos que los valores encontrados no son significativos y que, por tanto, las varianzas de las poblaciones son homogéneas. (Chi-cuadrado con 59 Gl. y α = .05 es 77.931).

A continuación, presentamos los resultados de los distintos análisis de varianza. En el capítulo siguiente, procederemos a un análisis y comentario de los mismos. Ahora bien, la complejidad del diseño y la extensión de la información proporcionada por el programa SPSS, nos impulsa a trazar un plan en la exposición de los datos con objeto de facilitar el manejo y la interpretación posterior de los mismos. Para una información más exhaustiva y completa acerca de la comunicación de resultados proporcionada por el SPSS remitimos al Anexo nº 6.

En el presente capítulo simplificaremos dicha información. Los datos que proporcionamos de cada una de las ocho variables dependientes son los siguientes:

a) Una Tabla Resumen de ANOVA.

b) Un cuadro del Análisis de Clasificación Múltiple (Tabla MCA).

c) Tablas de medias de efectos principales e interacciones significativas con su correspondiente representación gráfica.

d) Pruebas de contraste de medias (Scheffé).

TABLA MCA (Multiple Classification Analysis)

Ninguna de las tablas anteriores plantea problemas de interpretación, salvo el caso de la Tabla MCA. Antes de proceder a la comunicación de resultados haremos algunas puntualizaciones que faciliten sumanejo e interpretación. El programa SPSS, junto a las tablas resumen de ANOVA y la descripción de las medias y desviaciones típicas de las subpoblaciones de la muestra (Anexo 6), proporciona la tabla MCA (Análisis de Clasificación Múltiple). Es otro estadístico más para examinar los patrones de comportamiento seguidos por los efectos principales en relación con la variable dependiente.

Para interpretar correctamente las tablas MCA hemos de aclarar el significado de los datos que allí aparecen. En primer lugar, presenta la media (GRAND MEAN) de la variable dependiente y, a continuación una tabla con las medias obtenidas en los diversos niveles que integran cada factor. Las medias vienen expresadas como desviaciones de la media principal.

De esta forma, se aprecia mejor la maginitud del efecto en cada nivel.

En el cuadro, las medias para cada nivel aparecen en dos columnas diferentes a continuación de la columna de las N (número de sujetos por nivel). En la primera (UNADJUSTED DEV´N ETA), los valores-desviación de la media se presentan desajustados en cada nivel en relación con los valores medios de los restantes niveles dentro de cada factor. En la segunda columna (ADJUSTED FOR INDEPENDENTS DEV´N BETA), los valores medios (de nuevo expresados como desviaciones de la media principal), se presentan ajustados. Indican que se ha realizado un ajuste de las medias para cada categoría o nivel en función de los restantes niveles. Respecto a las diferencias que puede resultar de la comparación de una y otra columna, lo importante es observar si a consecuencia del ajuste se producen en las medias cambios sustanciales. De ser así, el aumento o disminución ocasionado nos estaría indicando el grado de relación entre los factores. Una de las posibilidades que encierran las puntuaciones de la tabla MCA es precisamente poder examinar la naturaleza de los cambios que se producen en los efectos de una variable cuando introducimos otras como control.

Junto a cada una de las columnas de medias, aparece en la tabla un nuevo dato. Nos referimos a los coeficientes Eta y Beta para cada uno de los factores. Se trata de una nueva forma de mostrar los resultados del análisis de varianza por lo que a efectos principales se refiere.

El análisis de varianza tiene por objeto someter a verificación hipótesis estadísticas sobre la significación de medias. Sin embargo, también es importante la estimación de

las varianzas componentes: cada fuente de varianza contribuye con un porcentaje dado a la varianza total resultando de utilidad, por consiguiente, conocer la proporción de varianza atribuible a cada factor en la varianza total. Esta es la información que proporciona el programa SPSS en la tabla MCA con los coeficientes Eta y Beta.

Así pues, junto a la columna de medias desajustadas aparece un valor para cada factor: el coeficiente Eta. El cuadrado de Eta indica la proporción de varianza explicada por cada factor sin que estén controlados los restantes factores. Del mismo modo, el coeficiente beta, también llamado coeficiente de regresión parcial, indica la relación de correlación entre los distintos niveles o categorías de un factor, cuando este se encuentra ajustado o controlado por los otros factores. Pongamos un ejemplo: en la tabla MCA correspondiente a los NOMBRES, los coeficientes Beta para los cuatro factores son de .67, .08, .08 y .07 para la Edad, el Sexo, la Zona y la Clase social, respectivamente. Si elevamos al cuadro el primero ($.67^2 = .45$) el resultado nos indica la proporción de varianza explicada por la Edad, la contribución del factor edad en la varianza observada en los Nombres. Si repetimos la operación para los restantes factores obtendríamos igualmente el porcentaje de variación correspondiente a cada factor en la varianza total observada en los nombres, manteniendo constantes los efectos de los restantes factores.

Al final de la tabla MCA tenemos el MULTIPLE R y el MULTIPLE R SQUARED (R^2). El R^2 o coeficiente de Determinación indica la varianza determinada en la variable dependiente debida a la combinación de los efectos de las cuatro variables independientes. Expresa la correlación más alta posible entre las variables independientes y la variable dependiente.

8.2.1.- ANALISIS DE VARIANZA: 1ª Variable Dependiente (V5) NOMBRES.-

```
* * * * * * * * * * A N A L Y S I S   O F   V A R I A N C E * * * * * * * * * *
                V5        NOMBRES
            BY  V1        EDAD
                V2        SEXO
                V3        ZONA
                V4        CLASE SOCIAL
* * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * *
                                        SUM OF                    MEAN        SIGNIF
SOURCE OF VARIATION                    SQUARES      DF          SQUARE     F     OF F

MAIN EFFECTS                          419459.359    8        52432.420 235.246   .001
   V1                                 407630.207    4       101907.552 457.224   .001
   V2                                   6515.542    1         6515.542  29.233   .001
   V3                                   4526.766    1         4526.766  20.310   .001
   V4                                   3895.282    2         1947.641   8.738   .001

2-WAY INTERACTIONS                      8716.566   21          415.075   1.862   .010
   V1        V2                         1837.995    4          459.499   2.062   .062
   V1        V3                          301.149    4          200.287    .899   .999
   V1        V4                         2679.994    8          334.999   1.503   .150
   V2        V3                         1062.987    1         1062.987   4.769   .027
   V2        V4                         1750.615    2          875.307   3.927   .020
   V3        V4                          975.807    2          487.904   2.189   .110

3-WAY INTERACTIONS                      7589.340   22          344.970   1.549   .050
   V1        V2        V3                110.344    4           27.586    .124   .999
   V1        V2        V4               3721.571    8          465.196   2.087   .034
   V1        V3        V4               3215.646    8          401.931   1.804   .071
   V2        V3        V4                377.704    2          188.852    .847   .999

4-WAY INTERACTIONS                      1747.973    8          218.497    .980   .999
   V1        V2        V3               1747.969    8          218.496    .980   .999
             V4

EXPLAINED                             437513.234   59         7415.479  33.271   .001

RESIDUAL                              469392.547 2106          222.883

TOTAL                                 906905.781 2165          418.894
   2166 CASES WERE PROCESSED.
      0 CASES (    .0 PCT) WERE MISSING.
```

* * * M U L T I P L E C L A S S I F I C A T I O N A N A L Y S I S * * *

	V5	NOMBRES
BY	V1	EDAD
	V2	SEXO
	V3	ZONA
	V4	CLASE SOCIAL

* * * \

V 5 - NOMBRES.- Tabla de Medias.-

- EFECTOS PRINCIPALES:

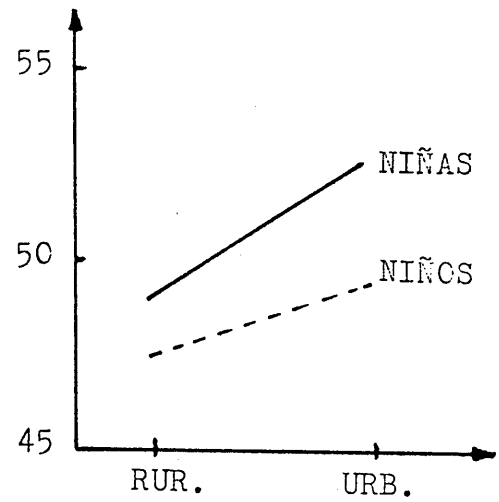
VARIABLE	NIVEL	N	$\bar{X}$
V1	SEIS	402	30.72
	SIETE	326	34.49
	OCHO	380	46.31
	NUEVE	510	57.64
	DIEZ	548	66.40
V2	NIÑOS	1086	47.66
	NIÑAS	1080	51.13
V3	URBANA	1316	50.62
	RURAL	850	47.48
V4	ALTA	174	53.66
	MEDIA	418	50,27
	BAJA	1574	48.68

- INTERACCIONES DE PRIMER ORDEN:

a) V2 - V3: SEXO - ZONA.

V2 - V3	URBANA		RURAL	
	N	$\bar{X}$	N	$\bar{X}$
NIÑOS	661	48.41	425	46.46
NIÑAS	655	52.60	425	48.90

Gráfica de la Interacción Sexo - Zona:

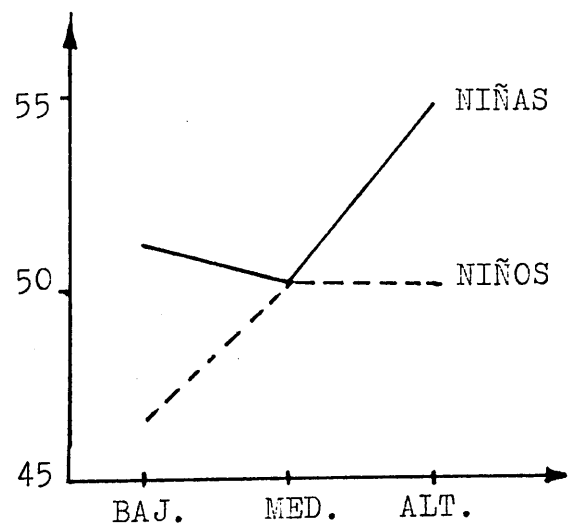


Gráf. 1

b) V2 - V4: SEXO - CLASE SOCIAL.

V2 - V4	ALTA N	ALTA $\bar{X}$	MEDIA N	MEDIA $\bar{X}$	BAJA N	BAJA $\bar{X}$
NIÑOS	87	50.04	235	49.96	764	46.62
NIÑAS	87	54.64	183	49.87	810	51.06

Gráfica de la Interacción Sexo - Clase Social:



Gráf. 2

- INTERACCIONES DE 2º ORDEN:

V1 - V2 - V4: EDAD - SEXO - CLASE SOCIAL.

a) 6 años:

V2 - V4	ALTA N	ALTA $\bar{X}$	MEDIA N	MEDIA $\bar{X}$	BAJA N	BAJA $\bar{X}$
NIÑOS	24	30.46	53	31.92	127	27.27
NIÑAS	16	33.06	35	30.71	147	33.32

b) 7 años:

V2 - V4	ALTA N	ALTA $\bar{X}$	MEDIA N	MEDIA $\bar{X}$	BAJA N	BAJA $\bar{X}$
NIÑOS	10	44.60	32	37.75	121	33.07
NIÑAS	13	33.54	33	34.85	117	34.83

c) 8 años:

V2 - V4	ALTA N	ALTA $\bar{X}$	MEDIA N	MEDIA $\bar{X}$	BAJA N	BAJA $\bar{X}$
NIÑOS	15	45.93	41	52.02	134	42.64
NIÑAS	20	55.70	32	50.65	138	45.85

d) 9 años:

V2 - V4	ALTA N	ALTA $\bar{X}$	MEDIA N	MEDIA $\bar{X}$	BAJA N	BAJA $\bar{X}$
NIÑOS	23	59.39	49	58.63	183	55.02
NIÑAS	29	71.07	44	58.70	182	57.69

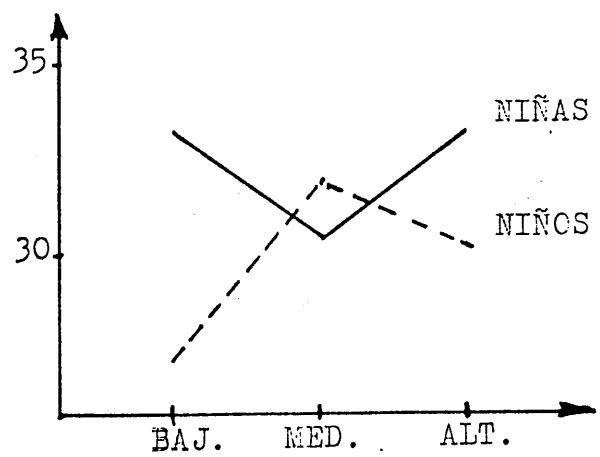
e) 10 años:

V2 - V4	ALTA N	ALTA $\bar{X}$	MEDIA N	MEDIA $\bar{X}$	BAJA N	BAJA $\bar{X}$
NIÑOS	15	74.06	60	63.92	199	62.16
NIÑAS	9	68.22	39	69.18	226	68.83

Gráficas de las Interacciones de 2º Orden:

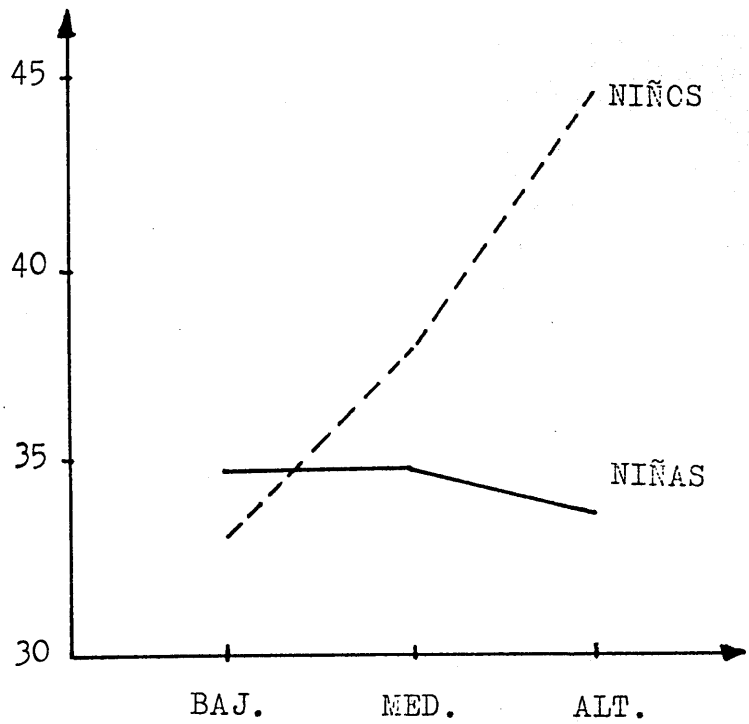
V1 - V2 - V4.

a) 6 años:



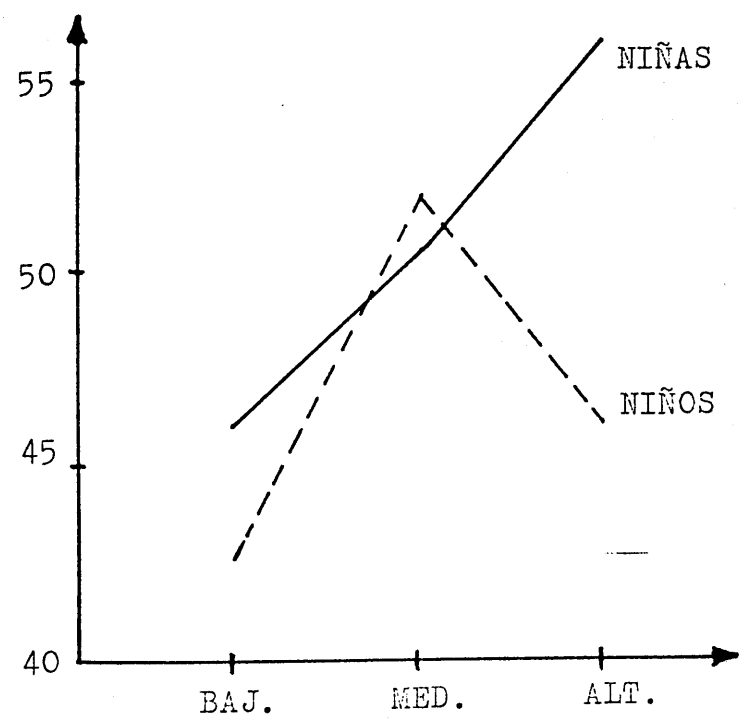
Gráf. 3

b) 7 años:



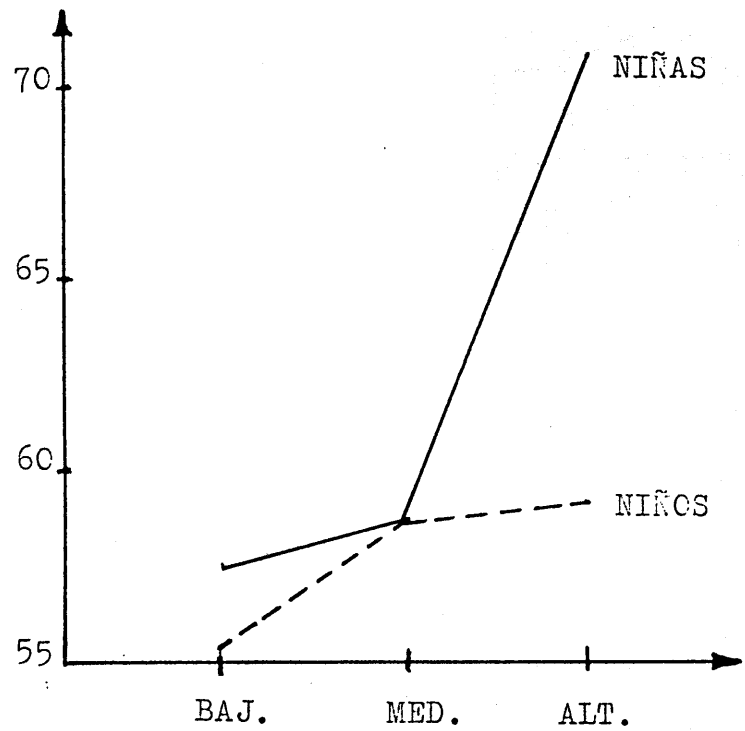
Gráf. 4

c) 8 años:



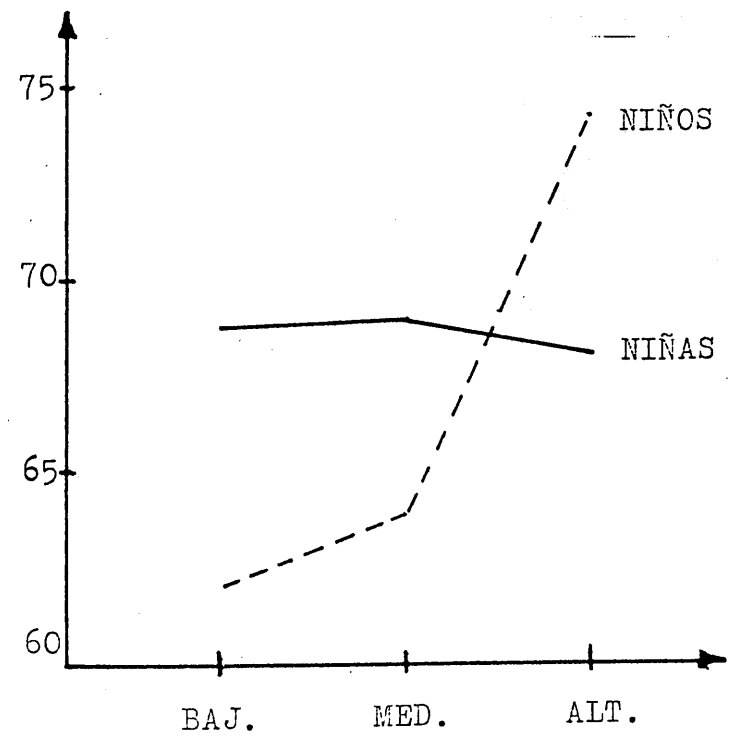
Gráf. 5

d) 9 años:



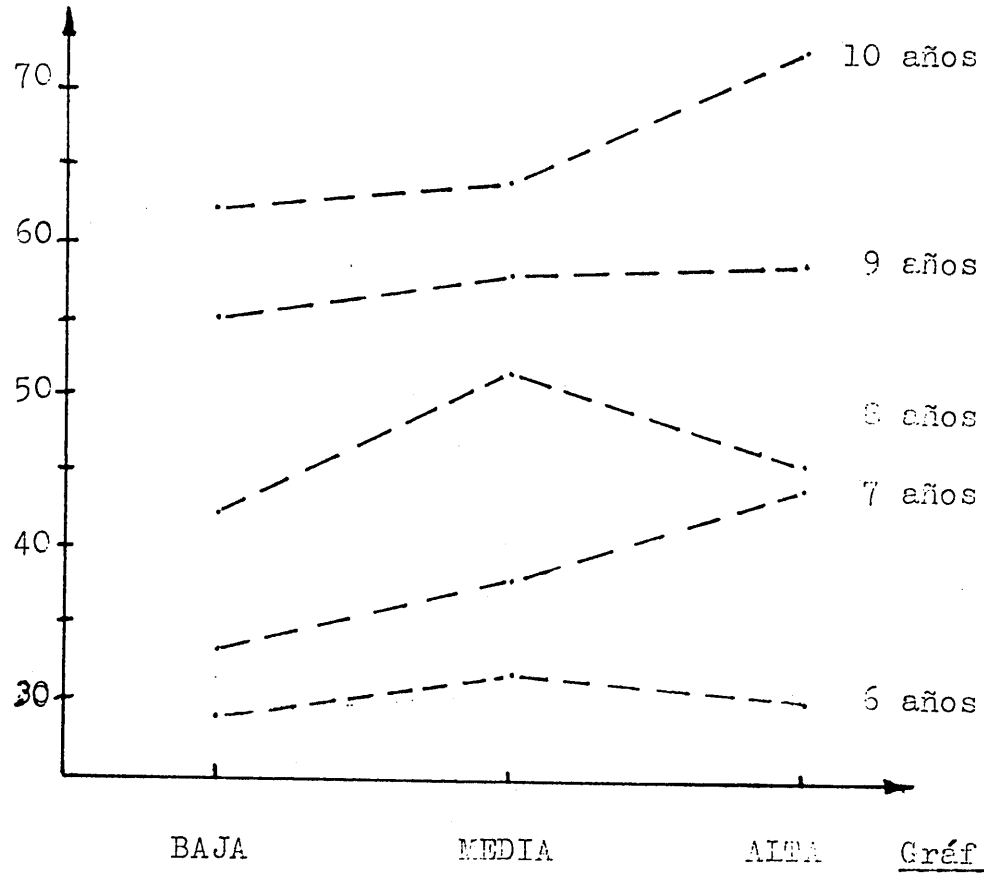
Gráf. 6

e) 10 años:



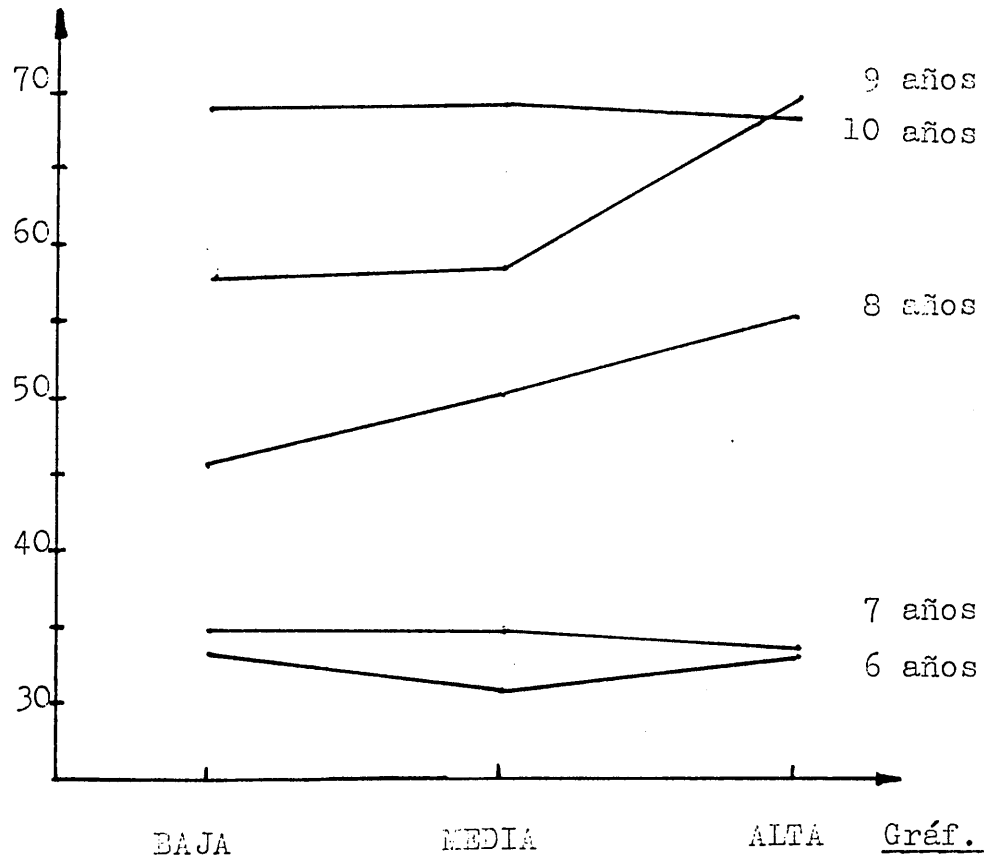
Gráf. 7

NIÑOS DE LOS CINCO CURSOS:



Gráf. 8

NIÑAS DE LOS CINCO CURSOS:



Gráf. 9

8.2.2.- ANALISIS DE VARIANZA: 2º Variable Dependiente (V6) ADJETIVOS.-

```
* * * * * * * * * * A N A L Y S I S   O F   V A R I A N C E * * * * * * * * * *
                V6        ADJETIVOS
            BY  V1        EDAD
                V2        SEXO
                V3        ZONA
                V4        CLASE SOCIAL
* * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * *
                                        SUM OF                    MEAN          SIGNIF
SOURCE OF VARIATION                    SQUARES      DF          SQUARE     F     OF F

MAIN EFFECTS                          28290.264      8        3536.283  87.761   .001
   V1                                 26132.535      4        6533.134 162.135   .001
   V2                                   972.246      1         972.246  24.129   .001
   V3                                    57.664      1          57.664   1.431   .236
   V4                                  1583.586      2         791.793  19.650   .001

2-WAY INTERACTIONS                     1826.089     21          86.957   2.158   .002
   V1        V2                         434.755      4         108.689   2.697   .029
   V1        V3                         244.297      4          61.074   1.516   .194
   V1        V4                         781.417      8          97.677   2.424   .013
   V2        V3                         239.677      1         239.677   5.949   .014
   V2        V4                         105.220      2          52.610   1.306   .270
   V3        V4                          36.661      2          18.330    .455   .999

3-WAY INTERACTIONS                     1039.438     22          47.247   1.173   .262
   V1        V2        V3               193.816      4          48.454   1.202   .307
   V1        V2        V4               628.219      8          78.527   1.949   .049
   V1        V3        V4               310.551      8          38.819    .963   .999
   V2        V3        V4                11.670      2           5.835    .145   .999

4-WAY INTERACTIONS                      185.350      8          23.169    .575   .999
   V1        V2        V3               185.350      8          23.169    .575   .999
             V4

EXPLAINED                              31341.142     59         531.206  13.163   .001

RESIDUAL                               84360.055   2106          40.294

TOTAL                                 115701.196   2165          53.673
   2166 CASES WERE PROCESSED.
      0 CASES (    .0 PCT) WERE MISSING.
```

* * * M U L T I P L E C L A S S I F I C A T I O N A N A L Y S I S * * *

V6 ADJETIVOS

BY V1 EDAD

V2 SEXO

V3 ZONA

V4 CLASE SOCIAL

* *

GRAND MEAN = 8.74

VARIABLE + CATEGORY	N	UNADJUSTED DEV'N	UNADJUSTED ETA	ADJUSTED FOR INDEPENDENTS DEV'N	ADJUSTED FOR INDEPENDENTS BETA	ADJUSTED FOR INDEPENDENTS + COVARIATES DEV'N	ADJUSTED FOR INDEPENDENTS + COVARIATES BETA
V1							
1 SEIS	402	-4.60		-4.88			
2 SIETE	326	-3.38		-3.41			
3 OCHO	380	-.62		-.84			
4 NUEVE	510	1.90		1.87			
5 DIEZ	548	4.33		4.44			
			.47		.48		
V2							
1 NINO	1086	-.64		-.67			
2 NINA	1080	.64		.67			
			.09		.09		
V3							
1 URBANA	1316	.27		.14			
2 RURAL	850	-.42		-.22			
			.05		.02		
V4							
1 ALTA	174	1.19		1.65			
2 MEDIA	413	1.18		1.41			
3 BAJA	1574	-.45		-.56			
			.10		.12		
MULTIPLE R SQUARED				.243			
MULTIPLE R				.493			

V 6 - ADJETIVOS.- Tabla de Medias.-

- EFECTOS PRINCIPALES:

VARIABLE	NIVEL	N	$\bar{X}$
V1	SEIS	402	3.86
	SIETE	326	5.33
	OCHO	380	7.90
	NUEVE	510	10.61
	DIEZ	548	13.81
V2	NIÑOS	1086	8.07
	NIÑAS	1080	9.41
V3(*)	URBANA	1316	8.88
	RURAL	850	8.52
V4	ALTA	174	10.39
	MEDIA	418	10.15
	BAJA	1574	8.18

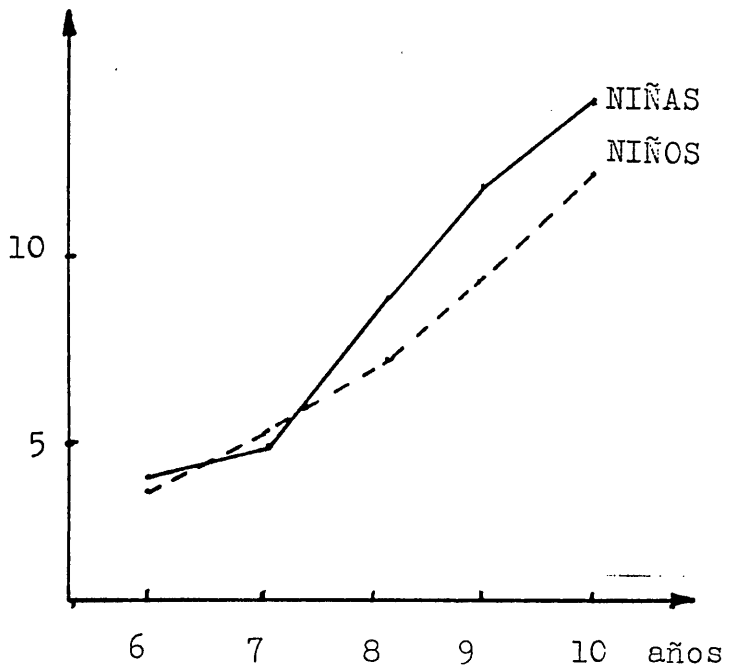
(*) El Efecto Principal de Zona no es significativo.

- INTERACCIONES DE PRIMER ORDEN:

a) V1 - V2: EDAD - SEXO.

V1-V2	SEIS		SIETE		OCHO		NUEVE		DIEZ	
	N	$\bar{X}$	N	$\bar{X}$	N	$\bar{X}$	N	$\bar{X}$	N	$\bar{X}$
NIÑOS	204	3.82	163	5.40	190	7.27	255	9.48	274	12.21
NIÑAS	198	4.06	163	5.32	190	8.58	255	11.82	274	13.94

Gráfica de la Interacción Edad - Sexo:

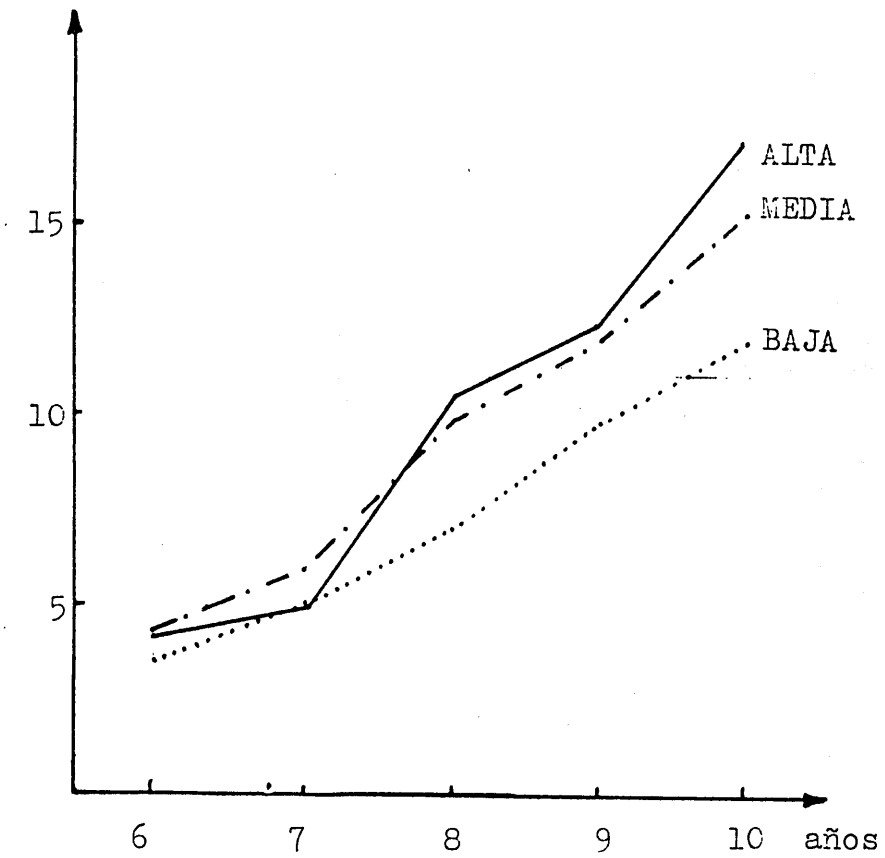


Gráf. 10

b) V1 - V4: EDAD - CLASE SOCIAL.

V1-V4	SEIS N	SEIS $\bar{X}$	SIETE N	SIETE $\bar{X}$	OCHO N	OCHO $\bar{X}$	NUEVE N	NUEVE $\bar{X}$	DIEZ N	DIEZ $\bar{X}$
ALTA	40	4.40	23	5.17	35	10.68	52	12.32	24	17.42
MEDIA	88	4.42	65	6.06	73	9.77	93	11.98	99	15.54
BAJA	274	3.72	238	5.18	272	7.07	365	10.07	425	12.25

Gráfica de la Interacción Edad - Clase Social:

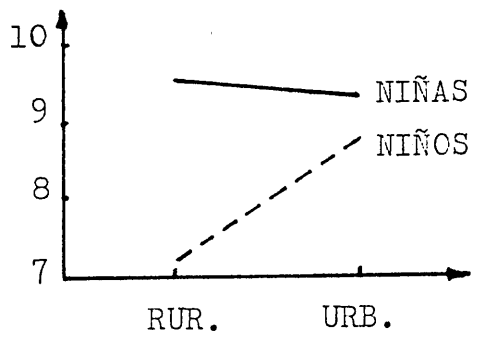


Gráf. 11

c) V2 - V3: SEXO - ZONA.

V2 - V3	URBANA N	$\bar{X}$	RURAL N	$\bar{X}$
NIÑOS	661	8.70	425	7.18
NIÑAS	655	9.33	425	9.46

Gráfica de la Interacción Sexo - Zona:



Gráf. 12

8.2.3.- ANALISIS DE VARIANZA: 3ª Variable Dependiente (V7) VERBOS.-

```
* * * * * * * * * * A N A L Y S I S   O F   V A R I A N C E * * * * * * * * * *
                V7        VERBOS
            BY  V1        EDAD
                V2        SEXO
                V3        ZONA
                V4        CLASE SOCIAL
* * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * *
```

SOURCE OF VARIATION			SUM OF SQUARES	DF	MEAN SQUARE	F	SIGNIF OF F
MAIN EFFECTS			81933.362	8	10241.670	116.130	.001
V1			79820.573	4	19955.143	226.271	.001
V2			1384.013	1	1384.013	15.693	.001
V3			208.620	1	208.620	2.366	.120
V4			1258.862	2	629.431	7.137	.001
2-WAY INTERACTIONS			5963.543	21	283.978	3.220	.001
V1	V2		1329.594	4	332.398	3.769	.005
V1	V3		946.888	4	236.722	2.684	.030
V1	V4		2249.021	8	281.128	3.188	.001
V2	V3		669.637	1	669.637	7.593	.006
V2	V4		273.124	2	136.562	1.548	.211
V3	V4		321.977	2	160.989	1.825	.159
3-WAY INTERACTIONS			3378.590	22	153.572	1.741	.018
V1	V2	V3	539.661	4	134.915	1.530	.189
V1	V2	V4	1215.783	8	151.973	1.723	.088
V1	V3	V4	1483.078	8	185.385	2.102	.032
V2	V3	V4	90.433	2	45.217	.513	.599
4-WAY INTERACTIONS			911.595	8	113.949	1.292	.242
V1	V2	V3	911.595	8	113.949	1.292	.242
	V4						
EXPLAINED			92187.090	59	1562.493	17.717	.001
RESIDUAL			185730.617	2106	88.191		
TOTAL			277917.707	2165	128.368		

2166 CASES WERE PROCESSED.
0 CASES (.0 PCT) WERE MISSING.

* * * M U L T I P L E C L A S S I F I C A T I O N A N A L Y S I S * * *

V7 VERBOS

BY V1 EDAD

V2 SEXO

V3 ZONA

V4 CLASE SOCIAL

* *

GRAND MEAN = 17.71

VARIABLE + CATEGORY	N	UNADJUSTED DEV'N	UNADJUSTED ETA	ADJUSTED FOR INDEPENDENTS DEV'N	ADJUSTED FOR INDEPENDENTS BETA	ADJUSTED FOR INDEPENDENTS + COVARIATES DEV'N	ADJUSTED FOR INDEPENDENTS + COVARIATES BETA
V1							
1 SEIS	402	-9.45		-9.51			
2 SIETE	326	-4.56		-4.59			
3 OCHO	380	-1.28		-1.30			
4 NUEVE	510	3.47		3.41			
5 DIEZ	548	7.31		7.44			
			.53		.54		
V2							
1 NIÑO	1086	-.81		-.80			
2 NINA	1080	.81		.80			
			.07		.07		
V3							
1 URBANA	1316	.28		.27			
2 RURAL	850	-.44		-.41			
			.03		.03		
V4							
1 ALTA	174	1.67		2.48			
2 MEDIA	418	.06		.41			
3 BAJA	1574	-.20		-.38			
			.04		.07		

MULTIPLE R SQUARED .295

MULTIPLE R .543

V 7 - VERBOS.- Tabla de Medias.-

- EFECTOS PRINCIPALES:

VARIABLE	NIVEL	N	$\bar{X}$
V1	SEIS	402	8.2
	SIETE	326	13.12
	OCHO	380	16.41
	NUEVE	510	21.12
	DIEZ	548	25.15
V2	NIÑOS	1086	16.91
	NIÑAS	1080	18.51
V3(*)	URBANA	1316	17.98
	RURAL	850	17.30
V4	ALTA	174	20.19
	MEDIA	418	18.12
	BAJA	1574	17.33

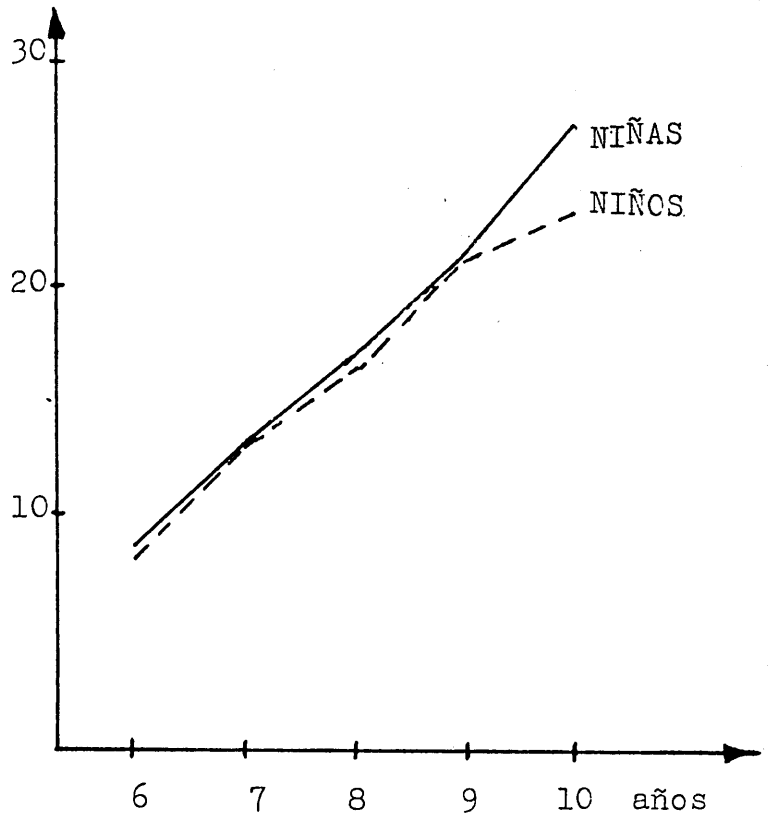
(*) El Efecto Principal de Zona no es significativo.

- INTERACCIONES DE PRIMER ORDEN:

a) V1 - V2: EDAD - SEXO.

V1-V2	SEIS		SIETE		OCHO		NUEVE		DIEZ	
	N	$\bar{X}$	N	$\bar{X}$	N	$\bar{X}$	N	$\bar{X}$	N	$\bar{X}$
NIÑOS	204	7.75	163	13.00	190	15.59	255	21.14	274	23.02
NIÑAS	198	8.79	163	13.31	190	17.27	255	21.26	274	27.03

Gráfica de la Interacción Edad - Sexo.



Gráf. 13

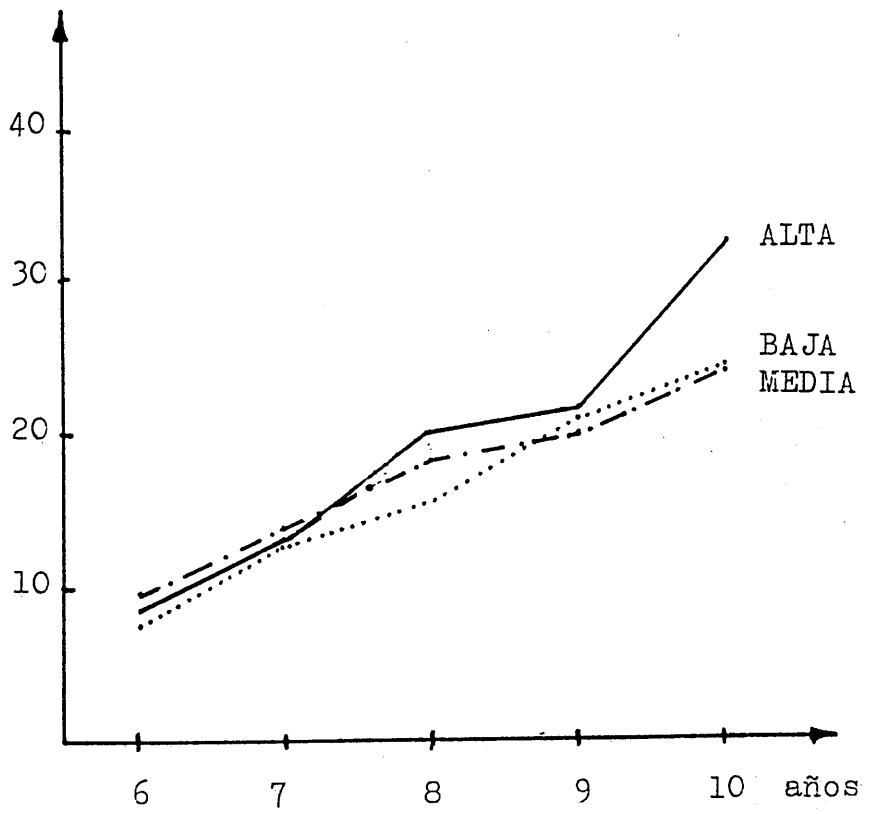
b) V1 - V3: EDAD - ZONA.

V1-V3	SEIS N	SEIS $\bar{X}$	SIETE N	SIETE $\bar{X}$	OCHO N	OCHO $\bar{X}$	NUEVE N	NUEVE $\bar{X}$	DIEZ N	DIEZ $\bar{X}$
URBANA	242	9.48	224	13.87	222	16.88	312	20.94	316	25.48
RURAL	160	6.42	102	11.56	158	15.79	198	21.56	232	24.62

c) V1 - V4: EDAD - CLASE SOCIAL.

V1-V4	SEIS N	SEIS $\bar{X}$	SIETE N	SIETE $\bar{X}$	OCHO N	OCHO $\bar{X}$	NUEVE N	NUEVE $\bar{X}$	DIEZ N	DIEZ $\bar{X}$
ALTA	40	8.72	23	13.48	35	21.03	52	22.40	24	33.87
MEDIA	88	9.83	65	13.63	73	18.44	93	20.53	99	24.46
BAJA	274	7.69	238	12.99	272	15.83	365	21.17	425	24.65

Gráfica de la Interacción Edad - Clase Social:

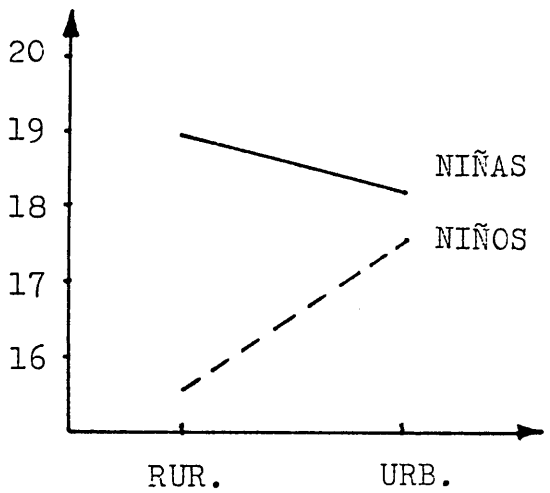


Gráf. 14

d) V2 - V3: SEXO - ZONA.

V2 - V3	URBANA N	$\bar{X}$	RURAL N	$\bar{X}$
NIÑOS	661	17.72	425	15.64
NIÑAS	655	18.27	425	18.91

Gráfica de la Interacción Sexo - Zona:



Gráf. 15

- INTERACCIONES DE 2º ORDEN:

V1 - V3 - V4: EDAD - ZONA - CLASE SOCIAL.

a) 6 años:

V3 - V4	ALTA N	$\bar{X}$	MEDIA N	$\bar{X}$	BAJA N	$\bar{X}$
URBANA	36	9.11	71	10.28	135	9.15
RURAL	4	5.25	17	7.94	139	6.28

b) 7 años:

V3 - V4	ALTA N	ALTA $\bar{X}$	MEDIA N	MEDIA $\bar{X}$	BAJA N	BAJA $\bar{X}$
URBANA	20	12.80	55	13.69	149	14.08
RURAL	3	18.00	10	13.30	89	11.16

c) 8 años:

V3 - V4	ALTA N	ALTA $\bar{X}$	MEDIA N	MEDIA $\bar{X}$	BAJA N	BAJA $\bar{X}$
URBANA	30	21.46	62	19.06	130	14.79
RURAL	5	18.40	11	14.91	142	15.77

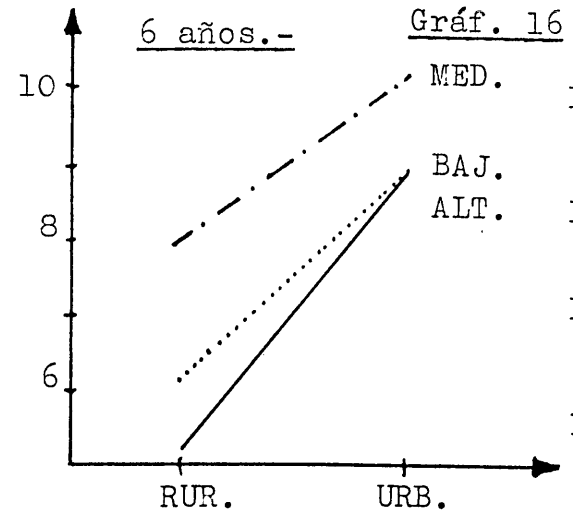
d) 9 años:

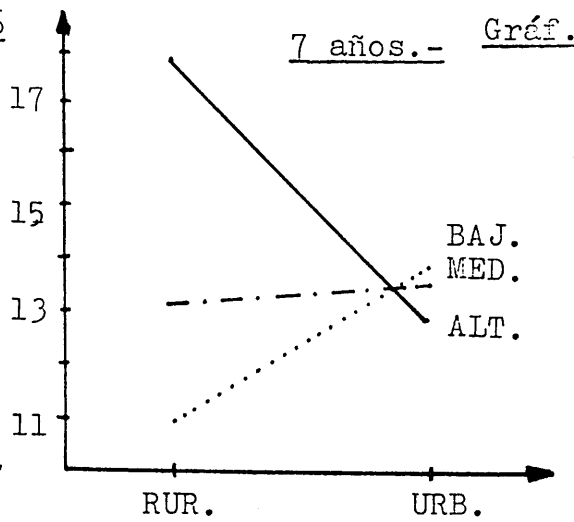
V3 - V4	ALTA N	ALTA $\bar{X}$	MEDIA N	MEDIA $\bar{X}$	BAJA N	BAJA $\bar{X}$
URBANA	49	21.86	83	20.20	180	21.02
RURAL	3	31.33	10	23.20	185	21.32

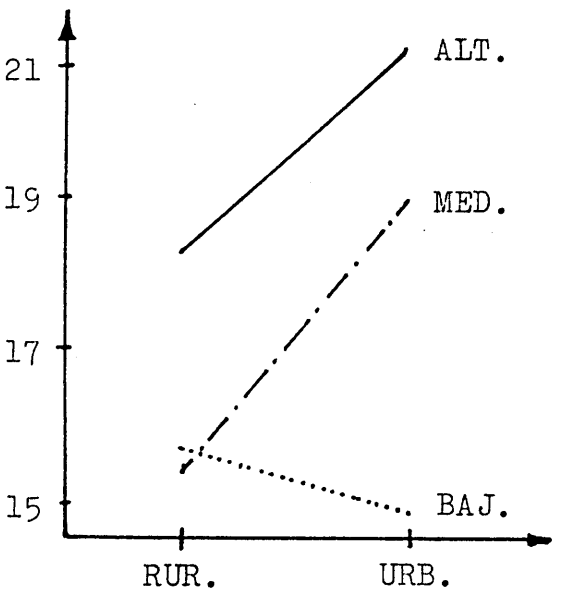
e) 10 años:

V3 - V4	ALTA N	ALTA $\bar{X}$	MEDIA N	MEDIA $\bar{X}$	BAJA N	BAJA $\bar{X}$
URBANA	22	31.86	87	24.32	207	25.04
RURAL	2	56.00	12	25.50	218	24.28

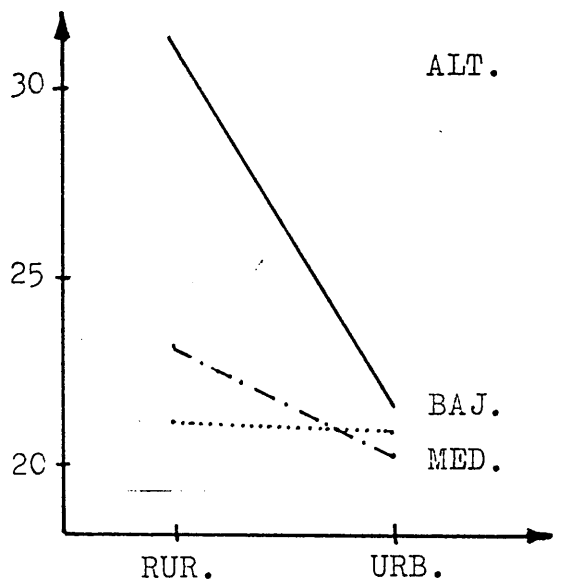
Gráficas de las Interacciones de 2º Orden:



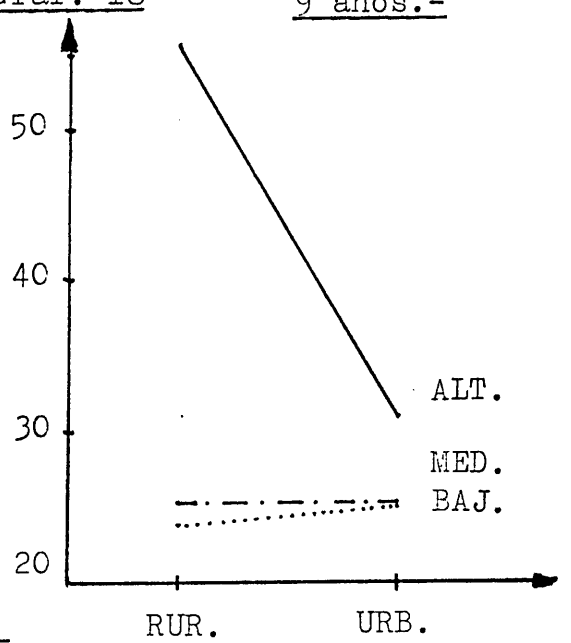




8 años.- Gráf. 18



9 años.- Gráf. 19



10 años.- Gráf. 20

8.2.4.- ANALISIS DE VARIANZA: 4ª Variable Dependiente (V8) ADVERBIOS.-

```
* * * * * * * * * * A N A L Y S I S   O F   V A R I A N C E * * * * * * * * * *
                V8        ADVERBIOS
            BY  V1        EDAD
                V2        SEXO
                V3        ZONA
                V4        CLASE SOCIAL
* * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * *
                                        SUM OF                    MEAN        SIGNIF
SOURCE OF VARIATION                     SQUARES      DF         SQUARE     F    OF F

MAIN EFFECTS                            20867.775     8        2608.472 120.786   .001
   V1                                   20030.474     4        5007.618 231.879   .001
   V2                                     653.326     1         653.326  30.252   .001
   V3                                        .705     1            .705    .033   .999
   V4                                     438.613     2         219.306  10.155   .001

2-WAY INTERACTIONS                       1444.737    21          68.797   3.186   .001
   V1        V2                          377.786     4          94.447   4.373   .002
   V1        V3                          397.043     4          99.261   4.596   .001
   V1        V4                          281.240     8          35.155   1.628   .111
   V2        V3                            5.777     1           5.777    .268   .999
   V2        V4                          219.871     2         109.936   5.091   .006
   V3        V4                          110.360     2          55.180   2.555   .076

3-WAY INTERACTIONS                        944.737    22          42.943   1.988   .004
   V1        V2        V3                168.666     4          42.167   1.953   .098
   V1        V2        V4                423.088     8          52.886   2.449   .012
   V1        V3        V4                445.563     8          55.695   2.579   .009
   V2        V3        V4                 33.850     2          16.925    .784   .999

4-WAY INTERACTIONS                        136.279     8          17.035    .789   .999
   V1        V2        V3                136.279     8          17.035    .789   .999
             V4

EXPLAINED                               23393.527    59         396.500  18.360   .001

RESIDUAL                                45480.824  2106          21.596

TOTAL                                   68874.352  2165          31.813
   2166 CASES WERE PROCESSED.
      0 CASES (    .0 PCT) WERE MISSING.
```

* * * M U L T I P L E C L A S S I F I C A T I O N A N A L Y S I S * * *

V8 ADVERBIOS

BY V1 EDAD

V2 SEXO

V3 ZONA

V4 CLASE SOCIAL

* *

GRAND MEAN = 6.59

VARIABLE + CATEGORY	N	UNADJUSTED DEV'N	UNADJUSTED ETA	ADJUSTED FOR INDEPENDENTS DEV'N	ADJUSTED FOR INDEPENDENTS BETA	ADJUSTED FOR INDEPENDENTS + COVARIATES DEV'N	ADJUSTED FOR INDEPENDENTS + COVARIATES BETA
V1							
1 SEIS	402	-4.60		-4.64			
2 SIETE	326	-2.24		-2.23			
3 OCHO	380	-.84		-.86			
4 NUEVE	510	1.53		1.50			
5 DIEZ	548	3.86		3.93			
			.54		.54		
V2							
1 NINO	1086	-.55		-.55			
2 NINA	1080	.56		.55			
			.10		.10		
V3							
1 URBANA	1316	.00		-.02			
2 RURAL	850	-.01		.02			
			.00		.00		
V4							
1 ALTA	174	.93		1.49			
2 MEDIA	418	-.09		.20			
3 BAJA	1574	-.08		-.22			
			.05		.08		

MULTIPLE R SQUARED .303

MULTIPLE R .550

V 8 - ADVERBIOS.- Tabla de Medias.-

- EFECTOS PRINCIPALES:

VARIABLE	NIVEL	N	$\bar{X}$
V1	SEIS	402	1.95
	SIETE	326	4.36
	OCHO	380	5.73
	NUEVE	510	8.09
	DIEZ	548	10.52
V2	NIÑOS	1086	6.04
	NIÑAS	1080	7.14
V3(*)	URBANA	1316	6.57
	RURAL	850	6.61
V4	ALTA	174	8.08
	MEDIA	418	6.79
	BAJA	1574	6.37

(*) El Efecto Principal de Zona no es significativo.

- INTERACCIONES DE PRIMER ORDEN:

a) V1 - V2: EDAD - SEXO.

V1-V2	SEIS		SIETE		OCHO		NUEVE		DIEZ	
	N	$\bar{X}$	N	$\bar{X}$	N	$\bar{X}$	N	$\bar{X}$	N	$\bar{X}$
NIÑOS	204	1.91	163	4.2	190	4.93	255	7.80	274	9.33
NIÑAS	198	2.08	163	4.49	190	6.57	255	8.45	274	11.57

V 8 - ADVERBIOS.- Tabla de Medias.-

- EFECTOS PRINCIPALES:

VARIABLE	NIVEL	N	$\bar{X}$
V1	SEIS	402	1.95
	SIETE	326	4.36
	OCHO	380	5.73
	NUEVE	510	8.09
	DIEZ	548	10.52
V2	NIÑOS	1086	6.04
	NIÑAS	1080	7.14
V3(*)	URBANA	1316	6.57
	RURAL	850	6.61
V4	ALTA	174	8.08
	MEDIA	418	6.79
	BAJA	1574	6.37

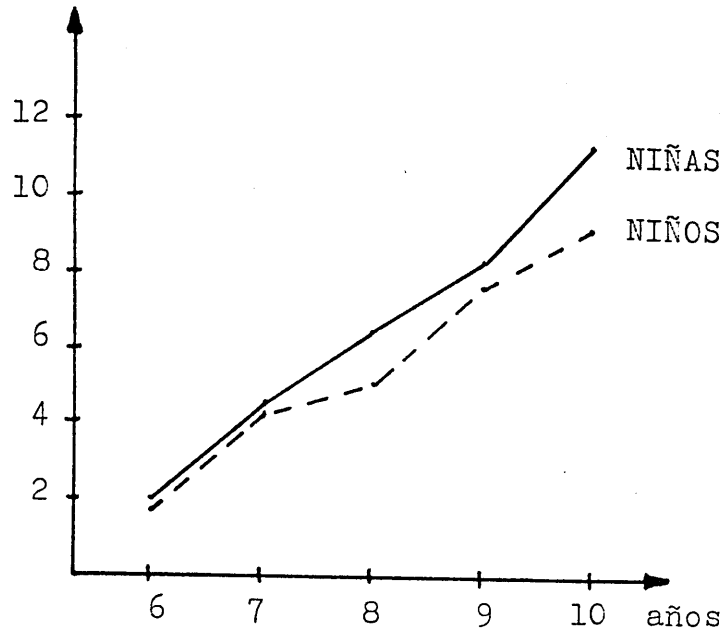
(*) El Efecto Principal de Zona no es significativo.

- INTERACCIONES DE PRIMER ORDEN:

a) V1 - V2: EDAD - SEXO.

V1-V2	SEIS		SIETE		OCHO		NUEVE		DIEZ	
	N	$\bar{X}$	N	$\bar{X}$	N	$\bar{X}$	N	$\bar{X}$	N	$\bar{X}$
NIÑOS	204	1.91	163	4.2	190	4.93	255	7.80	274	9.33
NIÑAS	198	2.08	163	4.49	190	6.57	255	8.45	274	11.57

Gráfica de la Interacción Edad - Sexo:

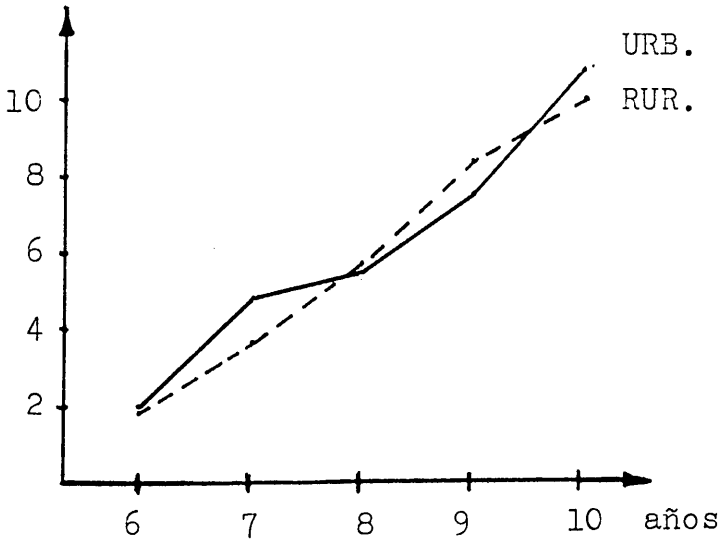


Gráf. 21

b) V1 - V3: EDAD - ZONA.

V1-V3	SEIS N	SEIS $\bar{X}$	SIETE N	SIETE $\bar{X}$	OCHO N	OCHO $\bar{X}$	NUEVE N	NUEVE $\bar{X}$	DIEZ N	DIEZ $\bar{X}$
URB.	242	2.04	224	4.72	222	5.66	312	7.73	316	10.93
RUR.	160	1.91	102	3.53	158	5.87	198	8.73	232	9.78

Gráfica de la Interacción Edad - Zona:

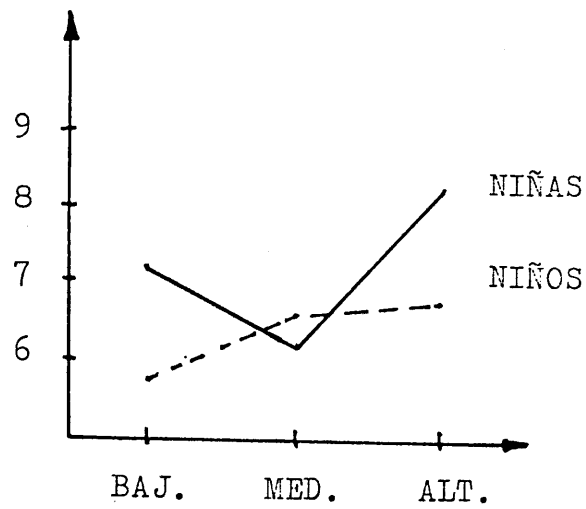


Gráf. 22

c) V2 - V4: SEXO - CLASE SOCIAL.

V2 - V4	ALTA N	$\bar{X}$	MEDIA N	$\bar{X}$	BAJA N	$\bar{X}$
NIÑOS	87	6.80	235	6.68	764	5.75
NIÑAS	87	8.24	183	6.27	810	7.22

Gráfica de la Interacción Sexo - Clase Social:



Gráf. 23

- INTERACCIONES DE 2º ORDEN:

a) V1 - V2 - V4: EDAD - SEXO - CLASE SOCIAL.

1) 6 años:

V2 - V4	ALTA N	$\bar{X}$	MEDIA N	$\bar{X}$	BAJA N	$\bar{X}$
NIÑOS	24	2.29	53	2.47	127	1.60
NIÑAS	16	1.56	35	2.34	147	2.07

2) 7 años:

V2 - V4	ALTA N	$\bar{X}$	MEDIA N	$\bar{X}$	BAJA N	$\bar{X}$
NIÑOS	10	7.60	32	5.16	121	3.67
NIÑAS	13	3.54	33	4.70	117	4.54

3) 8 años:

V2 - V4	ALTA N	$\bar{X}$	MEDIA N	$\bar{X}$	BAJA N	$\bar{X}$
NIÑOS	15	5.40	41	6.22	134	4.48
NIÑAS	20	8.30	32	5.81	138	6.49

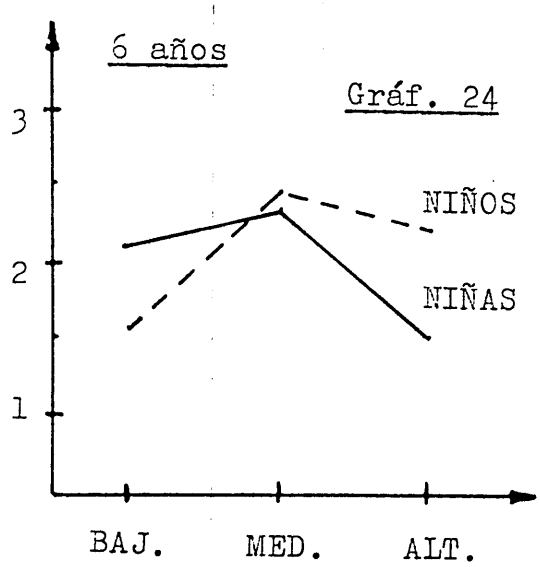
4) 9 años:

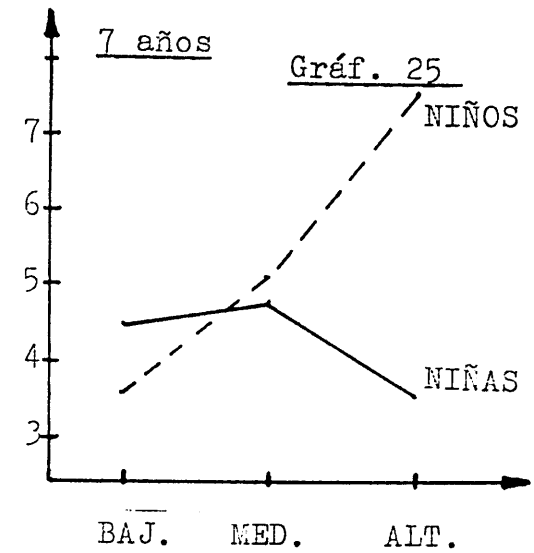
V2 - V4	ALTA N	$\bar{X}$	MEDIA N	$\bar{X}$	BAJA N	$\bar{X}$
NIÑOS	23	8.13	49	8.57	183	7.55
NIÑAS	29	11.65	44	6.59	182	8.38

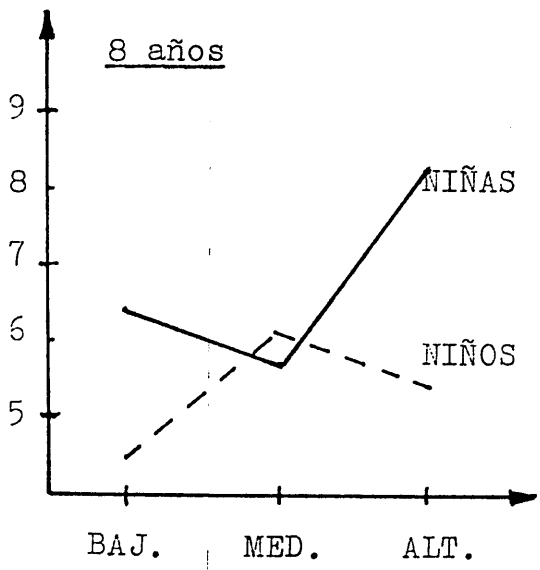
5) 10 años:

V2 - V4	ALTA N	$\bar{X}$	MEDIA N	$\bar{X}$	BAJA N	$\bar{X}$
NIÑOS	15	12.87	60	9.98	199	8.86
NIÑAS	9	15.78	39	11.13	226	11.48

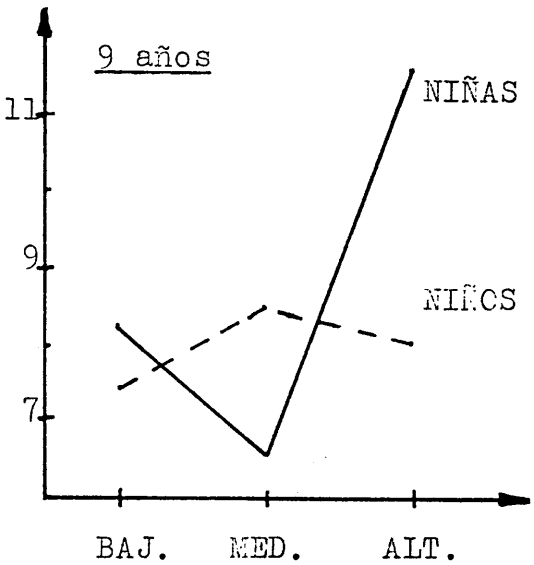
Gráficas de las Interacciones de 2º Orden:



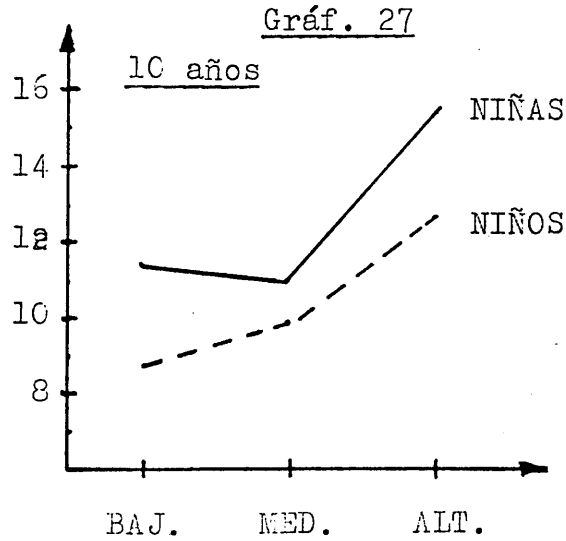




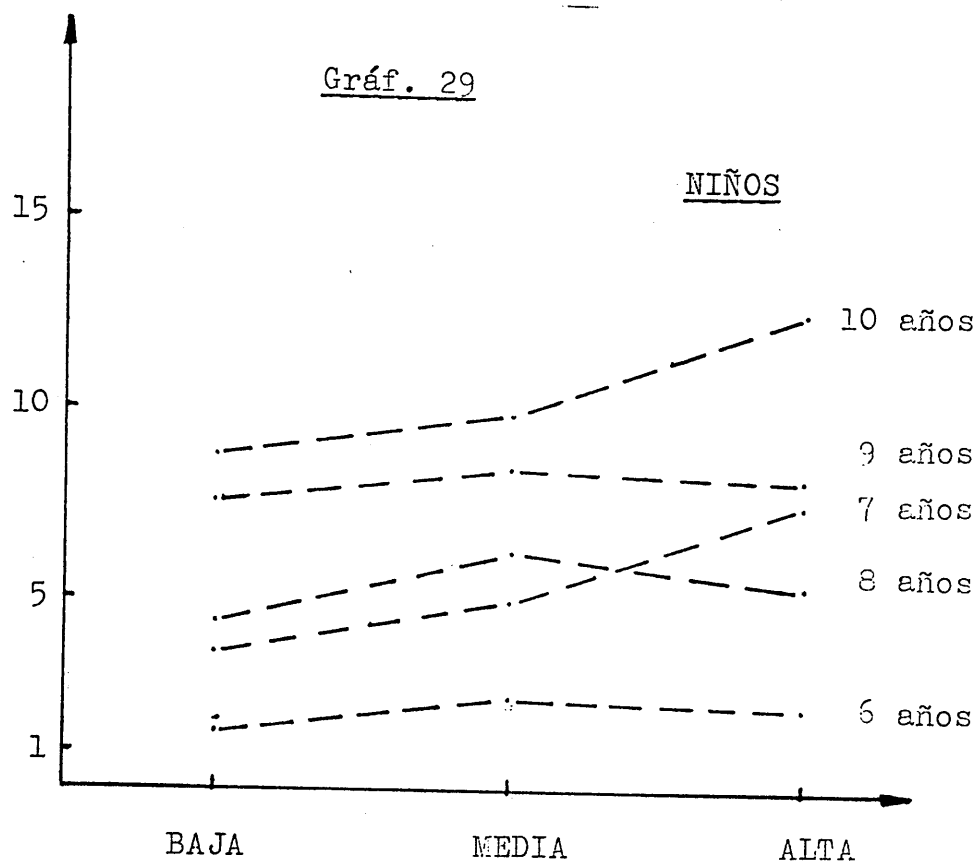
Gráf. 26

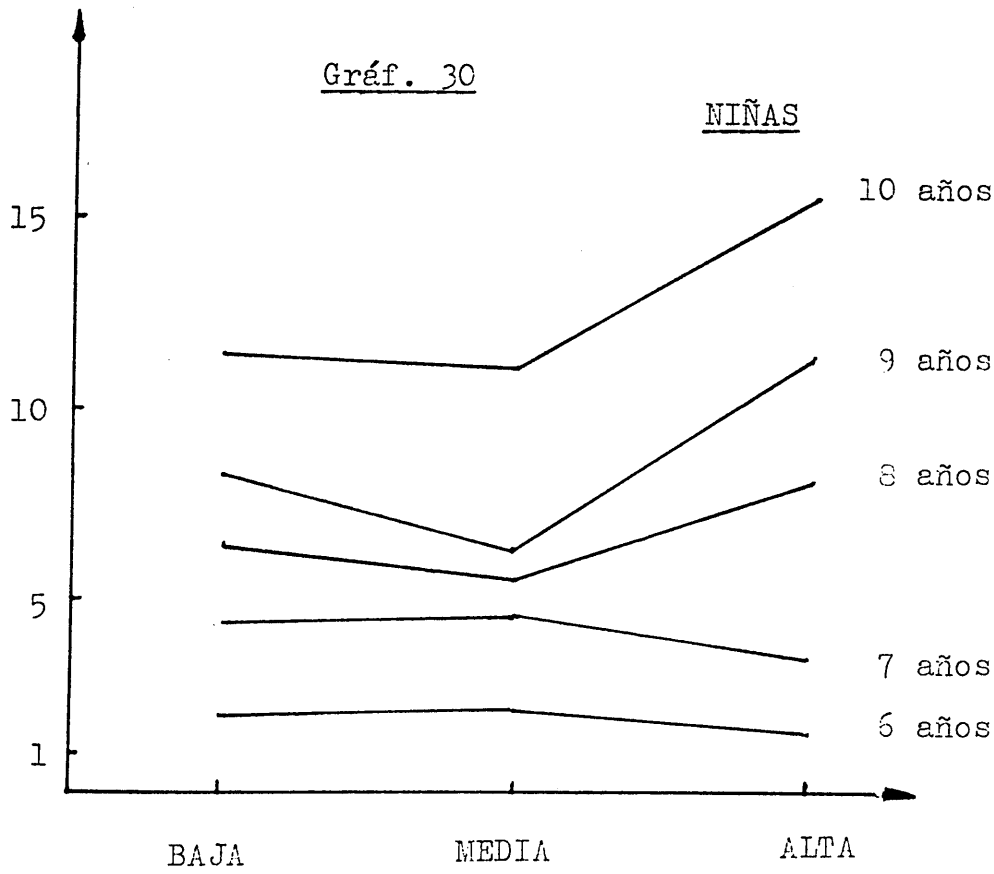


Gráf. 27



Gráf. 28


Gráf. 29
NIÑOS
15
10
5
1
10 años
9 años
7 años
8 años
6 años
BAJA
MEDIA
ALTA


Gráf. 30
NIÑAS
15
10
5
1
10 años
9 años
8 años
7 años
6 años
BAJA
MEDIA
ALTA

b) V1 - V3 - V4: EDAD - ZONA - CLASE SOCIAL.

1) 6 años:

V3 - V4	ALTA N	$\bar{X}$	MEDIA N	$\bar{X}$	BAJA N	$\bar{X}$
URBANA	36	2.03	71	2.49	135	1.81
RURAL	4	1.75	17	2.12	139	1.89

2) 7 años:

V3 - V4	ALTA N	$\bar{X}$	MEDIA N	$\bar{X}$	BAJA N	$\bar{X}$
URBANA	20	5.45	55	4.87	149	4.56
RURAL	3	4.33	10	5.20	89	3.31

3) 8 años:

V3 - V4	ALTA N	$\bar{X}$	MEDIA N	$\bar{X}$	BAJA N	$\bar{X}$
URBANA	30	7.13	62	5.98	130	5.17
RURAL	5	6.60	11	6.36	142	5.81

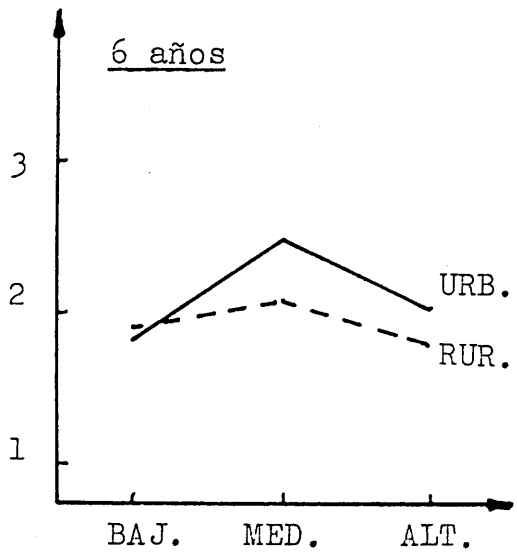
4) 9 años:

V3 - V4	ALTA N	$\bar{X}$	MEDIA N	$\bar{X}$	BAJA N	$\bar{X}$
URBANA	49	9.45	83	7.23	180	7.50
RURAL	3	20.67	10	11.00	185	8.42

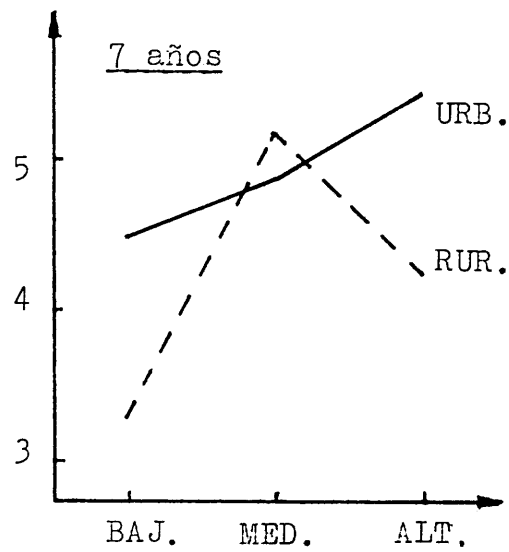
5) 10 años:

V3 - V4	ALTA N	ALTA $\bar{X}$	MEDIA N	MEDIA $\bar{X}$	BAJA N	BAJA $\bar{X}$
URBANA	22	13.45	87	10.45	207	10.87
RURAL	2	19.50	12	10.33	218	9.67

Gráfica de las Interacciones de 2º Orden:

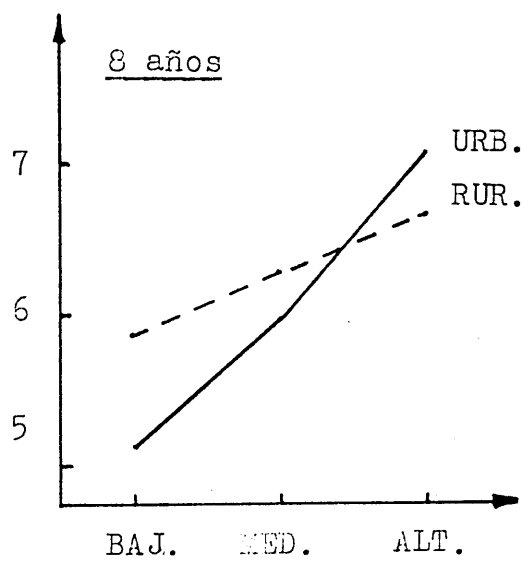


Gráf. 31

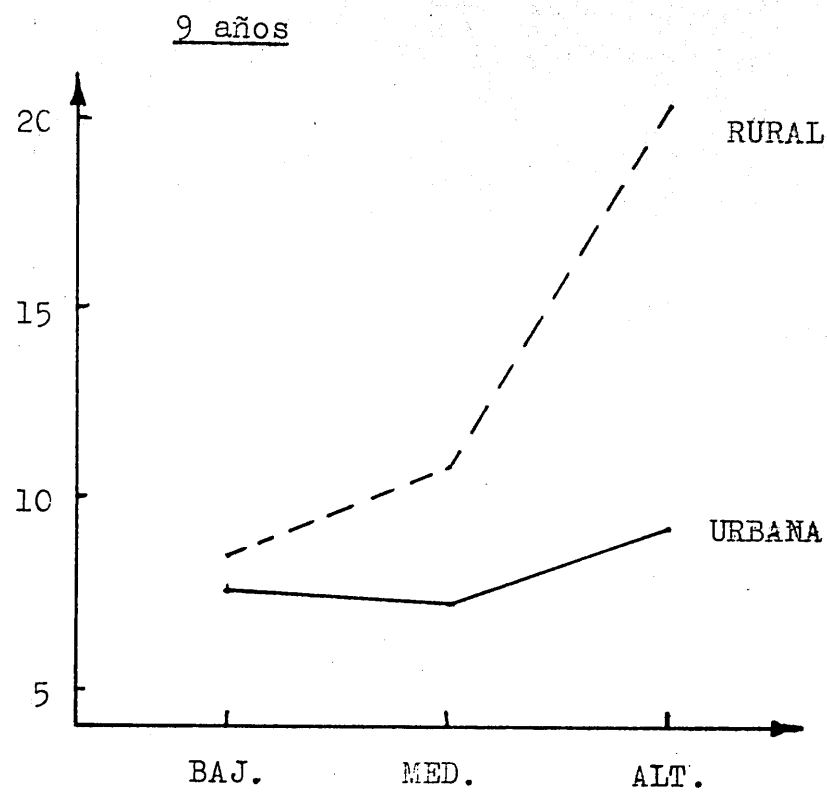


Gráf. 32

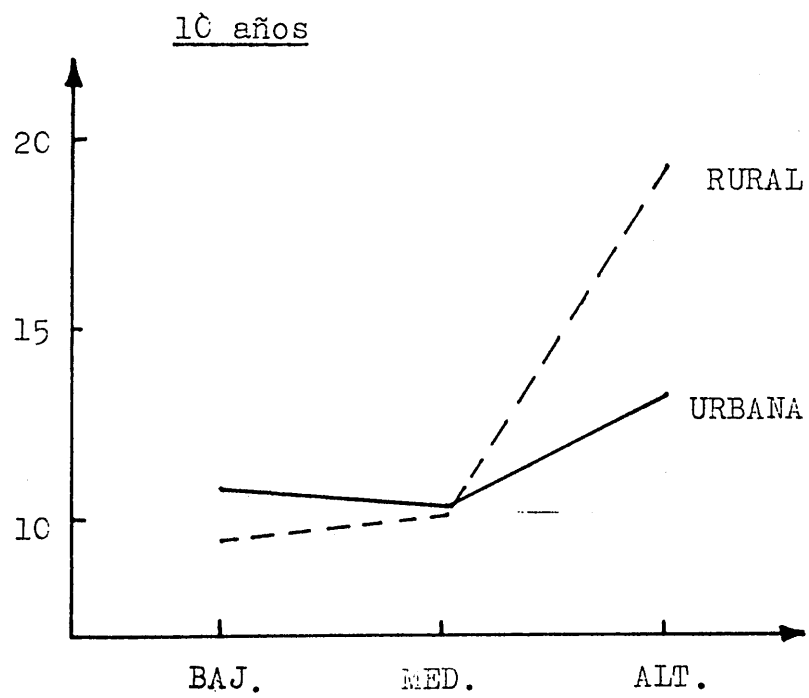
Gráf. 33



Gráf. 34



Gráf. 35



8.2.5.- ANALISIS DE VARIANZA: 5ª Variable Dependiente (V9) PRONOMBRES.-

```
* * * * * * * * * * * A N A L Y S I S   O F   V A R I A N C E * * * * * * * * * *
                V9        PRONOMBRES
            BY  V1        EDAD
                V2        SEXO
                V3        ZONA
                V4        CLASE SOCIAL
* * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * *
```

SOURCE OF VARIATION			SUM OF SQUARES	DF	MEAN SQUARE	F	SIGNIF OF F
MAIN EFFECTS			17388.835	8	2173.604	62.811	.001
V1			15570.227	4	3892.557	112.483	.001
V2			1216.303	1	1216.303	35.147	.001
V3			280.245	1	280.245	8.098	.005
V4			379.866	2	189.943	5.489	.004
2-WAY INTERACTIONS			2282.713	21	108.701	3.141	.001
V1	V2		573.542	4	143.386	4.143	.003
V1	V3		528.399	4	132.100	3.817	.004
V1	V4		1041.441	8	130.180	3.762	.001
V2	V3		54.132	1	54.132	1.564	.208
V2	V4		63.752	2	31.876	.921	.999
V3	V4		56.423	2	28.211	.815	.999
3-WAY INTERACTIONS			2123.863	22	96.539	2.790	.001
V1	V2	V3	1307.032	4	326.758	9.442	.001
V1	V2	V4	275.391	8	34.424	.995	.999
V1	V3	V4	557.679	8	69.710	2.014	.041
V2	V3	V4	24.300	2	12.150	.351	.999
4-WAY INTERACTIONS			493.778	8	61.722	1.784	.075
V1	V2	V3	493.778	8	61.722	1.784	.075
	V4						
EXPLAINED			22289.188	59	377.783	10.917	.001
RESIDUAL			72879.688	2106	34.606		
TOTAL			95168.877	2165	43.958		

2166 CASES WERE PROCESSED.
0 CASES (.0 PCT) WERE MISSING.

* * * M U L T I P L E C L A S S I F I C A T I O N A N A L Y S I S * * *

V9 PRONOMBRES

BY V1 EDAD

V2 SEXO

V3 ZONA

V4 CLASE SOCIAL

* *

GRAND MEAN = 7.49

VARIABLE + CATEGORY	N	UNADJUSTED DEV'N	UNADJUSTED ETA	ADJUSTED FOR INDEPENDENTS DEV'N	ADJUSTED FOR INDEPENDENTS BETA	ADJUSTED FOR INDEPENDENTS + COVARIATES DEV'N	ADJUSTED FOR INDEPENDENTS + COVARIATES BETA
V1							
1 SEIS	402	-4.62		-4.63			
2 SIETE	326	-1.11		-1.15			
3 OCHO	380	-.52		-.52			
4 NUEVE	510	1.39		1.35			
5 DIEZ	546	3.11		3.19			
			.40		.41		
V2							
1 NINO	1086	-.77		-.75			
2 NINA	1080	.77		.75			
			.12		.11		
V3							
1 URBANA	1316	.30		.31			
2 RURAL	850	-.46		-.48			
			.06		.06		
V4							
1 ALTA	174	1.17		1.41			
2 MEDIA	418	-.35		-.26			
3 BAJA	1574	-.04		-.09			
			.06		.06		

MULTIPLE R SQUARED	.183
MULTIPLE R	.427

V 9 - PRONOMBRES.- Tabla de Medias.-

- EFECTOS PRINCIPALES:

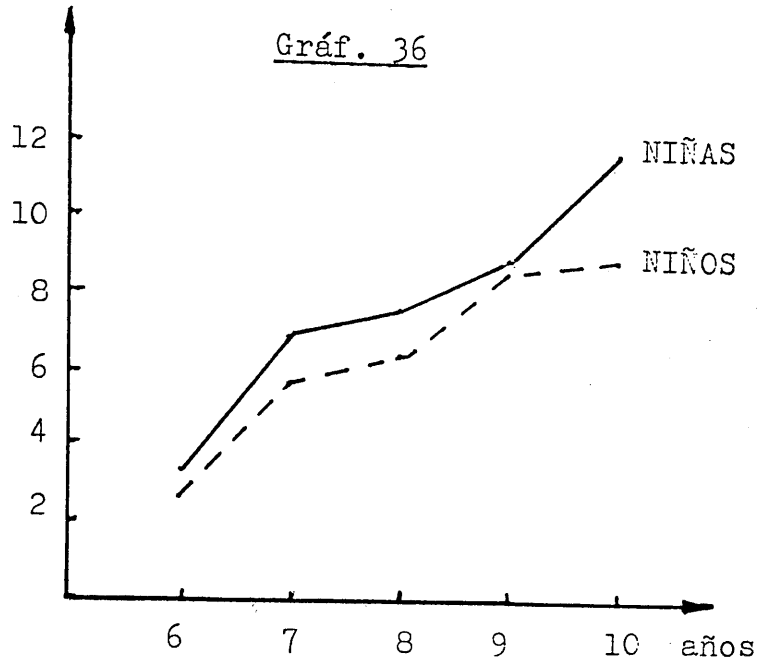
VARIABLE	NIVEL	N	$\bar{X}$
V1	SEIS	402	2.86
	SIETE	326	6.34
	OCHO	380	6.97
	NUEVE	510	8.84
	DIEZ	548	10.68
V2	NIÑOS	1086	6.74
	NIÑAS	1080	8.24
V3	URBANA	1316	7.80
	RURAL	850	7.01
V4	ALTA	174	8.90
	MEDIA	418	7.23
	BAJA	1574	7.40

- INTERACCIONES DE PRIMER ORDEN:

a) V1 - V2: EDAD - SEXO.

V1-V2	SEIS		SIETE		OCHO		NUEVE		DIEZ	
	N	$\bar{X}$	N	$\bar{X}$	N	$\bar{X}$	N	$\bar{X}$	N	$\bar{X}$
NIÑOS	204	2.44	163	5.78	190	6.45	255	8.51	274	9.01
NIÑAS	198	3.32	163	6.99	190	7.50	255	9.25	274	12.19

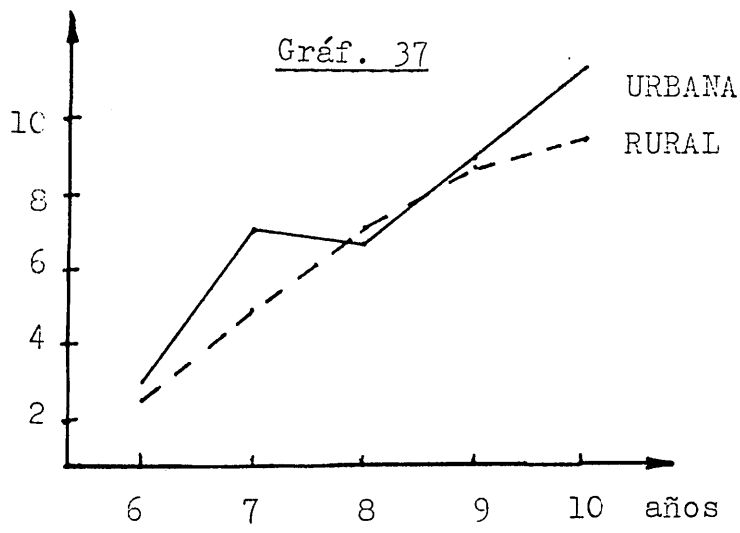
Gráfica de la Interacción Edad - Sexo:



b) V1 - V3: EDAD - ZONA.

V1-V3	SEIS N	SEIS $\bar{X}$	SIETE N	SIETE $\bar{X}$	OCHO N	OCHO $\bar{X}$	NUEVE N	NUEVE $\bar{X}$	DIEZ N	DIEZ $\bar{X}$
URBANA	242	3.11	224	7.06	222	6.85	312	8.92	316	11.44
RURAL	160	2.52	102	4.91	158	7.14	198	8.81	232	9.46

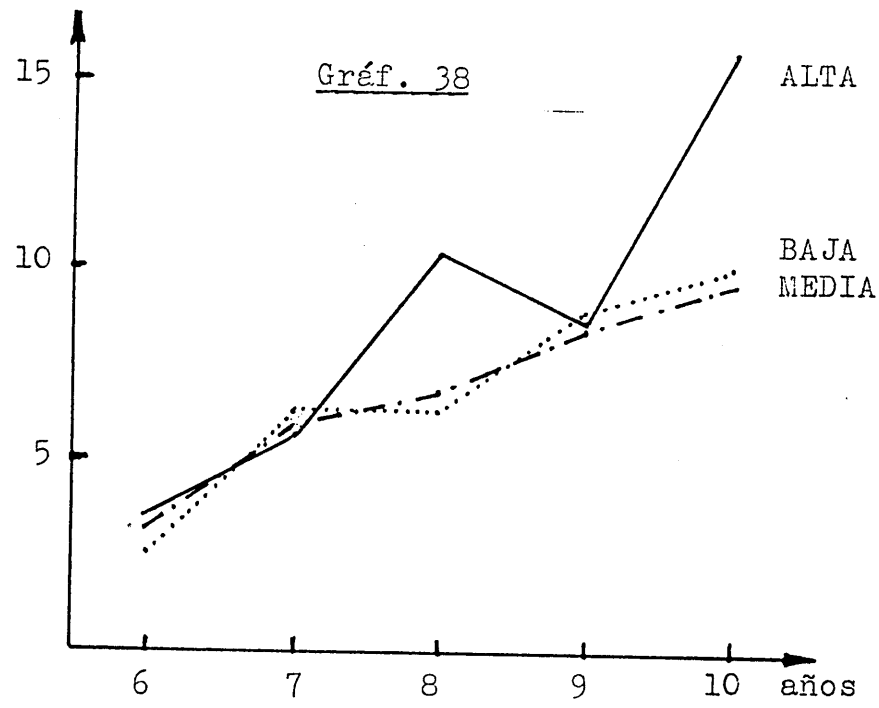
Gráfica de la Interacción Edad - Zona:



c) V1 - V4: EDAD - CLASE SOCIAL.

V1-V4	SEIS N	SEIS $\bar{X}$	SIETE N	SIETE $\bar{X}$	OCHO N	OCHO $\bar{X}$	NUEVE N	NUEVE $\bar{X}$	DIEZ N	DIEZ $\bar{X}$
ALTA	40	3.45	23	6.17	35	10.77	52	8.81	24	16.33
MEDIA	88	3.18	65	6.26	73	6.99	93	8.46	99	10.11
BAJA	274	2.69	238	6.44	272	6.48	365	9.00	425	10.39

Gráfica de la Interacción Edad - Clase Social:



- INTERACCIONES DE 2º ORDEN:

a) V1 - V2 - V3: EDAD - SEXO - ZONA.

1) 6 años:

V2 - V3	URBANA N	URBANA $\bar{X}$	RURAL N	RURAL $\bar{X}$
NIÑOS	124	2.69	80	2.05
NIÑAS	118	3.54	80	2.99

2) 7 años:

V2 - V3	URBANA N	$\bar{X}$	RURAL N	$\bar{X}$
NIÑOS	112	6.36	51	4.49
NIÑAS	112	7.75	51	5.33

3) 8 años:

V2 - V3	URBANA N	$\bar{X}$	RURAL N	$\bar{X}$
NIÑOS	111	6.72	79	6.06
NIÑAS	111	6.98	79	8.23

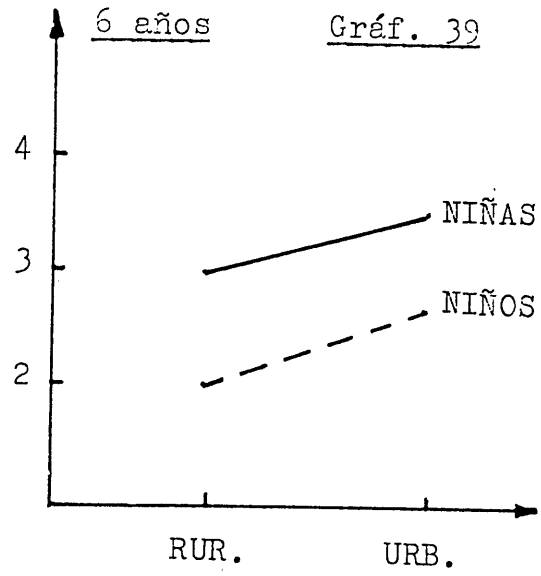
4) 9 años:

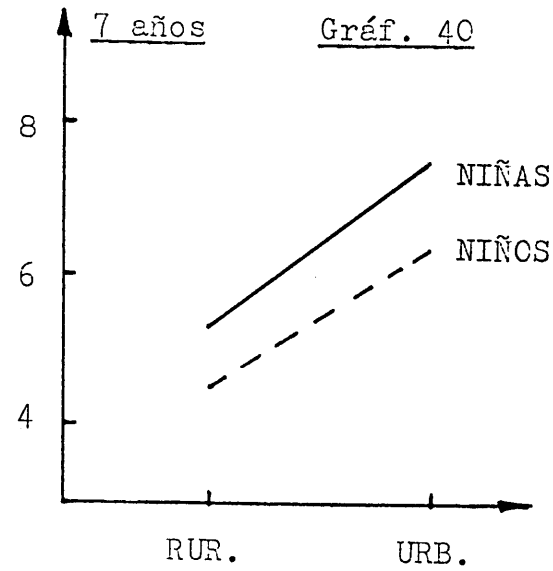
V2 - V3	URBANA N	$\bar{X}$	RURAL N	$\bar{X}$
NIÑOS	156	9.52	99	6.92
NIÑAS	156	8.33	99	10.71

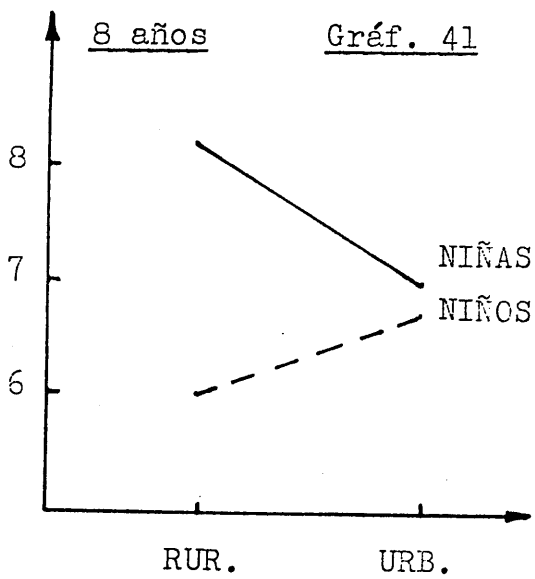
5) 10 años:

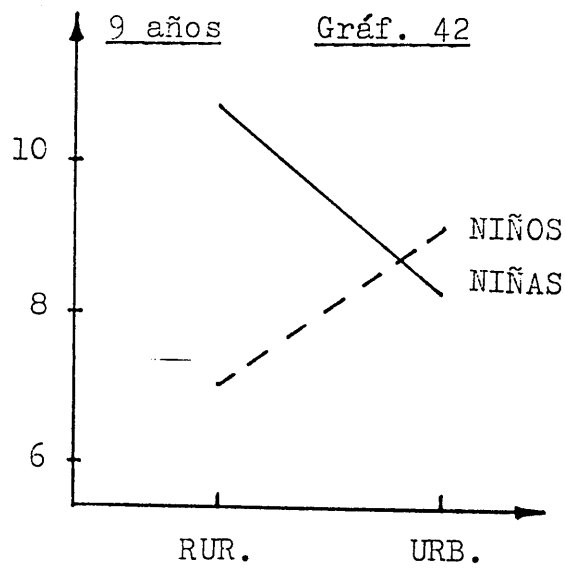
V2 - V3	URBANA N	$\bar{X}$	RURAL N	$\bar{X}$
NIÑOS	158	9.15	116	8.82
NIÑAS	158	13.73	116	10.10

Gráfica de la Interacción de 2º Orden: a):

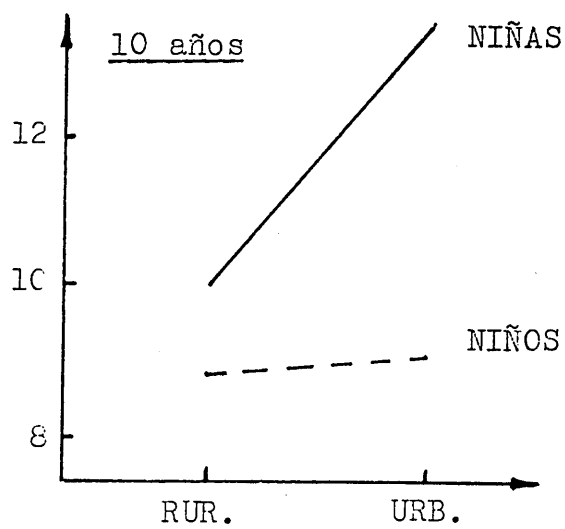


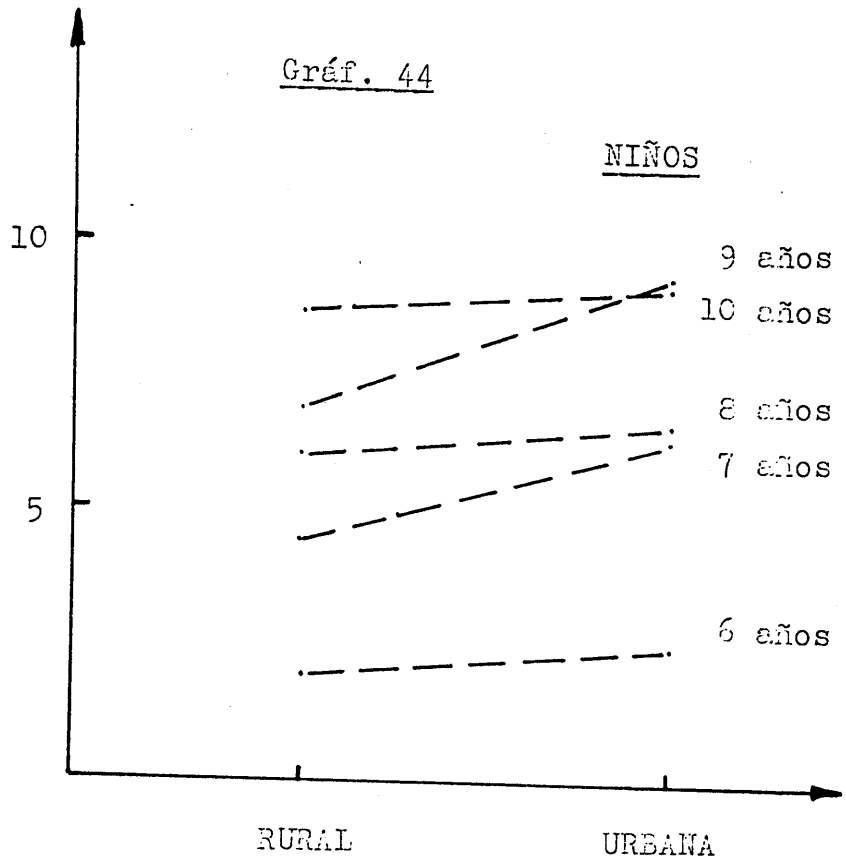


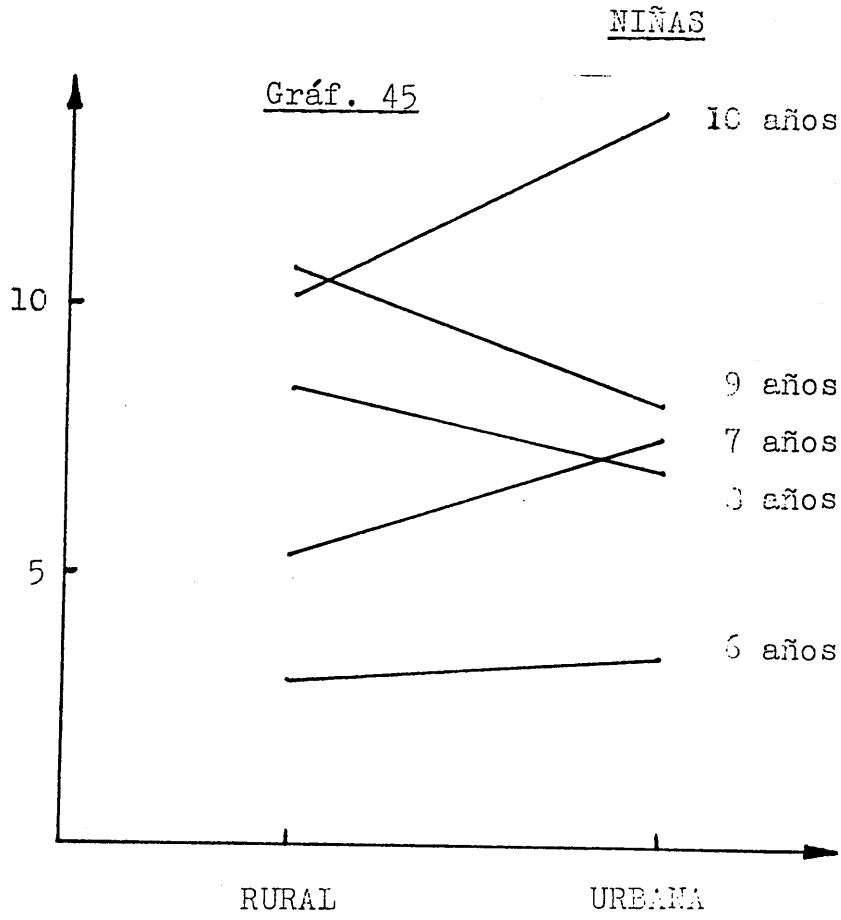




Gráf. 43




Gráf. 44
NIÑOS
10
5
9 años
10 años
8 años
7 años
6 años
RURAL
URBANA


NIÑAS
Gráf. 45
10
5
10 años
9 años
7 años
8 años
6 años
RURAL
URBANA

b) V1 - V3 - V4: EDAD - ZONA - CLASE SOCIAL.

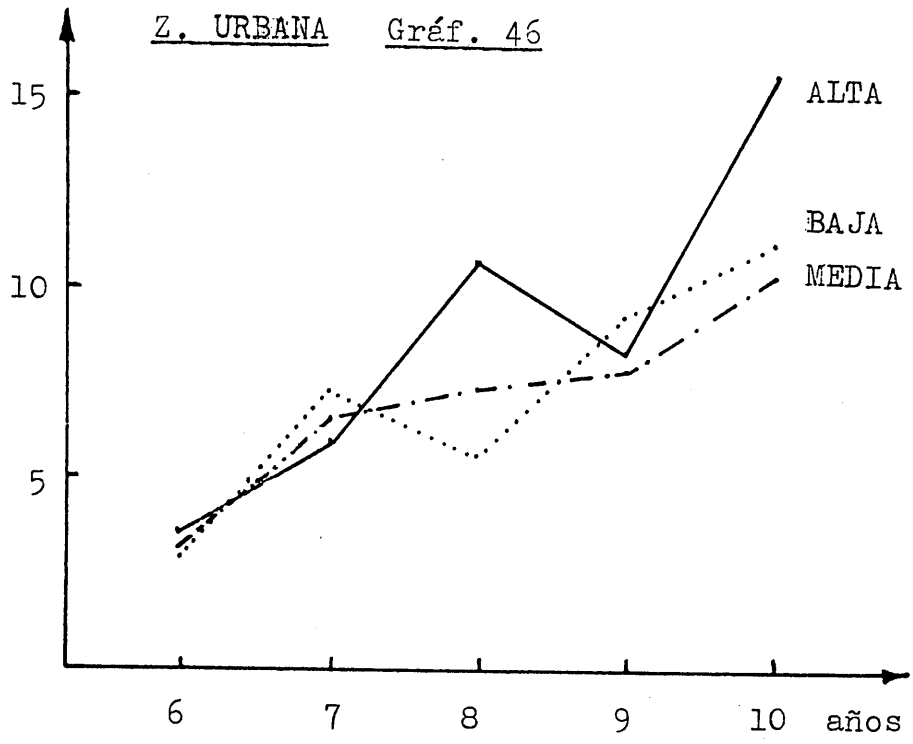
1) Zona URBANA:

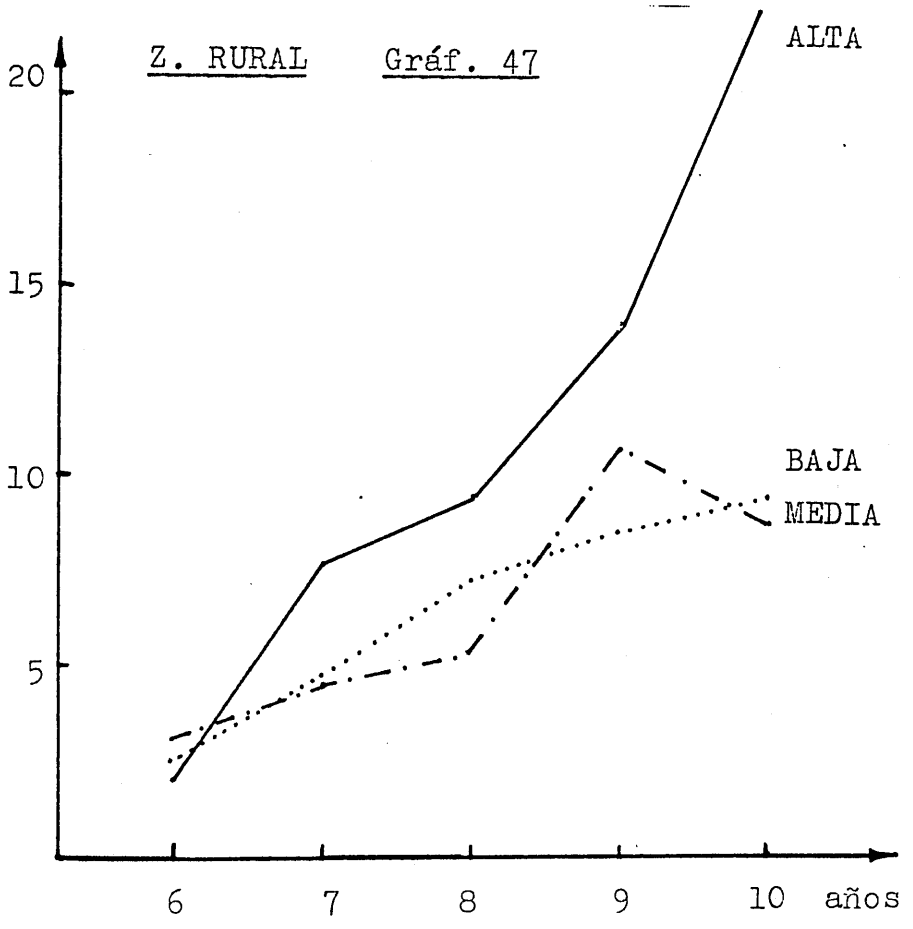
V3-V1	SEIS N	$\bar{X}$	SIETE N	$\bar{X}$	OCHO N	$\bar{X}$	NUEVE N	$\bar{X}$	DIEZ N	$\bar{X}$
ALTA	36	3.61	20	5.95	30	10.97	49	8.49	22	15.73
MEDIA	71	3.20	55	6.62	62	7.31	83	8.08	87	10.30
BAJA	135	2.92	149	7.37	130	5.68	180	9.43	207	11.46

2) Zona RURAL:

V3-V1	SEIS N	$\bar{X}$	SIETE N	$\bar{X}$	OCHO N	$\bar{X}$	NUEVE N	$\bar{X}$	DIEZ N	$\bar{X}$
ALTA	4	2.00	3	7.67	5	9.60	3	14.00	2	23.00
MEDIA	17	3.12	10	4.30	11	5.18	10	11.60	12	8.75
BAJA	139	2.46	89	4.89	142	7.21	185	8.59	218	9.38

Gráfica de la Interacción de 2º Orden: b):





8.2.6.- ANALISIS DE VARIANZA: 6ª Variable Dependiente (V10) ARTICULOS.-

```
* * * * * * * * * * A N A L Y S I S   O F   V A R I A N C E * * * * * * * * * *
                V10      ARTICULOS
            BY  V1       EDAD
                V2       SEXO
                V3       ZONA
                V4       CLASE SOCIAL
* * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * *
```

SOURCE OF VARIATION			SUM OF SQUARES	DF	MEAN SQUARE	F	SIGNIF OF F
MAIN EFFECTS			41326.033	8	5165.754	105.722	.001
V1			40572.590	4	10143.147	207.589	.001
V2			4.221	1	4.221	.086	.999
V3			653.137	1	653.137	13.367	.001
V4			332.402	2	166.201	3.401	.033
2-WAY INTERACTIONS			2728.204	21	129.914	2.659	.001
V1	V2		209.705	4	52.426	1.073	.369
V1	V3		1233.530	4	308.383	6.311	.001
V1	V4		1023.325	8	127.916	2.618	.008
V2	V3		42.239	1	42.239	.864	.999
V2	V4		126.895	2	63.447	1.299	.272
V3	V4		233.389	2	116.695	2.388	.090
3-WAY INTERACTIONS			1697.008	22	77.137	1.579	.042
V1	V2	V3	115.118	4	28.780	.589	.999
V1	V2	V4	788.871	8	98.609	2.018	.041
V1	V3	V4	267.276	8	33.409	.684	.999
V2	V3	V4	353.992	2	176.996	3.622	.026
4-WAY INTERACTIONS			352.527	8	44.066	.902	.999
V1	V2	V3	352.527	8	44.066	.902	.999
	V4						
EXPLAINED			46103.771	59	781.420	15.993	.001
RESIDUAL			102902.539	2106	48.862		
TOTAL			149006.311	2165	68.825		

```
   2166 CASES WERE PROCESSED.
      0 CASES (    .0 PCT) WERE MISSING.
```

* * * M U L T I P L E C L A S S I F I C A T I O N A N A L Y S I S * * *

V10 ARTICULOS

BY V1 EDAD

V2 SEXO

V3 ZONA

V4 CLASE SOCIAL

* *

GRAND MEAN = 12.85

VARIABLE + CATEGORY	N	UNADJUSTED DEV'N	UNADJUSTED ETA	ADJUSTED FOR INDEPENDENTS DEV'N	ADJUSTED FOR INDEPENDENTS BETA	ADJUSTED FOR INDEPENDENTS + COVARIATES DEV'N	ADJUSTED FOR INDEPENDENTS + COVARIATES BETA
V1							
1 SEIS	402	-6.73		-6.77			
2 SIETE	326	-3.64		-3.73			
3 OCHO	380	-.68		-.67			
4 NUEVE	510	2.97		2.94			
5 DIEZ	548	4.81		4.91			
			.52		.52		
V2							
1 NINO	1086	-.05		-.04			
2 NINA	1080	.05		.04			
			.01		.01		
V3							
1 URBANA	1316	.45		.47			
2 RURAL	850	-.69		-.73			
			.07		.07		
V4							
1 ALTA	174	.98		1.25			
2 MEDIA	418	.24		.25			
3 BAJA	1574	-.17		-.20			
			.04		.05		
MULTIPLE R SQUARED				.277			
MULTIPLE R				.527			

V 10 - ARTICULOS.- Tabla de Medias.-

- EFECTOS PRINCIPALES:

VARIABLE	NIVEL	N	$\bar{X}$
V1	SEIS	402	6.08
	SIETE	326	9.12
	OCHO	380	12.18
	NUEVE	510	15.79
	DIEZ	548	17.76
V2(*)	NIÑOS	1086	12.81
	NIÑAS	1080	12.89
V3	URBANA	1316	13.32
	RURAL	850	12.12
V4	ALTA	174	14.10
	MEDIA	418	13.10
	BAJA	1574	12.65

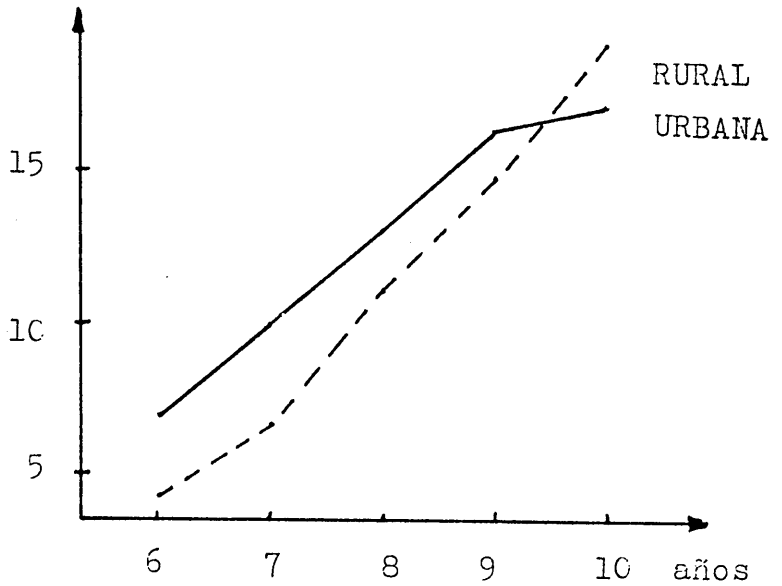
(*) El Efecto Principal del Sexo no es significativo.

- INTERACCIONES DE PRIMER ORDEN:

a) V1 - V3: EDAD - ZONA.

V1-V3	SEIS		SIETE		OCHO		NUEVE		DIEZ	
	N	$\bar{X}$	N	$\bar{X}$	N	$\bar{X}$	N	$\bar{X}$	N	$\bar{X}$
URB.	242	7.32	224	10.11	222	12.93	312	16.37	316	17.32
RUR.	160	4.28	102	7.22	158	11.08	198	14.94	232	18.11

Gráfica de la Interacción Edad - Zona:

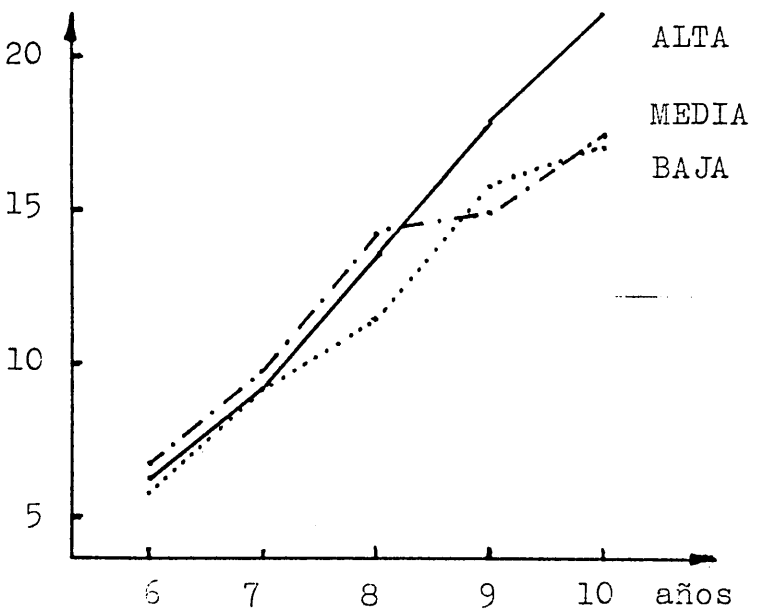


Gráf. 48

b) V1 - V4: EDAD - CLASE SOCIAL.

V1-V4	SEIS N	SEIS $\bar{X}$	SIETE N	SIETE $\bar{X}$	OCHO N	OCHO $\bar{X}$	NUEVE N	NUEVE $\bar{X}$	DIEZ N	DIEZ $\bar{X}$
ALTA	40	6.37	23	9.04	35	13.86	52	17.88	24	21.96
MEDIA	88	7.10	65	9.86	73	14.45	93	15.10	99	17.64
BAJA	274	5.76	238	9.04	272	11.34	365	15.71	425	17.64

Gráfica de la Interacción Edad - Clase Social:



Gráf. 49

- INTERACCIONES DE 2º ORDEN:

a) V2 - V3 - V4: SEXO - ZONA - CLASE SOCIAL.

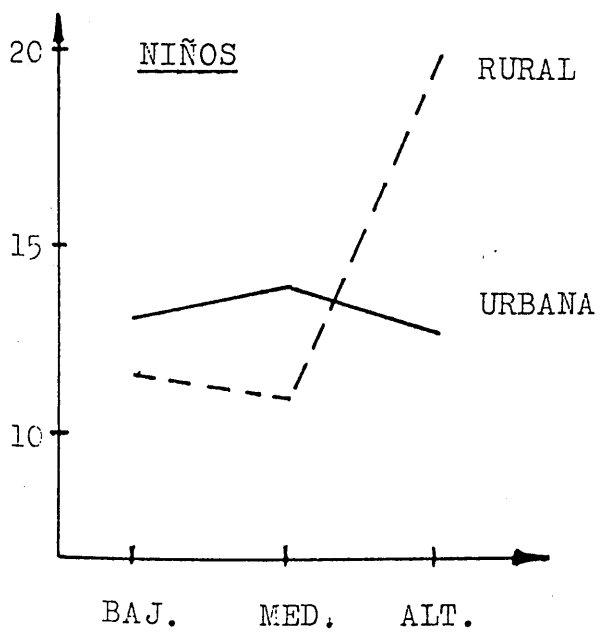
1) Niños:

V3 - V4	ALTA N	ALTA $\bar{X}$	MEDIA N	MEDIA $\bar{X}$	BAJA N	BAJA $\bar{X}$
URBANA	79	12.76	205	13.97	377	13.24
RURAL	8	20.25	30	11.03	387	11.73

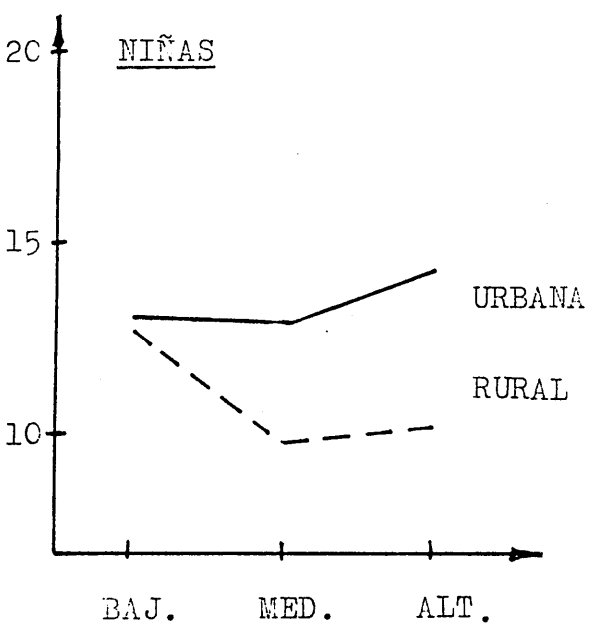
2) Niñas:

V3 - V4	ALTA N	ALTA $\bar{X}$	MEDIA N	MEDIA $\bar{X}$	BAJA N	BAJA $\bar{X}$
URBANA	78	14.65	153	12.92	424	12.99
RURAL	9	10.22	30	9.93	386	12.72

Gráficas de la Interacción de 2º Orden: a):



Gráf. 50



Gráf. 51

b) V1 - V2 - V4: EDAD - SEXO - CLASE SOCIAL.

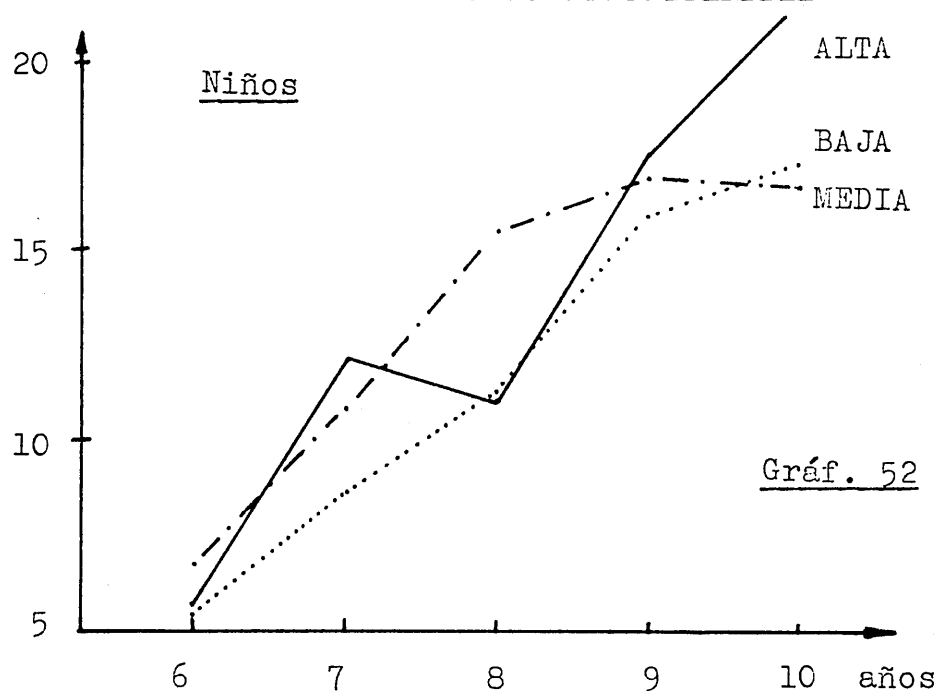
1) Niños:

V4-V1	SEIS N	SEIS $\bar{X}$	SIETE N	SIETE $\bar{X}$	OCHO N	OCHO $\bar{X}$	NUEVE N	NUEVE $\bar{X}$	DIEZ N	DIEZ $\bar{X}$
ALTA	24	5.54	10	12.20	15	11.07	23	17.69	15	22.80
MEDIA	53	6.79	32	10.87	41	15.63	49	17.06	60	16.83
BAJA	127	5.20	121	8.67	134	11.29	183	15.76	199	17.21

2) Niñas:

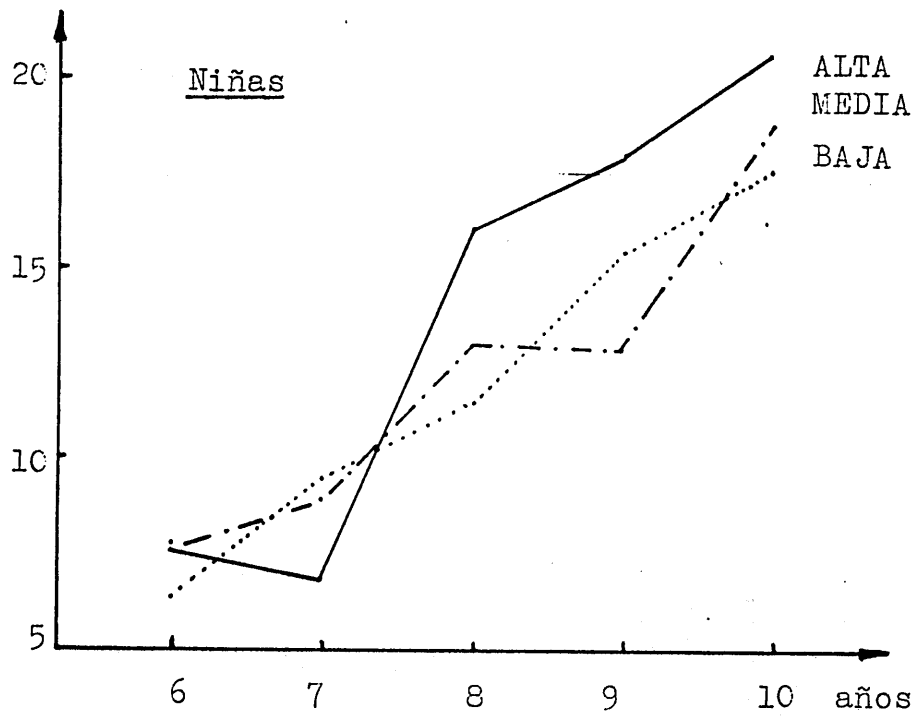
V4-V1	SEIS N	SEIS $\bar{X}$	SIETE N	SIETE $\bar{X}$	OCHO N	OCHO $\bar{X}$	NUEVE N	NUEVE $\bar{X}$	DIEZ N	DIEZ $\bar{X}$
ALTA	16	7.62	13	6.61	20	15.95	29	18.03	9	20.55
MEDIA	35	7.54	33	8.88	32	12.94	44	12.91	39	18.87
BAJA	147	6.24	117	9.42	138	11.38	182	15.65	226	17.61

Gráficas de la Interacción de 2º Orden: b):



Gráf. 52

Gráf. 53



8.2.7.- ANALISIS DE VARIANZA: 7ª VARIABLE Dependiente (V11) PREPOSICIONES.-

```
* * * * * * * * * * * A N A L Y S I S   O F   V A R I A N C E * * * * * * * * * *
                V11        PREPOSICIONES
            BY  V1         EDAD
                V2         SEXO
                V3         ZONA
                V4         CLASE SOCIAL
* * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * *
```

SOURCE OF VARIATION			SUM OF SQUARES	DF	MEAN SQUARE	F	SIGNIF OF F
MAIN EFFECTS			39487.512	8	4935.939	135.146	.001
V1			39354.027	4	9838.507	269.378	.001
V2			69.894	1	69.894	1.914	.163
V3			29.763	1	29.763	.815	.999
V4			296.425	2	148.212	4.058	.017
2-WAY INTERACTIONS			949.016	21	45.191	1.237	.208
V1	V2		92.273	4	23.068	.632	.999
V1	V3		209.225	4	52.306	1.432	.220
V1	V4		449.256	8	56.157	1.538	.138
V2	V3		64.424	1	64.424	1.764	.181
V2	V4		60.101	2	30.051	.823	.999
V3	V4		136.859	2	68.430	1.874	.151
3-WAY INTERACTIONS			1126.682	22	51.213	1.402	.101
V1	V2	V3	175.747	4	43.937	1.203	.307
V1	V2	V4	294.421	8	36.803	1.008	.428
V1	V3	V4	707.656	8	88.457	2.422	.013
V2	V3	V4	.297	2	.148	.004	.999
4-WAY INTERACTIONS			164.759	8	20.595	.564	.999
V1	V2	V3	164.759	8	20.595	.564	.999
	V4						
EXPLAINED			41727.969	59	707.254	19.365	.001
RESIDUAL			76917.505	2106	36.523		
TOTAL			118645.474	2165	54.802		

2166 CASES WERE PROCESSED.
0 CASES (.0 PCT) WERE MISSING.

* * * M U L T I P L E C L A S S I F I C A T I O N A N A L Y S I S * * *

V11 PREPOSICIONES

BY V1 EDAD

V2 SEXO

V3 ZONA

V4 CLASE SOCIAL

* *

GRAND MEAN = 9.90

VARIABLE + CATEGORY	N	UNADJUSTED DEV'N	UNADJUSTED ETA	ADJUSTED FOR INDEPENDENTS DEV'N	ADJUSTED FOR INDEPENDENTS EETA	ADJUSTED FOR INDEPENDENTS + COVARIATES DEV'N	ADJUSTED FOR INDEPENDENTS + COVARIATES BETA
V1							
1 SEIS	402	-6.49		-6.52			
2 SIETE	326	-3.21		-3.22			
3 OCHO	380	-1.15		-1.16			
4 NUEVE	510	2.29		2.27			
5 DIEZ	548	5.33		5.39			
			.57		.58		
V2							
1 NINO	1086	-.19		-.18			
2 NINA	1080	.19		.18			
			.03		.02		
V3							
1 URBANA	1316	.08		.10			
2 RURAL	850	-.13		-.16			
			.01		.02		
V4							
1 ALTA	174	.48		1.14			
2 MEDIA	418	.05		.30			
3 BAJA	1574	-.06		-.20			
			.02		.05		

MULTIPLE R SQUARED .333

MULTIPLE R .577

V 11 - PREPOSICIONES.- Tabla de Medias.-

- EFECTOS PRINCIPALES:

VARIABLE	NIVEL	N	$\bar{X}$
V1	SEIS	402	3.38
	SIETE	326	6.68
	OCHO	380	8.74
	NUEVE	510	12.17
	DIEZ	548	15.29
V2(+)	NIÑOS	1086	9.72
	NIÑAS	1080	10.08
V3(*)	URBANA	1316	10.00
	RURAL	850	9.74
V4	ALTA	174	11.04
	MEDIA	418	10.20
	BAJA	1574	9.70

(*) El Efecto Principal de Zona no es significativo.

(+) El Efecto Principal del Sexo no es significativo.

* * * * * * * * * * A N A L Y S I S O F V A R I A N C E * * * * * * * * * *

V12 CONJUNCIONES
BY V1 EDAD
V2 SEXO
V3 ZONA
V4 CLASE SOCIAL

SOURCE OF VARIATION			SUM OF SQUARES	DF	MEAN SQUARE	F	SIGNIF OF F
MAIN EFFECTS			28430.031	8	3553.754	101.467	.001
V1			27473.051	4	6868.263	196.102	.001
V2			539.196	1	539.196	15.395	.001
V3			106.010	1	106.010	3.027	.078
V4			105.025	2	52.513	1.499	.222
2-WAY INTERACTIONS			1574.540	21	73.073	2.086	.003
V1	V2		272.430	4	68.107	1.945	.099
V1	V3		340.153	4	85.038	2.428	.045
V1	V4		689.227	8	86.153	2.460	.012
V2	V3		29.327	1	29.327	.837	.999
V2	V4		75.775	2	37.888	1.082	.340
V3	V4		79.394	2	39.697	1.133	.323
3-WAY INTERACTIONS			2132.453	22	96.930	2.768	.001
V1	V2	V3	338.221	4	84.555	2.414	.046
V1	V2	V4	953.883	8	119.235	3.404	.001
V1	V3	V4	868.824	8	108.603	3.101	.002
V2	V3	V4	38.072	2	19.036	.544	.999
4-WAY INTERACTIONS			169.489	8	21.186	.605	.999
V1	V2	V3	169.489	8	21.186	.605	.999
	V4						
EXPLAINED			32266.513	59	546.890	15.615	.001
RESIDUAL			73760.266	2106	35.024		
TOTAL			106026.778	2165	48.973		

2166 CASES WERE PROCESSED.
0 CASES (.0 PCT) WERE MISSING.

* * * M U L T I P L E C L A S S I F I C A T I O N A N A L Y S I S * * *

V12 CONJUNCIONES

BY V1 EDAD

V2 SEXO

V3 ZONA

V4 CLASE SOCIAL

* *

GRAND MEAN = 2.46

VARIABLE + CATEGORY	N	UNADJUSTED DEV'N	UNADJUSTED ETA	ADJUSTED FOR INDEPENDENTS DEV'N	ADJUSTED FOR INDEPENDENTS BETA	ADJUSTED FOR INDEPENDENTS + COVARIATES DEV'N	ADJUSTED FOR INDEPENDENTS + COVARIATES BETA
V1							
1 SEIS	402	-6.46		-6.46			
2 SIETE	326	-1.76		-1.71			
3 OCHO	380	-.09		-.11			
4 NUEVE	510	2.65		2.63			
5 DIEZ	548	3.38		3.39			
			.51		.51		
V2							
1 NINO	1086	-.52		-.50			
2 NINA	1080	.53		.50			
			.08		.07		
V3							
1 URBANA	1316	-.25		-.19			
2 RURAL	850	.39		.29			
			.04		.03		
V4							
1 ALTA	174	.00		.62			
2 MEDIA	418	-.74		-.30			
3 BAJA	1574	.20		.01			
			.05		.03		

MULTIPLE R SQUARED .268

MULTIPLE R .518

V 12 - CONJUNCIONES.- Tabla de Medias.-

- EFECTOS PRINCIPALES:

VARIABLE	NIVEL	N	$\bar{X}$
V1	SEIS	402	2.00
	SIETE	326	6.75
	OCHO	380	8.35
	NUEVE	510	11.09
	DIEZ	548	11.83
V2	NIÑOS	1086	7.96
	NIÑAS	1080	8.96
V3(+)	URBANA	1316	8.27
	RURAL	850	8.75
V4(*)	ALTA	174	9.08
	MEDIA	418	8.16
	BAJA	1574	8.47

(+) El Efecto Principal de Zona no es significativo:

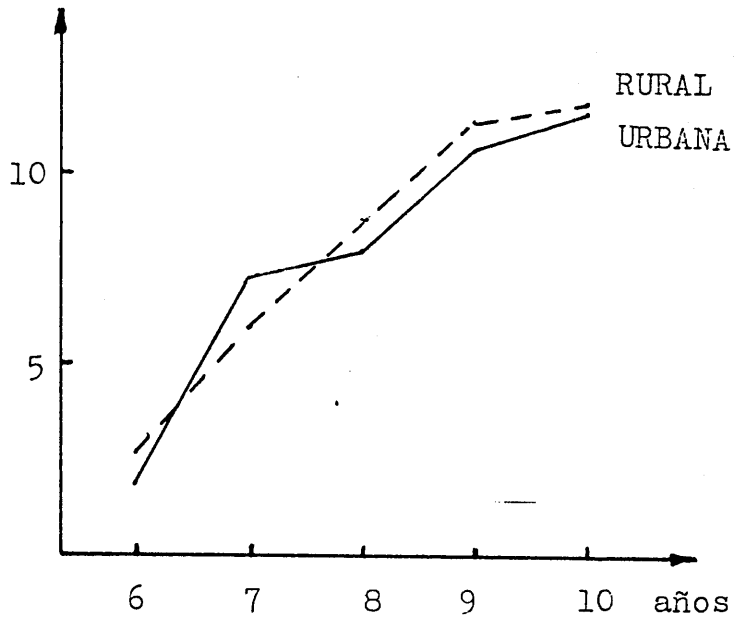
(*) El Efecto Principal de Clase Social no es significativo.

- INTERACCIONES DE PRIMER ORDEN:

a) V1 - V3: EDAD - ZONA.

V1-V3	SEIS N	SEIS $\bar{X}$	SIETE N	SIETE $\bar{X}$	OCHO N	OCHO $\bar{X}$	NUEVE N	NUEVE $\bar{X}$	DIEZ N	DIEZ $\bar{X}$
URB.	242	1.63	224	7.11	222	8.00	312	10.73	316	11.66
RUR.	160	2.54	102	5.79	158	8.89	198	11.69	232	12.09

Gráfica de la Interacción Edad - Zona:

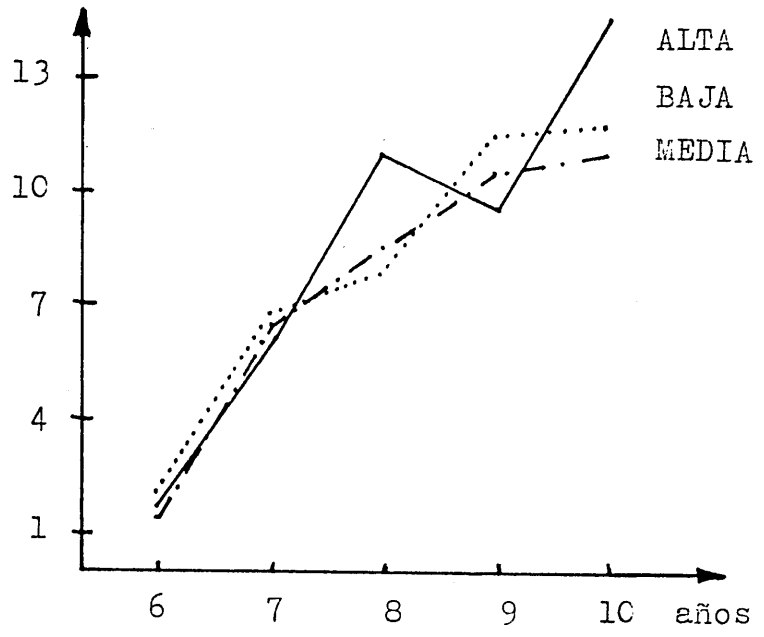


Gráf. 54

b) V1 - V4: EDAD - CLASE SOCIAL.

V4-V1	SEIS N	SEIS $\bar{X}$	SIETE N	SIETE $\bar{X}$	OCHO N	OCHO $\bar{X}$	NUEVE N	NUEVE $\bar{X}$	DIEZ N	DIEZ $\bar{X}$
ALTA	40	1.82	23	6.30	35	11.00	52	9.85	24	14.87
MEDIA	88	1.41	65	6.51	73	8.56	93	10.40	99	10.99
BAJA	274	2.20	238	6.79	272	7.98	365	11.46	425	11.87

Gráfica de la Interacción Edad - Clase Social:



Gráf. 55

- INTERACCIONES DE 2º ORDEN:

a) V1 - V2 - V3: EDAD - SEXO - ZONA.

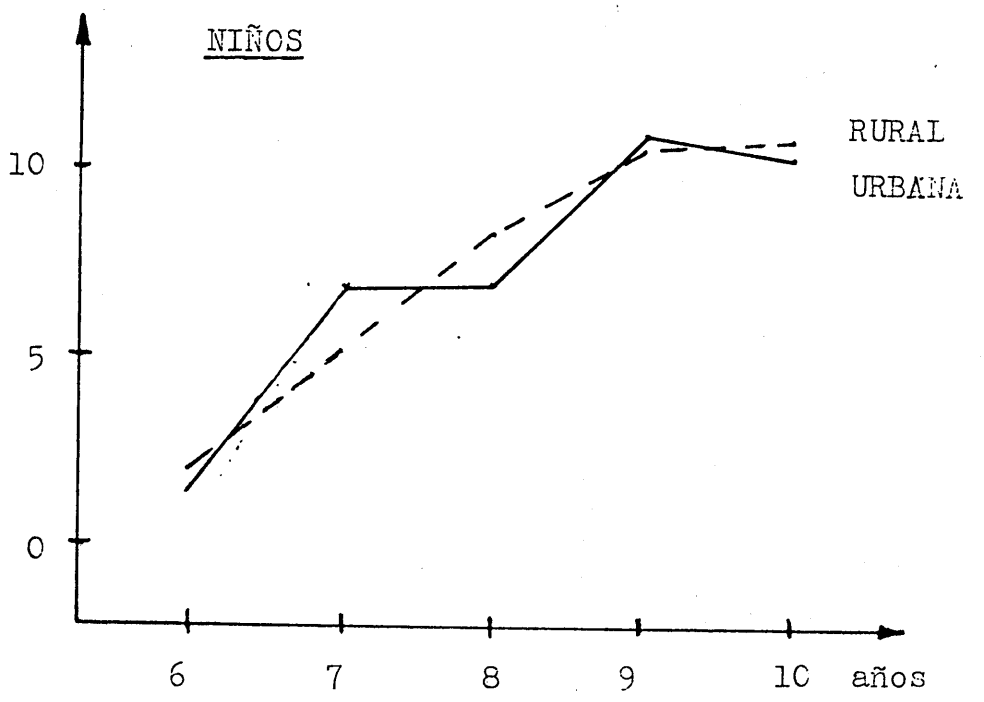
1) Niños:

V3-V1	SEIS N	SEIS $\bar{X}$	SIETE N	SIETE $\bar{X}$	OCHO N	OCHO $\bar{X}$	NUEVE N	NUEVE $\bar{X}$	DIEZ N	DIEZ $\bar{X}$
URB.	124	1.31	112	6.97	111	7.00	156	11.33	158	10.69
RUR.	80	2.01	51	5.21	79	8.35	99	10.70	116	11.13

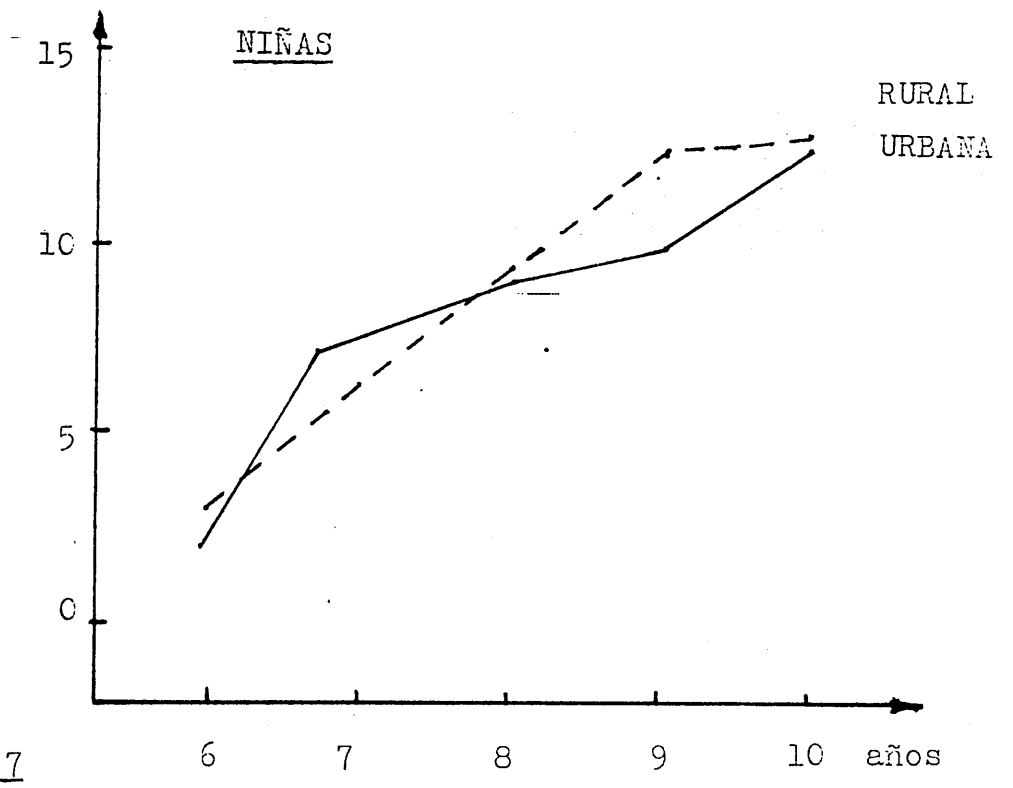
2) Niñas:

V3-V1	SEIS N	SEIS $\bar{X}$	SIETE N	SIETE $\bar{X}$	OCHO N	OCHO $\bar{X}$	NUEVE N	NUEVE $\bar{X}$	DIEZ N	DIEZ $\bar{X}$
URB.	118	1.97	112	7.25	111	9.02	156	10.12	158	12.63
RUR.	80	3.06	51	6.37	79	9.40	99	12.69	116	13.04

Gráficas de la Interacción de 2º Orden: a):



Gráf. 56



Gráf. 57

b) V1 - V2 - V4: EDAD - SEXO - CLASE SOCIAL.

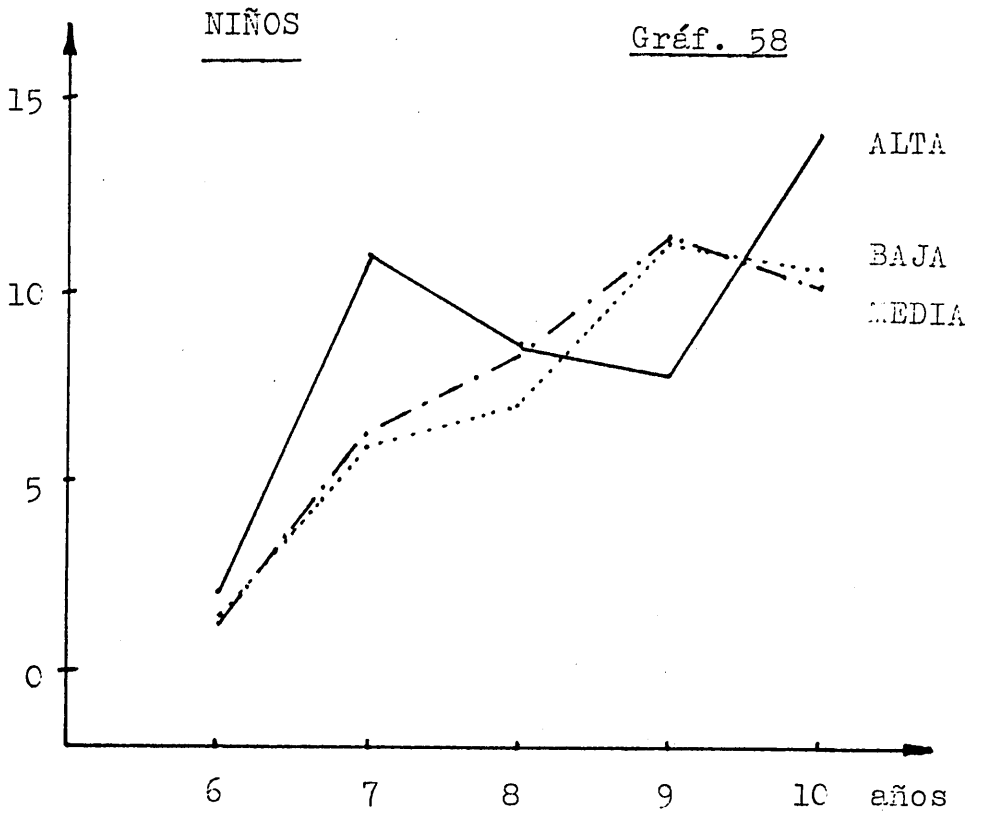
1) Niños:

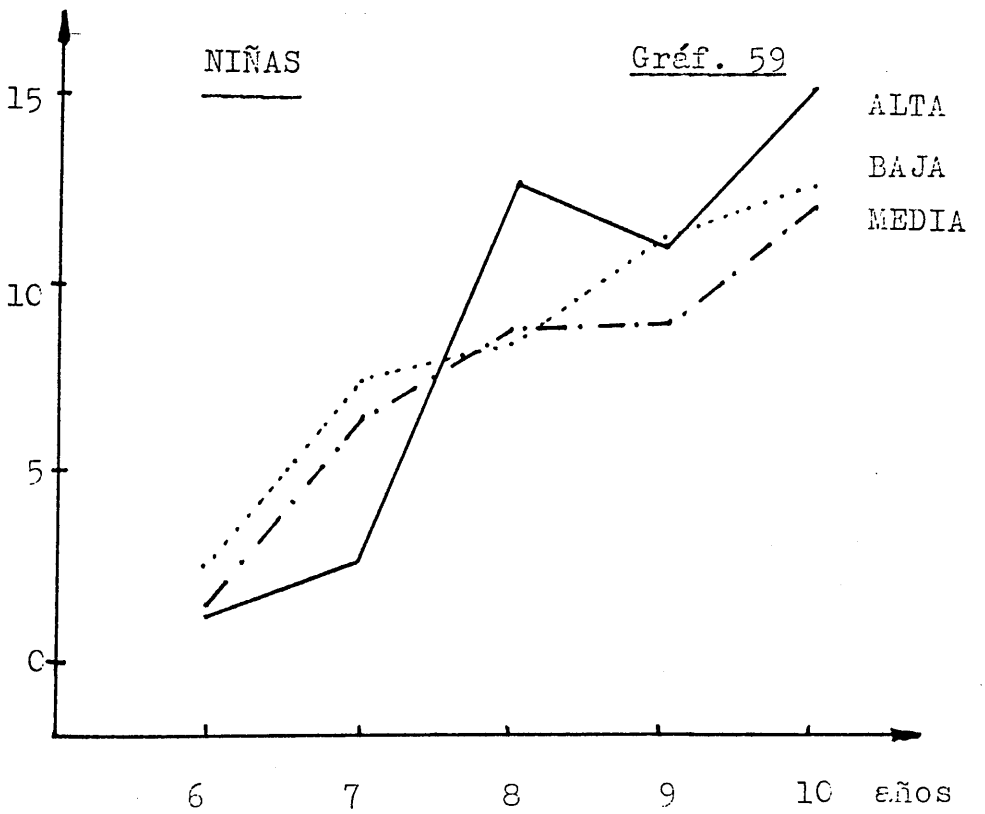
V4-V1	SEIS N	SEIS $\bar{X}$	SIETE N	SIETE $\bar{X}$	OCHO N	OCHO $\bar{X}$	NUEVE N	NUEVE $\bar{X}$	DIEZ N	DIEZ $\bar{X}$
ALTA	24	2.12	10	11.00	15	8.6	23	7.87	15	14.67
MEDIA	53	1.24	32	6.53	41	8.36	49	11.55	60	10.18
BAJA	127	1.62	121	6.02	134	7.20	183	11.37	199	10.80

2) Niñas:

V4-V1	SEIS N	SEIS $\bar{X}$	SIETE N	SIETE $\bar{X}$	OCHO N	OCHO $\bar{X}$	NUEVE N	NUEVE $\bar{X}$	DIEZ N	DIEZ $\bar{X}$
ALTA	16	1.37	13	2.69	20	12.80	29	11.41	9	15.22
MEDIA	35	1.66	33	6.48	32	8.81	44	9.11	39	12.20
BAJA	147	2.71	117	7.59	138	8.74	182	11.55	226	12.81

Gráficas de la Interacción de 2º Orden: b):





c) V1 - V3 - V4: EDAD - ZONA - CLASE SOCIAL.

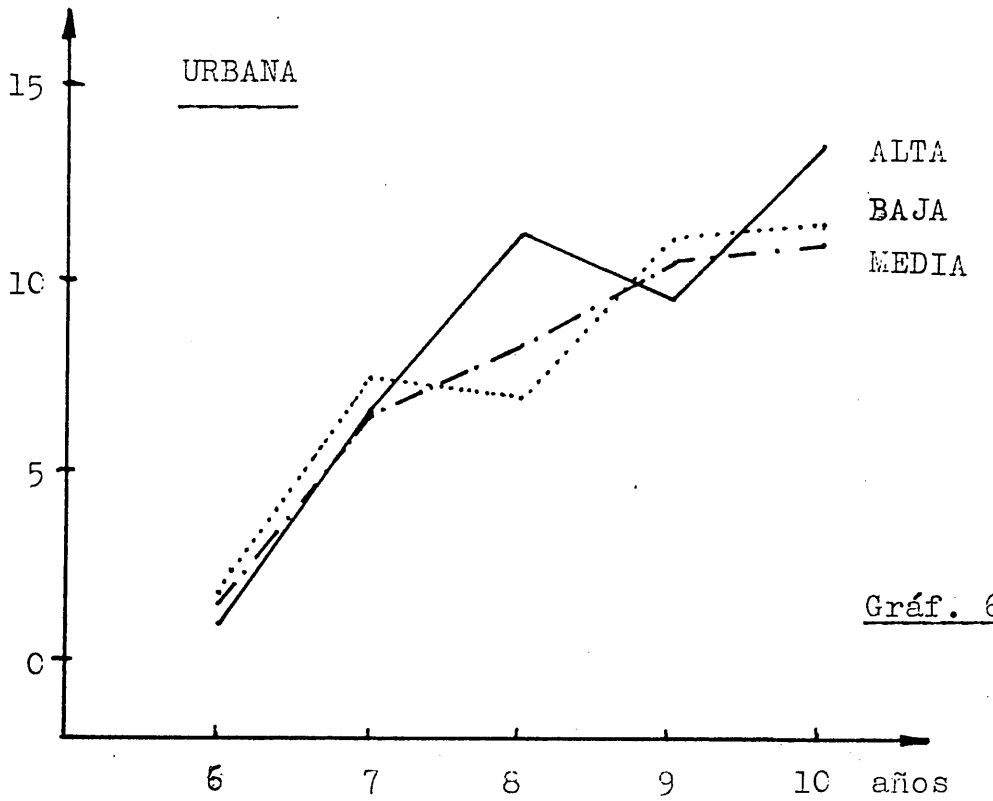
1) Zona URBANA:

V4-V1	SEIS N	SEIS $\bar{X}$	SIETE N	SIETE $\bar{X}$	OCHO N	OCHO $\bar{X}$	NUEVE N	NUEVE $\bar{X}$	DIEZ N	DIEZ $\bar{X}$
ALTA	36	1.78	20	6.35	30	11.33	49	9.43	22	13.41
MEDIA	71	0.97	55	6.34	62	8.74	83	10.43	87	10.94
BAJA	135	1.94	149	7.50	130	6.89	180	11.22	207	11.77

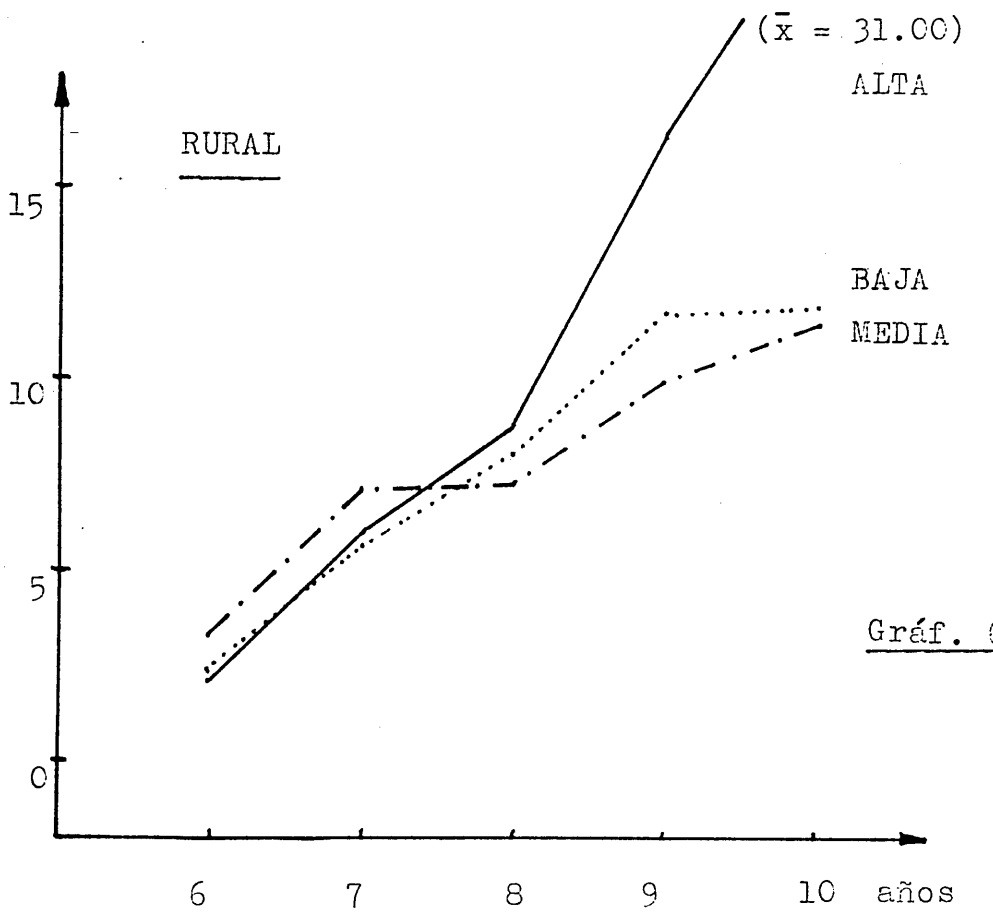
2) Zona RURAL:

V4-V1	SEIS N	SEIS $\bar{X}$	SIETE N	SIETE $\bar{X}$	OCHO N	OCHO $\bar{X}$	NUEVE N	NUEVE $\bar{X}$	DIEZ N	DIEZ $\bar{X}$
ALTA	4	2.25	3	6.00	5	9.00	3	16.67	2	31.00
MEDIA	17	3.23	10	7.40	11	7.54	10	10.10	12	11.25
BAJA	139	2.46	89	5.61	142	8.99	185	11.70	218	11.96

Gráficas de la Interacciones de 2º Orden: c):



Gráf. 60



Gráf. 61

CAPITULO IX

IX.- EVOLUCION CUANTITATIVA DEL VOCABULARIO. ANALISIS DIFERENCIAL (II). DISCUSION DE RESULTADOS Y CONCLUSIONES.

9.1.- Análisis de la 1ª Variable Dependiente: NOMBRES.-

El análisis de varianza (Véase tabla 8.2.1.) muestra la existencia de diferencias significativas en la producción de sustantivos, debidas alos efectos de los cuatro factores o variables independientes. Las cuatro F de efectos principales son significativas al .001. Sin embargo, la proporción de varianza explicada por cada uno de los cuatro factores es muy desigual (tabla MCA de Nombres). Corresponde el mayor porcentaje (45%) a la Edad, siendo de menor consideración los porcentajes correspondientes al Sexo (0.6%), a la Zona (0.6%) y a la Clase Social (0.5%). R^2 = .463.

Un análisis desglosado por factores arroja las siguientes conclusiones:

1º) El efecto principal significativo del factor Edad indica que el rendimiento verbal que los sujetos manifiestan en la frecuencia de uso que hacen de los sustantivos se ajusta a un incremento de tendencia creciente con el aumento de los años de escolarización. Entre los cinco niveles de edad las diferencias resultaron significativas. Las mayores diferencias aparecen entre los 7 y 8 años y los 8 y nueve años. La menor diferencia se detecta entre los seis y siete años.

2º) En cuanto al Sexo, las niñas son superiores a los niños. Ahora bien, la existencia de interacciones significativas de primero y segundo orden en las que interviene la variable Sexo (V2-V3 / V2-V4 / V1-V2-V4) nos obliga a introducir una serie de matices en la interpretación global que acabamos de hacer del efecto principal de dicha variable.

a) En la representación analítica de la interacción V2-V3 (Gráf. 1) observamos que la superioridad de las niñas se mantiene tanto en la zona urbana como en la zona rural, siendo más acusadas las diferencias entre niños y niñas de zona urbana. Por otra parte, los promedios nos indican también que el nivel que manifiestan los niños de zona urbana es semejante al de las niñas de zona rural. De la misma manera, si tenemos en cuenta la clase social (Gráf. 2) observamos que incluso los niños pertenecientes a la clase social más alta están por debajo de la media de las niñas de clase social baja.

b) Ahora bien, el comportamiento entre niños y niñas no es tan evidente si ponemos en juego un nuevo factor: la edad (V1-V2-V4). Así pues, sólo en las clases bajas la superioridad de las niñas sobre los niños se mantiene en todas las edades ocursos. En la clase media, por el contrario, hasta los nueve años, los niños son superiores a las niñas. A esta edad se establece un equilibrio que posteriormente se decanta a favor de las niñas en el último curso. Tampoco en la clase alta podemos confirmar el patrón predominante del sexo femenino: mientras en los cursos extremos (1º y 5º) el predominio corresponde a los niños, en los cursos intermedios la superioridad es de las niñas.

3º) Si tomamos en consideración el factor de Zona observamos (Gráf. 1) que los sujetos pertenecientes a la zona Urbana son superiores a los de zona Rural.

4º) La Clase Social también introduce diferencias significativas. En su mayor parte ya han sido comentadas en apartados anteriores. Hemos de añadir aquí, no obstante, que sólo los índices globales de rendimiento de los sujetos de clase alta son superiores a los de clase media y éstos a su vez se imponen sobre los de clase baja. La diferencia de medias entre las dos últimas es menor que la que se observa entre la clase alta y media. Cuando tomamos en consideración otros factores como la edad o el sexo la jerarquía deja de existir. Esto es debido, por una parte, al descenso en el rendimiento que se produce en las niñas de clase media respecto a las de clase baja (particularmente en 1º y 2º) y, por otra, al que se observa también en los niños de clase alta, sobre todo de 1º y 3º, respecto a los de clase media.

9.2.- Análisis de la 2ª Variable Dependiente: ADJETIVOS.-

En el cuadro resumen de ANOVA y en la tabla MCA para Adjetivos (Cfr. 8.2.2.) encontramos algunas diferencias en relación con las observadas en los Nombres referidas al efecto que producen las variables independientes, particularmente, la Zona.

1º) El coeficiente de determinación en Adjetivos es considerablemente más bajo (R^2 = .243), lo cual nos indica que el porcentaje de varianza explicado por la combinación de los cuatro factores o variables independientes apenas si alcanza la cuarta parte de la varianza total. En coincidencia con los Nombres, continúa siendo el factor Edad el que contribuye en mayor medida a la explicación de la varianza observada (23 %).

2º) El efecto principal de Zona no resultó significativo (F = 1.435; $p < .230$ N.S.). Los sujetos de zona Urbana, aunque manifiestan en el promedio una ligera superioridad sobre los de zona Rural, no difieren significativamente en la frecuencia de uso que hacen del adjetivo.

3º) Existe una relación positiva entre la edad y el incremento en el uso del adjetivo. Las diferencias entre los distintos cursos se incrementan progresivamente.

4º) El índice de rendimiento por Sexo es claramente

favorable a las Niñas. Sin embargo, si realizamos un análisis más minucioso por edades (Interacción V1-V2) observamos que las diferencias sólo se dan realmente a partir de los 7 años (Gráf. 10) y aumentan desde ese momento con el transcurso del tiempo. La superioridad de las niñas sobre los niños, según la zona, también es evidente (Gráf. 12), siendo más acusadas las diferencias en la zona Rural que en la Urbana.

5º) En cuanto a la Clase Social, la jerarquía Alta-Media-Baja sólo es posible observarla a partir de los 8 años. Hasta entonces el predominio corresponde a la clase media. No obstante, las diferencias entre la clase alta y la media no resultan significativas más que a los 10 años. Por el contrario, si comparamos la clase media con la baja los promedios difieren significativamente ya desde los 6 años y aumentan progresivamente con la edad.

9.3.- Análisis de la 3ª Variable Dependiente: VERBOS.-

Como en el caso anterior de los Adjetivos, el efecto de Zona no resultó significativo ($F = 2.366$; $p < .120$ N.S.). La desigualdad de medias entre zonas no es suficientemente alta como para que el efecto de la zona tenga una influencia real en la variable criterio (Verbos). Los efectos principales de las restantes variables son significativos ($p < .001$) así como también varias de las interacciones de primero y segundo orden, por consiguiente:

1º) Los cinco niveles de edad manifiestan entre sí diferencias significativas: la producción de verbos se incrementa significativamente con la edad de los sujetos. La mayor diferencia ocurre entre los 6 y 7 años.

2º) El rendimiento de las niñas es superior al de los niños; las diferencias son mínimas a los 7 y 9 años. Se acentúan sobre todo en el último curso, es decir, a los 10 años (Gráf. 13).

- La interacción Sexo - Zona (V2-V3, Gráf. 15) muestra igualmente que tanto en la zona Urbana como en la Rural la superioridad corresponde a las niñas siendo mayores las diferencias en la segunda que en la primera.

3º) El efecto de Zona, como hemos comentado al principio, no es significativo. Sin embargo, hemos de añadir

que, dada la existencia de interacciones significativas en las que interviene el factor zona (V2-V3 y V1-V3-V4), nos vemos obligados a matizar la interpretación previa.

- En efecto, en la gráfica 15 podemos constatar cómo en función del sexo se produce una inversión de zonas: en los niños la zona urbana es superior a la rural, mientras que en las niñas ocurre lo contrario, aunque en este segundo caso las diferencias ((entre niñas)) no sean significativas.

- Por otra parte, si consideramos la interacción de segundo orden ((Gráfs. 16 a 20)) observamos que:

a)) en las clases bajas, con el transcurso del tiempo tiende a desaparecer la superioridad de la zona urbana sobre la rural, evidente en los dos primeros cursos.

b)) En la clase media ocurre algo semejante a partir de los 9 años.

c)) La clase alta, sin embargo, manifiesta un patrón de comportamiento más irregular en función de la zona: en 6 y 8 años la zona urbana supera a la rural; en los restantes niveles el prodominio corresponde a la zona rural.

4º)) Existe una relación positiva entre las distintas clases sociales y la frecuencia de uso de los Verbos. No obstante, las diferencias interclases que aparecen en los dos primeros niveles o cursos en ningún caso son significativas, de

aquí podemos deducir que a los 6 y 7 años el equilibrio entre las distintas clases sociales no permite establecer ningún tipo de jerarquización definida (Gráf. 14).

El efecto de la clase social en el uso de los Verbos se inicia a partir de los 8 años. A esta edad, se establece una jerarquía social: la clase alta supera en rendimiento a la clase media y ésta se impone, a su vez, sobre la clase baja. A los nueve años se acortan las diferencias y al mismo tiempo se produce una igualación entre la clase media y baja. Esta equiparación entre las dos clases inferiores se mantiene también a los diez años, momento en el que se acentúa el mayor distanciamiento de la clase alta (Gráf. 14).

9.4.- Análisis de la 4ª Variable Dependiente: ADVERBIOS.-

El estudio de la cuarta variable dependiente, Adverbios, arroja unos resultados coincidentes en gran medida con los obtenidos en el estudio de variables anteriores, sobre todo en el caso de los adjetivos y los verbos.

Los coeficientes de regresión parcial (Beta), en función del orden con que han sido incorporados los factores a la ecuación de regresión, indican que la variable que contribuye en mayor medida a la explicación de la varianza observada en el uso de los Adverbios corresponde a la Edad (29.16%). La cantidad de varianza atribuible a las demás variables independientes apenas si es importante (Sexo: 1%; Zona: 0%; Clase Social: 0.6%). Ciertamente, estos procentajes cambiarían si el orden de entrada se hubiese alterado. La contribución relativa de cada una de las cuatro variables a la varianza total observada en la variable dependiente podría cambiar con una simple inversión del orden de entrada de los datos correspondientes a cada variable en la ecuación de regresión. De aquí que no tenga mucha importancia la consideración de este dato en nuestro caso. Por otra parte, hemos de añadir que el cambio de orden de las variables no originaría diferencias en el R^2 final y, por consiguiente en la predicción global. Para los adverbios el R^2 es de .550 (Tabla MCA de Adverbios).

En el análisis de varianza (Tabla 8.2.4.) los efectos principales significativos corresponden a la Edad, el Sexo y la Clase Social. Los tres al nivel de confianza del .001 . El factor zona no resultó significativo.

Los resultados obtenidos indican:

1º) Con la edad y el nivel de escolarización se incrementa progresivamente el uso del adverbio en la producción verbal escrita.

2º) El rendimiento medio es más alto en las niñas que en los niños. Esta superioridad no se manifiesta por igual en los cinco niveles de edad estudiados. Así, por ejemplo, podemos constatar que la diferencia de medias a los 6 y 7 años no resultan significativas mientras que en las restantes edades sí lo son (Gráf. 21). Esto nos indica que hasta los 8 años no podemos afirmar que se dé realmente, en los Adverbios, una superioridad en la frecuencia de uso por razones de sexo, en este caso, del sexo femenino. Tampoco creemos que a los 8 años concurran circunstancias especiales en el sentido de que sean las determinantes de unas diferencias que con anterioridad carecen de valor significativo. La explicación del hecho creemos que reside en el mayor dominio que los sujetos adquieren a partir del segundo año de escolarización de las técnicas de expresión escrita. Hasta entonces, esta dificultad afecta por igual a unos y otras e impide que se manifiesten con claridad las diferencias.

3º) La ausencia de significación en el factor zona se debe no tanto a la inexistencia de diferencias entre ambas, la zona rural y la zona urbana, cuanto a que se contrarrestan los efectos que producen en diferentes edades y ésto hace que en conjunto las diferencias desaparezcan. En efecto, si analizamos la interacción V1-V3 (Gráf. 22) observamos que a los 6 y 8 años las diferencias no son significativas; a los 7 y 10 años la zona Urbana se impone sobre la zona rural, mientras que

a los 9 años ocurre lo contrario, es decir, la preponderancia de la zona Rural sobre la urbana.

- Por otra parte, el predominio alternan-te de zonas antes comentado es aún más complejo si consideramos un nuevo factor: la clase social (Interacción V1-V3-V4. Gráfs. 31 a 35).

a) En la clase social alta, los sujetos de zona Urbana muestran un mayor rendimiento que los rurales en los tres primeros cursos (6, 7 y 8 años). A partir de entonces, el efecto se invierte: pasan a dominar los de zona rural (9 y 10 años).

b) En la clase media, sucede algo semejante aunque en sentido inverso: mientras que en los cursos intermedios (2º, 3º y 4º) el predominio corresponde a la zona rural, en el primero y último (6 y 10 años) se impone la zona urbana (en este segundo caso las diferencias no son significativas).

c) Por último, en la clase baja, la alternancia en el predominio de las zonas es casi un efecto de la edad. Así pues, a los 6 años domina la zona rural (diferencias N.S.), a los 7 la zona Urbana, a los 8 y 9 la Rural, y a los 10 la Urbana.

4º) El efecto de la clase social en la frecuencia de uso de los Adverbios establece un patròn de comportamiento semejante al observado en los Nombres.

- Los índices de rendimiento que manifiestan los su-

jetos de clase alta son en conjunto superiores a los que manifiesta la clase media que, a su vez, se imponen sobre los de clase baja. No obstante, las diferencias entre las dos últimas clases sociales no son significativas. La causa de esta aproximación entre la clase media y baja se debe al descenso en el rendimiento que acusan las niñas de clase media (Gráf. 23). Esta última circunstancia unida a la ausencia de diferencias significativas entre la clase media y alta en los niños dificulta la posibilidad de establecer cualquier tipo de jerarquía entre las tres clases sociales.

Este hecho lo podemos constatar en las gráficas 29 y 30. Sólo en los niños de 7 y 10 años se observa con cierta claridad un orden definido en el que predómina la clase alta sobre la media y ésta sobre la baja. Por lo demás, las diferencias sociales entre niños apenas si tienen importancia, mientras que en las niñas es claro el predominio que mantiene la clase alta, sobre todo a partir de los 8 años.

9.5.- Análisis de la 5ª Variable Dependiente: PRONOMBRES.-

El estudio de los datos presentados en la tabla resumen de ANOVA de esta variable (Cfr. 8.2.5.) muestra la existencia de efectos altamente significativos en los cuatro factores o variables independientes (Los niveles de significación son de .001, .001, .005, .004 para Edad, Sexo, Zona y Clase Social, respectivamente).

En cuanto a la proporción de varianza atribuíble a cada factor, observamos (Tabla MCA) que es la Edad quien asume el porcentaje más alto (16.8 %), seguida del Sexo (1.2 %) y de la Zona y la Clase Social (ambas con el 0.4 %). R^2 = .183 .

Del análisis de resultados extraemos las siguientes inferencias:

1º) La edad va asociada con un incremento progresivo en la frecuencia de uso de los Pronombres. La mayor diferencia se produce entre los grupos de 6 y 7 años.

2º) Las niñas se muestran claramente superiores a los niños. Esta superioridad se observa en los cinco cursos (Interacción V1-V2) presentándose las mayores diferencias entre niñas y niños en el último curso (10 años).

3º) Por Zonas, se confirma también la superioridad de las niñas con una única excepción: a los 9 años, los niños de zona urbana superan a las niñas de la misma zona. En la Gráf. 42, tenemos precisamente en un claro ejemplo de inter-

acción cruzada, la confirmación analítica de lo que acabamos de expresar, es decir, la inversión que se produce del efecto del sexo según consideremos una u otra zona.

Por otra parte, en la interacción V1-V2-V4 (Gráfs. 39 a 43) podemos constatar además, que hasta los 9 años, las diferencias favorables a las niñas son más acusadas en la zorural que en la urbana. En el último curso (10 años), ocurre el fenómeno contrario: disminuye la superioridad del sexo femenino en la zona rural y aumenta en la zona urbana.

4º) En cuanto a la Zona, el efecto que esta tiene sobre el rendimiento en la frecuencia de uso de los Pronombres, manifiesta un predominio de la zona urbana sobre la rural. Esta superioridad, sin embargo, no se aprecia por igual en los cinco cursos, siendo 2º y 5º (7 y 10 años, respectivamente) los cursos donde las diferencias adquieren mayores proporciones. (Gráf. 37).

5º) La procedencia social de los sujetos también afecta al uso del Pronombre. No obstante, es difícil observar un orden jerárquico establecido entre las tres clases sociales.

- En la interacción V1-V4 (Gráf. 38) observamos que las únicas diferencias consistentes en los promedios las introduce la clase alta y no en todos los cursos (sólo en 3º y 5º). Entre la clase media y baja no se aprecian diferencias significativas en ninguno de los cinco niveles.

- El análisis de los perfiles de las tres clases sociales, por zonas (Gráfs. 46 y 47), matiza aún más estas

las dos conclusiones precedentes. Así pues, mientras que el perfil de la zona Urbana es una copia exacta del perfil general (comparar gráficas 46 y 38), en la zona Rural la superioridad de la clase alta se inicia a los 7 años y se acentúa con el transcurso del tiempo. Tanto en una zona como en otra, las clases media y baja siguen un mismo patrón, se alternan en el predominio pero en cualquier caso las diferencias siempre son menores que las que les separan de la clase alta.

9.6.- Análisis de la 6ª Variable Dependiente: ARTICULOS.-

En el estudio esta variable dependiente encontramos algunas diferencias no observadas hasta ahora en los análisis de las variables precedentes. Nos referimos a la ausencia del efecto principal del Sexo ($F= .086$; $p < .999$ N.S.), lo cual nos está indicando que , en conjunto, los promedios en la cantidad de artículos empleados por niños y niñas no difieren significativamente.

En relación con los restantes factores, los resultados del ANOVA (Cfr. 8.2.6.) son estos:

1º) En cuanto al factor Edad, se repite la constante te que venimos observando en todas las categorías gramaticales: la producción de artículos se incrementa progresivamente con el transcurso de los años. Corresponde el mayor aumento a la edad de 9 años.

2º) El incremento es más uniforme en la zona urbana que en la zona rural. El índice de producción de la primera supera al de la segunda en todas las edades, salvo a los 10 años donde la zona rural se impone a la urbana de manera significativa. En la gráfica 48, se observa la tendencia hacia una disminución progresiva de las diferencias entre zonas que culmina a los 10 años con el cambio ya comentado, es decir, la superioridad de la zona rural.

3º) Este cambio se debe fundamentalmente al alto rendimiento en el uso de los artículos demostrado por los ni-

ños de clase alta de zona rural (Gráf. 50), ya que en las niñas cualquiera que sea su clase social, se mantiene el predominio de la zona urbana sobre la rural (Gráf. 51).

4º) El efecto de la clase Social también resultó significativo. En efecto, podemos constatar que los sujetos provenientes de las clases sociales más bajas presentan una producción de artículos menor en promedio que los correspondientes a las clases medias y altas.

5º) Sin embargo, la existencia de interacciones significativas en las que interviene la variable clase social (V1-V4 / V1-V2-V4 y V2-V3-V4) introduce necesariamente diversos matices en la interpretación del efecto principal.

a) Durante los primeros años (6, 7 y 8) las clases superiores (media y alta) no presentan diferencias significativas. Sólo a partir de los 9 años se produce un fuerte distanciamiento de la clase alta. (Gráf. 49).

b) Tampoco la clase baja aparece desfasada en los dos primeros cursos (6 y 7 años). Por tanto, las diferencias entre clases apenas si tienen importancia en un principio. Por otra parte, se observa también que tanto en los niños como en las niñas (Gráfs. 52 y 53) el incremento de artículos responde a una pauta uniforme con el paso de una edad a otra. En cambio, en las clases media y alta el crecimiento es mucho más irregular.

c) Las diferencias entre la clase media y baja que aumentan progresivamente con la edad en las niñas, tienden a desaparecer en los últimos cursos (9 y 10 años) (Gráf. 52). En los niños, apenas si tienen importancia en ninguno de los cursos, dado que los promedios de una y otra clase social son muy similares.

9.7.- Análisis de la 7ª Variable Dependiente: PREPOSICIONES.-

Sólo los efectos principales del factor Edad y del factor Clase Social resultaron significativos en el ANOVA (Cfr. 8.2.7.) de las Preposiciones. Ni en función del Sexo, ni de la Zona, se observaron diferencias en los promedios de preposiciones utilizadas por los sujetos. Tampoco resultaron significativas ninguna de las interacciones.

En consonancia, pues, con los datos estas son las principales inferencias:

1º) Con la edad se incrementa el uso de las preposiciones. Si contrastamos la diferencia de medias entre un curso y otro, podemos concluir también que el incremento es bastante uniforme dado que en todos los casos oscila en torno a los tres puntos, excepto en el paso de los 6 a los 7 años donde la diferencia resulta algo menor.

2º) Entre las clases sociales observamos que se establece la jerarquía esperada, es decir, la superioridad de la clase alta sobre la clase media y de ésta sobre la clase baja, aunque estas últimas se encuentran más próximas entre si que las dos primeras.

3º) En la tabla MCA de Preposiciones observamos, como en casos anteriores, que la edad es el factor de mayor participación en el procentaje de la varianza explicada (R^2 = .333).

Por consiguiente, de este breve análisis sobre la categoría gramatical de las Preposiciones podemos extraer la conclusión de que entre todas las categorías, a tenor de los análisis precedentes, parece ser la que goza de una mayor estabilidad en el sentido de que sólo se muestra sensible al efecto, lógico desde el punto de vista evolutivo, de la edad y, en menor medida, al efecto de las clases sociales. La diferencia de medias que presentan los niveles de los restantes factores (Sexo y Zona) no son significativas aunque se apunta la tendencia de una mayor preponderancia de las niñas, en cuanto al sexo, y de la zona urbana tal como se observa en la mayor parte de las categorías gramaticales.

9.8.- Análisis de la 8ª Variable Dependiente: CONJUNCIONES.-

Las conjunciones, del mismo modo que las Preposiciones consideradas en el punto anterior, sirven, según la definición que de ellas hace la Gramática de la Academia, para denotar el enlace entre dos o más palabras en una oración. Sin embargo, a pesar de ser consideradas como palabras que carecen por sí mismas de un significado completo y que cumplen una función nexiva como aquellas (las preposiciones), el análisis de varianza muestra unos resultados bien distintos (Cfr. 8.2.8.).

En primer lugar, constatamos que el efecto principal del factor clase Social no resulta significativo, siendo ésta la única variable dependiente en que esto ocurre.Así pues, el orden jerárquico preestablecido en una de nuestras hipótesis acerca del predominio de las clases sociales aquí no se cumple. En otros términos, por efecto de la clase social los sujetos no presentan diferencias significativas en la frecuencia de uso de las Conjunciones. No obstante, y a pesar de que el efecto de clase social resulte no significativo, la presencia de este factor en combinación con otros en interacciones significativas (V1-V4 / V1-V2-V4 / V1-V3-V4) precisará de un comentario que más adelante realizaremos.

En segundo lugar, tampoco resulta significativo el efecto de Zona aunque, al igual que la clase social participe en interacciones significativas. La ausencia de significación en este factor resulta, no obstante, menos extraña si tomamos en consideración lo que ocurre en otras categorías gramaticales con este mismo factor.

En cuanto a un estudio más pormenorizado de la tabla ANOVA, estos son los aspectos más importantes a destacar:

1º) A lo largo de los cinco cursos se incrementa progresivamente el uso de las conjunciones. En el cuadro de medias (efecto principal de V1) observamos que la mayor diferencia ocurre entre los 6 y 7 años. Esta diferencia no creemos que responda al hecho real de que tanto los niños como las niñas, sea cual sea su clase social o la zona de donde provengan, desconozcan las conjunciones o a que durante el segundo año de escolarización se produzca un aprendizaje importante en este sentido, al que pudiesemos atribuir la diferencia. Como ya hemos comentado en alguna ocasión precedente, el cambio brusco que acusan los resultados entre los 6 y 7 años está condicionado por la dificultad para dominar, aunque sea rudimentariamente, la técnica de la escritura y, por consiguiente, el lenguaje escrito. Este fenómeno no ocurre de igual manera a nivel de lenguaje oral. A los 6 y 7 años, la expresión oral del niño manifiesta un dominio sobre estructuras sintácticas complejas que implican el uso frecuente de preposiciones y conjunciones pero que en modo alguno son capaces de trasladar y expresar por escrito. Durante la fase de recuento nos llamó precisamente la atención la escasa o nula proporción de conjunciones que manejaban los niños de 6 años y la pobreza que presentaban en cuanto a la variedad. Para la mayor parte de los sujetos la única conjunción utilizada era la conjunción coordinante "y".

2º) Teniendo en cuenta que la interacción Edad-Sexo es marginal ($F = 1.945$; $p < .099$) podemos confirmar que la superioridad del sexo femenino aparece en determinadas edades (6, 8 y 10 años) donde los contrastes de medias resultaron significativos y favorables a las niñas.

3º) Los sujetos de zona Urbana no difieren de los rurales en el uso de las conjunciones ($F = 3.027$; $P < .078$) sin embargo, observamos que el índice de probabilidad está muy próximo al nivel de significación del .05. Lo extraño es que en el caso de las conjunciones la hipótesis de partida no sólo no se cumple sino que la tendencia que se advierte manifiesta un cierto dominio de la zona Rural sobre la urbana (véase la gráf. 54 de la interacción Edad-Zona). Analíticamente, se confirma, por tanto, aunque las diferencias no sean muy acusadas, que el índice de rendimiento de los sujetos rurales es superior al de los urbanos, salvo a los 7 años.

a) En la interacción de segundo orden (V1-V2-V3) donde se encuentra implicado un nuevo factor (el Sexo), tenemos oportunidad de clarificar el sentido de la tendencia anteriormente expuesta. Constatamos que la superioridad de la zona rural es debida, en gran parte, al patrón de comportamiento establecido por las niñas en los distintos cursos. Ciertamente, en la gráf. 57, observamos que el perfil de la zona rural se impone en todos los niveles al de la zona urbana (con excepción del caso ya comentado de los 7 años). En los niños, por el contrario, el patrón es más irregular y alternante. Además, las diferencias de medias entre zonas en los distintos cursos son menos acusadas que en las niñas.

b) También, en términos generales podríamos afirmar que el predominio rural se manifiesta en las distintas clases sociales si bien la interpretación de las gráficas resulta más compleja.

- En la clase baja, el predominio corresponde a la

zona rural en todos los cursos, salvo a los 7 años. No obstante, las diferencias tienden a desaparecer en los dos últimos cursos (4º y5º).

- En la clase media la superioridad de la zona rural sólo aparece con claridad en los dos primeros cursos. A partir de entonces (8 años) y al igual que en la clase baja, las diferencias dejan de ser significativas.

- En la clase alta, el patrón se invierte por completo. En los primeros cursos predomina la igualdad de zonas, mientras que en los últimos se impone con claridad la zona rural. Tal vez estas irregularidades se deban al bajo número de sujetos en las celdillas.

4º) El incremento observado por cursos en la producción de conjunciones también resulta significativo en relación con las clases sociales, aunque la jerarquía en rendimiento establecida en las hipótesis (A>M>B), no se cumpla realmente en este caso.

El incremento de los índices de rendimiento que se observan con el transcurso de los años escolares se da asímismo en todas las clases sociales, con la excepción de la clase alta a los nueve años en la que se produce un fuerte descenso en relación con el rendimiento observado a los ocho años (Gráf. 55).

Este descenso se acusa antes (8 años) y es más importante en los niños (Gráf. 58) que en las niñas (Gráf. 59) de clase alta.

9.9.- CONCLUSIONES.-

Tras el análisis pormenorizado de cada una de las ocho categorías gramaticales, procederemos a la exposición de las conclusiones más relevantes que se derivan de nuestra investigación.

En el presente trabajo nos habíamos propuesto dos objetivos: de una parte, la determinación del vocabulario usual del niño de Andalucía Oriental de 6 a 10 años; de otra, analizar el incremento cuantitativo que sufre el vocabulario en función de una serie de factores o variables atributo tales como la Edad, el Sexo, la Zona y la Clase social de los 2.166 sujetos que componen la muestra.

9.9.1.- EL VOCABULARIO USUAL.-

El Vocabulario Usual determinado está compuesto por 5.750 vocablos resultantes del recuento de 258.983 palabras. En los anexos aparecen las dos modalidades de presentación al uso en la mayor parte de los diccionarios de frecuencias:

a) Listado de Vocablos ordenados alfabéticamente (Anexo nº 1).

b) Listado de Vocablos ordenados en forma decreciente según su frecuencia (Anexo nº 2).

c) El Vocabulario Común. Está compuesto por 1.068 vocablos. Aparece, ordenado alfabéticamente, en el Anexo nº 3. Extracta la parte del vocabulario usual que posee una mayor estabilidad y consistencia ya que los

vocablos que contiene son aquellos que permanecen en uso a lo largo de todo el período evolutivo estudiado.

d) En el Anexo nº 4, se registran los vocabularios diferenciales y específicos correspondientes a cada uno de los niveles de las cuatro variables de estratificación utilizadas en la selección de la muestra. En total se presentan 16 vocabularios.

9.9.2.- EL INCREMENTO CUANTITATIVO DEL VOCABULARIO: DISCUSION DE RESULTADOS Y CONCLUSIONES.

En relación con el segundo objetivo de nuestra investigación, el análisis del desarrollo cuantitativo del vocabulario, la presente discusión de resultados y conclusiones pretende, en primer lugar, presentar una interpretación de los datos en términos comparativos entre unas variables dependientes y otras, dado que cada una de ellas responde a un aspecto parcial resultante de la subclasificación de la producción verbal expresada por los sujetos en los textos escritos. En segundo lugar, pretendemos asímismo argumentar críticamente los patrones de comportamiento a los que responde el desarrollo cuantitativo del vocabulario en función de la edad, el sexo, la zona y la clase social.

En consonancia con los datos obtenidos y el análisis por categorías realizado en la primera parte de este capítulo, podemos establecer las siguientes conclusiones:

1) En las ocho categorías gramaticales el factor o variable independiente de mayor peso específico en la explicación de la varianza observada es la edad (nivel de escolarización).

En las distintas tablas MCA se observa que el coeficiente Beta más alto siempre va asociado a esta variable.

El resultado es lógico y confirma la primera de nuestras hipótesis. El incremento del vocabulario se debe en mayor medida a la edad y a la experiencia y aprendizaje que con ella van asociados, que al hecho de ser hombre o mujer, de habitar en una zona determinada o pertenecer a una u otra clase social. Y esto es aún más evidente si tenemos en cuenta la importancia e influencia de la escolarización, como elemento homogeneizador, en el proceso de adquisición del lenguaje. No es necesario precisar que esta valoración no resta importancia a la influencia de las restantes variables. Sin embargo, el efecto que producen en el aumento del vocabulario, aunque sea significativo, es, en magnitud, cuantitativamente menor.

2) El análisis de los ocho coeficientes de determinación (R^2) indica que la varianza determinada por la acción conjunta de las cuatro variables independientes en ningún caso supera el 50%. Los coeficientes más altos aparecen en Nombres (.463), Preposiciones (.333) y Adverbios (.303), y los más bajos en Pronombres (.183), Adjetivos (.243) y Conjunciones (.268).

3) Edad y Vocabulario.-

Todas las investigaciones sobre el desarrollo cuantitativo del vocabulario coinciden en la constatación del ritmo acelerado con que se incrementa el vocabulario con el transcurso de los años. Este incremento lógico en las primeras etapas consiguientes a la aparición de las primeras palabras, continúa con el inicio de la escolarización y se mantiene a lo lago de los cinco cursos estudiados.

El Vocabulario de los niños de 6 años (1.676 vocablos) resulta casi triplicado a los 10 años (4.369 vocablos). No obstante podríamos decir, utilizando una expresión de la física, que el ritmo de crecimiento se mantiene uniformemente acelerado hasta los 9 años, momento en el que se inicia una desaceleración en el crecimiento. La expresión del incremento, en porcentajes (véase CUADRO A. 1. del Anexo 1º), así lo confirma: el vocabulario de 6 años se ve incrementado en un 19.45% a los 7 años, en un 29.22% a los 8, en un 41.75% a los 9 y en un 19.14% a los 10.

Este mismo incremento se aprecia en los valores medios que adquiere el rendimiento de los distintos grupos de edad en las ocho categorías gramaticales. En todos los casos, las diferencias de medias, al contrastar los grupos entre sí, resultaron significativas.

No debe extrañar que en algunas categorías gramaticales por ejemplo, en Nombres y Adjetivos, la menor diferencia se produzca entre los dos primeros niveles (6 y 7 años), mientras que en otras (por ejemplo, Verbos, Conjunciones y Pronombres) son estos mismos niveles los que presentan la mayor discrepancia. A los 6 años, la cantidad de palabras que el niño es capaz de escribir de forma espontánea está constituída fundamentalmente por nombres y, en menor proporción, por adjetivos y verbos. Las frases son breves y sintácticamente de poca complejidad, dada la escasa frecuencia con que hacen uso de palabras nexo o relacionales, como las Preposiciones y Conjunciones. A los 7 años, la producción verbal escrita se hace más compleja hasta el punto de que en no pocas categorías (Verbos, Conjunciones, Pronombres y Adverbios) la mayor diferencia en los promedios se da entre los 6 y 7 años. Este fenómeno no debe resultar contradictorio, si tenemos en cuenta que la dificultad para expresarse por escrito disminuye pro-

gresivamente a medida que el niño adquiere un mayor dominio de la escritura y, evidentemente, esto se consigue con los años de escolarización. Así pues, la magnitud del efecto de la edad en los primeros cursos se encuentra mediatizada, más que en los últimos, por el procedimiento utilizado para la recogida del vocabulario ya que supone en el niño un dominio de la escritura que no tiene todavía.

En relación con los restantes cursos, hemos de consignar que el mayor incremento se produce en 4º de E.G.B. (9 años) y, a partir de aquí, el ritmo de aumento decrece.

a) 6 años:

A los seis años, la superioridad de las niñas sobre los niños en la cantidad de palabras emitidas se detecta en todas las categorías gramaticales, si bien en el caso de los Adjetivos, Adverbios y Preposiciones las diferencias no son significativas.

Por Zonas, observamos que a esta edad la zona urbana se impone sobre la zona rural, excepto en el caso de las Conjunciones donde tanto las niñas como los niños rurales muestran un rendimiento superior. (La superioridad urbana en Verbos y Preposiciones no es significativa. Tampoco lo es para las niñas en la categoría de Adjetivos).

En las clases sociales no se observa todavía un patrón definido de comportamiento que permita establecer entre ellas un orden jerárquico consistente ya sea en la dirección de nuestra hi como en cualquiera otra. No obstante, podemos confirmar que en la producción de palabras con significación propia (Nombres, Adjetivos, Verbos y Adverbios)(grupo que asume casi el 70% de frecuencias) se da un predominio de la clase media sobre la alta y la baja, estando las dos primeras muy próximas entre sí.

En el resto de categorías gramaticales se alterna este patrón con el que postulaba nuestra hipótesis, es decir, superioridad de la clase alta sobre la media y de ésta sobre la baja. En las Conjunciones, el efecto de clase social no resultó significativo.

b) 7 años:

A los 7 años, las diferencias en función del sexo, apenas si son significativas en ninguna de las categorías gramaticales. Por consiguiente, predomina la igualdad en la producción verbal de niños y niñas. Sólo en el caso de Conjunciones y Pronombres, los valores medios resultaron significativos a favor de las niñas.

Si atendemos a la variable zona, a esta edad la zona urbana resulta superior en todas las categorías (en Adjetivos y Preposiciones las diferencias no son significativas) a la zona rural.

En cuanto a la clase social, tampoco a esta edad existe un patrón definido. La jerarquía de clases sociales que hipotetizamos sólo aparece en los niños en cuatro de las ocho categorías (Nombres, Adverbios, Artículos y Preposiciones). En las niñas y en general, en el resto de las categorías, cuando aparecen diferencias significativas suelen ser favorables a la clase media (Adjetivos) o bien a ésta y a la clase baja, entre sí muy igualadas sobre todo en las niñas (Nombres, Adverbios y Artículos) o, lo que es más frecuente, que no aparezcan diferencias significativas entre las distintas clases sociales como ocurre en el caso de los Verbos, Pronombres y Conjunciones.

Por tanto, en relación con la clase social, no podemos afirmar (como tampoco lo podíamos hacer en el nivel anterior) que

exista una preponderancia de cualquiera de las tres clases sociales o se pueda establecer una jerarquía entre ellas que se repita de forma constante en las distintas categorías verbales. Sin embargo, hemos de destacar dos aspectos que no se aprecian en el curso anterior. Por un lado, se produce un mayor equilibrio entre las clases sociales, es decir, tiende a desaparecer el predominio de la clase media que se detectaba en el curso anterior. Y en segundo lugar, se apunta ya, aunque sólo sea en los niños un orden jerárquico consonante a nuestra hipótesis que se afianzará en los cursos posteriores.

c) 8 años:

A esta edad el comportamiento diferencial de niños y niñas aparece mucho más definido. La superioridad de las niñas es más generalizada y estable. Sólo caben como excepciones el caso de las preposiciones en las que el efecto del sexo no resulta significativo ni en esta ni en ninguna otra edad, y la inversión del patrón que se observa en las clases medias en las categorías de Nombres y Artículos donde los niños resultan ser superiores a las niñas.

Al igual que ocurre con el Sexo, también a esta edad la jerarquía entre las clases sociales se perfila con mayor nitidez: las clases superiores, casi siempre la alta, se imponen sobre las demás. La estructura A M B sólo se ve distorsionada por las niñas en la categoría de Nombres (B=M>A) y asímismo en los sujetos de zona rural (M>B=A) en la categoría de Preoposiciones.

Más extraño resulta el efecto de Zona de acuerdo con lo establecido en la hipótesis 3ª ya que la tendencia que se apunta en los dos primeros cursos, es decir, el mayor rendimiento de los

sujetos de zona urbana tiende a reducirse a los 8 años. En efecto, la supremacía de la zona urbana sólo aparece en Nombres, Verbos y Artículos que son,no obstante, las categorías que aportan mayor número de frecuencias, sin embargo, en las restantes categorías desaparecen las diferencias e incluso,en las Conjunciones, el predominio corresponde a la zona rural.

d) 9 años:

Cuando aparecen diferencias significativas por efecto del sexo siempre es a favor de las chicas. Así sucede en Nombres, Adjetivos, Adverbios, Pronombres y Artículos (en este último caso, solamente entre sujetos de clase social alta). En las restantes categorías las diferencias en el contraste de los promedios no son significativas, aunque en su mayor parte siguen siendo favorables a las niñas. La única excepción se produce en los sujetos de clase media en la categoría de los Artículos donde los niños son claramente superiores.

Si examinamos el comportamiento de las distintas clases sociales en esta edad, observamos que la influencia sociocultural del ambiente en el que el niño se desenvuelve incide en el desarrollo del vocabulario. La jerarquía A M B aparece en seis de las ocho categorías con ligeras alternancias asociativas de la clase media con la baja en determinadas categorías, por ejemplo, los Verbos (A>B=M). Sólo en el caso de las Conjunciones se altera de manera significativa el orden establecido (B>M>A). En los pronombres la clase social no introduce diferencias significativas.

En relación con la zona, observamos que las diferencias

entre zona urbana y zona rural tienden a contrarrestarse. Así pues, mientras en la categoría de Nombres, Adjetivos y Artículos el predominio corresponde a la zona urbana, en Verbos, Adverbios y Conjunciones se impone la zona rural. En Pronombres y Preposiciones las diferencias no resultaron significativas. Continúa, pues, la tendencia apuntada en la edad anterior, en el sentido de una reducción progresiva del efecto de zona en el incremento cuantitativo del vocabulario.

e) 10 años:

A los 10 años, la cantidad de palabras emitidas por las niñas resultó ser significativamente mayor a la expresada por los niños. Las únicas excepciones que hemos de reseñar se limitan a un pequeño grupo de 15 niños de clase social alta que mostraron un rendimiento superior en comparación con el que arrojaba el grupo de niñas (8) de la misma clase social. Esto ocurre tanto en la categoría de normbres como de artículos. Las diferencias en la frecuencia de uso de las preposiciones no resultaron dignificativas como tampoco lo fueron en los cursos precedentes.

Las diferencias entre zonas vuelven a aparecer con más intensidad que en los dos años anteriores (8 y 9 años), sobre todo en relaciòn con el grupo de vocablos con significación propia (salvo el caso de los Adjetivos) y en los pronombres. En estos casos la preponderancia corresponde a la zona urbana. La zona rural supera a la urbana con claridad únicamente en el caso de los artículos y, en menor grado, en las conjunciones.

Es difícil precisar la causa de este cambio ya que la tendencia observada en los dos cursos anteriores apuntaba hacia una desaparición progresiva del predominio de la zona urbana so-

bre la rural y, en último término, hacia la desaparición del efecto de zona, lo cual hubiera supuesto el rechazo de nuestra hipótesis 3ª. Sin embargo, los datos no confirman dicha tendencia sino la que postula nuestra hipótesis.

Respecto a la influencia de las diferencias ambientales representadas por las distintas clases sociales, en la frecuencia de uso del vocabulario, comprobamos que a los 10 años en todas las categorías gramaticales se impone la jerarquía A>M>B. La única excepción aparece en los nombres, donde las niñas imponen el patrón M=B>A. Se observa, por otra parte, que se da una mayor aproximación entre la clase media y baja (por ejemplo, en Verbos, Pronombres, Preposiciones y Conjunciones) que entre las clases media y alta donde las diferencias siempre resultaron significativas.

4) Sexo y Vocabulario:

Respecto a las diferencias entre niños y niñas, los datos corroboran que la producción verbal espontánea de las niñas es significativamente superior a la de los niños en todas las edades y en la casi totalidad de las categorías gramaticales (sólo en el caso de artículos y preposiciones las diferencias dejan de tener valor significativo.) En esto coincidimos con la mayor parte de investigaciones que han estudiado el efecto del sexo en el desarrollo cuantitativo del vocabulario.

Sin embargo, hay otro aspecto importante en relación con las diferencias entre niños y niñas que hemos de destacar. La mayor locuacidad de las niñas, implícita en la afirmación anterior, no se corresponde con el hecho de que posean un vocabulario más rico. En efecto, si consideramos globalmente el número de vocablos distintos que utilizan unos y otras, constatamos que el vocabulario más rico y extenso corresponde a los niños, aunque

utilicen menor cantidad de palabras para expresarse que las niñas. Tanto si comparamos la razón P/V de ambos (26.18 para niños y 32.34 para niñas) como si contrastamos el total de vocablos que integran el vocabulario diferencial (1.539 vocablos en los niños y 1.061 en las niñas), la mayor riqueza léxica corresponde a los niños.

5) Zona y Vocabulario:

La zona es la variable que produce un efecto menos definido en el desarrrollo cuantitativo del vocabulario. En la mayor parte de las categorías gramaticales apenas si resultan significativas las diferencias entre zonas.

Desde un punto de vista evolutivo, tampoco se observa en la influencia de la zona una tendencia definida. Así, mientras en los dos primeros cursos aparece un predominio de la zona urbana sobre la rural, éste posteriormente se diluye en 3º y 4º, para manifestarse de nuevo con cierta claridad en el último curso.

Resulta extraño que sea en los dos primeros cursos donde aparezca con mayor nitidez la superioridad de la zona urbana cuando en estas mismas edades la magnitud del efecto de los restantes factores apenas si es perceptible debido a la dificultad con la que el niño se enfrenta de tenerse que expresar a través de un medio (la escritura) que aún no domina. La explicación de este fenómeno tal vez resida en la mayor oportunidad de escolarización temprana de que gozan los niños que habitan en zonas urbanas. Es sabido que la asistencia a la escuela y las posibilidades de recibir una educación preescolar están más al alcance del niño urbano que del niño rural. Esto le permite acceder a la

E.G.B. con un mayor grado de madurez que propicia y facilita el aprendizaje de la lectura y la escritura. Esta sería la causa de que las diferencias entre zonas se detecten en los primeros cursos. Más tarde, cuando se adquiere un cierto dominio de la escritura (recordamos que la recogida de datos se hizo a través de textos escritos) y como consecuencia también del efecto homogeneizador de la enseñanza, las diferencias entre zonas tienden a desaparecer.

Sin embargo, a pesar de las consideracio-nes antes expuestas, los datos corroboran globalmente nuestra hipótesis acerca de la superioridad de la producción verbal de la zona urbana frente a la rural. En efecto, si observamos, por un lado, que aquellas categorías donde las diferencias son claramente favorables a la zona urbana tienen, además, un mayor peso específico en el total de frecuencias del vocabulario, como es el caso de los Nombres, Pronombres, Artículos y Adjetivos (en estos últimos la superioridad urbana sólo se detecta en los niños) y, por otro lado, que en el resto de categorías (salvo Adverbios y Preposiciones) los niveles de significación indican que la probabilidad de superioridad de la zona urbana está por encima del 80%, hemos de concluir,de acuerdo con lo que postulaba nuestra hipótesis, que la producción verbal de los sujetos de zona urbana es superior a la de los sujetos de zona rural. Superioridad que no consiste solo en escribir mayor cantidad de palabras sino que comporta también una mayor riqueza y variedad léxica.

6) Vocabulario y Clase Social:

Las influencias socioculturales y familiares, representadas en las tres clases sociales, también ejercen un efecto posi-

tivo en la producción verbal de los sujetos. En todas las categorías gramaticales (con la única excepción de las conjunciones) se confirma el orden jerárquico que postula nuestra hipótesis, es decir, la superioridad de la clase alta sobre la media y de ésta sobre la clase baja. Las distancias entre ellas son más acusadas en los últimos cursos que en los primeros, lo que nos indica que con el transcurso del tiempo las diferencias sociales se incrementan. Ya comentamos, con anterioridad, que precisamente en el último curso (a los 10 años) es donde aparece más definida la jerarquía (A M B). Por otra parte, la aproximación entre la clase alta y media, frecuente en los primeros cursos, tiende a desaparecer al mismo tiempo que se reducen también las distancias entre ésta (la clase media) y la clase baja.

RESUMEN

El objetivo fundamental de este trabajo fue la determinación del vocabulario usual del niño de Andalucía Oriental, comprendido entre los 6 y 10 años de edad. La validez del mismo reside en la representatividad de la muestra de sujetos y del corpus de palabras de donde se extrae el vocabulario.

Seleccionamos 2.166 sujetos de los cuales se registraron 258.983 palabras (frecuencias). Del recuento de estas palabras se obtuvieron 5.750 vocablos.

En una segunda parte, comprobamos la influencia que ejercen la edad (nivel de escolarización), el sexo, la zona y la clase social en la producción verbal diferenciada en categorías gramaticales.

El crecimiento del vocabulario depende fundamentalmente de la edad o nivel de escolarización de los sujetos, sin embargo, los datos confirman también, la mayor precocidad de las niñas en el desarrollo, son más locuaces que los niños, la superioridad de los sujetos que habitan en zonas urbanas y la diversidad del rendimiento en función de la clase social, correspondiendo los índices más altos de producción verbal a los sujetos de las clases sociales más acomodadas.