

Protextos: un instrumento informático para el análisis del lenguaje escrito

Liliana Tolchinsky*

Ana Teberosky**

José Luis Rodríguez Illera*

*ICE. Universidad de Barcelona

**Universidad de Barcelona

Presentamos un instrumento informático para el análisis de textos escritos cuya elaboración resulta de la confluencia de una línea de investigación en psicolingüística y de una línea de desarrollo de herramientas informáticas para la enseñanza y el análisis de aspectos formales del lenguaje. Una vez definidas la aproximación y la metodología de análisis en las cuales se basa este instrumento, pasamos a su caracterización desde el punto de vista informático y funcional. A diferencia de otras aproximaciones que utilizan ordenadores para ayudar en el análisis de textos, la que sostiene este programa está basada en tres aspectos diferenciales: a) tener una metodología de análisis de referencia que incluye los aspectos de transcripción y configuración gráfica de los textos; b) apostar por una reducción al mínimo de la preparación y codificación previa del texto para su categorización, y c) apostar por una reducción al mínimo de las capacidades o destrezas que el analista debe poseer desde el punto de vista informático. Ejemplificamos la utilización del instrumento con uno de los posibles itinerarios de análisis.

Palabras clave: *Análisis del discurso, textos, instrumento informático.*

We introduce a software for the analysis of written texts. This software is the result of the confluence of a particular line of psycholinguistic research and a line of development of teaching and language-analysis computer tools. After defining the approach and the methodology of analysis upon which this software is based, we move on to its computational and functional nature. In contrast to other approaches where computers are used as aids in text analysis, the approach of this programme is based on three differential aspects: a) it has a defined theoretical framework; a methodology of analysis which includes aspects of transcription and graphic configuration of texts; b) it reduces to the minimum the preparation and pre-codification of the text for its categorisation and c) it reduces to the minimum the computer knowledge needed by the analyst. We provide an example of the software's use in one of the possible paths of analysis.

Key words: *Discourse Analysis, Texts, Software.*

El propósito de este trabajo es presentar un instrumento informático para el análisis de textos escritos que es el resultado de la confluencia de dos líneas de trabajo, una centrada en la investigación psicolingüística (Teberosky, 1989, 1991, 1992; Tolchinsky, 1991, 1992) y la otra en el desarrollo de herramientas informáticas para la enseñanza y el análisis de aspectos formales del lenguaje (Rodríguez Illera, 1990c; Rodríguez Illera y León, 1990).

La idea surgió en el marco de una serie de estudios cuyo objeto era constatar la realidad psicológica de las tipologías textuales. El interés se centraba en describir cómo se inscribe un tipo de discurso en otro tipo de discurso. En este caso, cómo el niño aprendía las buenas —y las malas— correspondencias entre el discurso familiar que domina y aquél que parcialmente conoce pero cuyas convenciones debe aprender.

Tres premisas han orientado estos trabajos. La primera se refiere al hecho de que no podemos abordar las características de los distintos tipos de texto sin tener en cuenta al sujeto. Es decir, sin tener en cuenta el conocimiento que permite a los escritores producir las expresiones lingüísticas que distinguen los textos y a los lectores interpretar y evaluar las diferencias. Estudiamos pues la «competencia literaria» de los hablantes considerándola como parte de su competencia lingüística, es decir de su capacidad «de saber y hacer cosas con el lenguaje» (Culler, 1975; Ryan, 1977; Blanche-Benveniste, 1986; Iser, 1976). La segunda premisa se refiere a la heterogeneidad de los datos lingüísticos disponibles. A partir de una sintaxis compartida, la lengua es un conjunto de dialectos cuya diferenciación sigue no sólo pautas geográficas y sociales sino pautas de uso y finalidad.

Ahora bien, incluir en la consideración de lo textual el conocimiento de los usuarios y relacionarse con la lengua como conjunto heterogéneo no libera al estudioso del lenguaje de tener que definir cuáles son las características distintivas de los distintos tipos de discurso, aquéllas que permiten su identificación y restringen su interpretación. Coincidimos con Eco (1992) en tomar una postura crítica¹ hacia la producción e interpretación de textos. Nuestra tercera premisa concierne a la noción de texto. Consideramos el texto como una unidad funcional de significado complejo, una predicación extendida que incluye la elaboración de conjuntos de frases por un proceso de composición y concatenación. Definido como tal, el texto es un conjunto de medios lingüísticos de expresión, realizado en un contexto social determinado y actualizado a través de una actividad cognitiva (Scinto, 1986). Por medio de su actividad cognitiva consciente e inconsciente el sujeto genera un esquema (un esquema intencional, no amorfo sino informado por su experiencia textual previa) que lo orienta en el cumplimiento de su objetivo comunicacional. El esquema unificador se desarrolla en unidades de información en cuya generación y concatenación se integran los niveles sintácticos, semánticos, ortográficos, pragmáticos y metapragmáticos del lenguaje escrito.

1. Eco propone una distinción entre interpretación *semántica* e interpretación *crítica*. La primera «es el resultado del proceso por el cual el destinatario, ante la manifestación lineal del texto la llena de significado. La interpretación crítica o semiótica es, en cambio, aquélla por la que se intenta explicar por qué *razones estructurales* el texto puede producir esas (u otras, alternativas) interpretaciones semánticas» (Eco, 1992, p. 36).

Nuestra tarea consiste en descubrir las expresiones lingüísticas que sirven para distinguir y definir los distintos tipos de texto que aparecen a distintos niveles del lenguaje escrito. Por ese motivo debemos contar con una metodología de análisis textual que atienda los diferentes niveles y con un instrumento informático que facilite la comparación de textos producidos por sujetos de distintas edades, en distintas situaciones enunciativas.

Protexos es, pues, un producto del trabajo interdisciplinario y un proyecto en proceso de elaboración, cuya completud dependerá del uso y del feedback de los investigadores interesados en análisis textuales.² En lo que sigue definiremos brevemente nuestra aproximación al análisis de la variación textual y las características más generales de la metodología de análisis que sirvió de base a la elaboración de este instrumento.

Distintas aproximaciones al análisis de la variación textual

En el amplio marco de las investigaciones realizadas para detectar las posibles bases de la variación lingüística, los investigadores suelen optar entre una aproximación macroscópica o una aproximación microscópica al análisis de la variación textual. Una aproximación macroscópica tiende a definir dimensiones generales a lo largo de las cuales supuestamente variarán los textos y que sirven para orientar el análisis y la búsqueda de construcciones o elementos lingüísticos detallados. La clásica diferenciación de Bernstein (1971) en términos de código elaborado vs. código restringido es un ejemplo de utilización de dimensiones de variación general para la interpretación de propiedades lingüísticas específicas.

Los análisis microscópicos proveen en cambio una descripción detallada de las funciones comunicativas de determinadas construcciones o de elementos lingüísticos específicos. Por ejemplo, el uso de pronombres de primera persona como marcadores de compromiso personal. Es importante definir tanto las construcciones y los elementos lingüísticos considerados como los propósitos comunicativos a los cuales sirven estas características lingüísticas.

Existen diversos trabajos realizados que se incluyen o bien en una aproximación macroscópica o bien en una microscópica. Las investigaciones de Shiffrin (1981) y Tannen (1982), son ejemplos de análisis microscópicos y la clásica de Carroll (1960), de aproximación macroscópica. El trabajo de Biber (1988) es un magnífico ejemplo de integración de ambas aproximaciones, las cuales poseen ventajas y limitaciones complementarias. El análisis microscópico es necesario si queremos descubrir cuál es la función comunicativa de determinada construcción o elemento; servirá a ese propósito verificar su ocurrencia en distintos tipos de texto. Además, este análisis complementa al de nivel macro como medio para identificar aquellas propiedades lingüísticas que son realmente pertinentes

2. Nuestra idea es que a partir de esta primera versión, los interesados en este tema puedan utilizar la aplicación, comprobar sus ventajas y sobre todo señalar sus limitaciones para poder mejorar el instrumento. La versión actual de *Protexos*, así como el manual de utilización están publicados en forma de libro acompañado de *software* (Tolchinsky, Teberosky y Rodríguez Illera, 1993).

para diferenciar entre géneros y facilita la interpretación funcional de las dimensiones generales. Este tipo de análisis no sirve para identificar las dimensiones generales de variación lingüística entre textos ya que su análisis está restringido a unas pocas expresiones o construcciones por texto. Por otra parte, la aproximación macroscópica es fundamental para identificar las dimensiones subyacentes que proveen de un marco de comparación entre distintos tipos de texto. Estas dos aproximaciones son mutuamente interdependientes. El análisis macro depende del micro para la identificación y la interpretación funcional de las propiedades lingüísticas que son más importantes para definir un tipo de texto. Al mismo tiempo, el análisis micro se beneficia del enmarque teórico que el análisis macro puede proporcionarle y sobre todo de la posibilidad de orientar la búsqueda de aquellas propiedades lingüísticas pertinentes (Biber, 1988, p. 62 y sig.).

El uso de *Protectos* facilita ambas aproximaciones a la variación textual. Por un lado será posible verificar la ocurrencia de una o varias expresiones lingüísticas en distintos tipos de texto y aclarar las funciones comunicativas que cumplen en ellos. Recordemos que los textos están relacionados por las características lingüísticas que sirven a las funciones, es decir que son similares en su forma lingüística. Los géneros, en cambio, están determinados por criterios externos de uso, más que de forma. Así, los textos de un género particular pueden diferir en su tipo (Biber, 1988). Por otro lado, a partir de la definición operacional de determinada dimensión en términos de un conjunto de propiedades lingüísticas, podremos determinar su co-ocurrencia. Pero aunque posible, sería antieconómico utilizar *Protectos* para verificar las ocurrencias de una expresión determinada. En general es aconsejable utilizar este instrumento para análisis más completos en los cuales se involucren varios aspectos y propiedades lingüísticas.

Metodología de análisis

Transcripción y configuración gráfica

El tema de la transcripción de los datos lingüísticos plantea siempre dificultades al investigador; estas dificultades parecen mayores cuando se trata de los datos de producción oral que en el caso de los textos. Parece mayor, decimos, porque estamos acostumbrados a ver los textos ya editados, sin «las escorias del borrador» (Gresillon y Lebrave, 1983). En cambio los textos de los niños o de los adultos poco letrados, no están editados en una versión pulida: generalmente presentan una ortografía incorrecta, errores o falta de puntuación, separación no convencional de las palabras, etc., que complican su lectura e interpretación. A pesar de ello, se trata de que la falta de convencionalidad no impida el acceso a la evaluación de la expresión lingüística. Además, la evolución de las convenciones en sí es también parte de la preocupación investigadora. Por esos motivos, las operaciones de transcripción y configuración gráfica del texto son fundamentales. Sin ellas no es posible establecer ningún tipo de comparación entre textos producidos por niños de distintas edades y de distintas lenguas. Ahora bien, nin-

guna de estas operaciones es neutra sino que dependen de la postura teórica y de los criterios de análisis lingüístico. Muchas veces las transcripciones no son tales sino adaptaciones con correcciones. Sin quererlo, el investigador corrige, porque la transcripción no es una tarea mecánica, sino una tarea que requiere cierto nivel de análisis de los datos, ciertas decisiones teóricas. Estas consideraciones nos llevaron a inclinarnos por la propuesta de análisis del discurso desarrollada por Blanche-Benveniste y colaboradores (Blanche-Benveniste, 1987, 1990, 1991). Esta propuesta provee, además de un análisis sintáctico, un método de transcripción y de puesta en configuración de los textos consistente con los criterios sintácticos de base.

En cuanto a la transcripción, se recomienda trabajar sobre dos versiones del mismo texto. Una de ellas es la transcripción del texto tal como lo produjo el niño, con todos sus errores de segmentación, puntuación, etc. La otra es una versión normalizada del texto; esto es, una transcripción de acuerdo con la ortografía convencional. Al trabajar sobre la versión normalizada es posible concentrarse en los aspectos sintácticos y discursivos, sin que los errores de ortografía perturben al investigador. Al mismo tiempo, la comparación entre ambas versiones facilita la detección de errores de ortografía, diferencias de segmentación de palabras, etc., para que sea posible estudiar la evolución de las convenciones entre sí.

Protectos trabaja con estas dos versiones, la normalizada y la original. El análisis sintáctico y semántico se realiza sobre la versión normalizada y el análisis de la compaginación, la ortografía y la puntuación se hace sobre la versión original.

La segunda operación, la configuración gráfica del texto, se realiza después de la transcripción de los textos y según los criterios de análisis sintáctico que explicaremos a continuación. Siguiendo la propuesta del *Groupe Aixois de Recherche en Syntaxe* (GARS) (1983) consideramos un tipo de representación gráfica de los textos, las *grilles*, que explotan dos principios:

a) La secuencia de palabras que es posible analizar en función de los emplazamientos y las relaciones sintácticas.

b) La secuencia de palabras que pertenecen a un mismo emplazamiento sintáctico y que mantienen entre sí ya no relaciones sintácticas sino relaciones paradigmáticas.

Ambos principios se representan gráficamente a través de la explotación de los ejes horizontales y verticales. En el eje horizontal, en cada línea se representa la serie concreta de palabras que pertenecen a un mismo sintagma y que mantienen entre sí relaciones de gobierno o de asociación. En el eje vertical se representan los «rellenos» de un mismo emplazamiento sintáctico para ubicar las palabras que mantienen entre sí relaciones paradigmáticas. Es así en el ejemplo siguiente,

| | | |
|-----------|------|---------|
| és un gos | molt | pacific |
| | i | graciós |

«graciós» ha sido emplazado en la misma columna que «pacífic» porque ambos pertenecen a la misma unidad sintáctica. Los emplazamientos en la misma posición sintáctica dan lugar a «listas». En las listas paradigmáticas se puede colocar una palabra completa, un fragmento de palabra o bien un sintagma a condición de que pertenezcan a la misma unidad sintáctica.

En la puesta en configuración la primera línea se reserva para el título que el niño le dio al texto; en las sucesivas líneas, las diferentes posiciones sintagmáticas que comprenden una secuencia gramatical. La línea se corta al interrumpirse una secuencia gramatical. La siguiente secuencia gramatical se escribe debajo y así sucesivamente. Como consecuencia, el texto queda configurado gráficamente en *grille*. Cada línea incluye todas las posiciones sintagmáticas realizadas en el texto y cada columna los elementos lexicales que ocupan la misma posición sintagmática, es decir, los paradigmas.

Este análisis sintáctico en *grilles* ofrece una primera aproximación formal al texto, una especie de «esqueleto» al cual se le pueden superponer otros análisis. El principio es el siguiente: la explotación de lo gramatical antes que lo discursivo (a través del análisis apoyado en las categorías y la distribución en los ejes sintagmáticos y paradigmáticos) y la explotación de la sintaxis antes que lo lexical (a través de una medida de gramaticalidad para decidir entre lo sintáctico y lo lexical).

Siendo éste un análisis de superficie, no se cambia ni el orden sintagmático ni el paradigmático utilizado por el autor del texto, como sería el caso en otros esquemas de análisis (Harris, 1952). Una vez puesto el texto en configuración, es posible observar el avance textual en el eje sintagmático y abordar cada uno de los paradigmas sin perder de vista el texto total. El texto puede avanzar por agregado de posiciones sintagmáticas o avanzar por reducción de posiciones sintagmáticas. Un texto también avanza por sustitución o repetición de elementos lexicales en cada uno de los paradigmas.

La confección de la *grille* es el punto clave a partir del cual se desarrollará el posterior análisis y el investigador debe ser especialmente cuidadoso en dicha confección. Para esta operación recomendamos la utilización de una hoja de cálculo.

Las categorías de análisis

Como principio general, los aspectos sintácticos y semánticos (campos) seleccionados para el análisis son, en su mayoría, indicadores potenciales de diferencias entre distintos tipos de textos. Por ejemplo, trabajos empíricos han demostrado que las construcciones sintagmáticas difieren por tipo de texto. Hemos incluido, pues, su posibilidad de categorización. Con el mismo criterio se ha incluido la posibilidad de categorizar la introducción y mantenimiento de la referencia, el tiempo, modo, aspecto y voz de los verbos; la categoría semántica y el rol de los predicados.

Se facilita, además, el análisis de las distintas categorías de palabra (nombres, adjetivos, preposiciones o verbos) y la posibilidad de calcular su relativa

co-ocurrencia. Aunque esta opción ha resultado ser poco fiable para evaluar la evolución sintáctica de los sujetos (Templin por ejemplo, demostró ya en 1957 que la proporción de estas categorías de palabra en niños de tres años es bastante similar a la de niños mucho mayores), la hemos considerado porque es otro de los indicadores potenciales de diferenciación entre tipos de textos. (Véase en el Apéndice I la lista de campos categorizados en sintaxis, semántica y ortografía.)

Protexotos propicia un análisis cualitativo inicial cuyo resultado puede ser luego exportado a paquetes estadísticos para su tratamiento cuantitativo. A pesar de ello hemos incluido unos pocos aspectos que pueden ser tratados inicialmente en forma cuantitativa. Por ejemplo, cálculo de número de palabras, letras y oraciones. Estos cálculos podrán eventualmente utilizarse para evaluar índices de dificultad de un texto o de complejidad sintáctica de las construcciones.

Unidades de análisis

Llegados a este punto nos parece esclarecedor hacer una disquisición teórica sobre el tema de las unidades de análisis del texto. No utilizamos la frase como unidad porque la frase es una unidad gráfica definida por criterios de puntuación: la mayúscula y el punto (Arrivé y otros, 1986). Nosotros sabemos que los niños no son expertos en el uso de la puntuación. Además también sabemos que es una unidad inoperante para el análisis del discurso (oral o escrito, si el escritor no es un experto, Blanche-Benveniste y Jeanjean, 1986). Pero sobre todo la frase no es una unidad lingüística sino una unidad del letrado (o más específicamente del tipógrafo, como lo señala Borrendonner, 1993). Tan pronto se sale de las unidades canónicas del tipo sintagma nominal + sintagma verbal, y se trabaja sobre las unidades atípicas es difícil decidir dónde termina la frase. En su lugar utilizamos dos tipos de unidades: una unidad de enunciación, el sintagma, definida desde el punto de vista sintáctico (pero que a nivel oral podría estar acompañado por una pauta entonativa, aunque es necesario advertir que las fronteras entonativas no siempre se corresponden exactamente con las fronteras sintácticas). Dentro del sintagma los elementos mantienen entre sí relaciones de gobierno o de asociación, ambos con selección y concordancia interna. La segunda unidad ha sido definida como bloque, los bloques son unidades mayores que el sintagma (aunque podrían coincidir en algunos casos) que se definen por su función comunicativa. Por ejemplo, en el caso de la conversación podrían corresponder a los turnos de habla. Pero en el caso del lenguaje escrito corresponden a unidades de información o de presuposición (Borrendonner, 1993). Por lo tanto para el análisis de los aspectos sintácticos y semánticos, el texto se dividirá en bloques y dentro de cada bloque podrán categorizarse unidades menores: sintagmas y palabras. La situación es algo diferente en ortografía. La ortografía estabiliza la escritura de los aspectos fonográficos de una lengua, por eso siempre se la considera en función de la norma. La difusión de la imprenta, por otra parte, ayudó a estabilizar también los aspectos (ideo)gráficos, tales como la separación entre palabras, la compaginación, el subrayado, el uso de puntuación, etc. Dado que se trata de estudiar la ortografía de los productores en relación a una reproduc-

ción autorizada, debemos considerar como unidades de referencia aquellas unidades gráficas que han sido estabilizadas y aceptadas por el uso ortográfico y tipográfico de la lengua, es decir la palabra, la frase y el párrafo.

Análisis «automático» de textos

Este tipo de análisis retoma la tradición de Harris (1952) y de Pêcheux (1969), más centrado en la actualidad en el campo de la lexicografía y del denominado análisis de datos textuales (Bécue y otros, 1992). Este tipo de enfoques, que había sido criticado históricamente por su excesiva ingenuidad (Maingueneau, 1976), se ha revitalizado con el uso de herramientas estadísticas e informáticas muy poderosas. La forma de proceder es, en lo esencial, un análisis de frecuencias de las distintas palabras que aparecen en un texto, además de un análisis de segundo nivel en el que tales palabras son colocadas en el interior de ejes de significación. Los resultados son muy variados, pero quizás uno de los más espectaculares es una representación en forma de «racimos» (*clusters*), y una representación de las distancias entre las distintas agrupaciones en un espacio plano. Esta técnica se puede aplicar a cualquier tipo de texto escrito, además de a otros (como, por ejemplo, las respuestas abiertas en encuestas).

Una de las principales ventajas de una aproximación lexicográfica es el escaso o nulo compromiso que mantiene con teorías lingüísticas. Es decir, un análisis como el que resulta de proceder a «descomponer» un texto mediante los programas estadísticos agrupados en paquetes como SPAD (Bécue y otros, 1992) es, en cierto sentido, anterior a cualquier trabajo de detalle sobre lingüística oracional, o sobre cualquier aproximación semiótica o narrativa. Anterior en cuanto puede hacerse rápidamente y proporciona una visión distinta a la que se obtiene con otras metodologías de más detalle, y anterior también por poder realizarse sin casi participación del analista.

Otra línea de investigación es la que se deriva del sistema CHILDES *Child Language Data Exchange System* (MacWhinney y Snow, 1990) y que, recientemente, se ha empezado a extender a textos escritos (García Hidalgo, 1991). El interés de un proyecto como CHILDES se evidencia a lo largo de los años en que se ha venido desarrollando, ya que supone una puesta en común de las transcripciones, realizadas por numerosos investigadores del lenguaje infantil.³

Los objetivos de este sistema son de carácter psicolingüístico y no lexicológico. Por tanto, hay un énfasis en los turnos de intervención (para la lengua oral), en aspectos morfológicos e incluso fonéticos, pero, muy especialmente, en aspectos sintácticos. Por otra parte, a pesar de que uno de los objetivos del sistema es la estandarización de herramientas y criterios para la transcripción-codificación, se trata de un enfoque abierto a las distintas preguntas de los investigadores que pueden crear nuevas categorías de análisis sobre la marcha.

3. Este sistema incluye una base de datos con las transcripciones CHILDES, una metodología para transcripción y codificación CHAT y un programa CLAN para el análisis de los archivos CHAT.

Entre las desventajas del CHILDES no es menor la complejidad codificativa, que requiere una gran dedicación por parte de los investigadores, así como un uso cercano del manual y un importante periodo previo de formación.

Descripción de Protexos

Protexos es un sistema de análisis de textos, que retoma algunas características del tipo de análisis desarrollado por otros programas informáticos anteriores, pero que parte de unos objetivos y propósitos muy diferentes.

En primer lugar, la unidad de análisis. Desde el punto de vista del programa informático esto se concreta en la definición de bloques. Un bloque es cualquier parte del texto, definida por el investigador de acuerdo con los criterios teóricos ya mencionados, que se caracteriza por ser continua (un bloque puede contener frases y párrafos distintos, con tal de que sean continuos).

En segundo lugar, las formas de «manejar» el texto. El texto siempre aparece en pantalla tal y como fue escrito, sin códigos ni otro tipo de descomposición o marcado. En el caso de análisis ortográficos se trabaja sobre el texto original (tal y como ha sido transcrito, con los errores cometidos); para el resto de análisis se utiliza una versión «normalizada» del texto, a la que simplemente se ha liberado de errores ortográficos y de puntuación —se trata, por tanto, de un texto legible—. Ambos textos permanecen siempre en pantalla durante todo el proceso de análisis.

En tercer lugar, el proceso de análisis. Denominado proceso de categorización, consiste en marcar en el texto, seleccionándolas, las unidades que van a ser categorizadas. Las categorías de análisis, distribuidas en tres grandes campos: sintaxis, semántica y ortografía, están permanentemente disponibles en forma de menú y submenú jerarquizados, que sólo deben ser señalados para que se produzca el proceso de categorización. El usuario no necesita conocer ningún tipo de comando especial o código para realizar todo el proceso. Las categorías de análisis son una lista cerrada, formada por 346 categorías.

En cuarto lugar, las prestaciones de visualización del sistema. *Protexos* guarda todas las categorizaciones hechas de manera doble: en el interior de cada subcategoría y en el del bloque de referencia. De esta forma, la visualización (y la rectificación en caso de error en la categorización) se realiza de manera natural: basta con elegir cualquier subcategoría o bloque para que aparezca el conjunto de las categorizaciones realizadas, y es suficiente con señalar una de ellas para que aparezca seleccionando el conjunto de caracteres que la forman en el interior del texto. Con este procedimiento la visualización del análisis es siempre contextualizada.

En quinto lugar, el sistema permite crear modelos de análisis. Un modelo de análisis es un subconjunto, elegido libremente, del total de categorías. La creación de un modelo de análisis es independiente del proceso de categorización en un doble sentido: se realizan en momentos (y pantallas) distintos y pueden ser aplicados sobre distintas categorizaciones. *Protexos* permite la creación de un máximo de 30 modelos, operación que se realiza con la misma filosofía que la catego-

rización —solamente señalando las subcategorías que se precisan del interior de una serie de listas—.

En sexto y último lugar, el sistema tiene unas opciones de impresión y de exportación de datos muy detalladas. Tanto la impresión, la grabación en disco de informes, como la exportación de los totales de frecuencias para cada subcategoría se realizan desde pantallas específicas, en las que el usuario sólo debe elegir entre opciones predefinidas. Todas las opciones incluyen la posibilidad de imprimir o exportar exclusivamente los datos relativos al modelo de análisis que se encuentre activo en ese momento.

Archivo Modos y Bloques Operaciones Sintaxis Semántica Ortografía

ProTextos

Joan/1/Nor Tipo de letra

1 Bloque 2 Tamaño de letra

El gos bonic i blanc
 En Miquel és un gos molt gandul i quan vol anar al parc ens fa fer que li portem al parc és un gos molt pacífic i graciós i quan s'enfada es fica a bordar i. Porta una corretja molt estranya en Miquel no és un gos com els altres es molt jove i té molt bona salut.

El gos vonic i blanc
 En Migel és un gos molt gandul i quan bol anar al parc ens fafer qué li portem al parc es un gos molt pasífic i graciós i quan senfada es fica a bordar i. Porta una corretxa molt estrany en Migel no es un gos quom els altres es molt jove i te molt bona salut.

Operaciones con Listas

oraciones: 2
 palabras: 64
 sílabas: 86
 caracteres: 284

a, 2
 al, 2
 otros, 1
 enar, 1
 blanc, 1
 bona, 1

Visualización del análisis

visualización
 Bloque 2

DV, En Miquel es un gos molt gandul, 22 to 53
 DV, es un gos molt pacífic i graciós, 111 to 142
 DV, i. Porta una corretja molt estranya, 177 to 212
 Indicativo presente, és, 32 to 33
 Perfectivo, és, 32 to 33
 Activo, és, 32 to 33
 Asertivo, és, 32 to 33

Figura 1. Pantalla de *Protectos* con texto original y normalizado.

En la Figura 1, una pantalla de *Protectos* donde pueden observarse los dos textos, el original y el normalizado, los resultados del cálculo de los rasgos formales y la visualización de la categorización realizada en uno de los bloques.

En la Figura 2, otra pantalla en la cual puede observarse el menú de sintaxis abierto.

En conjunto, *Protectos* representa un sistema completo de análisis de textos escritos, con capacidades de exportación de sus datos a un paquete estadístico para análisis cuantitativos, que se ha diseñado pensando en un marco teórico explícito y con la intención de simplificar al máximo desde el punto de vista informático.

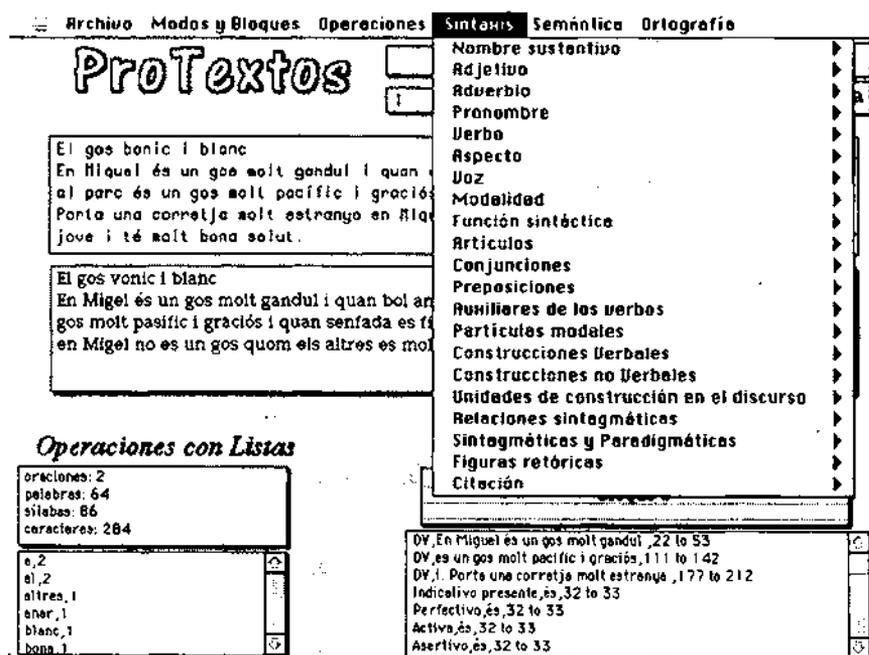


Figura 2. Pantalla de *Protexos* con menú de sintaxis.

Itinerarios de análisis

Es fundamental que el investigador defina sus preguntas y sus estrategias. Dado que el programa ofrece muchas posibilidades y categorías de análisis, la estrategia del investigador es la que organizará el orden y la búsqueda. Como ya dijimos *Protexos* dispone de la posibilidad de crear un modelo de análisis que incluya los aspectos y categorías seleccionados por el investigador para un proyecto específico. De esta manera no se tendrá que poner en juego el programa completo cada vez que se desarrolle un proyecto.

Presentaremos a continuación un itinerario posible de análisis, que sirva como ejemplo de utilización de *Protexos* en una investigación en curso (Teberosky, Tolchinsky, Arnau y Valeta, en prensa). Ésta se propone estudiar el desarrollo de la competencia literaria en niños de 8 y 9 años de edad, escolarizados en catalán pero que difieren en su lengua familiar. Se trata de averiguar cómo se manifiesta la competencia literaria en la producción escrita de textos que se identifican como «descripción» y cómo se ve influida por una programada intervención pedagógica. Además, muy especialmente, cómo la coincidencia o divergencia entre la lengua familiar y la lengua escolar afecta a la producción de textos.

Para cumplir con el primer objetivo, uno de los pasos podrá ser buscar la co-ocurrencia de ciertas características lingüísticas en los textos que estos niños han producido y determinar su relación con las supuestas características lingüísticas de un texto descriptivo. Tal como señalamos antes una co-ocurrencia de características lingüísticas no es aleatoria, está motivada por una influencia funcional subyacente, como puede ser, por ejemplo, el género.

Otro de los pasos podrá consistir en la verificación de la compaginación, la ortografía y la puntuación de los textos. Se podrá luego determinar la relación entre estos aspectos y la lengua de origen de los niños.

Para contextualizar algunas de las preguntas específicas que se pueden formular a partir de estos objetivos y ejemplificar la utilización del programa, se pueden consultar en el anexo algunas de las características de los textos descriptivos.

Análisis de aspectos sintácticos del texto descriptivo

A partir de una caracterización del tipo de texto, el investigador se podrá formular una serie de preguntas sobre las cuales basar su estrategia de análisis. Cualquiera que sea ésta el investigador deberá, obviamente, haber introducido todos los textos que quiera analizar. Debemos recordar que el investigador dispondrá en pantalla de la versión original y el texto normalizado, tal como se puede ver en la Figura 1. Pero, además, deberá tener visualmente disponible la *grille* del texto, ya que tanto los bloques como ciertas decisiones respecto a las construcciones dentro de cada bloque, deberán tomarse en función de la *grille*. En la Figura 3 aparece en *grille* el texto sobre el cual desarrollaremos nuestro ejemplo. Se trata de una descripción escrita por un niño de 8 años a partir de un texto modelo.

| | | | | | |
|----------------------|------|----------|------|--------------|-----------------|
| El gos bonic i blanc | | | | | |
| i | quan | vol | anar | al | parc |
| | | | | En Miquel | |
| | | | | ens | és |
| | | | | fa fer | un gos |
| | | | | és | que li portem |
| | | | | | al parc |
| | | | | | un gos |
| | | | | | molt pacífic |
| | | | | | i graciós |
| i | quan | s'enfada | | es | fica a bordar |
| i | | | | | porta |
| | | | | en Miquel no | una corretja |
| | | | | | un gos |
| | | | | | molt estranya |
| | | | | | com els altres |
| | | | | | és |
| | | | | | molt jove |
| | | | | | molt bona salut |
| i | | | | | té |

Figura 3. *Grille* del texto utilizado como ejemplo.

Procederemos ahora siguiendo un itinerario de análisis posible. Para comenzar el investigador podrá ir al menú «Operaciones» y en la categoría «Variables del texto» podrá hacer constar todos los datos del sujeto: nombre, edad, sexo, escuela, lengua de origen, etc., y definir la «situación de entrevista». Dentro del

menú «Operaciones» podrá solicitar la lista de palabras y el cálculo de otros rasgos formales del texto, tales como número de palabras, número de frases, etc.

A partir del menú «Modos y Bloques», procederá a definir los bloques. En el ejemplo concreto que nos ocupa, se ha definido el Bloque 1 para el título del texto: «*El gos bonic i blanc*». Los otros dos bloques han sido definidos en función de las dos unidades de información. La primera unidad, «*En Miquel és un gos gandul*», está caracterizada por la estructura «X es un Y», donde el nombre propio necesita un valor: *és un gos* y una propiedad: *gandul*. La segunda unidad, «*en Miquel no és un gos com els altres*», que está caracterizado por la estructura «X no es un Y», introduce la nueva información de negar una propiedad: la de *ser com els altres* (Fauconnier, 1984). Nótese que cada una de estas descripciones está organizada con un nombre propio y un artículo definido e introduce nuevos elementos en el espacio del discurso del escritor.

Una vez definidos los bloques, el investigador podrá comenzar la categorización. Recordemos que el punto de partida de esta metodología de análisis es siempre sintáctico, excepto en aquellos casos que el investigador esté interesado sólo en el análisis de la ortografía. En el ejemplo aquí presentado, imaginemos algunas de las preguntas que se pueden plantear a nivel sintáctico:

¿Qué construcción sintáctica tiene el título?

¿Qué construcciones del discurso aparecen en cada bloque?

¿Qué características de tiempo, modo, voz y aspecto tienen los verbos rectores?

¿Qué categorías del tiempo y del modo, de la voz, del aspecto y de la modalidad?

¿Qué categorías de nombres, de adjetivos y de formas adverbiales aparecen en el texto?

¿Qué construcciones paradigmáticas especiales aparecen en cada bloque?

Cada uno de estos aspectos está categorizado en el programa. El investigador comenzará pues por el bloque 1 (título) y buscará la categoría correspondiente en el campo «construcciones sintácticas», en este caso señalará la categoría «construcción nominal desnuda». Así continuará con el resto de las preguntas formuladas. Evidentemente la respuesta a estas preguntas lo ayudará a decidir en qué medida el texto del niño presenta las características atribuidas a un texto descriptivo. Además, al visualizar la categorización realizada de todos los campos para cada bloque, el investigador podrá observar todas las realizaciones lexicales correspondientes a cada una de las construcciones o de los elementos sintácticos categorizados.

Análisis de aspectos semánticos del texto descriptivo

El análisis de los aspectos semánticos seleccionados se realiza en los mismos bloques que han sido definidos para el análisis de los aspectos sintácticos. Lo primero que el investigador podrá definir es el tópico del texto. Como ésta

es una noción intuitiva difícil de formalizar, se ha decidido que la mejor opción en este tipo de trabajos es que el investigador se formule una palabra, un nominal preferentemente o una construcción nominal, que sirva para explicar «acerca de qué es el texto». Por ese motivo esta variable no aparece categorizada y puede introducirse manualmente. Al finalizar el proyecto se podrá obtener una lista de todos los tópicos que aparecen. En el ejemplo aquí presentado, imaginemos algunas de las preguntas que se pueden plantear a nivel semántico:

- ¿Qué estructuras textuales aparecen en cada bloque?
- ¿En qué sintagma aparece la primera mención del referente?
- ¿Cómo se mantiene esta referencia en los sucesivos sintagmas nominales?
- ¿Cuál es el rol semántico de los predicados?
- ¿Cuáles son las categorías semánticas de los distintos predicados?
- ¿Qué formas adverbiales aparecen?
- ¿Qué relaciones de cohesión lexical aparecen en cada bloque?

Cada uno de estos aspectos está categorizado en el «menú semántica». Al igual que en el «menú sintaxis», al visualizar la categorización el investigador podrá observar las realizaciones lexicales de las estructuras textuales que ha definido y categorizado, del sintagma que introduce el referente, de las sucesivas menciones, etc. Respecto de la última pregunta formulada, la visualización le permitirá observar en qué bloque aparecen relaciones de sinonimia, de hiperonimia, etc. Evidentemente, la respuesta a estas preguntas lo ayudará a decidir en qué medida el texto del niño presenta las características atribuidas a un texto descriptivo desde un punto de vista semántico.

Análisis de los aspectos ortográficos

El texto que nos sirve de ejemplo fue producido por un niño de 8 años y fue analizado también desde el punto de vista de su ortografía. Los dominios del menú «ortografía» se refieren tanto a aspectos estrictamente ortográficos como a aspectos de compaginación y formato y a las separaciones que en el texto dan lugar a las unidades gráficas de palabras, frases y párrafos. Además el instrumento contempla no sólo la posibilidad de categorizar los tipos de errores, sino también la operación realizada por el niño que dio lugar al error, definida ésta como: omisión, adición, sustitución o desplazamiento y además la categoría gramatical sobre la cual se realizó el error.

En el ejemplo que nos ocupa, el investigador comenzó por categorizar la presencia del título, el tipo de justificación (por la izquierda) y el número de párrafos (uno sólo). Una vez categorizados estos aspectos gráficos, el investigador podrá proceder a categorizar los errores de ortografía. Seleccionará entonces la palabra en la cual aparece el error y categorizará en el campo correspondiente la operación realizada por el niño. Posteriormente podrá categorizar, por ejemplo, la categoría gramatical a la que pertenece la palabra sobre la cual se cometió el error. Siguiendo sucesivamente con las distintas palabras podrá decidir qué aspectos le interesa categorizar en cada caso. Al finalizar la categorización de to-

das las palabras en las cuales aparecen errores podrá visualizarlas con la correspondiente categorización. Podrá comprobar así que para este texto, su autor mantuvo una correspondencia fonográfica de tipo alfabético, que realizó omisión de acentos y de apóstrofe y que sustituyó consonantes en nombres y adjetivos.

Algunas de las preguntas adicionales que se pueden plantear a nivel ortográfico son:

¿Qué errores de «segmentación» aparecen en el texto?

¿Qué errores de puntuación?

¿En qué lugar de la cadena sintagmática ocurren los errores de puntuación?

Evidentemente el investigador debe acompañar este análisis de un análisis sintáctico-semántico para relacionar la gramática del discurso con la gramática de las inscripciones gráficas.

Observaciones finales

Los itinerarios aquí presentados son una muestra para ejemplificar las posibilidades de uso y explotación del programa informático. Pero resaltaremos nuevamente que *Protexos* se encuentra en una fase inicial de elaboración. Pensamos que a partir de su uso en distintos proyectos de investigación y después de sucesivas evaluaciones podrá constituirse en una herramienta de trabajo útil que permita la comparación cualitativa y cuantitativa de textos producidos por sujetos de diferente edad, diferente historia lingüística y en circunstancias enunciativas con propósitos comunicativos diversos. Queda un enorme trabajo por hacer. Esperamos completar y reformar el programa a partir de las sugerencias y observaciones de todos los que tengan interés en probarlo.

Por otra parte es posible que a partir de la utilización de *Protexos* podamos llegar a constituir una base de datos de textos escritos. Esta base de datos podrá ser útil para confeccionar vocabularios básicos, para comparar la competencia textual de distintos sujetos, para evaluar la incidencia de variables socio-culturales o educativas, para evaluar las diferencias individuales, para definir criterios de evaluación y calidad textual. Con esta herramienta estamos trabajando en nuevas maneras de evaluación del desarrollo del lenguaje escrito. Probablemente podamos, además, contribuir a la elaboración de programas de intervención y de pautas de evaluación de textos escritos en el marco de la institución escolar.

REFERENCIAS

- Arrivé, M., Gadet, F. & Galniche, M. (1986). *La grammaire d'aujourd'hui: guide alphabétique de linguistique française*. Paris: Flammarion.
- Bécue, M., Lebart, I. y Rajadell, N. (1992). El análisis estadístico de datos textuales. La lectura según los escolares de enseñanza primaria. *Anuario de Psicología*, 55, 7-22.
- Biber, D. (1988). *Variation across speech and writing*. Cambridge: Cambridge University Press.
- Blanche-Benveniste, C. & Chervel, A. (1976). *L'Orthographe*. Paris: Maspéro.

- Blanche-Benveniste, C. & Jeanjean, C. (1986). *Le français parlé: Édition et transcription*. Paris: Didier Érudition.
- Blanche-Benveniste, C., Bilger, M., Rouget, Ch. & Van den Eynde, K. (1990). *Le français parlé: Études grammaticales*. Paris: Éditions du CNRS.
- Blanche-Benveniste, C. (1990). Un modèle d'analyse syntaxique «en grilles» pour les productions orales. *Anuario de Psicología*, 47, 11-28.
- Bernstein, B. (1971). *Class, Codes and Control. Vol. 1*. London: Routledge.
- Borrendonner, A. (1990). Pour une macro-syntaxe. *Travaux de linguistique*, 21, 25-36.
- Borrendonner, A. (1993). La phrase et les articulations du discours. *Le Français dans le monde*. Février-Mars, 20-26.
- Carroll, J.B. (1962) Language development in children. In S.W. Harris (Ed.) *Encyclopedia of Educational Research*. New York: Macmillan.
- Culler, J. (1975). *Structuralist Poetics*. London. (Trad. cast. *La política estructuralista*. Barcelona: Anagrama, 1979.)
- Eco, H. (1992). *Los límites de interpretación*. Barcelona: Lumen.
- Fauconnier, G. (1984). *Espaces mentaux*. Paris: Les Éditions de Minuit.
- García Hidalgo, M.I. (1991). Documento no publicado. Proyecto para una base internacional de escrituras infantiles. Universidad «la Sapienza» de Roma y CINVESTAV de México.
- Gresillon, A. & Lebrave, J.L. (1983). Avant-propos, Manuscrits-Ecriture, Production linguistique. *Langages*, 69, 5-10.
- Harris, Z. (1952). Discourse Analysis. *Language*, 28, 1-30.
- Iser, W. (1987). *El acto de leer*. Madrid: Taurus.
- Lebart, L. & Salem, A. (1988). *Analyse Statistique des Données Textuelles*. Paris: Dunod.
- MacWhinney, B. & Snow, C. (1990). The child language data exchange system: an update. *Journal of Child Language*, 17, 457-472.
- Rodríguez Illera, J.L. (1993). *Strings*. ICE. Universitat de Barcelona. Versión 1.5, 1993.
- Rodríguez Illera, J.L. y León M. (1990). La producció de software en educació Especial: estudi d'un cas. En J.L. Rodríguez Illera (Comp.). *Informàtica i Educació Especial* (pp. 79-99), ICE, Universitat de Barcelona.
- Ryan, M.L. (1977/1988). Hacia una teoría de la competencia genérica. En M.A. Garrido Gallardo (Comp.). *Teoría de los géneros literarios*. Madrid: Arcos/Libros.
- Scinto, L. (1986). *Written language and psychological development*. London: Academic Press.
- Shiffrin, D. (1981). Tense variations in narratives. *Language*, 57, 45-62.
- Tannen, D. (1982). Oral and written strategies in spoken and written narratives. *Language*, 58, 1-21.
- Templin, M. (1957). *Certain language skills in children: their development and interrelationships*. Institute of Child Welfare Monograph Series, 26. University of Minnesota Press.
- Teberosky, A. (1989). La escritura de textos narrativos. *Infancia y Aprendizaje*, 46, 17-36.
- Teberosky, A. (1991). Re-escribiendo noticias. *Anuario de Psicología*, 47, 43-63.
- Teberosky, A., Tolchinsky, I., Arnau, J. y Valeta, A. (en prensa). Composición de «descripciones». De qué se apropian los niños de los modelos literarios. En M. Siguan (Coord.). *Modelos de enseñanza de segundas lenguas*. Barcelona: ICE/Horsori.
- Tolchinsky, I., Teberosky, A. y Rodríguez Illera, J.L. (1993). *Protectos*. Barcelona: ICE, Universitat de Barcelona.
- Tolchinsky, I. (1991). La reproducción de relatos en niños entre cinco y siete años: Organización sintáctica y funciones narrativas. *Anuario de Psicología*, 47, 64-87.
- Tolchinsky, I. (1992). Calidad narrativa y contexto escolar. *Infancia y Aprendizaje*, 58, 83-106.

APÉNDICE I

LISTADO DE CAMPOS CATEGORIZADOS EN EL MENÚ SYNTAXIS, SEMÁNTICA Y ORTOGRAFÍA

| Menú | Análisis sintáctico |
|--------------|--------------------------|
| DOMINIO: | Categorías gramaticales: |
| SUB-DOMINIO: | Categorías léxicas: |
| CAMPO: | El nombre sustantivo |

| | |
|--------------|--|
| CAMPO: | El adjetivo |
| CAMPO: | El adverbio |
| CAMPO: | El pronombre |
| CAMPO: | El verbo |
| CAMPO: | Aspecto |
| CAMPO: | Voz |
| CAMPO: | Modalidad |
| CAMPO: | Función sintáctica |
| DOMINIO: | Categorías no léxicas |
| CAMPO: | Artículos |
| CAMPO: | Conjunciones |
| CAMPO: | Preposiciones |
| CAMPO: | Auxiliares de los verbos |
| CAMPO: | Partículas modales |
| DOMINIO: | Unidades de construcción |
| SUB-DOMINIO: | De construcción sintáctica: |
| CAMPO: | Construcciones verbales |
| CAMPO: | Construcciones no verbales |
| CAMPO: | Unidades de construcción en el discurso |
| DOMINIO: | Las relaciones |
| SUB-DOMINIO: | Relaciones sintagmáticas |
| CAMPO: | Sintagmáticas |
| CAMPO: | Sintagmáticas y paradigmáticas: |
| CAMPO: | Figuras retóricas |
| SUB-DOMINIO: | Relaciones enunciativas: |
| CAMPO: | Citación |
| Menú | Análisis semántico |
| DOMINIO: | Tópico |
| DOMINIO: | Unidades textuales |
| CAMPO: | Tipo de unidades textuales |
| CAMPO: | Relación con texto modelo |
| DOMINIO: | Cohesión textual |
| SUB-DOMINIO: | Referencia nominal |
| CAMPO: | Mención de los referentes centrales y secundarios |
| CAMPO: | Formas de los sintagmas nominales |
| CAMPO: | Función sintáctica del sintagma nominal en el cual se menciona el referente. |
| CAMPO: | Roles semánticos de los sintagmas nominales |
| CAMPO: | Posición respecto del verbo constructor |
| CAMPO: | Relación de co-referencia intrasintagmática |
| SUB-DOMINIO: | Referencia temporal/temporalidad |
| CAMPO: | Predicados |
| CAMPO: | Conectivos |
| CAMPO: | Adverbios y / construcciones adverbiales |
| CAMPO: | Conectivos lexicales |
| Menú | Organización gráfica y ortografía |
| DOMINIO: | Compaginación y formato |
| CAMPO: | Título |
| CAMPO: | Sub-títulos o intertítulos |
| CAMPO: | Párrafos |
| CAMPO: | Formato |
| CAMPO: | Subrayado, otros (no dibujo) |

| | |
|--------------|---|
| DOMINIO: | Ortografía de los textos |
| SUB-DOMINIO: | Correspondencia fonográfica |
| CAMPO: | La operación realizada por el niño |
| CAMPO: | La unidad sobre la cual se realizó la operación |
| CAMPO: | Categoría de la palabra |
| CAMPO: | Posición y tonicidad de la sílaba sobre la cual se realizó la operación |
| CAMPO: | La cantidad de sílabas de la palabra: |
| SUB-DOMINIO: | Acentuación |
| CAMPO: | La operación realizada por el niño |
| CAMPO: | La unidad sobre la cual se realizó la operación: |
| CAMPO: | Categoría de la palabra |
| CAMPO: | Patrón de tonicidad de la palabra en la cual se realizó la operación |
| CAMPO: | La cantidad de sílabas de la palabra: |
| CAMPO: | Tipo de marca |
| SUB-DOMINIO: | Segmentación |
| CAMPO: | La operación realizada por el niño |
| CAMPO: | Unidad sobre la cual se realizó la operación |
| CAMPO: | Tipo de segmentaciones |
| SUB-DOMINIO: | Léxico / morfosintáctico |
| CAMPO: | La operación realizada por el niño |
| CAMPO: | Unidad sobre la cual se realizó la operación |
| CAMPO: | Categoría de palabra sobre la cual se realizó la operación |
| DOMINIO: | Puntuación |
| CAMPO: | Operación realizada por el niño |
| CAMPO: | Tipo de signo |
| CAMPO: | Ubicación del signo en la cadena sintagmática |
| CAMPO: | Blancos |

APÉNDICE II

LA DESCRIPCIÓN COMO TIPO DE TEXTO

La descripción se opone a la narración, como tipo de texto, en el aspecto sintáctico y en el aspecto semántico. En el sintáctico, particularmente en lo que respecta a los dos tiempos del discurso: la descripción supone una interrupción de uno de los tiempos de la narración, el tiempo histórico y una focalización en el tiempo del discurso. Una de las funciones de la descripción es la de detener el tiempo de narración de los acontecimientos. En cuanto a lo que los teóricos del discurso han denominado «macroestructura» (van Dijk, 1977) la jerarquía de la descripción se constituye en el discurso. La descripción se relaciona con el nombrar, el definir y el calificar más que con el relatar.

Si la competencia literaria se refiere a un conocimiento acerca de los textos, conocer la descripción querrá decir aplicar un conjunto de reglas o principios obligatorios y opcionales que el productor y el lector comparten respecto a este tipo de textos. En el caso de la descripción, éstos deberán manifestarse en:

Las características de la persona y del verbo (modo, tiempo, aspecto y tipo de verbo). Es obligatorio el uso de modo indicativo, en tiempos del presente o del pasado inmediato. Aunque es más frecuente el uso de aspecto perfectivo, el aspecto es más opcional. La persona del relator también es opcional.

La organización sintagmática. El uso de verbos estativos obliga a la presencia de nominales y adjetivales en su valencia así como a cierto tipo de elementos asociados a la construc-

ción verbal. Éstos deberán cumplir una función de contextualización locativa y/o temporal, pero puntual —de ubicación discursiva— y sin constituir una secuencia cronológica.

La organización paradigmática. Es obligatorio nombrar y calificar. Es decir, deberán aparecer enumeraciones, comparaciones, series escalares o ejemplificaciones. Es opcional que estas alternativas lexicales aparezcan sobre una de las posiciones sintácticas produciendo el efecto de listas paradigmáticas sobre diferentes posiciones sintácticas.

El desarrollo de la cohesión textual. La cohesión textual se puede plantear desde la perspectiva de la introducción y mantenimiento de la referencia y de la temporalidad (o referencia temporal). La primera incluye toda la información correspondiente a las construcciones nominales y la segunda a las construcciones verbales. La descripción cumple ciertas restricciones respecto a estos dos aspectos. La primera mención de el o los referentes centrales puede realizarse por medio de «presentativos» incluyendo sintagmas nominales con determinantes indefinidos o definidos. Pero las menciones subsecuentes —que en las narraciones suelen realizarse por medio de cadenas anáforicas con utilización de pronombres o anáfora cero— se mantiene en las descripciones por el uso del mismo tipo de determinantes que se utilizó en el «presentativo». En cuanto al tipo de predicados de conectivos y de expresiones adverbiales que estructuran la referencia temporal, en la descripción encontraremos fundamentalmente predicados que representan estados y expresiones adverbiales no temporales.

La descripción también se caracteriza por la explotación de medios lexicales de cohesión. Estas «relaciones de sentido» entre elementos lexicales incluidos en el texto se verificarán tanto en el eje paradigmático como en el sintagmático. Se espera la presencia de relaciones de sinonimia, de hiperonimia, de equivalencias, de oposiciones y de repeticiones.

La descripción de la información. El tema de la descripción (y a veces algunas de las particularidades del referente) influye sobre la organización y sobre las unidades de contenido. Pero es opcional el tipo de organización y la cantidad, calidad y modo de progreso de las categorías y unidades de contenido.

