

UN REGISTRO LLAMADO ORACIÓN, UNA BASE DE DATOS LLAMADA TEXTO

Antonio García Velasco
Universidad de Málaga

Sumario

El paralelismo de los conceptos

TEXTO.....	BASE DE DATOS
ORACIÓN.....	REGISTRO
“CONSTITUYENTES ORACIONALES”	CAMPOS

responde a la necesidad de establecer las bases teóricas (o teórico-prácticas) para dos ámbitos diferentes de aplicación del análisis sintáctico-semántico: el computacional y el escolar.

En el primero, tales bases teóricas orientan un modo de procesamiento del lenguaje natural, del que se derivarían numerosas aplicaciones, entre las que hay que contar tanto la de utilización de nuevas tecnologías a la enseñanza de la lengua, como, de un modo más concreto, la del desarrollo de cierto tipo de material didáctico.

En el segundo, justifican una estrategia para una eficaz instrucción a los escolares en la comprensión lectora, en el comentario de textos y en el análisis sintáctico.

El anterior presupuesto explica las siguientes partes de este trabajo:

1. El texto como base de datos, la oración como registro: justificación.
2. Análisis sintáctico-semántico computacional: aplicaciones.
3. Estrategia didáctica para la comprensión lectora, el comentario de textos y el análisis sintáctico-semántico.

1. El texto como base de datos, la oración como registro: justificación

1.1. Introducción

El texto es concebido en la actualidad como la unidad de comunicación, es decir, como unidad poseedora de sentido integral, autosuficiencia y completitud. Kallmeyer, por ejemplo, lo define como “*la totalidad de las señales comunicativas que aparecen en una interacción comunicativa*”. Podemos, pues, idear el texto como un conjunto de elementos que completan y, a la vez, integran y encierran o delimitan un universo comunicativo. Los elementos de ese universo no está aislados unos de otros, por el contrario, aparecen organizados en subconjuntos y relacionados entre sí atendiendo a reglas de coherencia interna y

sintagmática y secuencialmente.

El texto puede, a veces, dividirse en párrafos o parágrafos (unidades relativamente autónomas tanto en el plano semántico como sintáctico y, en ocasiones, con marcas delimitadoras específicas). Por otra parte, el texto, como en su caso el párrafo, necesariamente, está constituido por una o más oraciones. Desde la perspectiva que adoptamos, es decir, al constituir un subconjunto de oraciones del texto, el párrafo no tiene ninguna relevancia y, en consecuencia, prescindimos de considerarlo. La oración, por el contrario, es unidad compo-nencial básica de los textos y será, y ha sido siempre para toda gramática, unidad de sentido, delimitada por entonación terminal y con independencia (relativa) sintáctica. Por tanto, dados las características de los elementos de la escala de la figura 1, sólo nos van a interesar las unidades texto, oración, proposición, sintagma y palabra.

1.2. Oraciones y texto, Registros y base de datos

Un texto, desde el punto de vista semántico, contiene un conjunto de informaciones sobre un determinado tema. Un conjunto de informaciones es una base de datos. Si una base de datos está formada por un conjunto de registros, cabe preguntarse cuáles son los registros de un texto. La respuesta es simple: las oraciones.

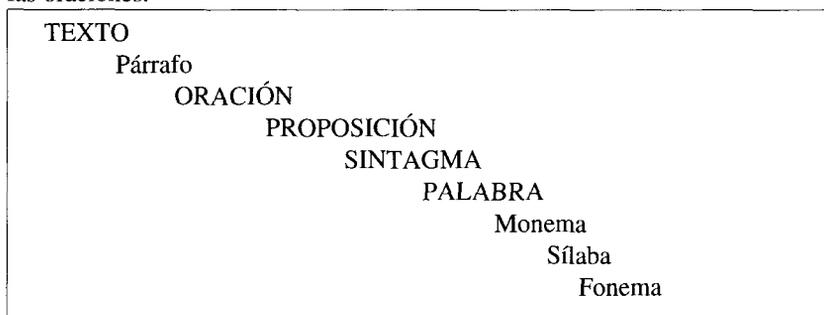


Figura 1

Si un registro está compuesto por campos, su equivalencia en el texto, o sea, la oración, debe estar compuesta por un conjunto de estructuras paralelas a tales campos. En efecto, es así. Pero precisemos, antes de hablar de tales campos, el carácter de estructura multiforme que tiene la oración. Una oración, en efecto, puede aparecer con las formas que, utilizando los símbolos de vocabulario no-terminal, se indican en la figura 2.

Estas estructuras oracionales, que pertenecen a oraciones unitarias o a oraciones bimembres, si bien muestran una reescritura sencilla, pueden desarrollarse con gran complejidad, ya que, dadas las posibilidades de recursividad, la complejidad de los sintagmas, llega, en ocasiones, a extraordinaria. Pero lo que ahora nos interesa es constatar que esta variedad de estructuras se corresponde

O -> Nom	O = Oración
O -> Vrb	Nom = Nombre
O -> Pro	Vrb = Verbo
O -> Adj	Adj = Adjetivo
O -> Adv	Adv = Adverbio
O -> Inf	Inf = Infinitivo
O -> Grd	Grd = Gerundio
O -> Prt	Prt = Participio
O -> Nom + Vrb	Pro = Pronombre
O -> Nom + SV	SN = Sintagma nominal
O -> Nom + Pro	SV = Sintagma verbal
O -> Pro + Vrb	SAdj = Sintagma adjetival
O -> Pro + SV	SAdv = Sintagma advercial
O -> SN	SInf = Sint. de Infinitivo
O -> SV	SGrd = Sint. de Gerundio
O -> SAdj	SPrt = Sint. de Participio
O -> SAdv	
O -> SInf	
O -> SGrd	
O -> SPrt	
O -> SN + SV	
O -> Pro + Nom	
O -> Pro + SN	
O -> Pro + Adj	
O -> Pro + SAdj	
O -> Pro + Adv	
O -> Pro + SAdv	

Subfigura 2.1

Figura 2.

con el mayor o menor número de informaciones que puede contener una oración y que este número de informaciones se va a corresponder con el contenido o vacío de los campos correspondientes. Volveremos sobre este asunto.

Una oración es, como queda dicho, un registro, registro para el que podemos establecer los siguientes campos: Hecho, Actor, Objeto, Modo, Tiempo, Causa, Lugar, Finalidad y Relación. Véase la figura 3, donde el contenido de algunos de estos campos se establece a partir de la información proporcionada por la oración *“Ayer empezó su funcionamiento el alumbrado de la travesía de Cártama Estación, en la carretera MA-402, junto con una serie de mejoras realizadas en la zona, para favorecer la seguridad vial, según informó el alcalde de Cártama”*.

Registro: Oración citada.
 Campos:
 Actor: *el alumbrado de la travesía de Cártama Estación*
 Hecho: *empezó*
 Objeto: *su funcionamiento*
 Modo: *junto con una serie de mejoras realizadas en la zona según informó el alcalde de Cártama*
 Tiempo: *ayer*
 Causa:
 Lugar: *en la carretera MA-402*
 Finalidad: *para favorecer la seguridad vial*
 Relación:

Figura 3

No toda oración tiene respuesta (contenido) para cada uno de los campos y, por otra parte, una oración puede tener más de una respuesta para alguno de los campos y respuesta implícita (en el contexto o en la situación extralingüística) para otros. Por ejemplo, para *Causa*: “peligro por mala iluminación” es contenido implícito. Además, en una oración pueden aparecer estructuras marginales que no sean respuesta a ninguno de sus campos, por ejemplo, pueden aparecer interjecciones y estructuras con función de vocativo.

1.3. Justificación

El establecimiento previo de la información que podemos esperar en una oración y la distribución de la misma en campos concretos está en función de los dos objetivos que ya quedan apuntados en el **Sumario** del presente trabajo: el análisis lingüístico computacional y la instrucción a los escolares en la comprensión lectora, el comentario de texto y el análisis sintáctico. Se parte de la idea de que en todo análisis lingüístico, sea oracional o textual, hemos de tener muy clara la distinción entre *enunciado*, *estructura*, *reescritura* y *función*.

El texto, la oración, la proposición, el sintagma, la palabra en concreto nos proporciona el *enunciado* o enunciados. La *estructura* vendrá proporcionada por un elemento terminal o simple (Nombre, Adjetivo, Verbo...) o por un elemento definido (Sintagma, Proposición...). La *reescritura* nos proporciona la composición de la estructura (elementos terminales o simples que la integran) y, por último, la *función* vendrá determinada por el campo que el elemento informativo oracional correspondiente completa.

En la oración citada anteriormente, por ejemplo, el enunciado *el alumbrado de la travesía de Cártama Estación* tiene una estructura definida (Sintagma nominal), que se reescribe como Determinativo + Nombre + SPrp (Sintagma

preposicional), este *SPrp* consta de Preposición (*Prp*) y Sintagma nominal formado por Determinativo + Nombre +, de nuevo, *SPrp*. Este *SPrp* consta de *Prp* + *SN* y, por último, este *SN* aparece formado por Nombre + Nombre.

La información que proporciona corresponde al campo Actor. En la figura 4 aparece el análisis de manera más formalizada.

Enunciado	<i>el alumbrado de la travesía de Cártama Estación</i>
Estructura	<i>SN</i>
Reescritura	<i>Det + Nom + SPrp; SPrp -> Prp + SN;</i> <i>SN -> Det + Nom + SPrp; SPrp -> Prp + SN;</i> <i>SN -> Nom + Nom</i>
Función	"Completar el campo Actor"

Figura 4

Resulta obvio que la función "Completar el campo Actor" se corresponde en la terminología habitual del análisis sintáctico con la función Sujeto y, por otra parte, a nadie escapa que los campos establecidos anteriormente para el registro Oración tienen un fácil paralelismo con funciones tales como Núcleo del Predicado, Complemento Directo y/o Suplemento, Complementos Circunstanciales diversos, Complemento Indirecto, etc. Siempre estaremos a tiempo de usar la terminología convencional, pero, ahora, no resulta procedente teniendo en cuenta las propuestas que trato de presentar en este trabajo.

En los puntos 2 y 3 quedará completa la justificación que en 1.3 estamos enunciando.

2. Análisis sintáctico-semántico computacional: aplicaciones

¿Es posible realizar un análisis sintáctico, semántico o sintáctico-semántico por medio de un ordenador? Muchos esfuerzos investigadores están y han estado en el empeño de conseguirlo. Y verdaderamente, en el marco de ciertos diseños de gramáticas restringidas, se han alcanzado logros importantes. Pero el lenguaje natural ofrece resistencias muy persistentes a ser analizado directamente por medio de una computadora. Ello es debido, sobre todo, a su complejidad: vocabulario extensísimo, estructuras sintagmáticas y, por tanto oracionales, imprevisibles, sumo grado de recursividad de tal modo que, incluso, una estructura superior puede formar parte componencial de una inferior (una proposición, por ejemplo, está formada por sintagmas que, a su vez, podrían estar compuestos por proposiciones y sucesivamente)... El autor de este trabajo, desarrolla en la actualidad el programa SINTAGMA con el objetivo de llegar al análisis de textos y oraciones por medio del ordenador.

Las bases teóricas de las que parte el citado programa contemplan las consideraciones siguientes:

1. Las expuestas en el apartado anterior, es decir, el texto es una base de datos y las oraciones son registros compuestos por los campos enunciados.
2. El análisis de un texto debe pasar por la descomposición de cada una de sus oraciones en estructuras (simples o definidas) que cumplimenten sus campos correspondientes.
3. Importa más el delimitador de una determinada estructura (sintagma, proposición...) que la complejidad de la misma.
4. Es preciso encontrar en cada caso las estructuras que pueden dar respuesta a cada campo y definir el procedimiento para que el programa las reconozca y las asigne inequívocamente.

No es momento de exponer aquí los detalles de los diversos algoritmos que hasta ahora reúne el citado programa. Baste que conste que estamos investigando en este campo y la utilidad que reporta para el desarrollo de SINTAGMA la concepción del texto como base de datos y la de la oración como registro.

Las aplicaciones de un programa como SINTAGMA para el procesamiento del lenguaje natural son innegables:

- resúmenes de textos
- valoraciones estilísticas
- caracterizaciones sintácticas
- desarrollo de ejercicios escolares y material didáctico de manera inmediata y de acuerdo con el nivel de los alumnos
- etc.

La meta inmediata es que genere automáticamente ejercicios del tipo EERF (Enunciado, Estructura, Reescritura y Función) de la figura 5, que el usuario podrá imprimir y llevar a clase para que sean resueltos por sus alumnos o, en el caso de alumnos, resolver por medio del ordenador.

Ejercicio

Texto:

La primavera fría, la luz todavía pobre, retrasan el brote de las plantas; aceleran, sólo medianamente, la circulación de las savias y mixtifican la alegría de estos domingos de Abril.

La gente, que se hiela en los toros, que transita de prisa y abrigada por las calles y paseos, marcha, que no flanea todavía, hacia los sitios cubiertos, defendidos del frío por la calefacción o la aglomeración.

Por mi parte, yo me he refugiado en casa y abierto al azar un libro sobre mi mesa. Es *El cartero del rey*, de Tagore, el gran poeta indio, el más alto poeta contemporáneo, insuperablemente traducido al español por Zenobia Camprubí de Jiménez.

Cuestión:

Teniendo en cuenta el texto anterior, completa los siguientes apartados de acuerdo con los datos que aparecen en los mismos:

1	ENUNCIADO	La primavera fría, la luz todavía pobre
	ESTRUCTURA	
	REESCRITURA	
	FUNCION	
2	ENUNCIADO	
	ESTRUCTURA	SAdv
	REESCRITURA	
	FUNCION	“Completar el campo Modo”
3	ENUNCIADO	
	ESTRUCTURA	
	REESCRITURA	Prp + Det + Nom
	FUNCION	
4	ENUNCIADO	
	ESTRUCTURA	
	REESCRITURA	
	FUNCION	“Completar el campo Objeto”

Figura 5

Como se deduce del enunciado del este ejercicio, se trata de que, en cada caso, se complete el cuadro de acuerdo con los datos que figuran en la casilla o casillas. Se pretende con ello que el alumno demuestre que distingue entre *Enunciados*, *Estructuras*, *Reescrituras* y *Funciones* y sabe localizar en un texto (lectura discriminatoria y orientada a una finalidad concreta) la respuesta a cada uno de tales apartados.

3. Estrategias didácticas para la comprensión lectora, el comentario de textos y el análisis sintáctico-semántico

El establecimiento de campos, en el sentido que venimos considerando, tiene el objetivo de crear expectativas para la recepción de informaciones. Tales expectativas constituyen una estrategia -creemos adecuada- para potenciar en los alumnos la atención, la competencia decodificadora y el procesamiento mental intelectual de las informaciones que les llegan a través de la lectura o el oído.

Enunciamos, brevemente, las secuencias de enseñanza/aprendizaje que podrían seguirse:

Secuencia 1ª: reorganización personal de la información recibida.

1. Preparación de fichas similares a las de la figura 3, en las que consten los rotulos *Oración* y los correspondientes a los campos establecidos. Con espacios para ser completados por los alumnos.
2. Se les proporciona el texto de trabajo y tantas fichas, al menos, como oraciones tenga el mismo.
3. El ejercicio consiste en rellenar una ficha por cada oración.

Secuencia 2ª: manipulación de la información “reorganizada”.

1. Con las fichas como material de consulta:
 - elaborar informes sobre actores, hechos, lugares, modos, tiempos...
 - establecer paralelismos entre las “informaciones” aportadas por cada oración.
 - comparar oraciones (fichas) para deducir interrelaciones oracionales (observación de enlaces gramaticales, léxicos, fraseológicos, textuales...)

Secuencia 3ª: sacar conclusiones.

1. Comentario de texto: relación de la “forma de decir” con los contenidos (estructura interna del texto): a la vista de los campos con o sin datos, de los que contengan más información, de la naturaleza de ésta (nivel metafórico, redundancias, connotaciones diversas...), llegar, si ello fuera posible, a la lectura total.

Secuencia 4ª: reflexión metalingüística.

1. Análisis morfosintáctico de las oraciones y del texto: relacionando el enunciado propuesto en cada campo con la estructura, reescritura y función. Esta función ha de presentarse ahora siguiendo la terminología gramatical convencional: Sujeto y Predicado, Núcleos y Complementos.

4. Conclusión.

Ciertamente me gustaría presentar el desarrollo práctico de los enunciados propuestos. Pero la extensión limitada de este tipo de trabajos no permite, a estas alturas, otra cosa que poner la palabra fin.

BIBLIOGRAFÍA:

- GONZALO CUEVAS AGUSTÍN (1975), *Teoría de la información, codificación y lenguajes*. Inst. de Informática. Madrid.
- NOAM A. CHOMSKY (1957), *Estructuras sintácticas*. Siglo XXI. Madrid.
- (1965), *Aspectos de la teoría de la sintaxis*. Aguilar. Madrid.
- CHARLES J. FILLMORE (1968), *The case for Case. Universals in linguistic theory*. New York.

Holt.

-(1971), *Verbs of judging: An exercise in semantic description. Studies in Linguistic Semantics*. New York. Holt.

RICHARD FROST (1986), *Bases de datos y sistemas expertos. Ingeniería del conocimiento*. Ed. Díaz de Santos. Madrid.

A. GARCÍA VELASCO, *Análisis computacional de los sintagmas de un texto y de su relación con el verbo de la oración. Aplicaciones al procesamiento del lenguaje natural*. Actas del VIII Congreso de Lenguajes Naturales y Lenguajes Formales, Univ. de Barcelona. Septiembre, 1992.

WILLIAM M. GEVARTER (1985), *Máquinas inteligentes*. Ed. Díaz de Santos. Madrid.

RALPH GRISHMAN (1986), *Introducción a la lingüística computacional*. Ed. Visor, Lingüística y conocimiento. Madrid.

CARL KIRSCHNER (1981), *Semántica generativa del español*. Ed. Almar. Salamanca.

SANCHÍS LLORCA y GALÁN PASCUAL (1986), *Compiladores: teoría y construcción*. Paraninfo. Madrid.

ROBERT C. SCHANK (1984), *El ordenador inteligente*. Antoni Bosch editor. Barcelona.

HERBERT SCHILDT (1987), *Utilización de C en inteligencia artificial*. Osborne/McGraw-Hill. Madrid.

K. SPARCK JONES y Y. WILKS (1983), *Automatic natural language parsing*. Chichester, West Sussex, Ellis Horwood Ltd. y New York, Halsted Press.

NIKLAUS WIRTH (1986), *Compiladores*. Ed. Rueda. Madrid.