

# Curso

## Tratamiento de lenguaje natural

### **DURACIÓN:**

80 horas.

### **AUTOR DEL PROGRAMA:**

Dr. Alexander Gelbukh, Dr. Igor Bolshakov

### **OBJETIVO GENERAL**

Presentar a nivel detalle los conceptos principales de sintaxis. Analizar los sistemas de análisis automático sintáctico.

### **BIBLIOGRAFÍA**

1. I. Bolshakov, A. Gelbukh. Computational linguistics. CIC-IPN, 2004, 187 p.
2. J. Allen. Natural Language Understanding. The Benjamin/Cummings Publishing Company, 1995, 654 p.
3. R. Grishman. Computational Linguistics. Cambridge University Press, 1986.
4. D. Crystal. A dictionary of Linguistics and Phonetics. Blackwell Publishers, 1991.
5. The Handbook of Linguistics. Blackwell Publishers, 2001, 824 p.
6. Oxford Handbook on computational linguistics. Oxford University Press, 2003, 808 p.
7. Manning C. and H. Schütze. Foundations of statistical natural language processing. MIT Press, 1999, 680 p.
8. Jurafsky, D. and J. Martin. Speech and language processing. Prentice Hall, 2000, 934 p.

### **CONTENIDO DEL CURSO**

#### **1. Un esbozo histórico sobre sintaxis**

- 1.1. El enfoque estructuralista en el inicio del siglo XX
- 1.2. Contribución inicial de Chomsky: la jerarquía de gramáticas generativas

#### **2. Gramáticas generativas simples**

- 2.1. Una sencilla gramática libre de contexto para secuencias de nombres hispánicos
- 2.2. Gramáticas libres de contexto para oraciones del español de complejidad variada
- 2.3. Gramáticas transformacionales

#### **3. Las investigaciones lingüísticas después de Chomsky**

- 3.1. Valencias sintácticas y semánticas.
- 3.2. Subcategorización y patrones de manejo
- 3.3. Restricciones y su representación formal
- 3.4. Gramática de Estructura de Frase dirigida por el Núcleo
- 3.5. La idea de unificación, ejemplos de sintaxis y morfología
- 4. Estructura y representación sintáctica**
  - 4.1. Tipos de representación sintáctica
    - 4.1.1. Estructuras de constituyentes inmediatos
    - 4.1.2. Estructuras en árboles de dependencia
  - 4.2. Relaciones sintácticas
  - 4.3. Valencias sintácticas
  - 4.4. Tipos de construcciones sintácticas
  - 4.5. Representación sintáctica
    - 4.5.1. La “maldición” de la ambigüedad sintáctica
  - 4.6. Niveles finos de representación
  - 4.7. Estructuras profundas y de superficie
  - 4.8. El papel del procesamiento previo
  - 4.9. Resumen de los métodos para el análisis sintáctico
    - 4.9.1. Gramáticas formales
    - 4.9.2. Clasificación de las gramáticas formales
    - 4.9.3. Gramáticas libres de contexto
    - 4.9.4. Gramática
      - 4.9.4.1. Gramáticas de base restringida
      - 4.9.4.2. Gramáticas de Estructura por Frases, gobernadas por una palabra
      - 4.9.4.3. Otros lenguajes para la descripción de gramáticas
- 5. Métodos y sistemas para el análisis sintáctico**
  - 5.1. Tareas y papel del módulo sintáctico
    - 5.1.1. Representación sintáctica
  - 5.2. Arquitecturas de los analizadores sintácticos
    - 5.2.1. Estructura general de un lenguaje natural y el lugar de la sintaxis
      - 5.2.1.1. Estructuras del lenguaje: observable, profunda e inmediata
      - 5.2.1.2. Interrelaciones entre las estructuras del lenguaje
      - 5.2.1.3. Ejemplos de lenguas específicas
        - 5.2.1.3.1. Inglés
        - 5.2.1.3.2. Ruso
        - 5.2.1.3.3. Español
    - 5.2.2. Descripción sintáctica basada en el orden de las palabras
    - 5.2.3. Descripción sintáctica basada en las características
    - 5.2.4. Descripciones sintácticas combinadas
  - 5.3. Descripción sintáctica basada en el orden de las palabras
    - 5.3.1. Las gramáticas formales de Chomsky
      - 5.3.1.1. Tipos de gramáticas formales
      - 5.3.1.2. Ejemplos de algoritmos para *parsing*
  - 5.4. El papel de los métodos matemáticos en el *parsing* sintáctico y, en general, en la lingüística aplicada

- 5.4.1. ¿La lingüística aplicada es una disciplina matemática?
- 5.4.2. El papel de la eficiencia
- 5.4.3. El papel de la fiabilidad de los resultados
- 5.5. Descripciones sintácticas combinadas
  - 5.5.1. Gramáticas aumentadas
  - 5.5.2. Gramáticas restringidas
  - 5.5.3. Gramáticas de estructuras de frases generalizadas y gobernadas por una palabra
- 5.6. Descripción sintáctica basada en las características
  - 5.6.1. Reglas sintácticas
  - 5.6.2. Reglas de preferencia
  - 5.6.3. Restricciones y filtros
  - 5.6.4. Ejemplo: ETAP-2
    - 5.6.4.1. Análisis presintáctico
    - 5.6.4.2. Algoritmo de análisis sintáctico
- 5.7. Métodos de análisis de sensibilidad estadística
  - 5.7.1. El problema de la ambigüedad sintáctica
  - 5.7.2. Gramáticas formales estadísticas
  - 5.7.3. Diccionarios estadísticos
  - 5.7.4. El primero de los *parsings* que resulte el mejor
  - 5.7.5. Asignación de pesos estadísticos a los elementos de la fuente
    - 5.7.5.1. Nociones básicas de estadística matemática
      - 5.7.5.1.1. Teorema de Bayes
      - 5.7.5.1.2. Ley de Zipf
    - 5.7.5.2. El uso de los pesos estadísticos para la resolución de ambigüedades
    - 5.7.5.3. Definición de los pesos estadísticos
    - 5.7.5.4. Algoritmo para determinar los pesos estadísticos
      - 5.7.5.4.1. El caso de una sola fuente
      - 5.7.5.4.2. Combinación de fuentes independientes de estimado
- 5.8. Diccionarios utilizados en el análisis sintáctico
  - 5.8.1. Diccionarios sintagmáticos
    - 5.8.1.1. Diccionario de patrones de gobierno
      - 5.8.1.1.1. Ejemplo: Diccionario de español
      - 5.8.1.1.2. Diccionario estadístico de patrones de gobierno
    - 5.8.1.2. Diccionarios de combinaciones de palabras
      - 5.8.1.2.1. Modismos
      - 5.8.1.2.2. Combinaciones libres de palabras
      - 5.8.1.2.3. Ejemplo: CrossLexica™
  - 5.8.2. Diccionarios paradigmáticos
    - 5.8.2.1. Thesauri
    - 5.8.2.2. Diccionarios de redes semánticas
      - 5.8.2.2.1. Rutas en una red semántica
      - 5.8.2.2.2. El uso de un diccionario de redes semánticas para la resolución de ambigüedades
  - 5.8.3. Diccionarios de funciones léxicas

5.9. Síntesis sintáctica

**6. Representación sintáctica normalizada**

6.1. Tareas de normalización sintáctica

6.2. Transformaciones sintácticas

6.2.1. Gramáticas transformacionales

6.3. Operaciones sintácticas basadas en el léxico

6.3.1. Aplicación de las funciones léxicas

6.3.2. Eliminación de los nudos semánticos vacíos

6.3.3. Inserción de los nudos omitidos

6.4. Construcciones elípticas

6.5. Representación sintáctica normalizada

6.6. Diccionarios usados en la normalización sintáctica

6.7. Síntesis de la representación sintáctica normalizada

**7. Appendix. Clases prácticas y ejercicios**

7.1. Análisis y síntesis sintáctico

7.1.1. Cómo determinar la estructura sintáctica de una oración en español

7.1.2. Implementación del software de un *parser* y generador de gramática formal

7.1.3. Compilación de una gramática sintáctica del español.