

## ANÁLISIS Y SIMPLIFICACIÓN DE ORACIONES

### Práctica 2: Conocimiento y Razonamiento Automatizado

(Adaptado para la asignatura de Algorítmica y Complejidad, Curso 2024-25)

#### 1. Objetivos

- Crear una Gramática de Cláusulas Definidas que permita analizar sintácticamente una oración en inglés y crear el árbol de constituyentes.
- Simplificar oraciones complejas, convirtiéndolas en oraciones simples.

Dado que la gramática de la lengua inglesa es compleja, se trabajará con una versión reducida de la misma para analizar un conjunto limitado de frases.

#### 2. Práctica a Entregar

La práctica consistirá en la creación de una gramática que valide y represente el árbol de constituyentes, así como en la simplificación de oraciones complejas de los siguientes tipos:

- Oración Simple (o).
- Oración Coordinada (oc).
- Oración Subordinada de Relativo (or).
- Oraciones Compuestas, es decir, combinaciones de oraciones simples, coordinadas y de relativo (ocm).

Además, se deben incluir reglas que soporten los siguientes tipos de grupos sintácticos:

- Grupo Nominal (gn).
- Grupo Adjetivo (gadj).
- Grupo Adverbial (gadv).
- Grupo Preposicional (gp).
- Grupo Verbal (gv).

Esto implica la definición de los siguientes tipos de términos:

- Determinantes (det).
- Nombres (n).
- Nombres propios (np).
- Verbos (v).
- Adjetivos (adj).
- Adverbios (adv).
- Conjunciones (conj).
- Preposiciones (prep).

La simplificación de oraciones complejas (coordinadas, subordinadas de relativo o compuestas) consistirá en descomponerlas en tantas oraciones simples como sea necesario.

#### Ejemplos de Simplificación:

- "JOSÉ estudia Filosofía, pero MARÍA estudia Derecho" se debe descomponer en: "JOSÉ estudia Filosofía" y "MARÍA estudia Derecho".
- "JOSÉ come y bebe mientras MARÍA lee" se debe descomponer en: "JOSÉ come", "JOSÉ bebe" y "MARÍA lee".
- "JOSÉ, que es muy alto, tiene el pelo rubio" se debe descomponer en: "JOSÉ es muy alto" y "JOSÉ tiene el pelo rubio".

- "JOSÉ, que es muy alto, tiene el pelo rubio, pero MARÍA es morena" se debe descomponer en: "JOSÉ es muy alto", "JOSÉ tiene el pelo rubio" y "MARÍA es morena".

El código deberá entregarse en **dos ficheros**:

1. **El programa principal.**
2. **La utilidad draw.pl**, que representará gráficamente los árboles de constituyentes.

### 3. Conjunto de Oraciones

La gramática deberá analizar y simplificar correctamente las siguientes oraciones:

1. **JOSÉ es moreno y MARÍA es alta.**
2. **JOSÉ estudia Filosofía, pero MARÍA estudia Derecho.**
3. **MARÍA toma un café mientras JOSÉ recoge la mesa.**
4. **JOSÉ toma café y lee el periódico.**
5. **JOSÉ y HÉCTOR comen patatas fritas y beben cerveza.**
6. **JOSÉ come patatas fritas, pero MARÍA prefiere paella, aunque HÉCTOR toma café e IRENE lee una novela.**
7. **IRENE canta y salta mientras JOSÉ estudia.**
8. **HÉCTOR come patatas fritas y bebe zumo mientras JOSÉ canta y salta, aunque MARÍA lee una novela.**
9. **JOSÉ, que es ágil, escala en el rocódromo por las tardes.**
10. **JOSÉ, que es muy delicado, come solamente manzanas rojas.**
11. **El procesador de textos, que es una herramienta bastante potente, sirve para escribir documentos.**
12. **El procesador de textos es una herramienta muy potente que sirve para escribir documentos, pero es bastante lento.**
13. **El ratón que cazó el gato era gris.**
14. **El hombre que vimos ayer era mi vecino.**

Se debe construir un conjunto de oraciones en inglés que constituya un conjunto de prueba inicial para el sistema, que incluya al menos las correspondientes a las anteriormente indicadas en español.

Para refrescar los conceptos de análisis sintáctico (tipos de sintagmas, tipos de oración, funciones de los sintagmas, ejemplos resueltos) se recomienda como referencia inicial la página: <https://sintaxis.org>. El analizador que incluye no siempre es capaz de analizar las oraciones propuestas en este enunciado, pero puede servir para resolver los ejemplos de la propia página web.

Luego se debe construir un conjunto de al menos 30 oraciones como mínimo en un contexto que el profesor indicará para cada grupo. Los grupos con más de tres integrantes deben incrementar en 5 el número de oraciones por cada participante adicional.

Todas las traducciones debes ser traducidas con el programa <https://www.deepl.com> que se encuentra en forma gratuita en Google.

El contexto de donde vengas las oraciones debe estar debidamente referenciado en el trabajo.

#### 4. Mejoras del Programa

##### A. Posibles Mejoras

1. **Preprocesamiento de la Oración:**
  - Tokenización avanzada y normalización del texto.
2. **Flexión de Palabras:**
  - Manejo de plurales, género y tiempos verbales.
3. **Asignación de Funciones:**
  - Identificación de roles como Complemento Directo, Indirecto, etc.
4. **Traducción Básica Español-Inglés / Inglés-Español.**
5. **Optimización del Análisis Sintáctico:**
  - Algoritmos eficientes y manejo de errores.
6. **Visualización Mejorada de Árboles:**
  - Estilización e interactividad.

##### B. Metodología de Implementación de mejoras

1. Revisión teórica.
2. Propuesta de implementación.
3. Desarrollo del código en Prolog.
4. Pruebas y validación.
5. Análisis de resultados.
6. Conclusiones.
7. Documentación.

#### 5. Detalles de la Entrega

- **Equipo:** De 3 personas, con al menos dos mejoras implementadas. Los equipos con más de tres personas deben adicionar una mejora por cada integrante adicional.
- **Cada equipo debe presentar resultados parciales cada dos semanas (minutas) donde especifique tarea realizada, resultados obtenidos y próximo paso.**
- **Formato del archivo:**

El archivo .zip debe tener la siguiente estructura:

- apellido1Apellido2Apellido3.Nombre\_PLX.zip
  - |— main.pl
  - |— traductor.pl
  - |— draw.pl
  - |— Informe.pdf
- Use camelCase para el nombre del archivo

- **Informe (15-20 páginas):**
  - Portada.
  - Resumen.
  - Índice.
  - Introducción.
  - Objetivos.
  - Desarrollo: gramática, simplificación, mejoras, otros aspectos.
  - Resultados: análisis, simplificaciones, traducciones, mejoras, otros aspectos.
  - Conclusiones.
  - Bibliografía (**OJO: es importante**)
  - Anexos.

- **Entrega:**
  - Plataforma: Campus Virtual de la UAH.
  - Semana antes de PEC2
  - Importante: Las oraciones deben ser originales y no repetirse entre grupos y no pueden ser de semestres anteriores.

## 6. Aspectos de evaluación y rubrica

- Cada práctica se evalúa en base a 100 puntos y se promedia la nota.
- Luego se le aplica el porcentaje estimado.
- Cada práctica se evalúa según la rúbrica que el profesor presentará.
- Para aceptar la práctica por lo menos debe venir acompañada de dos minutas aceptadas por el profesor.
- El estudiante que no entregue un mínimo de dos minutas queda en suspenso.
- El profesor se reserva el derecho de consultar al estudiante y/o grupo cualquier aspecto sobre la implementación de la práctica durante la evaluación.
- El estudiante tiene derecho a defender su práctica.
- Para poder evaluar el informe de práctica, su implementación debe funcionar.
- La primera minuta se entrega a la siguiente práctica después de que el profesor la entregue y explique su contenido. Luego cada dos semanas se deben presentar resultados parciales que son las minutas. La última minuta puede ser la entrega de la práctica propiamente. Con la entrega de una minuta el estudiante debe anunciar el siguiente paso.

## 7. Preguntas Frecuentes: Práctica de Análisis y Simplificación de Oraciones

### 7.1. ¿Se deben analizar y simplificar las oraciones en español o en inglés?

- **Respuesta:**

Todo el análisis y la simplificación deben realizarse en inglés, tal como lo indican los objetivos de la práctica:

*"Crear una Gramática de Cláusulas Definidas que permita analizar sintácticamente una oración en Inglés."*

### 7.2. ¿Qué hacer con las oraciones en español?

Las oraciones en español del apartado "Conjunto de oraciones" son solo una referencia.

Deben traducirse al inglés para usarlas como conjunto de prueba inicial. Por ejemplo:

- Español: *"JOSÉ es moreno y MARÍA es alta."*
- Inglés: *"Joseph is dark-skinned and Mary is tall."*

La gramática, el análisis y la simplificación deben basarse únicamente en las versiones en inglés.

### 7.2. ¿Qué es una "versión reducida de la gramática inglesa"?

- **Respuesta:**

Es un subconjunto de las reglas gramaticales en inglés diseñado para esta práctica. Solo necesitas cubrir las siguientes estructuras:

  - **Oraciones simples:**

Ej.: *"John reads a book."*
  - **Oraciones coordinadas:**

Con conjunciones como *"and," "but," "or."*

Ej.: *"John reads a book and Mary drinks tea."*

- **Oraciones subordinadas de relativo:**  
Con pronombres como "who," "which," "that."  
Ej.: *"The man who wrote the book is famous."*
- **Oraciones compuestas:**  
Combinaciones de las anteriores.  
Ej.: *"John reads a book, but Mary drinks tea while Sarah writes."*

### 7.3. ¿Cómo se simplifican las oraciones complejas?

- **Respuesta:**  
La simplificación consiste en dividir oraciones complejas en varias oraciones simples.  
Algunos ejemplos:
  - **Oración coordinada:**  
Entrada: *"John reads a book and Mary drinks tea."*  
Salida:
    - *"John reads a book."*
    - *"Mary drinks tea."*
  - **Oración subordinada de relativo:**  
Entrada: *"The man who wrote the book is famous."*  
Salida:
    - *"The man is famous."*
    - *"The man wrote the book."*
  - **Oración compuesta:**  
Entrada: *"John reads a book, but Mary drinks tea while Sarah writes."*  
Salida:
    - *"John reads a book."*
    - *"Mary drinks tea."*
    - *"Sarah writes."*

### 7.4. ¿Qué elementos debe reconocer mi gramática?

Tu gramática debe manejar correctamente:

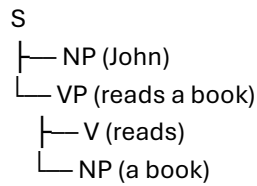
- **Grupos sintácticos:**
  - Grupo Nominal (GN): *"The man."*
  - Grupo Verbal (GV): *"Reads a book."*
  - Grupo Adjetival (GAdj): *"Very tall."*
  - Grupo Adverbial (GAdv): *"Quickly."*
  - Grupo Preposicional (GP): *"In the house."*
- **Términos básicos:**
  - Determinantes (*"The," "A"*)
  - Nombres comunes (*"Book," "Man"*) y propios (*"John," "Sarah"*)
  - Verbos (*"Reads," "Drinks"*)
  - Adjetivos (*"Tall," "Famous"*)
  - Adverbios (*"Quickly," "Silently"*)
  - Conjunciones (*"And," "But"*)
  - Preposiciones (*"In," "On"*)

### 7.5. ¿Cómo se representan los árboles de constituyentes?

- **Respuesta:**  
Un árbol de constituyentes refleja la estructura jerárquica de la oración. Por ejemplo:

Entrada: "John reads a book."

Árbol generado:



Utiliza la herramienta draw.pl para generar automáticamente los árboles.

## 7.6. ¿Cómo construir un conjunto de prueba inicial?

- **Respuesta:**

Traduce las oraciones del apartado "Conjunto de oraciones" al inglés. Ejemplos:

- Español: "JOSÉ estudia Filosofía, pero MARÍA estudia Derecho."  
inglés: "Joseph studies Philosophy, but Mary studies Law."

Estas traducciones se usarán para:

- Probar el análisis sintáctico.
- Verificar la representación de árboles de constituyentes.
- Evaluar la simplificación en oraciones simples.
- Implementar mejoras

## 7.7. ¿Qué mejoras puedo implementar en el sistema?

- **Respuesta:**

Algunas posibles mejoras incluyen:

1. **Preprocesamiento:**  
Dividir automáticamente las oraciones en palabras o tokens antes de analizarlas.
2. **Reconocimiento de flexiones gramaticales:**  
Identificar formas plurales y conjugaciones verbales.
3. **Traducción básica:**  
Crear un traductor básico de español a inglés como una funcionalidad adicional opcional.

## Resumen rápido

- **El análisis y simplificación:** Siempre en inglés.
- **Gramática:** Reducida, con reglas para oraciones simples, coordinadas, subordinadas y compuestas.
- **Oraciones en español:** Úsalas solo como referencia, tradúcelas al inglés para las pruebas.
- **Sistema:** Debe analizar, simplificar y generar árboles de constituyentes de las oraciones.

Si tienes más dudas, consulta nuevamente este manual. ¡Éxito en tu práctica!

Versión Revisada\_19Marzo2023