

# Análisis y simplificación de oraciones

Fco. Javier Bueno, Manuel de Buenaga

2021-2022

## 1. Objetivos

El objetivo de la práctica es doble:

- Por un lado, crear una Gramática de Cláusulas Definidas que permita analizar sintácticamente una oración en español y crear el árbol de constituyentes.
- Asegurarse de que las oraciones son gramaticalmente correctas en cuanto a **estructura** y **concordancia** entre los componentes de las mismas.

Dada la complejidad de la gramática de la lengua española se trabajará con una versión reducida de la misma que permita analizar un conjunto limitado de frases.

## 2. Concordancia

Uno de los problemas del análisis de oraciones es el de la agramaticalidad, es decir, que las secuencias de palabras o morfemas no se ajusten a las reglas de la gramática. Uno de los aspectos de dicha agramaticalidad es el de la **concordancia** de **género**, **número** y **persona**.

Según la RAE existen dos tipos de concordancia (en **negrita** se muestra la falta de concordancia):

### 1. Concordancia nominal (coincidencia de género y número).

- a) Es la que establece el sustantivo con el artículo o los adjetivos que lo acompañan:
  - *La blanca paloma* - *La blanca palomas*.
  - *Esos libros viejos* - ***Ese*** libros viejos.
- b) El pronombre con su antecedente o su consecuente:

- *A tus hijas las vi ayer* - *A tus hijas **los** vi ayer.*
  - *Les di tu teléfono a los chicos* - ***Le** di tu teléfono a los chicos.*
- c) El sujeto con el atributo, con el predicativo o con el participio del verbo de la pasiva perifrástica:
- *Juan es médico* - *Juan es **médicos**.*
  - *Ella estaba cansada* - *Ella estaba **cansado**.*
  - *Esas casas fueron construidas a principios de siglo* - *Esas casas **fue construida** a principios de siglo*

2. Concordancia verbal (coincidencia de número y persona). Es la que se establece entre el verbo y su sujeto:

- *Esos cantan muy bien* - ***Ese** cantan muy bien.*
- *La pelota la tiró el niño* - *La pelota la tiró **los niños**.*

A los ejemplos anteriores hay que añadir que, en ocasiones, los sustantivos pueden ser múltiples como en las siguientes oraciones:

- *Juan y María comen paella* - *Juan y María **come** paella.*
- *El canario de Juan y María canta bien* - *El canario de Juan y María **cantan** bien.*
- *Compré un pantalón y una corbata negros* - *Compré un pantalón y una corbata **negra**.*

### 3. Práctica a entregar

Una vez que se han realizado con éxito los ejercicios del documento de introducción a las gramáticas DCG, se puede abordar la creación de una **gramática** más compleja que **valide** y **represente el árbol de constituyentes** 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/o de relativo (ocm).

La validación de las oraciones implica **aplicar mecanismos de concordancia** allí donde sea necesario para respetar la gramaticalidad de la oración bajo estudio. De ese modo, el programa debe indicar que la estructura: *El*

*hombre come una manzana* es **correcta**, mientras que *El hombre come unas manzana* **no lo es**.

En cuanto a los grupos sintácticos, se deben incluir reglas que soporten los siguientes tipos:

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

lo que implica la definición de los siguientes tipos de términos:

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

## 4. Conjunto de oraciones

La gramática debe analizar y validar la concordancia de las siguientes oraciones tipo:

1. *El hombre grande come la manzana roja.*
2. *El hombre con un tenedor grande come la manzana roja.*
3. *Juan y María comen la manzana roja con un tenedor y un cuchillo.*
4. *Tú hablas muy rápidamente.*
5. *Ella hace la práctica de Juan.*

6. *El canario de Juan y María canta.*
7. *La blanca paloma alzó el vuelo.*
8. *Está muy lejos de Madrid.*
9. *Soy lento de reflejos.*
10. *Mañana iremos a la feria del cómic.*
11. *La esperanza de vida de un niño depende de su lugar de nacimiento.*
12. *El señor que vimos en la Universidad era mi profesor.*
13. *Juan, que es muy delicado, come solamente manzanas rojas.*
14. *Juan es moreno y María es alta.*
15. *Juan recoge la mesa mientras María toma un café.*
16. *Compré un pantalón y una corbata negros.*
17. *Juan y Héctor comen patatas fritas y beben cerveza.*
18. *Bebimos y bailamos toda la noche mientras tú estudiabas.*
19. *Esos niños cantan y saltan y sonrían alegres.*
20. *Creo que necesitáis la ayuda de un profesional.*
21. *El procesador de textos, que es una herramienta muy potente, sirve para escribir documentos.*
22. *El procesador de textos es una herramienta muy potente que sirve para escribir documentos pero es bastante lento.*

## 5. Mejoras del programa

Se valorará positivamente cualquier mejora de la gramática sobre lo expuesto anteriormente. Algunas ideas son las siguientes:

- Introducir algún preprocesamiento de la oración que agilice el proceso de análisis de una estructura sintáctica compleja.
- Implementar algún mecanismo para flexionar las palabras del diccionario, añadiendo el plural a las palabras en singular o el femenino para las palabras de género masculino (o viceversa).

- Adaptar el sistema desarrollado para utilizar el archivo *wn-data-spa.pl* con 57.766 hechos en Prolog con el léxico en español de la base de datos WordNet, de forma que se dé cobertura a un mayor espectro de frases para analizar. Especificar de forma estructurada las adaptaciones hechas y el nuevo conjunto de expresiones del idioma que es capaz de analizar el sistema.
- Cualquier otra que cuente con el visto bueno del profesor.

Para refrescar los conceptos de análisis sintáctico (tipos de sintagmas, tipos de oración, funciones de los sintagmas, ejemplos resueltos,...) se recomienda consultar 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.

## 6. Detalles de la entrega

La práctica se podrá realizar en equipos de hasta 3 personas (los mismos que en la práctica anterior).

La práctica se entregará en tres ficheros: la **gramática**, el **diccionario** y la utilidad **draw.pl** que deberá ser consultada por el primero para representar gráficamente los árboles de constituyentes de las oraciones analizadas y simplificadas.

Además del **código fuente del programa**, convenientemente estructurado y comentado, se deberá incluir una **breve memoria (PDF)** indicando los miembros del equipo, reparto de tareas entre los mismos, conjunto de frases analizadas/simplificadas con éxito, mejora realizada, errores y/o aspectos no implementados, así como las fuentes consultadas para la resolución de la práctica.

La entrega se realizará **a través de la plataforma** en un **único fichero .zip** antes de las **23:59 horas del día 25 de abril de 2022**. El nombre del fichero será NombreApellido1Apellido2.zip de uno de los integrantes del grupo. Una vez recibidos los trabajos y establecido el número de equipos, se publicará una fecha con la hora de defensa para cada equipo.

La entrega de prácticas **copiadas, total o parcialmente**, supondrá el **suspense** del laboratorio para todos los alumnos implicados.