

Lee el capítulo de Jurafsky & Martin (2008) Morphology and Finite-State Transducers

Con base en la presentación y la lectura del capítulo tal contesta el siguiente cuestionario:

1. ¿Qué es morfología y cuáles son la bases del análisis morfológico (morfología derivativa, flexiva y no concatenativa? Da ejemplos.

La morfología es el estudio de la forma en que se construyen las palabras a partir de unidades portadoras de significado o morfemas.

No concatenativa: Puede manipular el significado de las palabras sin necesidad de ampliarlas de manera escrita, por ejemplo *mütter* (madre en alemán), cambia su plural mediante diferencia en su pronunciación.

Concatenativa: Lo opuesto de la no-concatenativa.

Flexiva: Permite mantener el significado principal de las palabras variando por ejemplo su género o número: cantar, cantas, cantábamos.

Derivativa: Genera nuevos significados en las palabras que se crean. Por ejemplo: social, antisocial.

2. ¿Qué es un morfema y qué tipos de morfemas existen?

Es la unidad mínima de un lenguaje que contiene significado. Existen dos tipos de morfemas: los de raíz, que son los que contienen el mayor contenido de significado, mientras que, los adicionales o añadidos, simplemente diferencian a ciertas palabras con la misma raíz.

3. ¿Qué es stemming, lematización y tokenización? Da ejemplos.

Stemming: Reducción de palabras a su raíz.

Lematización: Reducción de palabras a su palabra raíz.

Tokenización: Ocultamiento de las palabras mediante codificación.

4. ¿Cuál es el algoritmo de stemming más usado y cómo funciona?

El Porter (1980), basado en una serie de reglas simples y editables en cascada.

5. ¿Qué es concordancia (agreement) y qué tipos de concordancia hay en español? Da ejemplos.

Es la relación de sentido común establecida entre adjetivos, verbos y determinantes. Por ejemplo, para nosotros como hispano-parlantes no tendría sentido escribir: Ayer jugarás muy bien.

6. ¿Qué rasgos (features) morfológicos se pueden agregar a verbos y sustantivos respectivamente en español?

A los adjetivos: género, número. A los verbos: desinencia (terminación según el tiempo de conjugación).

7. Siguiendo el ejemplo del autor, crea un autómata de estados finitos que produzca las siguientes palabras: computar, agradable, agradar, computacional, autoridad, autoritario, estacional, igualitario, computable. (Ignora cualquier cambio fonológico que resulte de las derivaciones.)



8. ¿Qué es el reconocimiento morfológico y qué necesita para llevarse a cabo?

Determinar si una palabra forma parte legítimamente de un lenguaje determinado. Se consigue adaptando un autómata finito que permita introducir palabras de un determinado lenguaje y este determine si estas son parte de el lenguaje o no.

9. ¿Cuál es la importancia de la detección y corrección de errores ortográficos y cuáles son los principales problemas en este ámbito? Da ejemplos en español.

Detectar errores ortográficos permite corregirlos antes de que un grupo de palabras pasen por un procedimiento de reconocimiento morfológico, evitando así cometer errores de no identificación a palabras que si corresponden a un determinado lenguaje.

çrus"podría no catalogarse como parte del español, siendo que la palabra çruz", si este era su verdadero significado, si forma parte de este.

"Mexico"podría ser identificado como una palabra en inglés y no en español.

10. ¿Qué es la distancia de edición mínima y cómo se mide?

La distancia de edición nos dice el coste mínimo de transformar una cadena en otra por medio de la inserción, eliminación o reemplazo de caracteres.

Se toman dos cadenas, se divide el número de caracteres no modificados de ambas cadenas entre la longitud de la cadena más larga.