

## Etiquetado gramatical

**ITALICA**  
**Universidad de Sevilla**  
José A. Troyano

## Índice

- **Introducción**
- Etiquetado con el modelo de Markov
- Etiquetado con el modelo de Markov oculto
- Etiquetado transformacional
- Precisión y aplicaciones

## Introducción

### Definición de etiquetado

Etiquetado gramatical (*Part-of-Speech tagging*):

asociar a cada palabra una categoría gramatical.

Se necesita:

- Conjunto de etiquetas (elegido cuidadosamente)
- Recurso (en las etapas de entrenamiento)
- Algoritmo de etiquetado

## Introducción

### Algunas etiquetas (conjuntos Brown/Penn)

Etiqueta	Categoría	Etiqueta	Categoría
AT	artículo	RB	adverbio
BEZ	la palabra <i>is</i>	RBR	adverbio comparativo
IN	preposición	TO	la palabra <i>to</i>
JJ	adjetivo	VB	verbo, forma base
JJR	adjetivo comparativo	VBD	verbo, pasado
MD	modal	VBG	verbo, presente-gerundio
NN	nombre común-singular	VBN	verbo, participio pasado
NNP	nombre propio	VBP	verbo, no 3ª singular
NNS	nombre plural	VBZ	verbo, 3ª singular
PERIOD	.:?!)	WDT	partícula <i>wh-</i>
PN	pronombre personal		

Introducción

## Ambigüedad en el etiquetado

The representative put chairs on the table

The representative put chairs on the table

AT NN VBD NNS IN AT NN

(el representante puso sillas sobre la mesa)

The representative put chairs on the table

AT JJ NN VBZ IN AT NN

(la opción de compra representativa preside sobre la mesa)

Introducción

## Información disponible: sintagmática

- Secuencias de etiquetas muy comunes

AT JJ NN

- Secuencias de etiquetas poco comunes o imposibles

AT JJ VBP

A new play

AT JJ ?

Precisión: 80%

Razón: En inglés hay muchas palabras que pueden tener más de una etiqueta gramatical (casi todos los nombres pueden actuar como verbos p.e. *to web our annual report*)

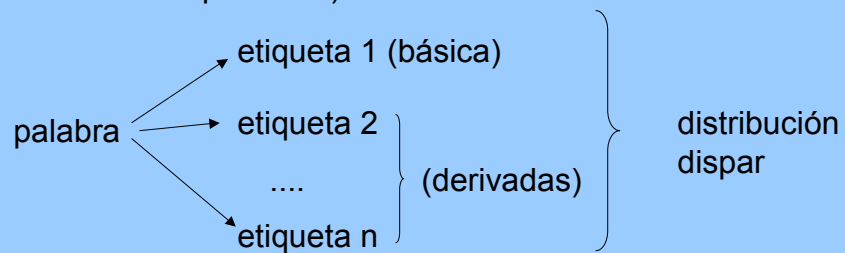
## Introducción

### Información disponible: léxica

Aprovechar la información que nos da la palabra que queremos etiquetar.

*flour* se utiliza más como nombre que como verbo

Un etiquetador ciego que asigne la categoría más común de cada palabra puede alcanzar el 90% (se puede tomar como límite inferior de precisión).



## Índice

- Introducción
- **Etiquetado con el modelo de Markov**
- Etiquetado con el modelo de Markov oculto
- Etiquetado transformacional
- Precisión y aplicaciones

Etiquetado con el modelo de Markov

## El modelo probabilístico

Horizonte limitado:

$$P(X_{i+1}=t_j|X_1,\dots,X_i) = P(X_{i+1}=t_j|X_i)$$

Invarianza en el tiempo:

$$P(X_{i+1}=t_j|X_i) = P(X_2=t_j|X_1)$$

Notación:

**Subíndices:** Palabras y etiquetas en una posición particular de la frase.

**Superíndices:** Tipos de palabras y etiquetas.

$$P(t_{i+1}^j|t_{i,1})=P(t_{i+1}^j|t_i) \quad \text{y} \quad P(t_{i+1}^j|t_{i,1})=P(t_2^j|t_1)$$

Etiquetado con el modelo de Markov

## Notación

<u>Expresión</u>	<u>Significado</u>
$w_i$	Palabra en la posición $i$
$t_i$	Etiqueta de $w_i$
$w_{i,i+m}$	Palabras de la posición $i$ a la $i+m$
$t_{i,i+m}$	Etiquetas de $w_{i,i+m}$
$w^j$	Palabra $j$ -ésima del lexicon
$t^j$	Etiqueta $j$ -ésima en el conjunto de etiquetas
$C(w^j)$	Número de ocurrencias de $w^j$ en el corpus de entrenamiento
$C(t^j)$	Número de ocurrencias de $t^j$ en el corpus de entrenamiento
$C(t^j,t^k)$	Número de ocurrencias de $t^j$ seguidas de $t^k$
$C(w^j:t^j)$	Número de ocurrencias de $w^j$ etiquetadas como $t^j$
$T$	Cardinalidad del conjunto de etiquetas
$W$	Cardinalidad del lexicon
$n$	Longitud de la frase

Etiquetado con el modelo de Markov

## Modelo visible o modelo oculto

El método que vamos a presentar para etiquetar es un modelo mixto:

- Modelo visible en el entrenamiento: la salida (etiquetas) se conoce perfectamente si estamos utilizando un corpus etiquetado.
- Modelo oculto en el etiquetado: en este proceso utilizamos precisamente las probabilidades estimadas en el entrenamiento.

Etiquetado con el modelo de Markov

## Parámetros del modelo de Markov

Parámetros del modelo de Markov oculto:

- Conjunto de estados (etiquetas)
- Alfabeto de salida (palabras)
- Probabilidades de estado inicial
- Probabilidades de transición ( $P(t^k|t^j)$ )
- Probabilidades de emisión de símbolos ( $P(w^i|t^j)$ )

Etiquetado con el modelo de Markov

### Probabilidad condicionada de etiquetas y palabras

Probabilidad de que  $t^k$  siga a  $t^j$

$$P(t^k|t^j) = \frac{C(t^j, t^k)}{C(t^j)}$$

Probabilidad de emisión de  $w^l$  dada  $t^j$

$$P(w^l|t^j) = \frac{C(w^l:t^j)}{C(t^j)}$$

Aplicación simple:

$$P(\text{NN}|\text{JJ}) = 0.45$$

$$P(\text{VBP}|\text{JJ}) = 0.0005$$

luego en la frase *new play* si se conoce que *new* es un adjetivo (JJ) es poco arriesgado concluir que *play* es un nombre (NN) en lugar de un verbo (VBP)

Etiquetado con el modelo de Markov

### Probabilidad de un etiquetado

Probabilidad de un etiquetado  $t_{1,n}$  dada una frase  $w_{1,n}$ :

$$P(t_{1,n}|w_{1,n}) = \frac{P(w_{1,n}|t_{1,n}) P(t_{1,n})}{P(w_{1,n})}$$

El mejor etiquetado:

$$\arg_{t_{1,n}} \max P(w_{1,n}|t_{1,n}) P(t_{1,n})$$

Etiquetado con el modelo de Markov  
**Probabilidad de un etiquetado (II)**

La expresión  $P(w_{1,n}|t_{1,n}) P(t_{1,n})$  se puede reducir aplicando las siguientes simplificaciones:

- La asunción de horizonte limitado del modelo de Markov

$$P(t_{1,n}) = P(t_n|t_{1,n-1})P(t_{n-1}|t_{1,n-2})\dots P(t_2|t_1) = P(t_n|t_{n-1})P(t_{n-1}|t_{n-2})\dots P(t_2|t_1)$$

- Independencia entre las palabras de la frase:

$$P(w_{1,n}|t_{1,n}) = \prod_{i=1,n} P(w_i|t_{1,n})$$

- La emisión de una palabra sólo depende de su etiqueta:

$$P(w_{1,n}|t_{1,n}) = \prod_{i=1,n} P(w_i|t_i)$$

Con todo ello, la expresión para determinar el mejor etiquetado es:

$$\arg_{t_{1,n}} \max \prod_{i=1,n} P(w_i|t_i) P(t_i|t_{i-1})$$

Etiquetado con el modelo de Markov  
**Algoritmo de entrenamiento**

**Para** todas las etiquetas  $t^i$  **hacer**

**Para** todas las etiquetas  $t^k$  **hacer**

$$P(t^k|t^i) = C(t^i, t^k) / C(t^i)$$

**Fin para**

**Fin para**

**Para** todas las etiquetas  $t^i$  **hacer**

**Para** todas las palabras  $w^l$  **hacer**

$$P(w^l|t^i) = C(w^l; t^i) / C(t^i)$$

**Fin para**

**Fin para**



Etiquetado con el modelo de Markov  
**Algoritmo de Viterbi: etiquetado (I)**

Evaluar la fórmula  $\prod_{i=1,n} P(w_i|t_i) P(t_i|t_{i-1})$  para todos los etiquetados posibles es un problema de orden exponencial sobre la longitud de la frase.

El algoritmo de Viterbi es una solución más eficiente. Se basa en la definición de dos funciones:

$\delta_i(j)$  = probabilidad de que la palabra  $w_i$  se etiquete con  $t_j$

$\psi_{i+1}(j)$  = etiqueta más probable para  $w_{i+1}$  dado que  $t_j$  es la etiqueta de  $w_{i+1}$

Etiquetado con el modelo de Markov  
**Algoritmo de Viterbi: etiquetado (II)**

**Comentario:** Inicialización

$\delta_1(\text{PUNTO}) = 1.0$

$\delta_1(t) = 0.0$  para  $t \neq \text{PUNTO}$

**Comentario:** Inducción

**Para**  $i=1..n$  **hacer**

**Para** todas las etiquetas  $t_j$  **hacer**

$\delta_i(t_j) = \max_{1 \leq k \leq T} [\delta_{i-1}(t_k) P(w_i|t_j) P(t_j|t_k)]$

$\psi_{i+1}(t_j) = \arg \max_{1 \leq k \leq T} [\delta_{i-1}(t_k) P(w_i|t_j) P(t_j|t_k)]$

**Fin para**

**Fin para**

**Comentario:** Etiquetado

$X_{n+1} = \arg \max_{1 \leq j \leq T} \delta_{n+1}(j)$

**Para**  $j=n..1$  **hacer**

$X_{n+1} = \psi_{j+1}(X_{j+1})$

**Fin para**

$P(X_1, \dots, X_n) = \max_{1 \leq j \leq T} \delta_{n+1}(j)$

Etiquetado con el modelo de Markov

## Palabras desconocidas

Son uno de los mayores problemas para los etiquetadores, las posibles soluciones son:

- Asumir que pertenecen a cualquier categoría sintáctica (con unas probabilidades tomadas de un diccionario).
- Utilizar pistas morfológicas para afinar dichas probabilidades.
- Utilizar otros indicadores como mayúsculas o guiones.

Etiquetado con el modelo de Markov

## Etiquetadores basados en trigramas

En el modelo presentado sólo se utiliza la información que da la etiqueta inmediatamente anterior (bigramas).

Una forma de ampliar el modelo es considerar más etiquetas. El etiquetado basado en trigramas utiliza dos etiquetas previas para etiquetar una tercera palabra.

En el siguiente ejemplo, *marked* es etiquetada de forma distinta en función de una palabra dos posiciones más atrás:

<u>is</u>	<u>clearly</u>	<u>marked</u>	<u>he</u>	<u>clearly</u>	<u>marked</u>
BEZ	RB	VBN	PN	RB	VBD

Etiquetado con el modelo de Markov

## Uso de memoria variable

Algunas veces el uso de más información (trigramas, tetragramas) puede despistar más que ayudar. Las comas, por ejemplo, suelen invalidar la información de un trigramas.

Posibles soluciones:

- Ponderar las probabilidades:

$$P(t_i|t_{i-1}) = \lambda_1 P_1(t_i) + \lambda_2 P_2(t_i|t_{i-1}) + \lambda_3 P_3(t_i|t_{i-1}, t_{i-2})$$

- Identificar manualmente los casos en los que el bigrama no es necesario y aplicar un trigramas (solución pseudo-gramatical).
- Utilizar un modelo variable que pueda ser entrenado en base a algún criterio.

## Índice

- Introducción
- Etiquetado con el modelo de Markov
- **Etiquetado con el modelo de Markov oculto**
- Etiquetado transformacional
- Precisión y aplicaciones

Etiquetado con el modelo de Markov oculto  
**Entrenamiento no supervisado (I)**

Si no tenemos datos de entrenamiento (corpus ya etiquetado), el modelo oculto de Markov puede servir para aprender las regularidades de las secuencias de etiquetas.

Una posibilidad es inicializar de forma aleatoria los parámetros del modelo de Markov.

Sin embargo podemos aprovechar la información de un diccionario para estimar dichos parámetros. El que más nos interesa estimar es la probabilidad de emisión de símbolos:

$$b_{jl}$$

En este caso es la probabilidad de emitir la palabra  $w^l$  dada la etiqueta  $t^j$

Etiquetado con el modelo de Markov oculto  
**Método de Jelinek**

Alfabeto de salida = palabras

$$b_{jl} = \frac{b_{jl}^* C(w^l)}{\sum_{w^m} b_{jm}^* C(w^m)}$$

$$b_{jl}^* = \begin{cases} 0 & \text{si } t^j \text{ no es una etiqueta válida para } w^l \\ 1/T(w^l) & \text{en otro caso} \end{cases}$$

Donde  $T(w^l)$  es el número de posibles etiquetas para  $w^l$   
(se consideran todas equiprobables)

Etiquetado con el modelo de Markov oculto

## Método de Kupiec

Alfabeto de salida = metapalabras (grupos de etiquetas)

$T$  = conjunto de posibles etiquetas

$L \in \text{Partes}(T)$

$u^L = \{w \mid t_i \in L \text{ y } t_i \text{ es una etiqueta válida para } w\}$

$$b_{jL} = \frac{b_{jL}^* C(u^L)}{\sum_{u^M} b_{jM}^* C(u^M)} \quad b_{jL}^* = \begin{cases} 0 & \text{si } j \notin L \\ 1/|L| & \text{en otro caso} \end{cases}$$

Etiquetado con el modelo de Markov oculto

## Algoritmos de entrenamiento y etiquetado

Entrenamiento: Algoritmo Forward-Backward

Etiquetado: Una vez que finalizamos el entrenamiento, disponemos de todos los elementos de un modelo de Markov oculto, así que podemos aplicar el algoritmo de Viterbi.

## Índice

- Introducción
- Etiquetado con el modelo de Markov
- Etiquetado con el modelo de Markov oculto
- **Etiquetado transformacional**
- Precisión y aplicaciones

Etiquetado transformacional

### Conceptos básicos

Las asunciones de Markov son demasiado fuertes para muchas de las propiedades del lenguaje natural.

Las flexibilizaciones de este modelo (como la interpolación entre unigramas, bigramas, ...) incrementan el número de parámetros haciéndolo más complejo.

El etiquetado transformacional aporta:

- Mayor información sintáctica
- Mayor flexibilidad al contemplar reglas de refinamiento del etiquetado.

Etiquetado transformacional  
**Conceptos básicos**

Para aplicar este etiquetado se necesita:

- Diccionario
- Corpus etiquetado
- Reglas de transformación

El proceso básico es:

- Estimar las etiquetas en base al diccionario
- Modificar el etiquetado en base a las reglas de transformación

Etiquetado transformacional  
**Reglas de transformación**

Son reglas de reescritura, por ejemplo:

<u>Etiqueta original</u>	<u>Nueva etiqueta</u>	<u>Condición de disparo</u>
NN	VB	la etiqueta anterior es TO
VBP	VB	una de las tres anteriores etiquetas es MD
JJR	RBR	la siguiente etiqueta es JJ
VBP	VB	una de las dos anteriores palabras no es <i>n't</i>

Tipos de reglas:

- Basadas en etiquetas
- Basadas en palabras
- Basadas en morfología

Etiquetado transformacional  
**Algoritmo de aprendizaje**

**Objeto:** Seleccionar las mejores reglas de transformación y determinar el orden de aplicación.

$C_0$  = corpus etiquetado según la frecuencia de un diccionario

**Para**  $k=1..n$  **hacer**

$v$  = la transformación  $u_i$  que minimiza  $E(u_i(C_k))$

**Si**  $(E(C_k) - E(v(C_k))) < \epsilon$  **entonces** salir **Fin si**

$C_{k+1} = v(C_k)$

$\tau_{k+1} = v$

**Fin para**

**Resultado:** Secuencia de transformaciones  $\tau_1, \dots, \tau_{k+1}$

Puede ser una alternativa de aprendizaje no supervisado al planteado con el modelo de Markov oculto.

Etiquetado transformacional  
**Conclusiones y relación con otros modelos**

- Es un método estadístico en el entrenamiento y totalmente simbólico en el etiquetado
- Es más flexible que el modelo de Markov. Permite incluir dependencias cercanas, lejanas, hacia delante y hacia atrás
- La especificación de las reglas de transformación es intuitiva.

Otros modelos:

Árboles de decisión: El etiquetado de un nodo influye en el etiquetado de los nodos a los que domina.

Autómatas finitos (transductores): Las reglas de transformación se implementan a través de autómatas (Roche, Schabes95).



## Índice

- Introducción
- Etiquetado con el modelo de Markov
- Etiquetado con el modelo de Markov oculto
- Etiquetado transformacional
- **Precisión y aplicaciones**

## Precisión y aplicaciones

### Otros modelos y otros lenguajes

El etiquetado es uno de los problemas más activos en los últimos años. Otras propuestas son:

- Redes neuronales
- Árboles de decisión
- Vecinos más cercanos
- Máxima entropía

La mayoría de los estudios son sobre el inglés:

- el orden de las palabras es bastante rígido lo que ayuda al etiquetado
- se suelen cambiar las categorías de las palabras sin inflexiones, esto dificulta

En general, se puede concluir que es un problema de dificultad similar para la mayor parte de los lenguajes

Precisión y aplicaciones

## Factores que influyen en la precisión

Precisión más habitual: 95%-97%

Los factores que influyen en ella son:

- Cantidad de texto de entrenamiento disponible
- El conjunto de etiquetas
- Diferencia (de ámbito) entre el corpus de entrenamiento (o el diccionario) y el de aplicación
- Tratamiento de palabras desconocidas

Precisión y aplicaciones

## Aplicaciones del etiquetado

### **Análisis sintáctico superficial:**

- Búsqueda de sintagmas a partir de bigramas
- Expresiones regulares sobre etiquetas
- Como paso previo al análisis completo (XTAG, chunk)

### **Extracción de información (Cardie 97)**

- Las etiquetas son categorías semánticas

### **Generación de respuestas**

Conclusión negativa: Los analizadores lexicalizados probabilísticos trabajan mejor con entradas sin etiquetar.