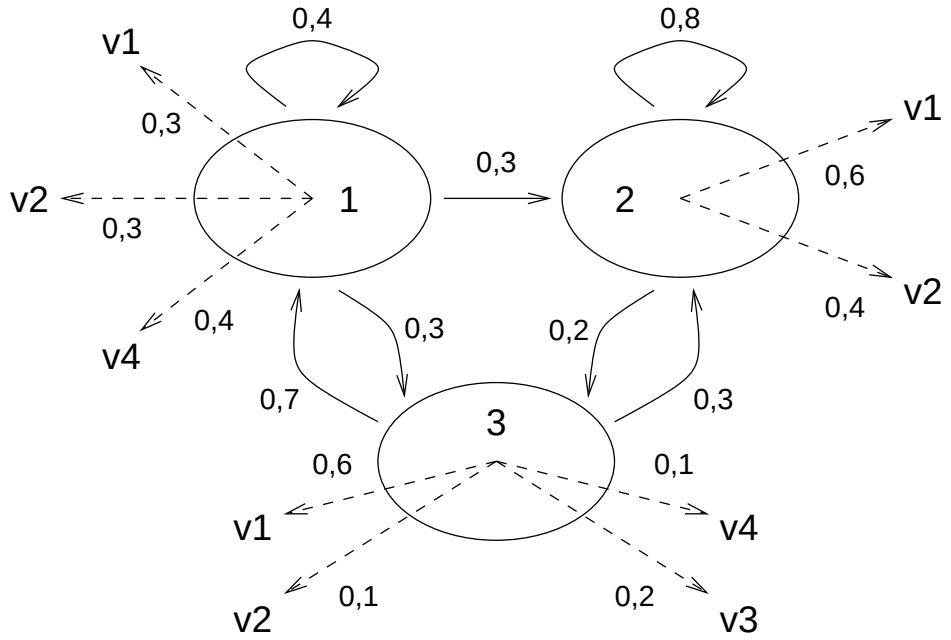
Modelos de Markov ocultos



- Propiedad del horizonte limitado

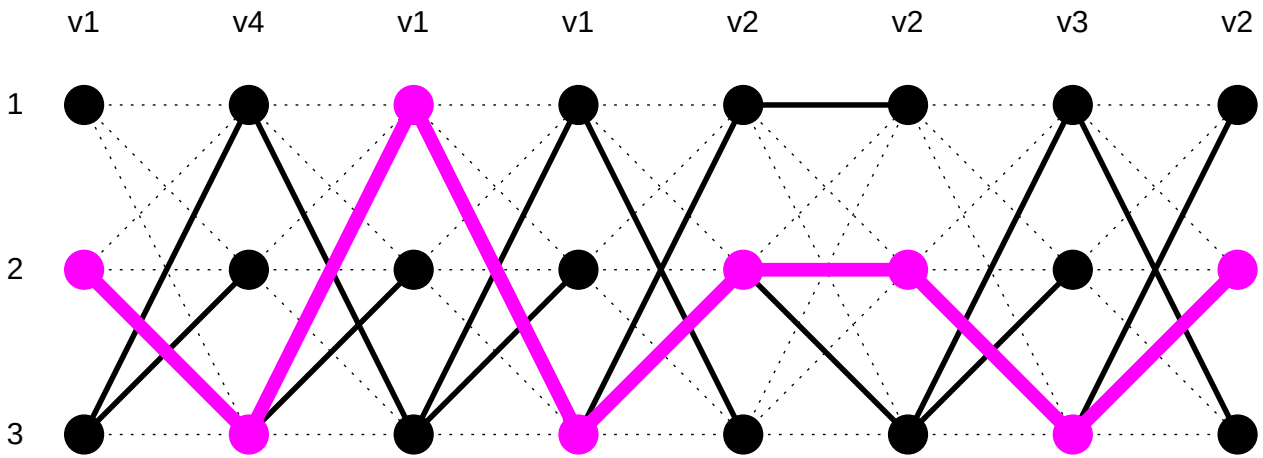
$$P(q_t = j | q_{t-1} = i, q_{t-2} = k, \ldots) = P(q_t = j | q_{t-1} = i)$$

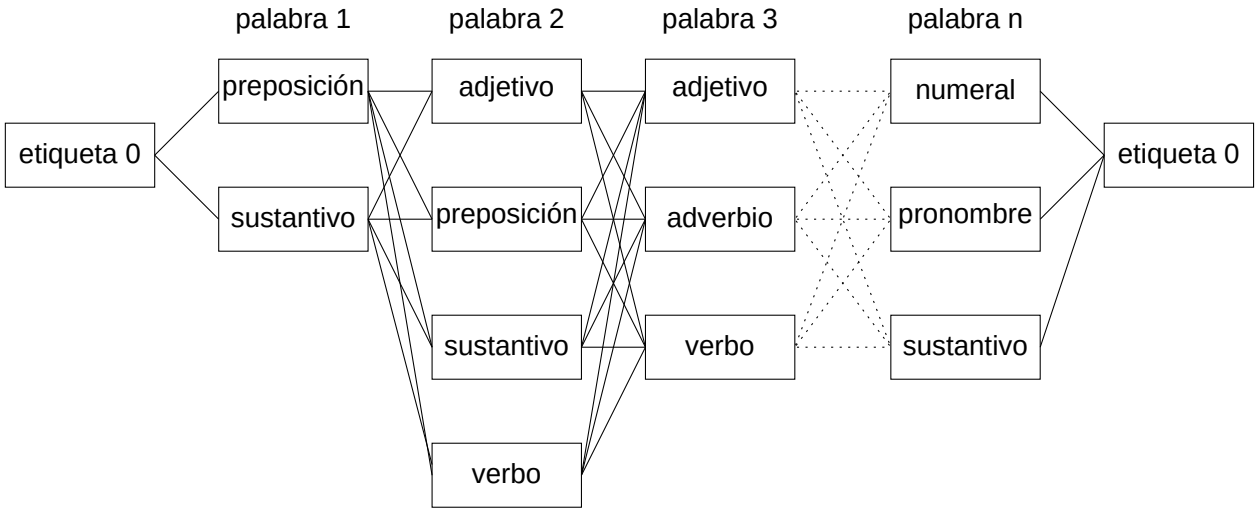
- Propiedad del tiempo estacionario

$$P(q_t = j | q_{t-1} = i) = P(j|i), \qquad \forall t$$

- Procesos de Markov doblemente estocásticos

Algoritmo de Viterbi



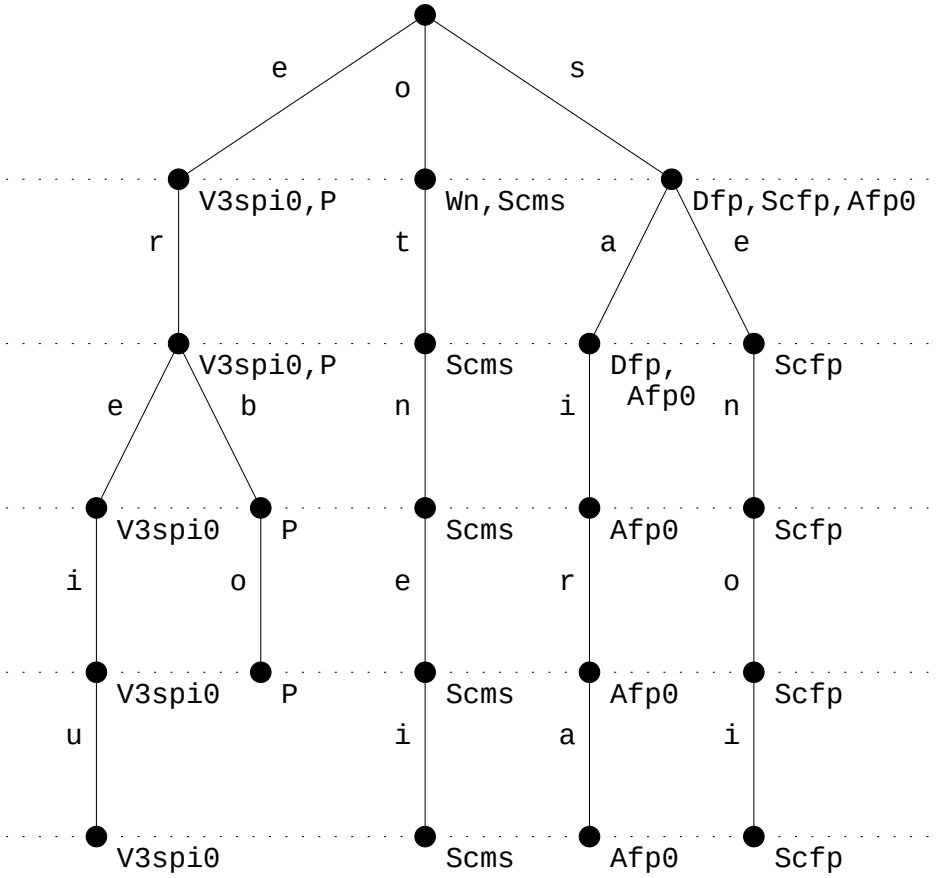


$$\hat{t}_{1,n} = \arg\max_{t_{1,n}} P(t_{1,n}|w_{1,n}) = \arg\max_{t_{1,n}} \prod_{i=1}^{n} [P(w_i|t_i) \times P(t_i|t_{i-2}, t_{i-1})]$$

Tratamiento de palabras desconocidas

Inferencias sucesivas de sufijos [Samuelsson 1993]:

⋮	⋮
no	Wn
requiere	V3spi0
conocimiento	Scms
sobre	P
las	Dfp
funciones	Scfp
suplementarias	Afp0
de	P
⋮	⋮



$$P(t|l_{n-i+1},\ldots,l_n) = \frac{\hat{p}(t|l_{n-i+1},\ldots,l_n) + \theta_i\, P(t|l_{n-i+2},\ldots,l_n)}{1+\theta_i}$$

Tratamiento de datos dispersos

Métodos de suavización de parámetros:

- Etiquetador TnT [Brants 2000]:
  - Interpolación lineal de borrado

$$\hat{p}(t_3|t_1,t_2) = \lambda_3\, f(t_3|t_1,t_2) + \lambda_2\, f(t_3|t_2) + \lambda_1\, f(t_3)$$

- Etiquetador Galena:
  - Interpolación lineal de borrado
  - Método *back-off* [Katz 1987]

$$\hat{p}(t_3|t_1,t_2) = \begin{cases} f(t_3|t_1,t_2) & \text{si } C(t_1,t_2,t_3) \geq K \\ \alpha\, Q_T(t_3|t_1,t_2) & \text{si } 1 \leq C(t_1,t_2,t_3) < K \\ \beta(t_1,t_2)\, \hat{p}(t_3|t_2) & \text{en caso contrario} \end{cases}$$

$$\hat{p}(t_3|t_2) = \begin{cases} f(t_3|t_2) & \text{si } C(t_2,t_3) \geq K \\ \alpha'\, Q_T(t_3|t_2) & \text{si } 1 \leq C(t_2,t_3) < K \\ \beta(t_2)\, f(t_3) & \text{en caso contrario} \end{cases}$$

Integración de diccionarios externos

- Método *adding-one* [Church 1988]:

$$\hat{p}(w^k|t^j) = \frac{C(w^k|t^j) + 1}{C(t^j) + K_j}$$

- Método Good-Turing:
  - Las palabras desconocidas se consideran *ceros reales* y las resuelve el adivinador
  - Las palabras presentes sólo en el diccionario se consideran *ceros no observados*
  - Este método asigna probabilidades distintas de cero a estos sucesos no observados
  - Las probabilidades serán menores que las de las palabras del corpus de entrenamiento
  - La integración considera conjuntos aislados de palabras sobre cada etiqueta
  - Es aplicable sólo cuando $n_0 \gg n_1 \gg n_2$
  - Fuera de esta situación, se debe utilizar la integración *adding-one*

Aprendizaje de etiquetas basado en transformaciones

Etiquetador BRILL [Brill 1994]:

- Generación automática de:
  - Reglas léxicas (para el tratamiento de palabras desconocidas)
  - Reglas contextuales (para la eliminación de ambigüedades)
    * `P Scms nexttag P`
    * `Scms Ams0 wdprevtag Scms receptor`
- Ventajas:
  - Las reglas de transformación presentan un alto grado de expresividad
- Desventajas:
  - La modificación manual de reglas puede provocar complejas interacciones
  - Los tiempos de entrenamiento son muy elevados
  - No se proporciona información probabilística sobre etiquetas y palabras

Modelos de máxima entropía

Etiquetador JMX [Ratnaparkhi 1996]:

- Entropía: función de mérito que permite comparar modelos probabilísticos

$$H(p) = -\sum_{h \in \mathcal{H}, t \in \mathcal{T}} p(h,t) \log(p(h,t)) \qquad E(f_j) = \sum_{h \in \mathcal{H}, t \in \mathcal{T}} p(h,t) f_j(h,t)$$

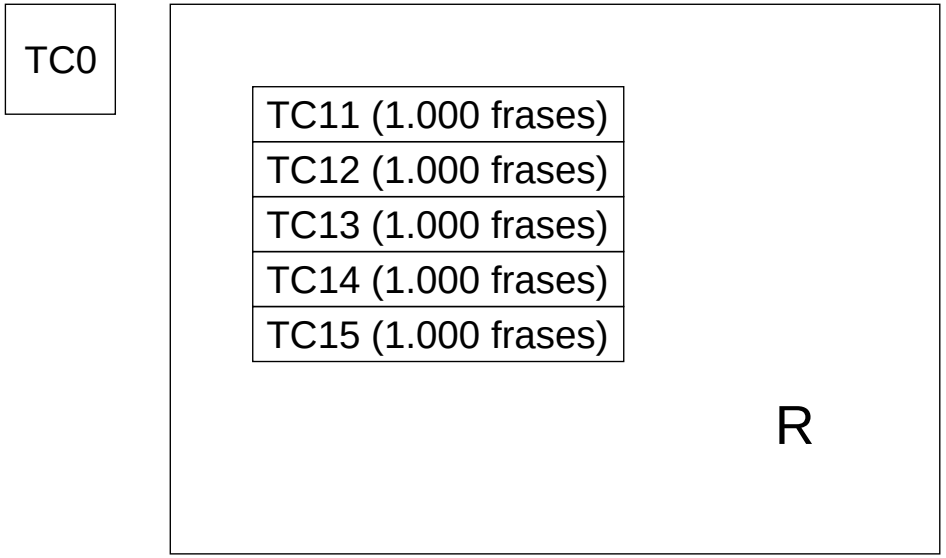
- Funciones de rasgos contextuales específicas para la etiquetación

$$f_j(h_i, t_i) = \begin{cases} 1 & \text{si } \textit{sufijo}(w_i) = \texttt{ing} \text{ y } t_i = \texttt{VBG} \\ 0 & \text{en caso contrario} \end{cases}$$

- Ventajas:
  - La representación del conocimiento lingüístico es muy flexible
  - Es posible la integración dentro de un marco probabilístico
- Desventajas:
  - Los tiempos de etiquetación son muy elevados

Metodología de evaluación

- Estrategia de experimentación:



- Contadores:

Desconocidas

$$\left.\begin{array}{l} OOVF+ \\ OOVF- \end{array}\right\} \#OOVF$$

No ambiguas

$$\left.\begin{array}{l} NAF+ \\ NAF- \end{array}\right\} \#NAF$$

Ambiguas

$$\left.\begin{array}{l} AF+ \\ AF- \end{array}\right\} \#AF$$

$$\Big\} \#F$$

- Índices de rendimiento:
  - Precisión:

$$S1 = \frac{(OOVF+)+(NAF+)+(AF+)}{\#F} \times 100$$

  - Decisión:

$$S2 = \frac{(OOVF+)+(AF+)}{\#F-\#NAF} \times 100$$

Análisis de resultados

S1						Índice	S2					
1.054	2.054	3.054	4.054	5.054	10.054	# frases	1.054	2.054	3.054	4.054	5.054	10.054
94,035	95,854	**96,910**	**97,466**	**97,730**	**98,032**	TnT	88,653	92,103	93,964	94,984	95,528	96,304
92,759	95,155	96,385	97,019	97,443	97,987	Brill	85,891	90,512	92,735	93,925	94,872	96,211
92,664	95,194	96,483	97,079	97,454	97,974	JMX	88,516	**92,269**	**94,166**	**95,120**	**95,730**	**96,676**
94,140	**95,915**	96,679	97,224	97,561	97,990	Galena	**88,881**	92,243	93,423	94,412	95,146	96,250