

Clase nº 13  
Año académico 2003

# **Introducción práctica a los modelos gráficos y su uso en ASR**

**6.345**

# Modelos gráficos para el ASR

- **Los HMM (modelos ocultos de Markov) (y la mayoría del resto de los modelos de ASR comunes) presentan algunos inconvenientes:**
  - **Fuertes suposiciones de independencia**
  - **Variable de estado simple por marco temporal**
- **Es posible que sea necesario modelar una estructura más compleja**
  - **Procesos múltiples (audio + video, discurso + ruido, múltiples corrientes de rasgos acústicos, rasgos articulatorios)**
  - **Dependencias entre estos procesos o entre observaciones acústicas**
- **Los modelos gráficos proporcionan:**
  - **Algoritmos generales para una gran clase de modelos**
    - ⇒ No es necesario escribir un código nuevo para cada modelo
  - **Un “lenguaje” con el que hablar sobre modelos estadísticos**

# Esquema

- **Primera mitad – introducción a los GM (modelos gráficos)**
  - Independencia e independencia condicional
  - **Redes bayesianas (BN)**
    - \* Definición
    - \* Principales problemas
  - **Modelos gráficos en general**
- **Segunda mitad – redes bayesianas dinámicas (DBN) para el reconocimiento de voz**
  - **Redes bayesianas dinámicas -- más allá de los HMM**
  - **Implementación de un decodificador de ASR /entrenamiento mediante DBN**
  - **DBN más complejas para el reconocimiento**
  - **GMTK (paquete de herramientas para el modelado gráfico)**

# Independencia (estadística)

- **Definición:** Dadas las variables aleatorias  $X$  e  $Y$

$$X \perp Y$$

 $\Leftrightarrow$ 

$$p(x | y) = p(x)$$

 $\Updownarrow$ 

$$p(x, y) = p(x)p(y)$$

 $\Leftrightarrow$  $\Updownarrow$ 

$$p(y | x) = p(y)$$

# Independencia condicional (estadística)

- **Definición:** Dadas las variables aleatorias  $X$ ,  $Y$  y  $Z$ ,

$$X \perp Y | Z$$

$$\Leftrightarrow p(x | y, z) = p(x | z)$$



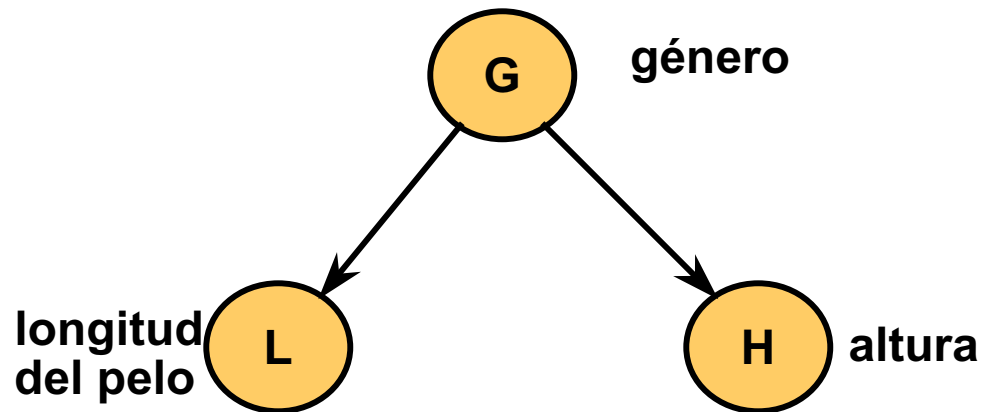
$$p(x, y | z) = p(x | z)p(y | z)$$

$$\Leftrightarrow p(y | x, z) = p(y | z)$$



## ¿Es la altura independiente de la longitud del pelo?

- Generalmente no
- Si se conoce el género, sí
- Este es el escenario de la “causa común”



$$p(h | l) \neq p(h)$$

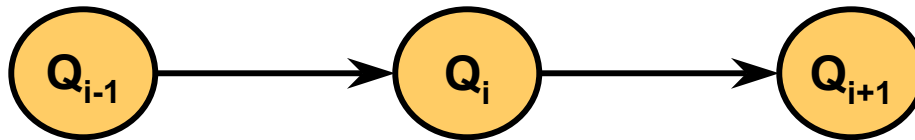
$$p(h | l, g) = p(h | g)$$

$$H \not\perp L$$

$$H \perp L | G$$

# ¿Es el futuro independiente del pasado (en un proceso de Markov)?

- Generalmente no
- Si se conoce el estado presente, entonces sí



$$p(q_{i+1} | q_{i-1}) \neq p(q_{i+1})$$

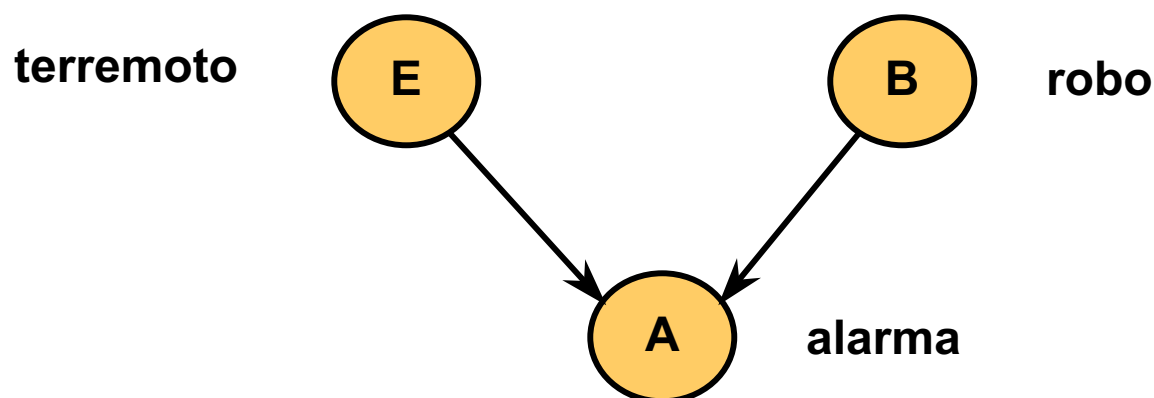
$$p(q_{i+1} | q_{i-1}, q_i) = p(q_{i+1} | q_i)$$

$$Q_{i+1} \not\perp Q_{i-1}$$

$$Q_{i+1} \perp Q_{i-1} | Q_i$$

# ¿Son los robos independientes de los terremotos?

- Generalmente sí
- Si se conoce el estado de la alarma, no
- Efecto de explicación convincente: el terremoto "encuentra una explicación convincente" del robo



$$p(b | e) = p(b)$$

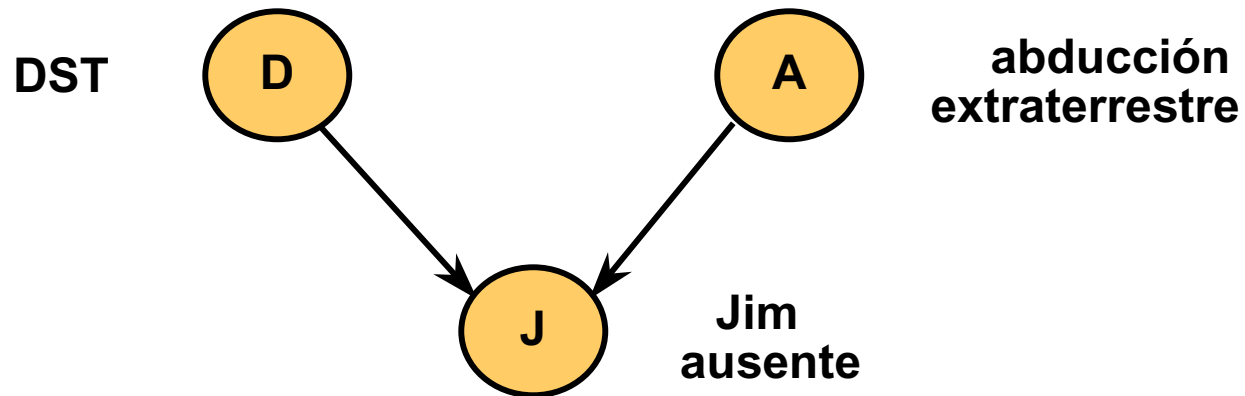
$$p(b | e, a) \neq p(b | a)$$

$$E \perp B$$

$$E \not\perp B | A$$

## ¿Son las abducciones extraterrestres independientes del horario de aprovechamiento diurno (DST)?

- **Generalmente sí**
- **Si Jim no aparece por clase, no**
- **De nuevo, el efecto de explicación convincente**



$$p(a | d) = p(a)$$

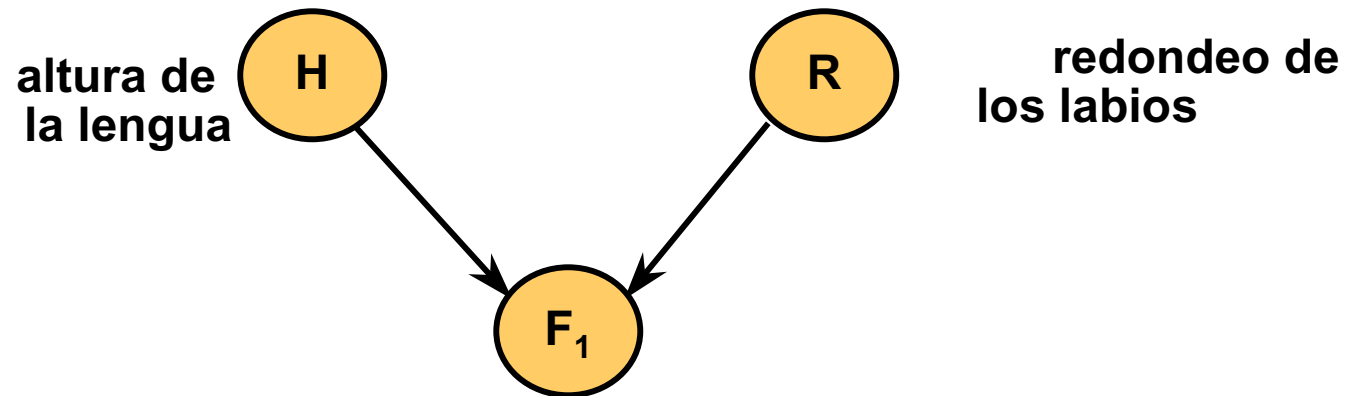
$$p(a | d, j) \neq p(a | j)$$

$$D \perp A$$

$$D \not\perp A | J$$

## ¿Es la altura de la lengua independiente del redondeo de los labios?

- **Generalmente sí**
- **Si  $F_1$  se conoce, no**
- **Aunque nuevamente, el efecto de explicación convincente...**



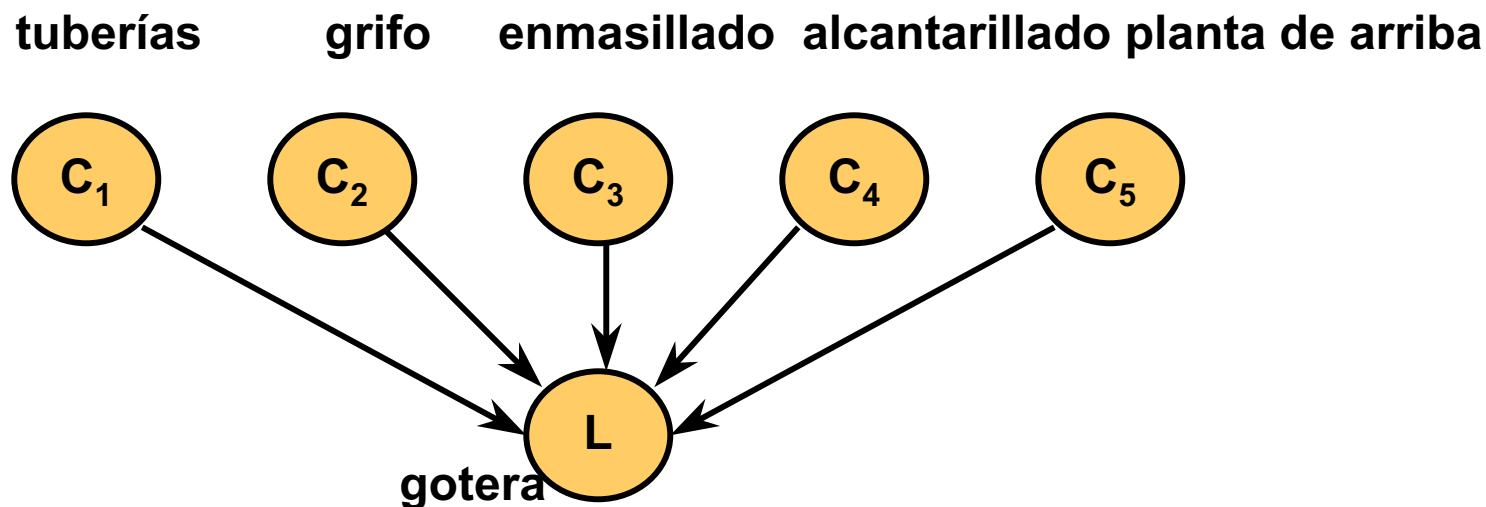
$$p(h | r) = p(h)$$

$$p(h | r, f_1) \neq p(h | f_1)$$

$$H \perp R$$

$$H \not\perp R | F_1$$

## Más explicaciones convincentes...



$$p(c_i | c_j) = p(c_i)$$

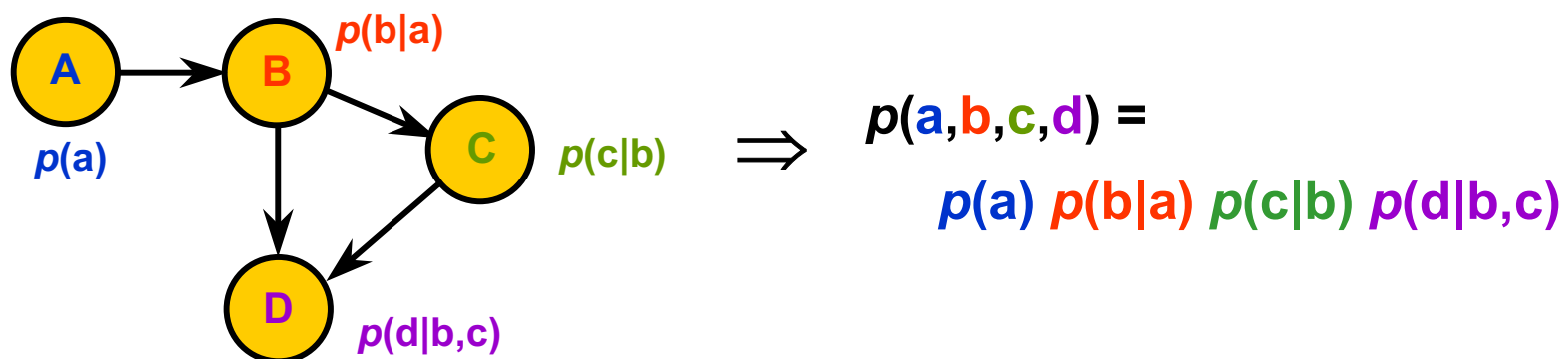
$$p(c_i | c_j, l) \neq p(c_i | l)$$

$$C_i \perp C_j \quad \forall i, j$$

$$C_i \not\perp C_j | L \quad \forall i, j$$

# Redes bayesianas

- Las transparencias anteriores son ejemplos de redes bayesianas simples
- Definición:
  - Grafo dirigido acíclico (DAG) con una correspondencia uno a uno entre nodo (vértices) y variables  $X_1, X_2, \dots, X_N$
  - Cada nodo  $X_i$  con padres  $pa(X_i)$  está asociado con la función de probabilidad local  $p_{X_i|pa(X_i)}$
  - La probabilidad de unión de todas las variables viene dada por el producto de las probabilidades locales, ej.,  $p(x_1, \dots, x_N) = \prod p(x_i|pa(x_i))$



- Una red bayesiana (BN) determinada representa una *familia* de distribuciones de probabilidad

# Redes bayesianas, continuación

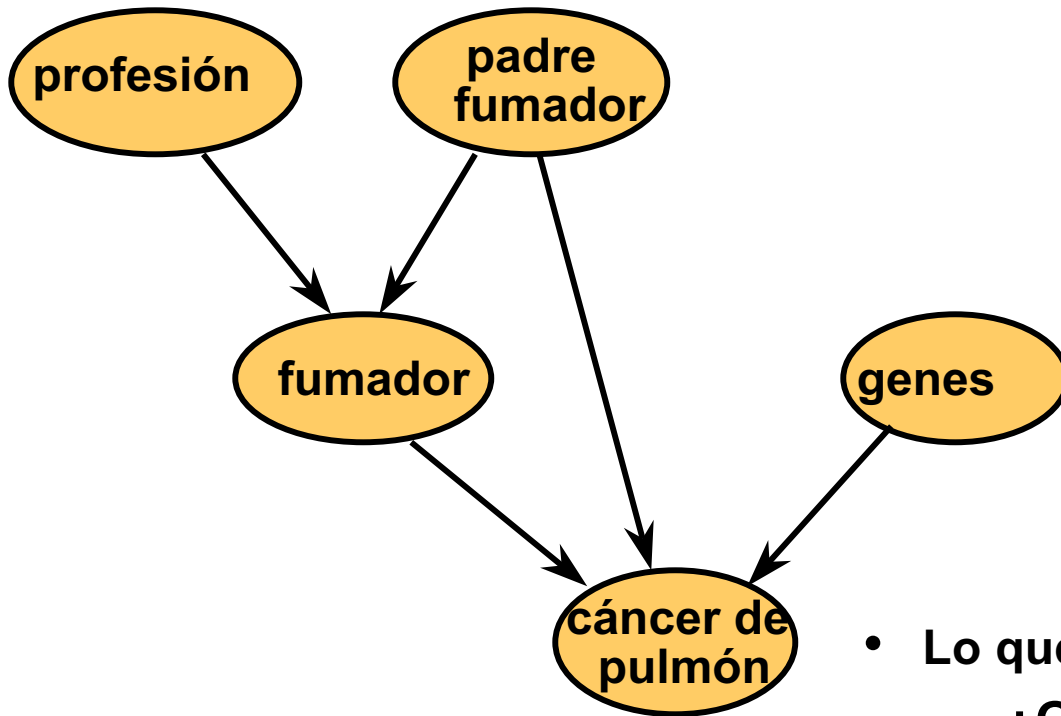
- Las puntas que faltan en el grafo corresponden a suposiciones de independencia
- La probabilidad de unión se puede siempre factorizar según la regla de la cadena:

$$p(a,b,c,d) = p(a) p(b|a) p(c|a,b) p(d|a,b,c)$$

- Sin embargo, al hacer algunas suposiciones de independencia, obtenemos una factorización **escasa**, ej., una con menos parámetros

$$p(a,b,c,d) = p(a) p(b|a) p(c|b) p(d|b,c)$$

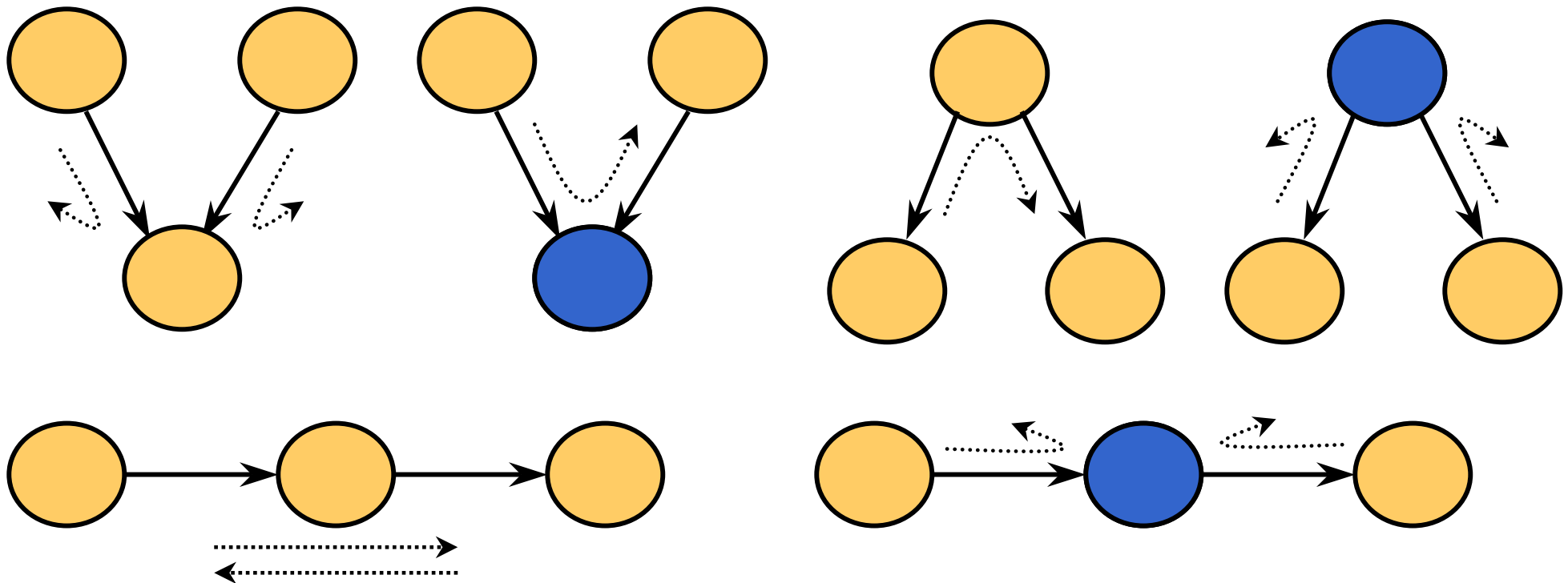
# Ejemplo en un entorno médico



- Lo que es posible que queramos saber:
  - ¿Qué suposiciones de independencia codifica este modelo?
  - ¿Qué es  $p(\text{cáncer de pulmón} \mid \text{profesión})$ ?  
¿ $p(\text{fumador} \mid \text{padre fumador}, \text{genes})$ ?
  - Dadas algunas de las variables, ¿cuáles son los valores más probables de otras?
  - ¿Cómo estimamos las probabilidades locales a partir de los datos?

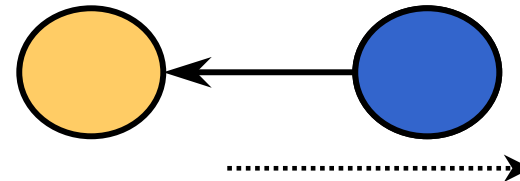
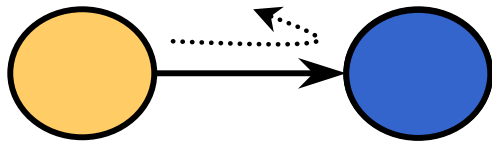
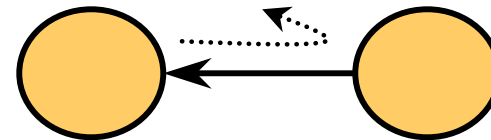
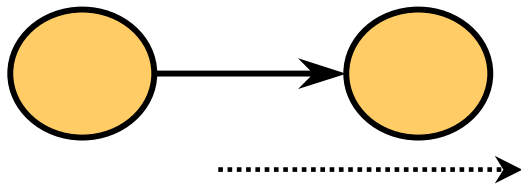
# Determinando independencias a partir de un grafo

- Existen varios modos...
- Algoritmo de la pelota de Bayes (“Pelota de Bayes : El tiempo pasado racional”, Schachter 1998)
  - Rebotar la pelota alrededor de un grafo según un conjunto de reglas
  - Dos nodos son independientes dados un conjunto de nodos observados, si una pelota no puede saltar de uno a otro

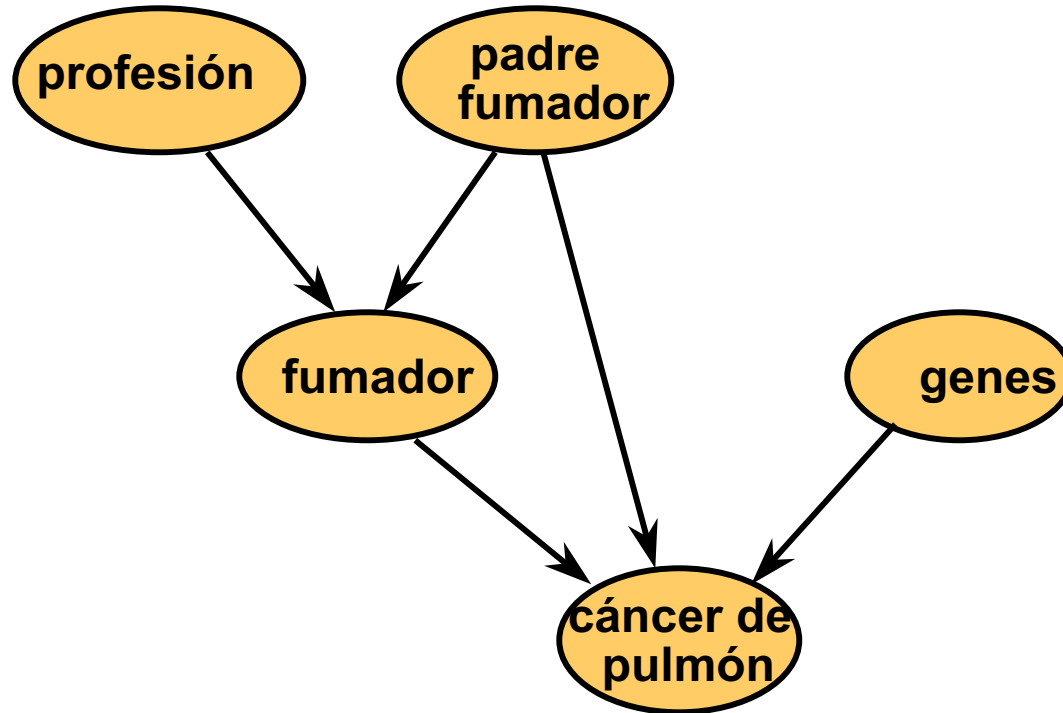


# Pelota de Bayes, continuación

- Condiciones limítrofes:



# Pelota de Bayes en un ejemplo médico



- **Según este modelo:**

- ¿Son los genes de una persona independientes de si tienen un padre que fume? ¿Y si sabemos que la persona tiene cáncer de pulmón?
- ¿Es el cáncer de pulmón independiente de la profesión dado que la persona es fumadora?
- (¿Tienen sentido las respuestas?)

# Inferencia

- **Definición:**
  - **Computación de la probabilidad de un subconjunto de las variables dado otro subconjunto**
- **La inferencia es una subrutina de:**
  - **Descodificación de Viterbi**

$$q^* = \operatorname{argmax}_q p(q|obs)$$

- **La estimación de máxima probabilidad de los parámetros de las probabilidades locales**

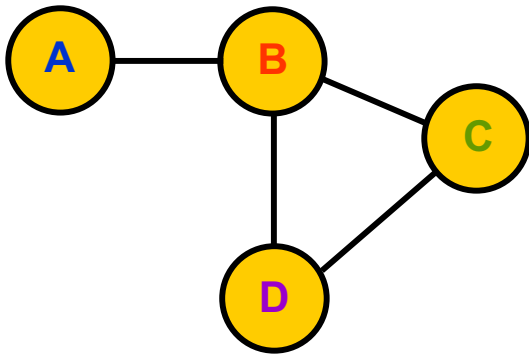
$$\lambda^* = \operatorname{argmax}_\lambda p(obs|\lambda)$$

# Modelos gráficos (GM)

- **En general, los GM representan familias de distribuciones de probabilidad mediante grafos**
  - dirigidos, ej. redes bayesianas
  - no dirigidos, ej. campos aleatorios de Markov
  - combinación, ej. grafos de cadena
- **Para describir una distribución *particular* con un GM, es necesario que especifiquemos:**
  - **Semántica:** red bayesiana, campo aleatorio de Markov, ...
  - **Estructura:** el mismo grafo
  - **Implementación:** la forma de las funciones locales (Gaussiana, tabla, ...)
  - **Parámetros** de las funciones locales (medias, covarianzas, entradas de tabla ...)
- **No todos los tipos de GM pueden representar todos los grupos de propiedades de independencia**

# Ejemplos de modelos gráficos no dirigidos: Campos aleatorios de Markov

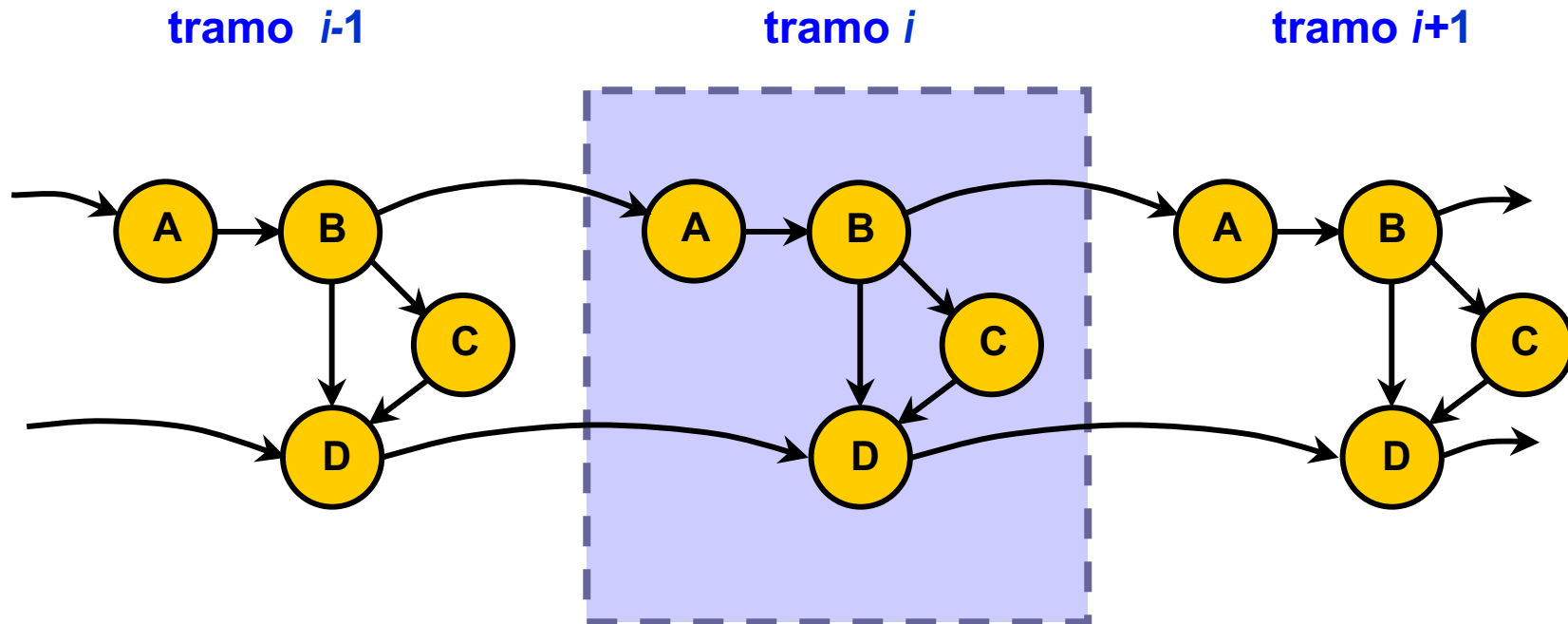
- **Definición:**
  - Grafo no dirigido
  - Función local (“potencial”) definida en cada clique máximo
  - Probabilidad de unión dada por el producto normalizado de potenciales
- **Las propiedades de independencia pueden deducirse mediante la separación de un grafo simple**



$$p(a, b, c, d) \propto \psi_{A,B}(a, b) \psi_{B,C,D}(b, c, d)$$

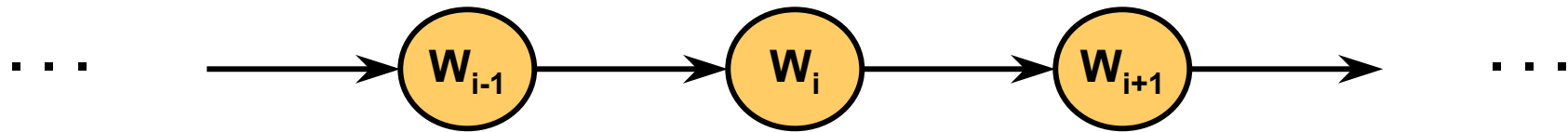
# Redes bayesianas dinámicas (DBN)

- Las BN constan de una estructura que se repite un número de veces (indefinido o dinámico)
  - Práctico para el modelado de series de tiempo (ej., discurso)

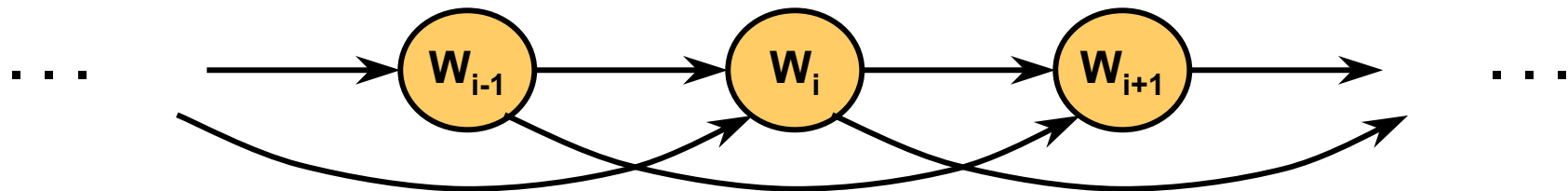


# Representación de DBN de modelos de lenguaje n-grama

- **Bigrama:**

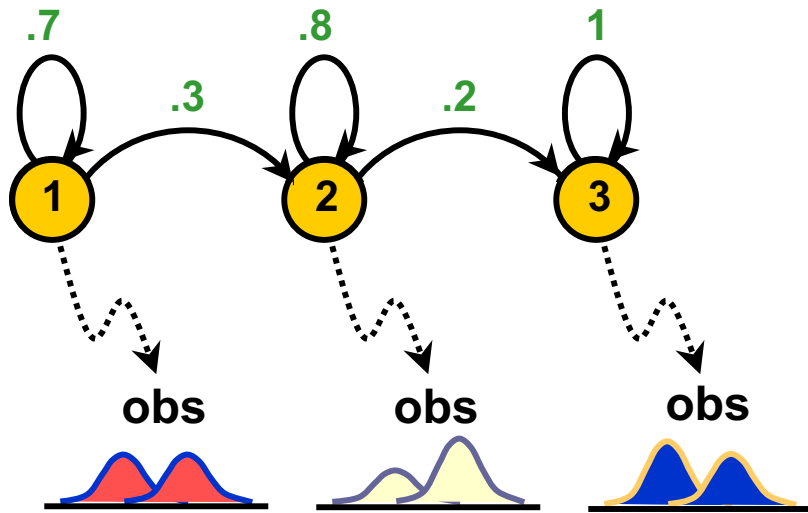


- **Trigrama:**

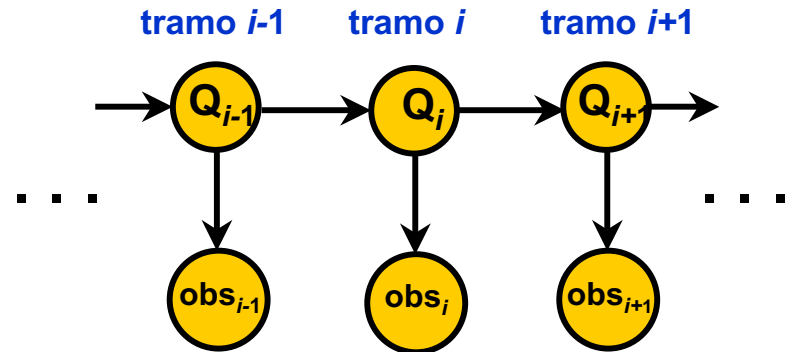


# Representación de HMM como DBN

**HMM**



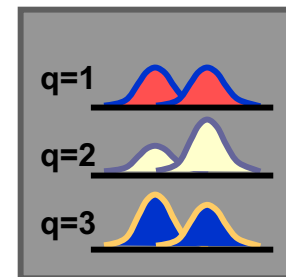
**DBN**







$P(q_i | q_{i-1})$

$q_i \backslash q_{i-1}$	1	2	3
1	.7	.3	0
2	0	.8	.2
3	0	0	1

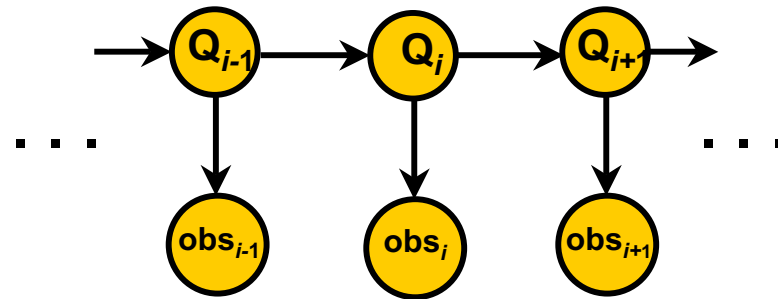
$P(obs_i | q_i)$



 = estado  
 = transición permitida

 = variable  
 = dependencia permitida

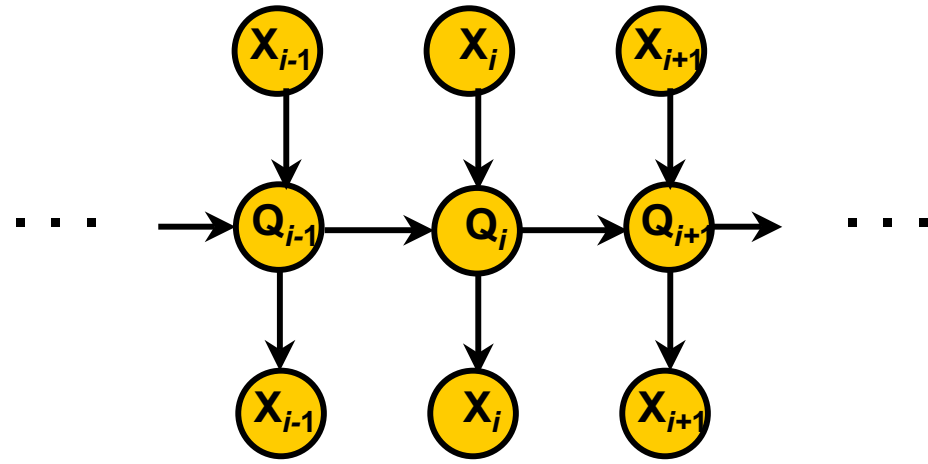
## Formulación de un ASR basado en HMM como un problema de GM



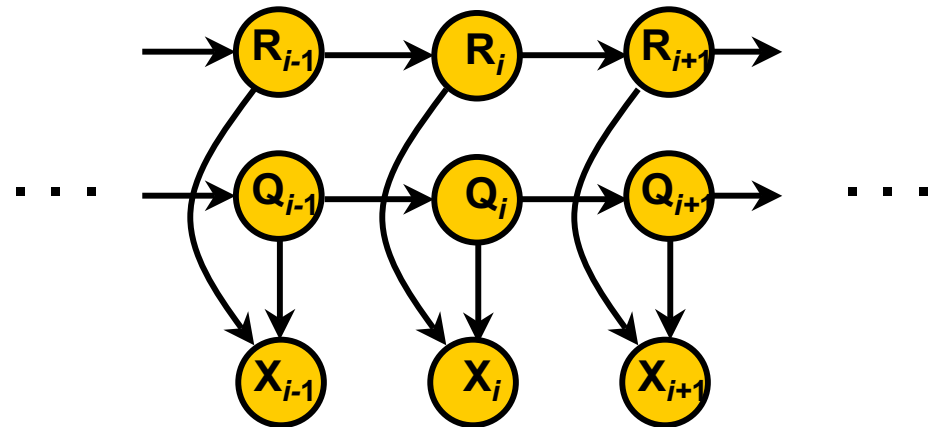
- **Descodificación de Viterbi** → encontrar los entornos más probables para todo  $Q_i$  dadas las observaciones acústicas  $\{obs_i\}$
- **Entrenamiento de Baum-Welch** → encontrar los entornos más probables para los parámetros de  $P(q_i|q_{i-1})$  y  $P(obs_i | q_i)$
- Ambos son casos especiales de los algoritmos de GM estándar para Viterbi y entrenamiento con EM

# Variaciones

- HMM de entrada-salida



- HMM factoriales

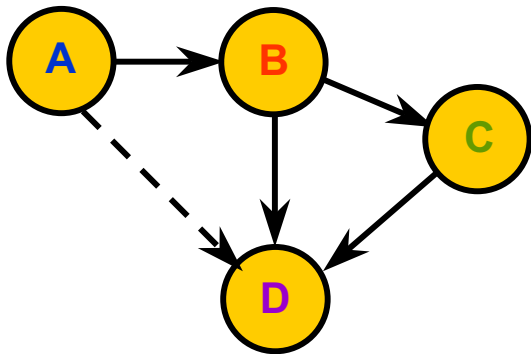


# Padres cambiantes

- **Definición:**

- Una variable  $X$  es un padre cambiante de variable  $Y$  si el valor de  $X$  determina los padres y/o implementación de  $Y$

- **Ejemplo:**

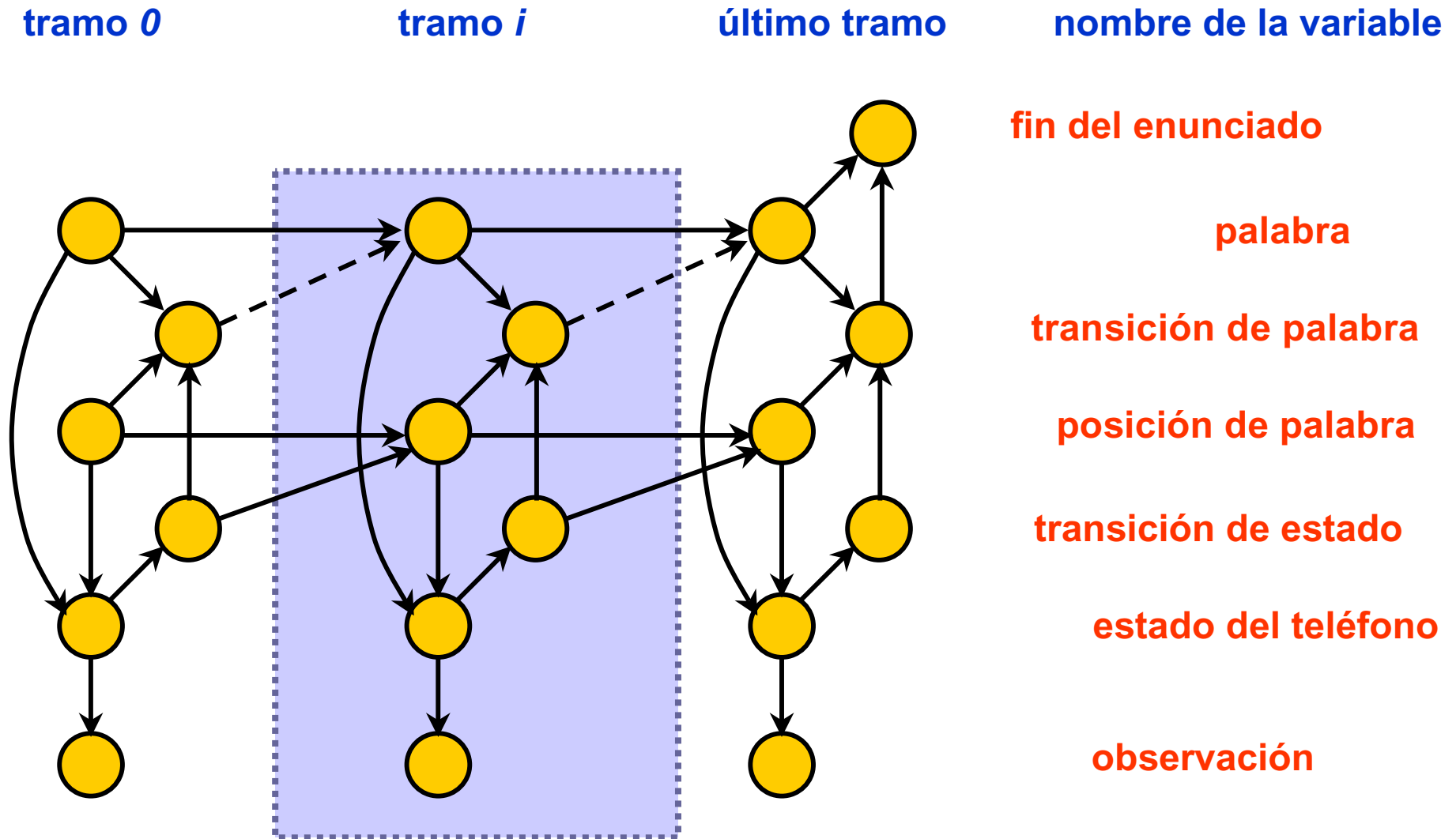


$A=0 \Rightarrow D$  tiene padre B con distribución gaussiana

$A=1 \Rightarrow D$  tiene padre C con distribución gaussiana

$A=2 \Rightarrow D$  tiene padre C con mezcla de distribución gaussiana

# Reconocimiento basado en HMM con una DBN



- ¿Qué modelo de lenguaje implementa este GM?

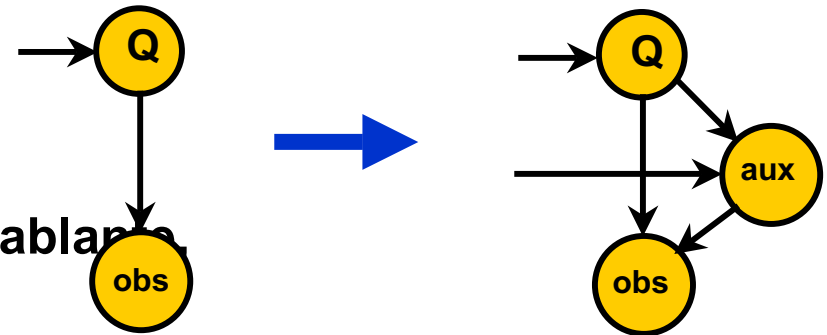
# Entrenamiento y pruebas de las DBN

- ¿Por qué son necesarias distintas estructuras para las pruebas de entrenamiento?  
¿No es exactamente lo mismo entrenar que hacer pruebas, pero con más de las variables observadas?
- No siempre
  - A menudo, durante el entrenamiento tenemos sólo información *parcial* sobre algunas de las variables, ej., la secuencia de palabras, pero no qué **tramo** va con cada palabra

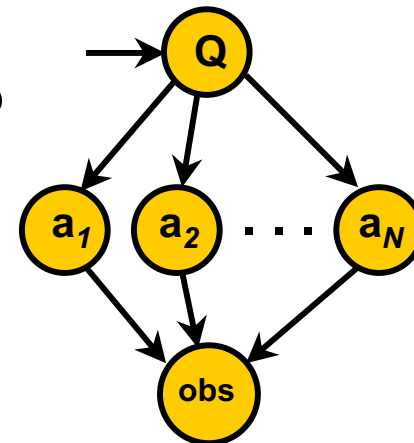
# Modelos GM más complejos para el reconocimiento

- **HMM + variables auxiliares (Zweig 1998, Stephenson 2001)**

- Agrupamiento de ruido
- Agrupamiento de hablante
- Dependencia de tono, velocidad de habla, etc.



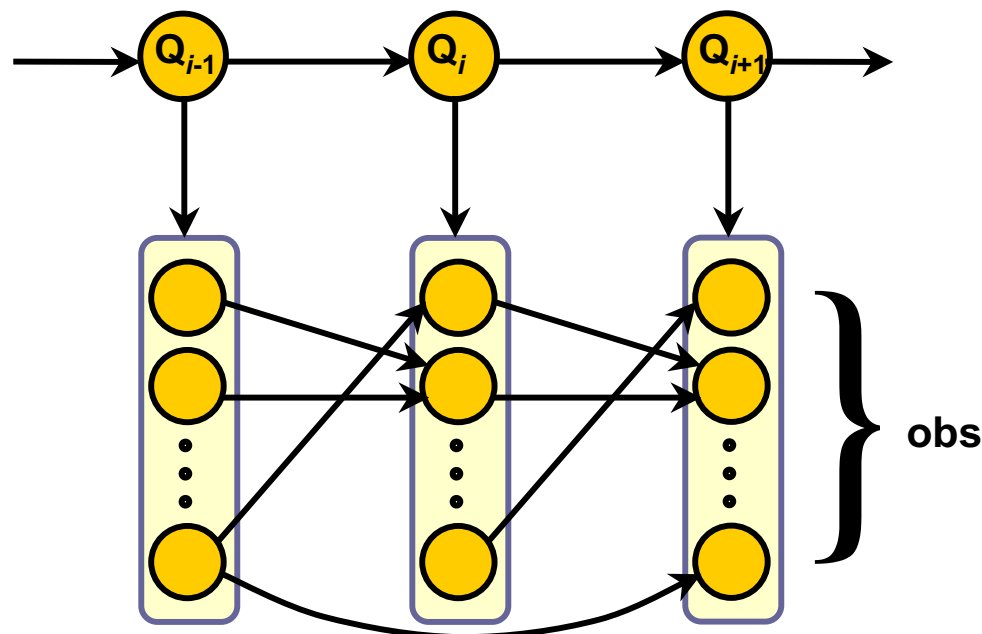
- **Modelado basado en rasgos / articulatorio**



- **Modelado multi-velocidad, reconocimiento de voz audiovisual (Nefian et al. 2002)**

# Modelando dependencias de inter-observación: Modelos de Markov enterrados (Bilmes 1999)

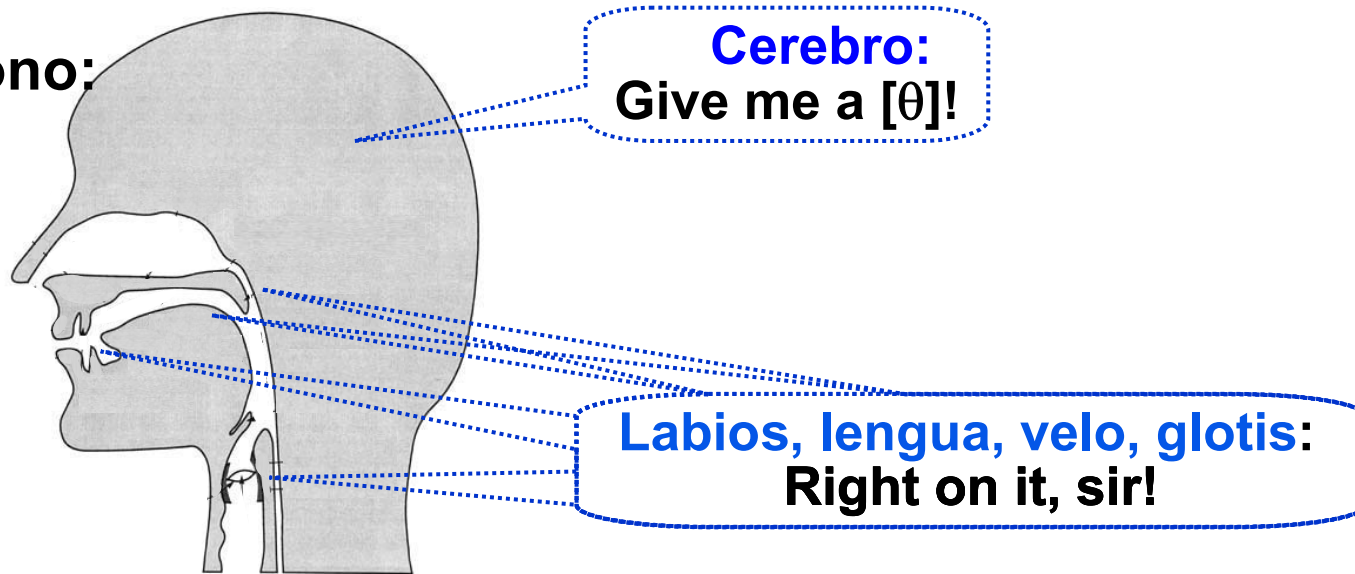
- Observe en primer lugar que la variable de observación es en realidad un vector de observaciones acústicas (ej., los MFCC (coeficientes cepstrales de frecuencia Mel))



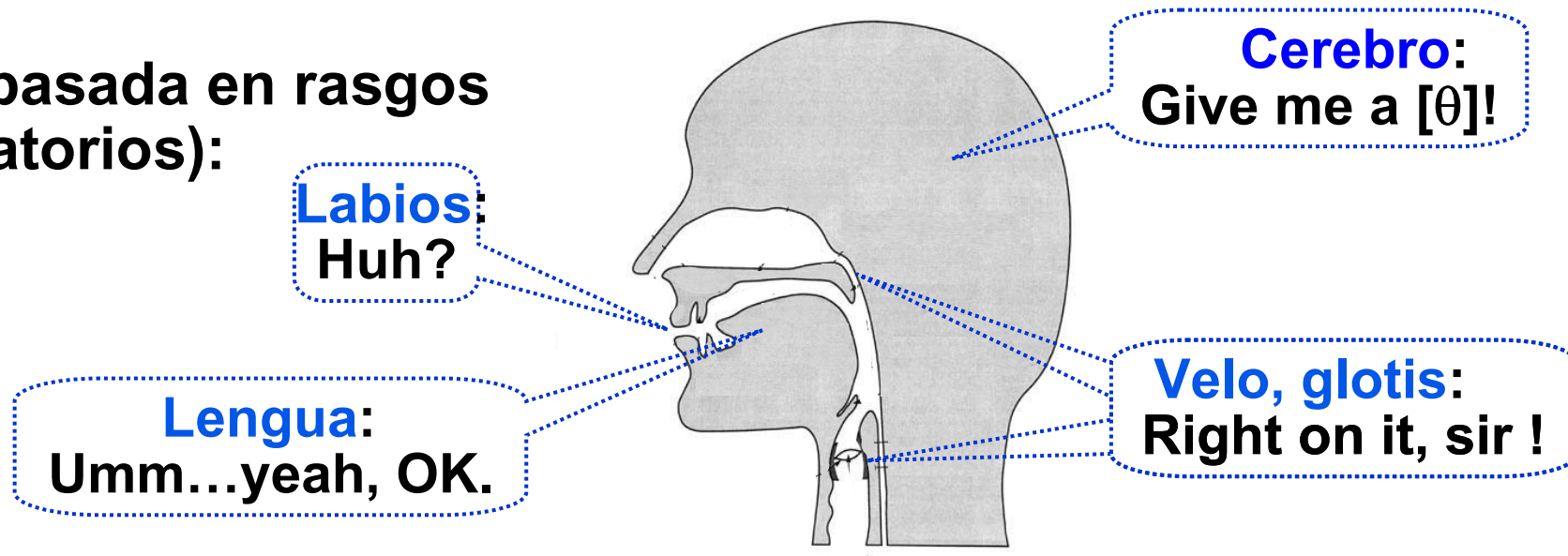
- Considere la adición de dependencias entre las observaciones
- Añada sólo aquellas que sean discriminativas con relación a la clasificación del estado/teléfono/palabra actual

# Modelado basado en rasgos

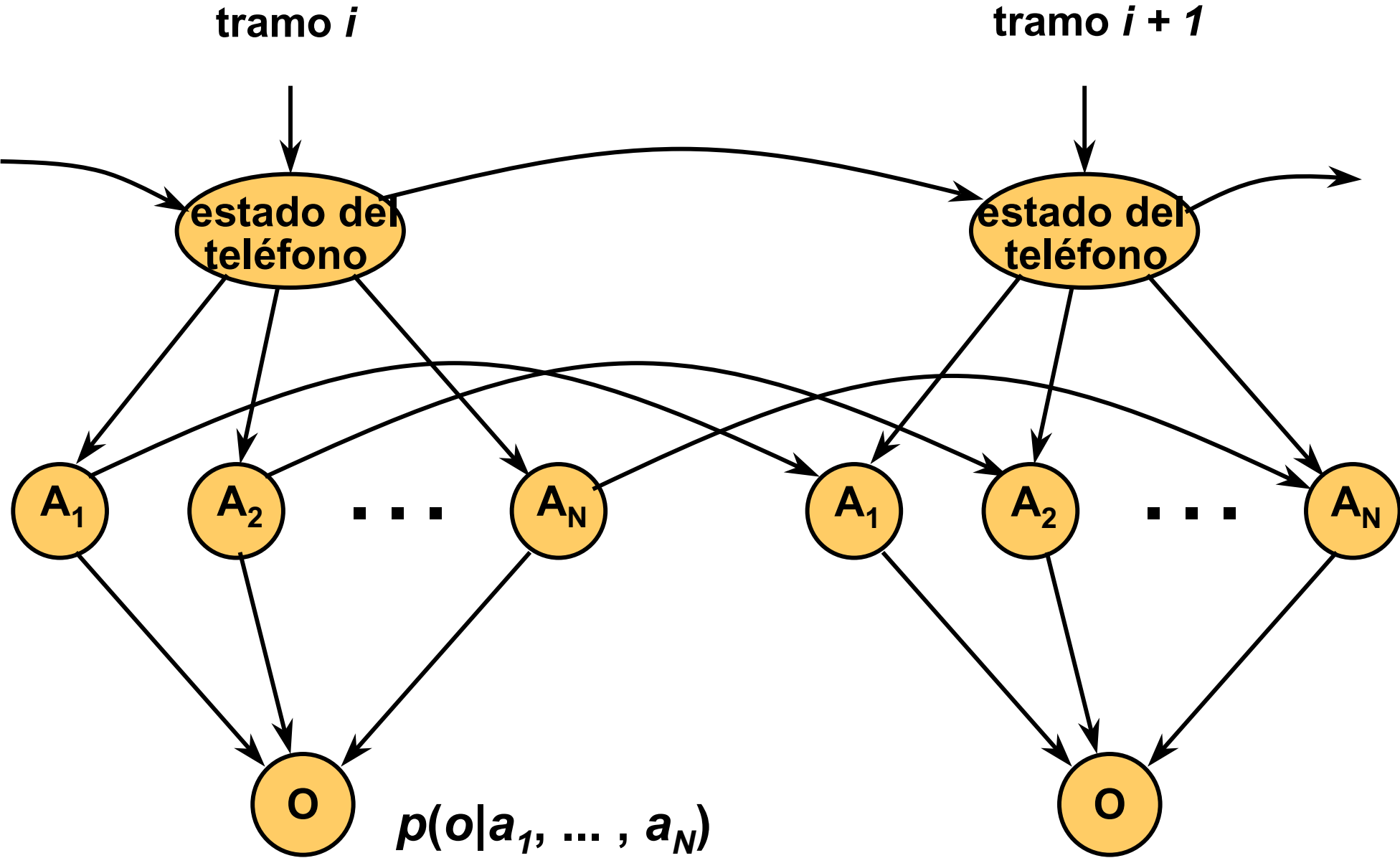
- **Visión de un teléfono:**



- **Visión basada en rasgos (articulatorios):**



# Una DBN basada en rasgos para el ASR



# **GMTK: Paquete de herramientas para el modelado gráfico (J. Bilmes y G. Zweig, ICASSP 2002)**

- **Paquete de herramientas para la especificación y computación con redes bayesianas dinámicas**
- **Los modelos se especifican mediante:**
  - **Archivo de estructura: define variables, dependencias, y una forma de distribuciones condicionales asociadas**
  - **Archivos del parámetro: especifica parámetros para cada distribución en el archivo de estructura**
- **Las distribuciones de variable pueden ser:**
  - **Mezcla de gaussianas + variantes**
  - **Tablas de probabilidad multidimensional**
  - **Tablas de probabilidad escasa**
  - **Determinísticas (árboles de decisión)**
- **Proporciona programas para el entrenamiento con EM, la descodificación de Viterbi y varias utilidades**

## Parte de un ejemplo de un archivo de estructura

```
variable : phone {
  type: discrete hidden cardinality NUM_PHONES;
  switchingparents: nil;
  conditionalparents: word(0), wordPosition(0) using
    DeterministicCPT("wordWordPos2Phone");
}

variable : obs {
  type: continuous observed OBSERVATION_RANGE;
  switchingparents: nil;
  conditionalparents: phone(0) using mixGaussian
    collection("global") mapping("phone2MixtureMapping");
}
```

# Algunas cuestiones...

- **Para algunas estructuras, la inferencia exacta puede resultar inviable desde un punto de vista computacional  $\Rightarrow$  algoritmos de inferencia aproximados**
- **La estructura no siempre se conoce  $\Rightarrow$  algoritmos de aprendizaje estructural**

# Referencias

- **J. Bilmes, “Graphical Models and Automatic Speech Recognition”, en *Mathematical Foundations of Speech and Language Processing*, Institute of Mathematical Analysis Volumes in Mathematics Series, Springer-Verlag, 2003.**
- **G. Zweig, *Speech Recognition with Dynamic Bayesian Networks*, Tesis doctoral, UC Berkeley, 1998.**
- **J. Bilmes, “What HMMs Can Do”, UWEETR-2002-0003, feb. 2002.**