

# Reconocimiento de Escritura Continua Off-Line

## Uso de Modelos de Lenguaje en RES

Alejandro H. Toselli

Departamento de Sistemas Informáticos y Computación  
Universidad Politécnica de Valencia



19 de febrero de 2008

# Temario

1

## Uso de Modelos de Lenguaje en RES

- Decodificación (Reconocimiento)
- Diferentes Niveles de Percepción
- Autómata Estocástico de Estados Finitos (AEEF)
- Modelado a Nivel Morfológico
- Modelado a Nivel Léxico
- Modelado a Nivel Sintáctico
- Modelo Integrado
- Modelos de Lenguaje de N-gramas
- Marco Teórico: Formulación del Problema de RES
- Mixtura de Gaussianas, GSF y WIP
- Métricas de Evaluación

2

## Aplicaciones

- Aplicación RNCN
- Aplicación RMFE
- Prototipo PAITYT

# Temario

1

## Uso de Modelos de Lenguaje en RES

- Decodificación (Reconocimiento)
- Diferentes Niveles de Percepción
- Autómata Estocástico de Estados Finitos (AEEF)
- Modelado a Nivel Morfológico
- Modelado a Nivel Léxico
- Modelado a Nivel Sintáctico
- Modelo Integrado
- Modelos de Lenguaje de N-gramas
- Marco Teórico: Formulación del Problema de RES
- Mixtura de Gaussianas, GSF y WIP
- Métricas de Evaluación

2

## Aplicaciones

- Aplicación RNCN
- Aplicación RMFE
- Prototipo PAITYT

# Decodificación (Reconocimiento)

El reconocimiento en sí se fundamenta en las siguientes conceptos:

- Modelado de los niveles de percepción: morfológico, léxico y sintáctico.
- Integración de estos niveles.
- Utilización de AEEFs: HMMs, gramáticas o autómatas.
- Algoritmo de Viterbi.

Desde el punto de vista del usuario:

Se puede ver como un proceso que toma como entrada imágenes de texto manuscrito y produce a la salida secuencia de palabras reconocidas.

Desde el punto de vista del modelado:

Dada una secuencia de observaciones  $x$ , cual es la secuencia de estados del modelo que con mayor probabilidad la ha producido.

# Decodificación (Reconocimiento)

El reconocimiento en sí se fundamenta en las siguientes conceptos:

- Modelado de los niveles de percepción: morfológico, léxico y sintáctico.
- Integración de estos niveles.
- Utilización de AEEFs: HMMs, gramáticas o autómatas.
- Algoritmo de Viterbi.

Desde el punto de vista del usuario:

Se puede ver como un proceso que toma como entrada imágenes de texto manuscrito y produce a la salida secuencia de palabras reconocidas.

Desde el punto de vista del modelado:

Dada una secuencia de observaciones  $x$ , cual es la secuencia de estados del modelo que con mayor probabilidad la ha producido.

# Decodificación (Reconocimiento)

El reconocimiento en sí se fundamenta en las siguientes conceptos:

- Modelado de los niveles de percepción: morfológico, léxico y sintáctico.
- Integración de estos niveles.
- Utilización de AEEFs: HMMs, gramáticas o autómatas.
- Algoritmo de Viterbi.

Desde el punto de vista del usuario:

Se puede ver como un proceso que toma como entrada imágenes de texto manuscrito y produce a la salida secuencia de palabras reconocidas.

Desde el punto de vista del modelado:

Dada una secuencia de observaciones  $x$ , cual es la secuencia de estados del modelo que con mayor probabilidad la ha producido.

# Temario

1

## Uso de Modelos de Lenguaje en RES

- Decodificación (Reconocimiento)
- **Diferentes Niveles de Percepción**
- Autómata Estocástico de Estados Finitos (AEEF)
- Modelado a Nivel Morfológico
- Modelado a Nivel Léxico
- Modelado a Nivel Sintáctico
- Modelo Integrado
- Modelos de Lenguaje de N-gramas
- Marco Teórico: Formulación del Problema de RES
- Mixtura de Gaussianas, GSF y WIP
- Métricas de Evaluación

2

## Aplicaciones

- Aplicación RNCN
- Aplicación RMFE
- Prototipo PAITYT

# Diferentes Niveles de Percepción

La eficacia en el reconocimiento radica en una intercooperación entre diferentes niveles de conocimiento:

- ① Nivel Morfológico: que dictamina acerca de la forma y estructura de los diferentes tipos de caracteres.
- ② Nivel Léxico: encargado de la formación de palabras con secuencias correctas de caracteres.
- ③ Nivel Sintáctico: encargado de la formación de sentencias con secuencias correctas de palabras.
- ④ Nivel Semántico: relacionado con la correcta interpretación de sentencias bien formadas.

Implica la necesidad de contar con:

- ① Modelos adecuados para cada uno de estos niveles.
- ② Modelos que puedan integrarse entre sí de una manera simple y natural.

# Diferentes Niveles de Percepción

La eficacia en el reconocimiento radica en una intercooperación entre diferentes niveles de conocimiento:

- ① Nivel Morfológico: que dictamina acerca de la forma y estructura de los diferentes tipos de caracteres.
- ② Nivel Léxico: encargado de la formación de palabras con secuencias correctas de caracteres.
- ③ Nivel Sintáctico: encargado de la formación de sentencias con secuencias correctas de palabras.
- ④ Nivel Semántico: relacionado con la correcta interpretación de sentencias bien formadas.

Implica la necesidad de contar con:

- ① Modelos adecuados para cada uno de estos niveles.
- ② Modelos que puedan integrarse entre sí de una manera simple y natural.

# Diferentes Niveles de Percepción

La eficacia en el reconocimiento radica en una intercooperación entre diferentes niveles de conocimiento:

- ① Nivel Morfológico: que dictamina acerca de la forma y estructura de los diferentes tipos de caracteres.
- ② Nivel Léxico: encargado de la formación de palabras con secuencias correctas de caracteres.
- ③ Nivel Sintáctico: encargado de la formación de sentencias con secuencias correctas de palabras.
- ④ Nivel Semántico: relacionado con la correcta interpretación de sentencias bien formadas.

Implica la necesidad de contar con:

- ① Modelos adecuados para cada uno de estos niveles.
- ② Modelos que puedan integrarse entre sí de una manera simple y natural.

# Diferentes Niveles de Percepción

La eficacia en el reconocimiento radica en una intercooperación entre diferentes niveles de conocimiento:

- ① Nivel Morfológico: que dictamina acerca de la forma y estructura de los diferentes tipos de caracteres.
- ② Nivel Léxico: encargado de la formación de palabras con secuencias correctas de caracteres.
- ③ Nivel Sintáctico: encargado de la formación de sentencias con secuencias correctas de palabras.
- ④ Nivel Semántico: relacionado con la correcta interpretación de sentencias bien formadas.

Implica la necesidad de contar con:

- ① Modelos adecuados para cada uno de estos niveles.
- ② Modelos que puedan integrarse entre sí de una manera simple y natural.

# Diferentes Niveles de Percepción

La eficacia en el reconocimiento radica en una intercooperación entre diferentes niveles de conocimiento:

- ① Nivel Morfológico: que dictamina acerca de la forma y estructura de los diferentes tipos de caracteres.
- ② Nivel Léxico: encargado de la formación de palabras con secuencias correctas de caracteres.
- ③ Nivel Sintáctico: encargado de la formación de sentencias con secuencias correctas de palabras.
- ④ Nivel Semántico: relacionado con la correcta interpretación de sentencias bien formadas.

Implica la necesidad de contar con:

- ① Modelos adecuados para cada uno de estos niveles.
- ② Modelos que puedan integrarse entre sí de una manera simple y natural.

# Temario

1

## Uso de Modelos de Lenguaje en RES

- Decodificación (Reconocimiento)
- Diferentes Niveles de Percepción
- **Autómata Estocástico de Estados Finitos (AEEF)**
  - Modelado a Nivel Morfológico
  - Modelado a Nivel Léxico
  - Modelado a Nivel Sintáctico
  - Modelo Integrado
  - Modelos de Lenguaje de N-gramas
  - Marco Teórico: Formulación del Problema de RES
  - Mixtura de Gaussianas, GSF y WIP
  - Métricas de Evaluación

2

## Aplicaciones

- Aplicación RNCN
- Aplicación RMFE
- Prototipo PAITYT

# Autómata Estocástico de Estados Finitos (AEEF)

Constituyen la base de los HMMs, e inclusive son utilizados para representar los modelos de léxico y los modelos de lenguaje de  $N$ -gramas.

Un AEEF  $\mathcal{T}$ , es una máquina de estados finitos definida por la séxtupla  $(Q, \Sigma, R, q_0, P, F)$ , donde:

$Q$  es un conjunto finito de estados,

$\Sigma$  es un conjunto finito de símbolos de entrada,

$R \subset Q \times \Sigma \times Q$  es un conjunto de transiciones de la forma  $(q, c, q')$   
para  $c \in \Sigma$  y  $q, q' \in Q$ ,

$q_0$  es el estado inicial ( $q_0 \in Q$ ),

$P$  función de probabilidad de transición  $P : R \rightarrow \mathbb{R}^+$ ,

$F$  función de probabilidad de estado final  $F : Q \rightarrow \mathbb{R}^+$ .

Debe cumplirse además:

$$F(q) + \sum_{\forall(p,c,p'):p=q} P(p, c, p') = 1 \quad \forall q \in Q$$

# Autómata Estocástico de Estados Finitos (AEEF)

Constituyen la base de los HMMs, e inclusive son utilizados para representar los modelos de léxico y los modelos de lenguaje de  $N$ -gramas.

Un AEEF  $\mathcal{T}$ , es una máquina de estados finitos definida por la séxtupla  $(Q, \Sigma, R, q_0, P, F)$ , donde:

$Q$  es un conjunto finito de estados,

$\Sigma$  es un conjunto finito de símbolos de entrada,

$R \subset Q \times \Sigma \times Q$  es un conjunto de transiciones de la forma  $(q, c, q')$   
para  $c \in \Sigma$  y  $q, q' \in Q$ ,

$q_0$  es el estado inicial ( $q_0 \in Q$ ),

$P$  función de probabilidad de transición  $P : R \rightarrow \mathbb{R}^+$ ,

$F$  función de probabilidad de estado final  $F : Q \rightarrow \mathbb{R}^+$ .

Debe cumplirse además:

$$F(q) + \sum_{\forall (p,c,p'): p=q} P(p, c, p') = 1 \quad \forall q \in Q$$

# Autómata Estocástico de Estados Finitos (AEEF)

Constituyen la base de los HMMs, e inclusive son utilizados para representar los modelos de léxico y los modelos de lenguaje de  $N$ -gramas.

Un AEEF  $\mathcal{T}$ , es una máquina de estados finitos definida por la séxtupla  $(Q, \Sigma, R, q_0, P, F)$ , donde:

$Q$  es un conjunto finito de estados,

$\Sigma$  es un conjunto finito de símbolos de entrada,

$R \subset Q \times \Sigma \times Q$  es un conjunto de transiciones de la forma  $(q, c, q')$   
para  $c \in \Sigma$  y  $q, q' \in Q$ ,

$q_0$  es el estado inicial ( $q_0 \in Q$ ),

$P$  función de probabilidad de transición  $P : R \rightarrow \mathbb{R}^+$ ,

$F$  función de probabilidad de estado final  $F : Q \rightarrow \mathbb{R}^+$ .

Debe cumplirse además:

$$F(q) + \sum_{\forall(p,c,p'):p=q} P(p, c, p') = 1 \quad \forall q \in Q$$

# Autómata Estocástico de Estados Finitos (AEEF)

Constituyen la base de los HMMs, e inclusive son utilizados para representar los modelos de léxico y los modelos de lenguaje de  $N$ -gramas.

Un AEEF  $\mathcal{T}$ , es una máquina de estados finitos definida por la séxtupla  $(Q, \Sigma, R, q_0, P, F)$ , donde:

$Q$  es un conjunto finito de estados,

$\Sigma$  es un conjunto finito de símbolos de entrada,

$R \subset Q \times \Sigma \times Q$  es un conjunto de transiciones de la forma  $(q, c, q')$  para  $c \in \Sigma$  y  $q, q' \in Q$ ,

$q_0$  es el estado inicial ( $q_0 \in Q$ ),

$P$  función de probabilidad de transición  $P : R \rightarrow \mathbb{R}^+$ ,

$F$  función de probabilidad de estado final  $F : Q \rightarrow \mathbb{R}^+$ .

Debe cumplirse además:

$$F(q) + \sum_{\forall (p,c,p'): p=q} P(p, c, p') = 1 \quad \forall q \in Q$$

# Autómata Estocástico de Estados Finitos (AEEF)

Constituyen la base de los HMMs, e inclusive son utilizados para representar los modelos de léxico y los modelos de lenguaje de  $N$ -gramas.

Un AEEF  $\mathcal{T}$ , es una máquina de estados finitos definida por la séxtupla  $(Q, \Sigma, R, q_0, P, F)$ , donde:

$Q$  es un conjunto finito de estados,

$\Sigma$  es un conjunto finito de símbolos de entrada,

$R \subset Q \times \Sigma \times Q$  es un conjunto de transiciones de la forma  $(q, c, q')$  para  $c \in \Sigma$  y  $q, q' \in Q$ ,

$q_0$  es el estado inicial ( $q_0 \in Q$ ),

$P$  función de probabilidad de transición  $P : R \rightarrow \mathbb{R}^+$ ,

$F$  función de probabilidad de estado final  $F : Q \rightarrow \mathbb{R}^+$ .

Debe cumplirse además:

$$F(q) + \sum_{\forall (p,c,p'): p=q} P(p, c, p') = 1 \quad \forall q \in Q$$

# Autómata Estocástico de Estados Finitos (AEEF)

Constituyen la base de los HMMs, e inclusive son utilizados para representar los modelos de léxico y los modelos de lenguaje de  $N$ -gramas.

Un AEEF  $\mathcal{T}$ , es una máquina de estados finitos definida por la séxtupla  $(Q, \Sigma, R, q_0, P, F)$ , donde:

$Q$  es un conjunto finito de estados,

$\Sigma$  es un conjunto finito de símbolos de entrada,

$R \subset Q \times \Sigma \times Q$  es un conjunto de transiciones de la forma  $(q, c, q')$  para  $c \in \Sigma$  y  $q, q' \in Q$ ,

$q_0$  es el estado inicial ( $q_0 \in Q$ ),

$P$  función de probabilidad de transición  $P : R \rightarrow \mathbb{R}^+$ ,

$F$  función de probabilidad de estado final  $F : Q \rightarrow \mathbb{R}^+$ .

Debe cumplirse además:

$$F(q) + \sum_{\forall(p,c,p'):p=q} P(p, c, p') = 1 \quad \forall q \in Q$$

# Autómata Estocástico de Estados Finitos (AEEF)

Constituyen la base de los HMMs, e inclusive son utilizados para representar los modelos de léxico y los modelos de lenguaje de  $N$ -gramas.

Un AEEF  $\mathcal{T}$ , es una máquina de estados finitos definida por la séxtupla  $(Q, \Sigma, R, q_0, P, F)$ , donde:

$Q$  es un conjunto finito de estados,

$\Sigma$  es un conjunto finito de símbolos de entrada,

$R \subset Q \times \Sigma \times Q$  es un conjunto de transiciones de la forma  $(q, c, q')$  para  $c \in \Sigma$  y  $q, q' \in Q$ ,

$q_0$  es el estado inicial ( $q_0 \in Q$ ),

$P$  función de probabilidad de transición  $P : R \rightarrow \mathbb{R}^+$ ,

$F$  función de probabilidad de estado final  $F : Q \rightarrow \mathbb{R}^+$ .

Debe cumplirse además:

$$F(q) + \sum_{\forall(p,c,p'):p=q} P(p, c, p') = 1 \quad \forall q \in Q$$

# Autómata Estocástico de Estados Finitos (AEEF)

Constituyen la base de los HMMs, e inclusive son utilizados para representar los modelos de léxico y los modelos de lenguaje de  $N$ -gramas.

Un AEEF  $\mathcal{T}$ , es una máquina de estados finitos definida por la séxtupla  $(Q, \Sigma, R, q_0, P, F)$ , donde:

$Q$  es un conjunto finito de estados,

$\Sigma$  es un conjunto finito de símbolos de entrada,

$R \subset Q \times \Sigma \times Q$  es un conjunto de transiciones de la forma  $(q, c, q')$  para  $c \in \Sigma$  y  $q, q' \in Q$ ,

$q_0$  es el estado inicial ( $q_0 \in Q$ ),

$P$  función de probabilidad de transición  $P : R \rightarrow \mathbb{R}^+$ ,

$F$  función de probabilidad de estado final  $F : Q \rightarrow \mathbb{R}^+$ .

Debe cumplirse además:

$$F(q) + \sum_{\forall(p,c,p'): p=q} P(p, c, p') = 1 \quad \forall q \in Q$$

# Temario

1

## Uso de Modelos de Lenguaje en RES

- Decodificación (Reconocimiento)
- Diferentes Niveles de Percepción
- Autómata Estocástico de Estados Finitos (AEEF)
- **Modelado a Nivel Morfológico**
- Modelado a Nivel Léxico
- Modelado a Nivel Sintáctico
- Modelo Integrado
- Modelos de Lenguaje de N-gramas
- Marco Teórico: Formulación del Problema de RES
- Mixtura de Gaussianas, GSF y WIP
- Métricas de Evaluación

2

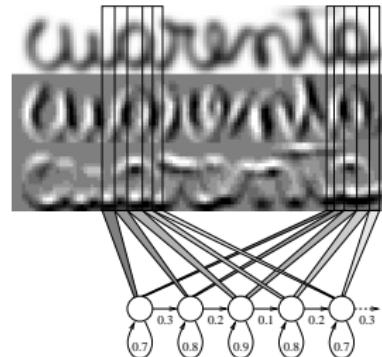
## Aplicaciones

- Aplicación RNCN
- Aplicación RMFE
- Prototipo PAITYT

# Modelado a Nivel Morfológico

La primer fuente de conocimiento (morfológico) se modelará con HMMs.

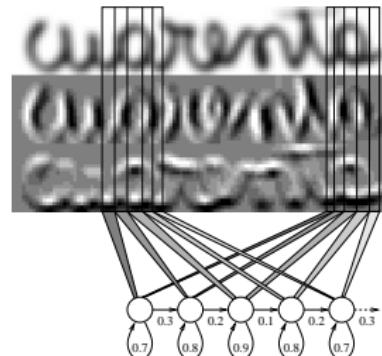
- Modelan la unidad básica de reconocimiento: clase de carácter.
- En base a una adecuada combinación de estas unidades básicas se construyen unidades de más alto nivel: palabras.
- Directamente podrían modelar palabras, si se contase con un vocabulario pequeño y un adecuado número de muestras de entrenamiento.
- Utilizar HMMs de caracteres, tiene la ventaja de utilizar como muestras de entrenamiento las letras embebidas en las palabras (o sentencias).



# Modelado a Nivel Morfológico

La primer fuente de conocimiento (morfológico) se modelará con HMMs.

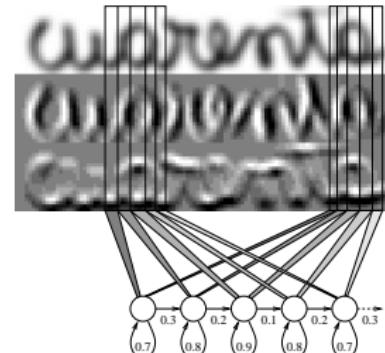
- Modelan la unidad básica de reconocimiento: clase de carácter.
- En base a una adecuada combinación de estas unidades básicas se construyen unidades de más alto nivel: palabras.
- Directamente podrían modelar palabras, si se contase con un vocabulario pequeño y un adecuado número de muestras de entrenamiento.
- Utilizar HMMs de caracteres, tiene la ventaja de utilizar como muestras de entrenamiento las letras embebidas en las palabras (o sentencias).



# Modelado a Nivel Morfológico

La primer fuente de conocimiento (morfológico) se modelará con HMMs.

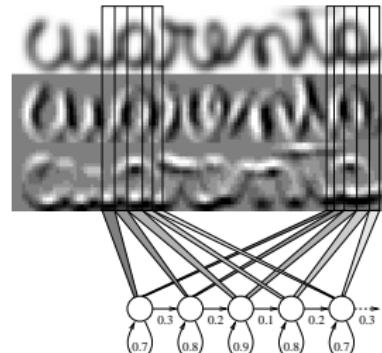
- Modelan la unidad básica de reconocimiento: clase de carácter.
- En base a una adecuada combinación de estas unidades básicas se construyen unidades de más alto nivel: palabras.
- Directamente podrían modelar palabras, si se contase con un vocabulario pequeño y un adecuado número de muestras de entrenamiento.
- Utilizar HMMs de caracteres, tiene la ventaja de utilizar como muestras de entrenamiento las letras embebidas en las palabras (o sentencias).



# Modelado a Nivel Morfológico

La primer fuente de conocimiento (morfológico) se modelará con HMMs.

- Modelan la unidad básica de reconocimiento: clase de carácter.
- En base a una adecuada combinación de estas unidades básicas se construyen unidades de más alto nivel: palabras.
- Directamente podrían modelar palabras, si se contase con un vocabulario pequeño y un adecuado número de muestras de entrenamiento.
- Utilizar HMMs de caracteres, tiene la ventaja de utilizar como muestras de entrenamiento las letras embebidas en las palabras (o sentencias).



# Temario

1

## Uso de Modelos de Lenguaje en RES

- Decodificación (Reconocimiento)
- Diferentes Niveles de Percepción
- Autómata Estocástico de Estados Finitos (AEEF)
- Modelado a Nivel Morfológico

### **● Modelado a Nivel Léxico**

- Modelado a Nivel Sintáctico
- Modelo Integrado
- Modelos de Lenguaje de N-gramas
- Marco Teórico: Formulación del Problema de RES
- Mixtura de Gaussianas, GSF y WIP
- Métricas de Evaluación

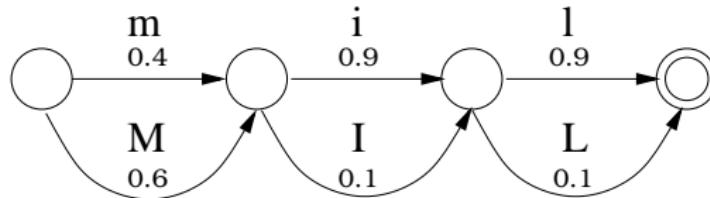
2

## Aplicaciones

- Aplicación RNCN
- Aplicación RMFE
- Prototipo PAITYT

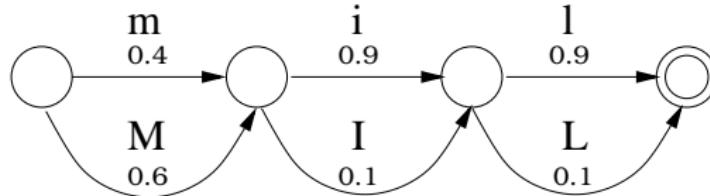
# Modelado a Nivel Léxico

- Las palabras están formadas por concatenación de caracteres.
- *Autómata estocástico de estados finitos (AEEF)* son utilizados para representar las posibles concatenaciones de caracteres individuales.
- Pueden tomar en cuenta todos las posibles formas de escribir una palabra: mayúsculas, minúsculas, con acentos o sin ellos, etc.
- Es posible modelar el espacio inter-palabra a este nivel, agregando al final un HMM de espacio en blanco.



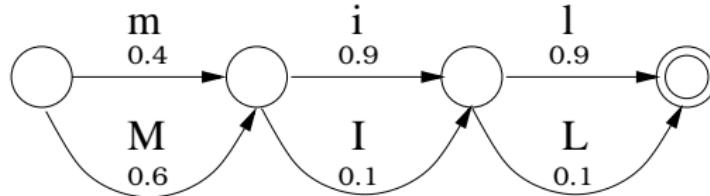
# Modelado a Nivel Léxico

- Las palabras están formadas por concatenación de caracteres.
- *Autómata estocástico de estados finitos* (AEEF) son utilizados para representar las posibles concatenaciones de caracteres individuales.
- Pueden tomar en cuenta todos las posibles formas de escribir una palabra: mayúsculas, minúsculas, con acentos o sin ellos, etc.
- Es posible modelar el espacio inter-palabra a este nivel, agregando al final un HMM de espacio en blanco.



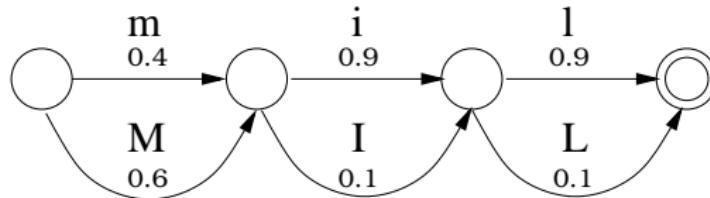
# Modelado a Nivel Léxico

- Las palabras están formadas por concatenación de caracteres.
- *Autómata estocástico de estados finitos* (AEEF) son utilizados para representar las posibles concatenaciones de caracteres individuales.
- Pueden tomar en cuenta todos las posibles formas de escribir una palabra: mayúsculas, minúsculas, con acentos o sin ellos, etc.
- Es posible modelar el espacio inter-palabra a este nivel, agregando al final un HMM de espacio en blanco.



# Modelado a Nivel Léxico

- Las palabras están formadas por concatenación de caracteres.
- *Autómata estocástico de estados finitos* (AEEF) son utilizados para representar las posibles concatenaciones de caracteres individuales.
- Pueden tomar en cuenta todos las posibles formas de escribir una palabra: mayúsculas, minúsculas, con acentos o sin ellos, etc.
- Es posible modelar el espacio inter-palabra a este nivel, agregando al final un HMM de espacio en blanco.



# Temario

1

## Uso de Modelos de Lenguaje en RES

- Decodificación (Reconocimiento)
- Diferentes Niveles de Percepción
- Autómata Estocástico de Estados Finitos (AEEF)
- Modelado a Nivel Morfológico
- Modelado a Nivel Léxico
- Modelado a Nivel Sintáctico**
- Modelo Integrado
- Modelos de Lenguaje de N-gramas
- Marco Teórico: Formulación del Problema de RES
- Mixtura de Gaussianas, GSF y WIP
- Métricas de Evaluación

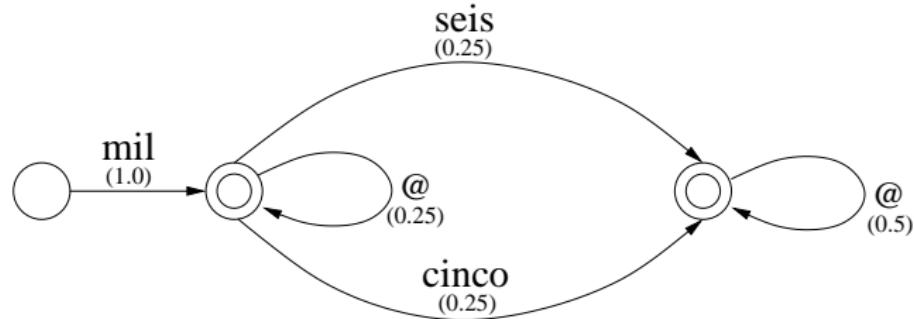
2

## Aplicaciones

- Aplicación RNCN
- Aplicación RMFE
- Prototipo PAITYT

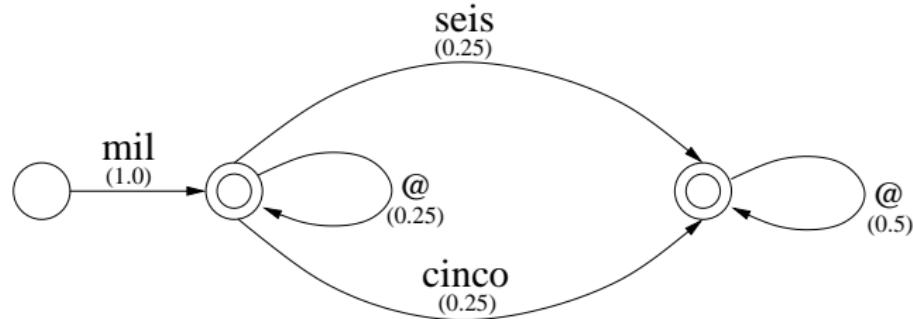
# Modelado a Nivel Sintáctico

- Las sentencias están formadas por concatenación de palabras.
- Se emplean modelos de lenguaje (ML) para modelar la concatenación correcta de palabras: “contexto”.
- Interesan aquellos lenguajes que puedan ser representados por un AEEF. Por ejemplo: modelos de lenguaje regulares, *N*-gramas, k-testables, etc.
- El espacio inter-palabra también puede ser modelado a este nivel.



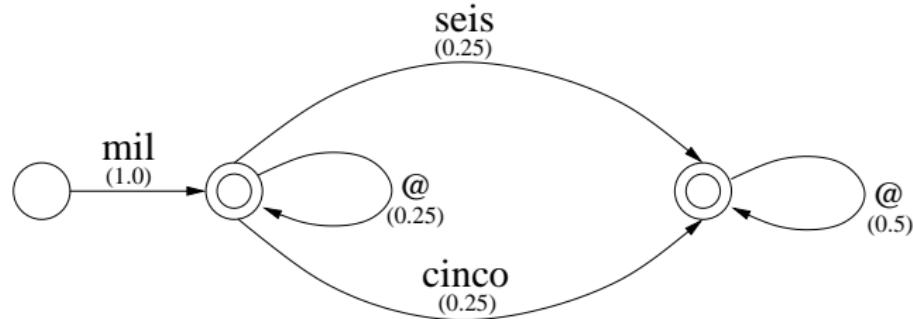
# Modelado a Nivel Sintáctico

- Las sentencias están formadas por concatenación de palabras.
- Se emplean modelos de lenguaje (ML) para modelar la concatenación correcta de palabras: “contexto”.
- Interesan aquellos lenguajes que puedan ser representados por un AEEF. Por ejemplo: modelos de lenguaje regulares,  $N$ -gramas,  $k$ -testables, etc.
- El espacio inter-palabra también puede ser modelado a este nivel.



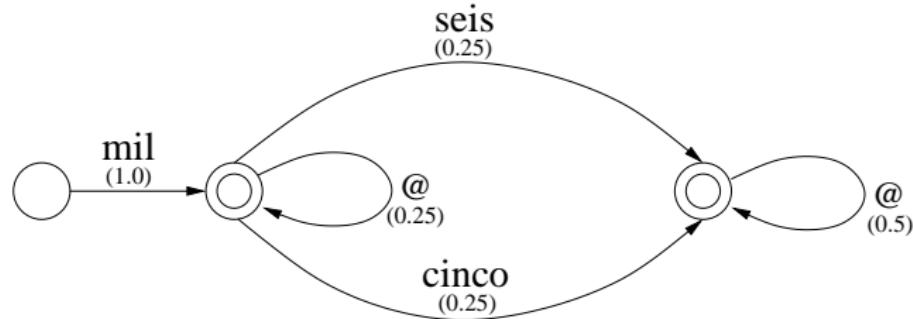
# Modelado a Nivel Sintáctico

- Las sentencias están formadas por concatenación de palabras.
- Se emplean modelos de lenguaje (ML) para modelar la concatenación correcta de palabras: “contexto”.
- Interesan aquellos lenguajes que puedan ser representados por un AEEF. Por ejemplo: modelos de lenguaje regulares, *N*-gramas, k-testables, etc.
- El espacio inter-palabra también puede ser modelado a este nivel.



# Modelado a Nivel Sintáctico

- Las sentencias están formadas por concatenación de palabras.
- Se emplean modelos de lenguaje (ML) para modelar la concatenación correcta de palabras: “contexto”.
- Interesan aquellos lenguajes que puedan ser representados por un AEEF. Por ejemplo: modelos de lenguaje regulares,  $N$ -gramas,  $k$ -testables, etc.
- El espacio inter-palabra también puede ser modelado a este nivel.



# Temario

1

## Uso de Modelos de Lenguaje en RES

- Decodificación (Reconocimiento)
- Diferentes Niveles de Percepción
- Autómata Estocástico de Estados Finitos (AEEF)
- Modelado a Nivel Morfológico
- Modelado a Nivel Léxico
- Modelado a Nivel Sintáctico
- **Modelo Integrado**
- Modelos de Lenguaje de N-gramas
- Marco Teórico: Formulación del Problema de RES
- Mixtura de Gaussianas, GSF y WIP
- Métricas de Evaluación

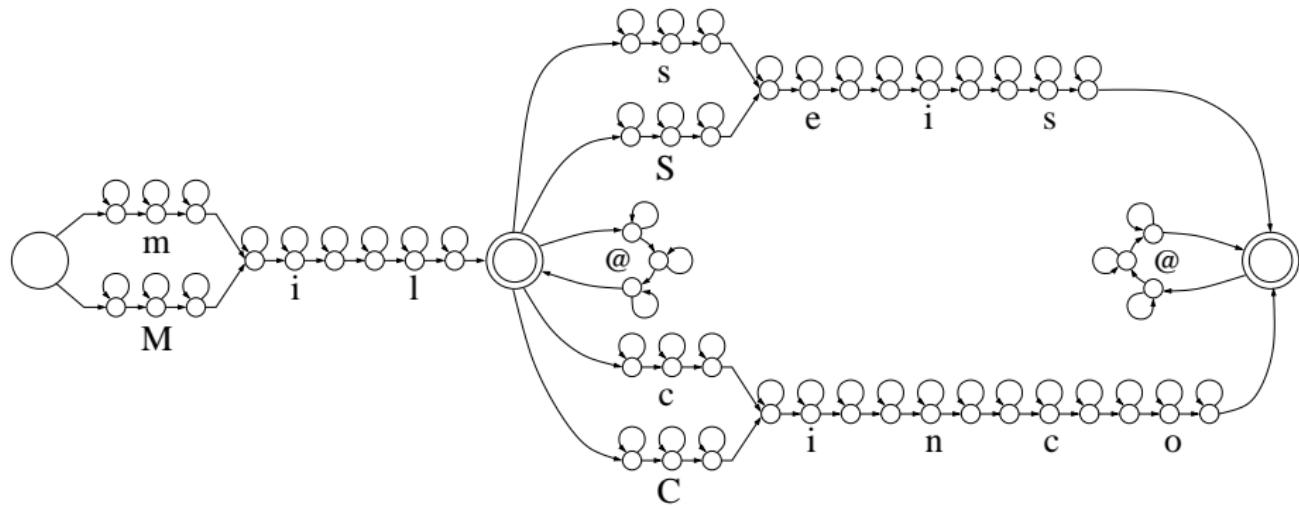
2

## Aplicaciones

- Aplicación RNCN
- Aplicación RMFE
- Prototipo PAITYT

# Modelo Integrado

Modelo global formado por ramas de series concatenadas de HMMs de caracteres, que acepta como entrada secuencias de vectores de características y produce en la salida secuencias de palabras (según se vea) reconocidas.



# Temario

1

## Uso de Modelos de Lenguaje en RES

- Decodificación (Reconocimiento)
- Diferentes Niveles de Percepción
- Autómata Estocástico de Estados Finitos (AEEF)
- Modelado a Nivel Morfológico
- Modelado a Nivel Léxico
- Modelado a Nivel Sintáctico
- Modelo Integrado

## ● Modelos de Lenguaje de N-gramas

- Marco Teórico: Formulación del Problema de RES
- Mixtura de Gaussianas, GSF y WIP
- Métricas de Evaluación

2

## Aplicaciones

- Aplicación RNCN
- Aplicación RMFE
- Prototipo PAITYT

# Modelos de Lenguaje

Los modelos de lenguaje (ML):

- Reflejan un conocimiento previo sobre la estructura del discurso o de qué es lo más probable que aparezca en un contexto determinado.
- Se usan como restricciones para reducir el espacio de búsqueda de la mejor secuencia de palabras que corresponden a una secuencia de vectores de características.
- Asignan una probabilidad a toda posible secuencia de palabras.

La probabilidad de observar una secuencia de  $\mathbf{w} = \langle w_1 \dots w_m \rangle$  de un vocabulario conocido  $\Sigma$  puede expresarse como:

$$P(\mathbf{w}) = P(w_1) \cdot \prod_{i=2}^m P(w_i | w_1 \dots w_{i-1})$$

donde  $P(w_i | w_1 \dots w_{i-1})$  es la probabilidad que, habiendo visto la *historia*  $w_1 \dots w_{i-1}$  de palabras, aparezca a continuación  $w_i$ .

La estimación de  $P(\mathbf{w})$  se hace impracticable cuando  $|\Sigma|$  es grande:  $|\Sigma|^{i-1}$  posibles historias.

# Modelos de Lenguaje

Los modelos de lenguaje (ML):

- Reflejan un conocimiento previo sobre la estructura del discurso o de qué es lo más probable que aparezca en un contexto determinado.
- Se usan como restricciones para reducir el espacio de búsqueda de la mejor secuencia de palabras que corresponden a una secuencia de vectores de características.
- Asignan una probabilidad a toda posible secuencia de palabras.

La probabilidad de observar una secuencia de  $\mathbf{w} = \langle w_1 \dots w_m \rangle$  de un vocabulario conocido  $\Sigma$  puede expresarse como:

$$P(\mathbf{w}) = P(w_1) \cdot \prod_{i=2}^m P(w_i | w_1 \dots w_{i-1})$$

donde  $P(w_i | w_1 \dots w_{i-1})$  es la probabilidad que, habiendo visto la *historia*  $w_1 \dots w_{i-1}$  de palabras, aparezca a continuación  $w_i$ .

La estimación de  $P(\mathbf{w})$  se hace impracticable cuando  $|\Sigma|$  es grande:  $|\Sigma|^{i-1}$  posibles historias.

# Modelos de Lenguaje

Los modelos de lenguaje (ML):

- Reflejan un conocimiento previo sobre la estructura del discurso o de qué es lo más probable que aparezca en un contexto determinado.
- Se usan como restricciones para reducir el espacio de búsqueda de la mejor secuencia de palabras que corresponden a una secuencia de vectores de características.
- Asignan una probabilidad a toda posible secuencia de palabras.

La probabilidad de observar una secuencia de  $\mathbf{w} = \langle w_1 \dots w_m \rangle$  de un vocabulario conocido  $\Sigma$  puede expresarse como:

$$P(\mathbf{w}) = P(w_1) \cdot \prod_{i=2}^m P(w_i | w_1 \dots w_{i-1})$$

donde  $P(w_i | w_1 \dots w_{i-1})$  es la probabilidad que, habiendo visto la *historia*  $w_1 \dots w_{i-1}$  de palabras, aparezca a continuación  $w_i$ .

La estimación de  $P(\mathbf{w})$  se hace impracticable cuando  $|\Sigma|$  es grande:  $|\Sigma|^{i-1}$  posibles historias.

# Modelos de Lenguaje

Los modelos de lenguaje (ML):

- Reflejan un conocimiento previo sobre la estructura del discurso o de qué es lo más probable que aparezca en un contexto determinado.
- Se usan como restricciones para reducir el espacio de búsqueda de la mejor secuencia de palabras que corresponden a una secuencia de vectores de características.
- Asignan una probabilidad a toda posible secuencia de palabras.

La probabilidad de observar una secuencia de  $\mathbf{w} = \langle w_1 \dots w_m \rangle$  de un vocabulario conocido  $\Sigma$  puede expresarse como:

$$P(\mathbf{w}) = P(w_1) \cdot \prod_{i=2}^m P(w_i | w_1 \dots w_{i-1})$$

donde  $P(w_i | w_1 \dots w_{i-1})$  es la probabilidad que, habiendo visto la *historia*  $w_1 \dots w_{i-1}$  de palabras, aparezca a continuación  $w_i$ .

La estimación de  $P(\mathbf{w})$  se hace impracticable cuando  $|\Sigma|$  es grande:  $|\Sigma|^{i-1}$  posibles historias.

# Modelos de Lenguaje

Los modelos de lenguaje (ML):

- Reflejan un conocimiento previo sobre la estructura del discurso o de qué es lo más probable que aparezca en un contexto determinado.
- Se usan como restricciones para reducir el espacio de búsqueda de la mejor secuencia de palabras que corresponden a una secuencia de vectores de características.
- Asignan una probabilidad a toda posible secuencia de palabras.

La probabilidad de observar una secuencia de  $\mathbf{w} = \langle w_1 \dots w_m \rangle$  de un vocabulario conocido  $\Sigma$  puede expresarse como:

$$P(\mathbf{w}) = P(w_1) \cdot \prod_{i=2}^m P(w_i | w_1 \dots w_{i-1})$$

donde  $P(w_i | w_1 \dots w_{i-1})$  es la probabilidad que, habiendo visto la *historia*  $w_1 \dots w_{i-1}$  de palabras, aparezca a continuación  $w_i$ .

La estimación de  $P(\mathbf{w})$  se hace impracticable cuando  $|\Sigma|$  es grande:  $|\Sigma|^{i-1}$  posibles historias.

# Modelos de $N$ -gramas - Definición

Se define una función  $\Phi_n : \Sigma^* \rightarrow \Sigma^{n-1}$  que clasifica en una misma clase de equivalencia todas aquellas cadenas que terminan con las mismas  $n-1$  palabras.  $P(\mathbf{w})$  puede aproximarse como:

$$P(\mathbf{w}) \approx \prod_{i=1}^m P(w_i | \Phi_n(w_1 \dots w_{i-1})) = \prod_{i=1}^m P(w_i | w_{i-n+1} \dots w_{i-1})$$

En las primeras  $n$  palabras de la cadena sucede que  $i-n \leq 0$ :

$$P(\mathbf{w}) \approx P(w_1) \cdot \prod_{i=1}^{n-1} P(w_i | w_1 \dots w_{i-1}) \cdot \prod_{i=n}^m P(w_i | w_{i-n+1} \dots w_{i-1})$$

La estimación de la probabilidad, que se dé una palabra  $w_i$ , habiendo ocurrido una historia determinada  $w_{i-n+1} \dots w_{i-1}$ , se calcula mediante la frecuencia relativa  $f(|)$ :

$$P(w_i | w_{i-n+1} \dots w_{i-1}) = f(w_i | w_{i-n+1} \dots w_{i-1}) = \frac{C(w_{i-n+1} \dots w_{i-1} w_i)}{C(w_{i-n+1} \dots w_{i-1})}$$

Se suelen utilizar valores de  $n$  reducidos (1,2 ó 3).

# Modelos de $N$ -gramas - Definición

Se define una función  $\Phi_n : \Sigma^* \rightarrow \Sigma^{n-1}$  que clasifica en una misma clase de equivalencia todas aquellas cadenas que terminan con las mismas  $n-1$  palabras.  $P(\mathbf{w})$  puede aproximarse como:

$$P(\mathbf{w}) \approx \prod_{i=1}^m P(w_i | \Phi_n(w_1 \dots w_{i-1})) = \prod_{i=1}^m P(w_i | w_{i-n+1} \dots w_{i-1})$$

En las primeras  $n$  palabras de la cadena sucede que  $i-n \leq 0$ :

$$P(\mathbf{w}) \approx P(w_1) \cdot \prod_{i=1}^{n-1} P(w_i | w_1 \dots w_{i-1}) \cdot \prod_{i=n}^m P(w_i | w_{i-n+1} \dots w_{i-1})$$

La estimación de la probabilidad, que se dé una palabra  $w_i$ , habiendo ocurrido una historia determinada  $w_{i-n+1} \dots w_{i-1}$ , se calcula mediante la frecuencia relativa  $f(|)$ :

$$P(w_i | w_{i-n+1} \dots w_{i-1}) = f(w_i | w_{i-n+1} \dots w_{i-1}) = \frac{C(w_{i-n+1} \dots w_{i-1} w_i)}{C(w_{i-n+1} \dots w_{i-1})}$$

Se suelen utilizar valores de  $n$  reducidos (1,2 ó 3).

# Modelos de $N$ -gramas - Definición

Se define una función  $\Phi_n : \Sigma^* \rightarrow \Sigma^{n-1}$  que clasifica en una misma clase de equivalencia todas aquellas cadenas que terminan con las mismas  $n-1$  palabras.  $P(\mathbf{w})$  puede aproximarse como:

$$P(\mathbf{w}) \approx \prod_{i=1}^m P(w_i | \Phi_n(w_1 \dots w_{i-1})) = \prod_{i=1}^m P(w_i | w_{i-n+1} \dots w_{i-1})$$

En las primeras  $n$  palabras de la cadena sucede que  $i-n \leq 0$ :

$$P(\mathbf{w}) \approx P(w_1) \cdot \prod_{i=1}^{n-1} P(w_i | w_1 \dots w_{i-1}) \cdot \prod_{i=n}^m P(w_i | w_{i-n+1} \dots w_{i-1})$$

La estimación de la probabilidad, que se dé una palabra  $w_i$ , habiendo ocurrido una historia determinada  $w_{i-n+1} \dots w_{i-1}$ , se calcula mediante la frecuencia relativa  $f(|)$ :

$$P(w_i | w_{i-n+1} \dots w_{i-1}) = f(w_i | w_{i-n+1} \dots w_{i-1}) = \frac{C(w_{i-n+1} \dots w_{i-1} w_i)}{C(w_{i-n+1} \dots w_{i-1})}$$

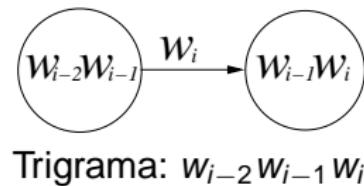
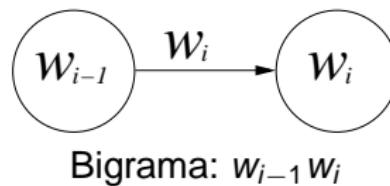
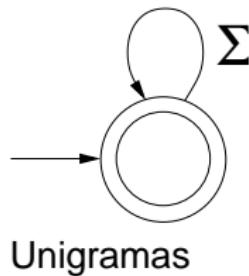
Se suelen utilizar valores de  $n$  reducidos (1,2 ó 3).

# Modelos de $N$ -gramas - Representación

Los  $N$ -gramas pueden ser representados por medio de AEEFs deterministas.

Un AEEF es determinista si dados  $(q, w_i, q'), (q, w_j, q') \in R$ , se cumple entonces que:  $w_i = w_j$ .

Ejemplos de AEEFs para representar 1-gramas, 2-gramas y 3-gramas respectivamente:



Para bigramas:  $q = w_{i-1}$  y  $q' = w_i$ .

Para trigramas:  $q = w_{i-2}w_{i-1}$  y  $q' = w_{i-1}w_i$ .

# Modelos de $N$ -gramas - Estimación y Suavizado

- Estimación se realiza mediante el cómputo de las frecuencias relativas:

$$P(w_i | w_{i-n+1} \dots w_{i-1}) = \frac{C(qw_i)}{C(q)} = \frac{C(w_{i-n+1} \dots w_{i-1} w_i)}{C(w_{i-n+1} \dots w_{i-1})}$$

- Modelo de  $N$ -gramas es completo si contempla todos los eventos posibles con una adecuada estimación de los parámetros correspondientes.
- **Problema:** eventos que no aparecen un número suficiente, o nada, de veces en el entrenamiento.

## Métodos de Suavizado

- Back-Off
- Interpolación

## Métodos de Descuento

- Good Turing
- Descuento Absoluto
- Witten Bell
- Lineal
- etc.

# Modelos de $N$ -gramas - Estimación y Suavizado

- Estimación se realiza mediante el cómputo de las frecuencias relativas:

$$P(w_i | w_{i-n+1} \dots w_{i-1}) = \frac{C(qw_i)}{C(q)} = \frac{C(w_{i-n+1} \dots w_{i-1} w_i)}{C(w_{i-n+1} \dots w_{i-1})}$$

- Modelo de  $N$ -gramas es completo si contempla todos los eventos posibles con una adecuada estimación de los parámetros correspondientes.
- **Problema:** eventos que no aparecen un número suficiente, o nada, de veces en el entrenamiento.

## Métodos de Suavizado

- Back-Off
- Interpolación

## Métodos de Descuento

- Good Turing
- Descuento Absoluto
- Witten Bell
- Lineal
- etc.

# Modelos de $N$ -gramas - Estimación y Suavizado

- Estimación se realiza mediante el cómputo de las frecuencias relativas:

$$P(w_i | w_{i-n+1} \dots w_{i-1}) = \frac{C(qw_i)}{C(q)} = \frac{C(w_{i-n+1} \dots w_{i-1} w_i)}{C(w_{i-n+1} \dots w_{i-1})}$$

- Modelo de  $N$ -gramas es completo si contempla todos los eventos posibles con una adecuada estimación de los parámetros correspondientes.
- **Problema:** eventos que no aparecen un número suficiente, o nada, de veces en el entrenamiento.

## Métodos de Suavizado

- Back-Off
- Interpolación

## Métodos de Descuento

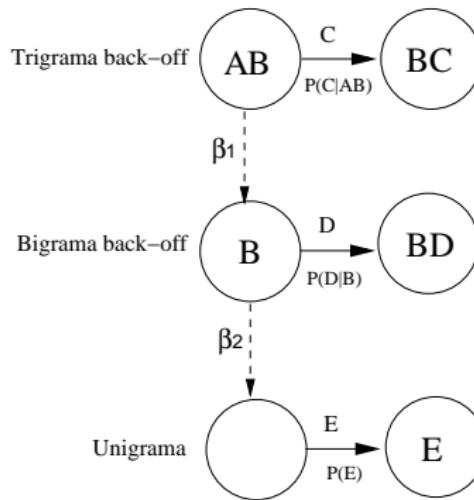
- Good Turing
- Descuento Absoluto
- Witten Bell
- Lineal
- etc.

# Modelos de $N$ -gramas - Suavizado

$$P(w_i | w_{i-n+1}^{i-1}) = \begin{cases} P^*(w_i | w_{i-n+1}^{i-1}) & \text{si } C(w_{i-n+1}^i) > 0 \\ \beta(w_{i-n+2}^i) P^*(w_i | w_{i-n+2}^{i-1}) & \text{caso contrario} \end{cases}$$

donde

- $P^*(.)$  es probabilidad con descuento.
- $\beta$  asegura la normalización de la probabilidad del modelo.



# Temario

1

## Uso de Modelos de Lenguaje en RES

- Decodificación (Reconocimiento)
- Diferentes Niveles de Percepción
- Autómata Estocástico de Estados Finitos (AEEF)
- Modelado a Nivel Morfológico
- Modelado a Nivel Léxico
- Modelado a Nivel Sintáctico
- Modelo Integrado
- Modelos de Lenguaje de N-gramas
- **Marco Teórico: Formulación del Problema de RES**
- Mixtura de Gaussianas, GSF y WIP
- Métricas de Evaluación

2

## Aplicaciones

- Aplicación RNCN
- Aplicación RMFE
- Prototipo PAITYT

# Marco Teórico: Formulación del Problema de RES

El problema de encontrar la secuencia más probable de palabras  $\hat{\mathbf{w}} = \langle w_1, w_2, \dots, w_n \rangle$  para una dada imagen representada por una  $\mathbf{x} = x_1^p = \langle x_1, x_2, \dots, x_p \rangle$ :

$$\begin{aligned}\hat{\mathbf{w}} &= \arg \max_{\mathbf{w}} Pr(\mathbf{w}|\mathbf{x}) \\ &= \arg \max_{\mathbf{w}} P(\mathbf{x}|\mathbf{w}) \cdot P(\mathbf{w})\end{aligned}$$

$P(\mathbf{x}|\mathbf{w})$  se approxima utilizando HMMs de caracteres, mientras que  $P(\mathbf{w})$  utilizando algún modelo de lenguaje.

$$\hat{\mathbf{w}} \approx \arg \max_{\mathbf{w}} P_{\text{HMM}}(\mathbf{x}|\mathbf{w}) \cdot P_{\text{ML}}(\mathbf{w})$$

# Marco Teórico: Formulación del Problema de RES

El problema de encontrar la secuencia más probable de palabras  $\hat{\mathbf{w}} = \langle w_1, w_2, \dots, w_n \rangle$  para una dada imagen representada por una  $\mathbf{x} = x_1^p = \langle x_1, x_2, \dots, x_p \rangle$ :

$$\begin{aligned}\hat{\mathbf{w}} &= \arg \max_{\mathbf{w}} Pr(\mathbf{w}|\mathbf{x}) \\ &= \arg \max_{\mathbf{w}} P(\mathbf{x}|\mathbf{w}) \cdot P(\mathbf{w})\end{aligned}$$

$P(\mathbf{x}|\mathbf{w})$  se aproxima utilizando HMMs de caracteres, mientras que  $P(\mathbf{w})$  utilizando algún modelo de lenguaje.

$$\hat{\mathbf{w}} \approx \arg \max_{\mathbf{w}} P_{\text{HMM}}(\mathbf{x}|\mathbf{w}) \cdot P_{\text{ML}}(\mathbf{w})$$

# Temario

1

## Uso de Modelos de Lenguaje en RES

- Decodificación (Reconocimiento)
- Diferentes Niveles de Percepción
- Autómata Estocástico de Estados Finitos (AEEF)
- Modelado a Nivel Morfológico
- Modelado a Nivel Léxico
- Modelado a Nivel Sintáctico
- Modelo Integrado
- Modelos de Lenguaje de N-gramas
- Marco Teórico: Formulación del Problema de RES
- Mixtura de Gaussianas, GSF y WIP**
- Métricas de Evaluación

2

## Aplicaciones

- Aplicación RNCN
- Aplicación RMFE
- Prototipo PAITYT

# Inconvenientes: Gaussianas de la Mixtura

- La probabilidad de emisión de un HMMs es un determinado estado:

$$p(x_t | s_t = q_j, \mathcal{M}) = b_j(x_t) = \sum_{m=1}^M c_{jm} \frac{1}{\sqrt{(2\pi)^d |\Sigma_{jm}|}} e^{-\frac{1}{2}(x_t - \mu_{jm})^T \Sigma_{jm}^{-1} (x_t - \mu_{jm})}$$

donde puede resultar que  $p(x_t | s_t = q_j, \mathcal{M}) >> 1$ . Esto sucede porque  $p(x_t | s_t = q_j, \mathcal{M})$  es una suma de densidades de probabilidad.

- Las Gaussianas se vuelven más “picudas” a medida que la varianza disminuye.
- Este efecto suele aparecer en el proceso de estimación de parámetros.
- Para suavizar este efecto se suelen poner cotas inferiores que no permiten que ninguna de ellas caiga por debajo de ese valor.

# Inconvenientes: Gaussianas de la Mixtura

- La probabilidad de emisión de un HMMs es un determinado estado:

$$p(x_t | s_t = q_j, \mathcal{M}) = b_j(x_t) = \sum_{m=1}^M c_{jm} \frac{1}{\sqrt{(2\pi)^d |\Sigma_{jm}|}} e^{-\frac{1}{2}(x_t - \mu_{jm})^T \Sigma_{jm}^{-1} (x_t - \mu_{jm})}$$

donde puede resultar que  $p(x_t | s_t = q_j, \mathcal{M}) >> 1$ . Esto sucede porque  $p(x_t | s_t = q_j, \mathcal{M})$  es una suma de densidades de probabilidad.

- Las Gaussianas se vuelven más “picudas” a medida que la varianza disminuye.
- Este efecto suele aparecer en el proceso de estimación de parámetros.
- Para suavizar este efecto se suelen poner cotas inferiores que no permiten que ninguna de ellas caiga por debajo de ese valor.

# Inconvenientes: Gaussianas de la Mixtura

- La probabilidad de emisión de un HMMs es un determinado estado:

$$p(x_t | s_t = q_j, \mathcal{M}) = b_j(x_t) = \sum_{m=1}^M c_{jm} \frac{1}{\sqrt{(2\pi)^d |\Sigma_{jm}|}} e^{-\frac{1}{2}(x_t - \mu_{jm})^T \Sigma_{jm}^{-1} (x_t - \mu_{jm})}$$

donde puede resultar que  $p(x_t | s_t = q_j, \mathcal{M}) >> 1$ . Esto sucede porque  $p(x_t | s_t = q_j, \mathcal{M})$  es una suma de densidades de probabilidad.

- Las Gaussianas se vuelven más “picudas” a medida que la varianza disminuye.
- Este efecto suele aparecer en el proceso de estimación de parámetros.
- Para suavizar este efecto se suelen poner cotas inferiores que no permiten que ninguna de ellas caiga por debajo de ese valor.

# Inconvenientes: Gaussianas de la Mixtura

- La probabilidad de emisión de un HMMs es un determinado estado:

$$p(x_t | s_t = q_j, \mathcal{M}) = b_j(x_t) = \sum_{m=1}^M c_{jm} \frac{1}{\sqrt{(2\pi)^d |\Sigma_{jm}|}} e^{-\frac{1}{2}(x_t - \mu_{jm})^T \Sigma_{jm}^{-1} (x_t - \mu_{jm})}$$

donde puede resultar que  $p(x_t | s_t = q_j, \mathcal{M}) >> 1$ . Esto sucede porque  $p(x_t | s_t = q_j, \mathcal{M})$  es una suma de densidades de probabilidad.

- Las Gaussianas se vuelven más “picudas” a medida que la varianza disminuye.
- Este efecto suele aparecer en el proceso de estimación de parámetros.
- Para suavizar este efecto se suelen poner cotas inferiores que no permiten que ninguna de ellas caiga por debajo de ese valor.

# GSF y WIP (1)

- Inconsistencia de los modelos: “scores” acústicos son mayores que uno ( $\log > 0$ ). Gaussianas con varianza pequeña.
- Conseguir un balance adecuado de los órdenes de magnitud entre  $P_{\text{HMM}}(\mathbf{x}|\mathbf{w})$  y  $P_{\text{ML}}(\mathbf{w})$ .
- Las transiciones del modelo de lenguaje son afectadas por la siguiente transformación:

$$P'(q, c, q') = \frac{[P(q, c, q')]^s}{\exp^w}$$

- s (Grammar Scale Factor) como w (Word Insertion Penalty) son establecidos en forma empírica.
- Usualmente las probabilidades del modelo de lenguaje se expresan en logaritmos (naturales), por lo que la transformación anterior queda:

$$\ln(P'(q, c, q')) = s \ln(P(q, c, q')) - w$$

# GSF y WIP (1)

- Inconsistencia de los modelos: “scores” acústicos son mayores que uno ( $\log > 0$ ). Gaussianas con varianza pequeña.
- Conseguir un balance adecuado de los órdenes de magnitud entre  $P_{\text{HMM}}(\mathbf{x}|\mathbf{w})$  y  $P_{\text{ML}}(\mathbf{w})$ .
- Las transiciones del modelo de lenguaje son afectadas por la siguiente transformación:

$$P'(q, c, q') = \frac{[P(q, c, q')]^s}{\exp^w}$$

- $s$  (Grammar Scale Factor) como  $w$  (Word Insertion Penalty) son establecidos en forma empírica.
- Usualmente las probabilidades del modelo de lenguaje se expresan en logaritmos (naturales), por lo que la transformación anterior queda:

$$\ln(P'(q, c, q')) = s \ln(P(q, c, q')) - w$$

# GSF y WIP (1)

- Inconsistencia de los modelos: “scores” acústicos son mayores que uno ( $\log > 0$ ). Gaussianas con varianza pequeña.
- Conseguir un balance adecuado de los órdenes de magnitud entre  $P_{\text{HMM}}(\mathbf{x}|\mathbf{w})$  y  $P_{\text{ML}}(\mathbf{w})$ .
- Las transiciones del modelo de lenguaje son afectadas por la siguiente transformación:

$$P'(q, c, q') = \frac{[P(q, c, q')]^s}{\exp^w}$$

- $s$  (Grammar Scale Factor) como  $w$  (Word Insertion Penalty) son establecidos en forma empírica.
- Usualmente las probabilidades del modelo de lenguaje se expresan en logaritmos (naturales), por lo que la transformación anterior queda:

$$\ln(P'(q, c, q')) = s \ln(P(q, c, q')) - w$$

# GSF y WIP (1)

- Inconsistencia de los modelos: “scores” acústicos son mayores que uno ( $\log > 0$ ). Gaussianas con varianza pequeña.
- Conseguir un balance adecuado de los órdenes de magnitud entre  $P_{\text{HMM}}(\mathbf{x}|\mathbf{w})$  y  $P_{\text{ML}}(\mathbf{w})$ .
- Las transiciones del modelo de lenguaje son afectadas por la siguiente transformación:

$$P'(q, c, q') = \frac{[P(q, c, q')]^s}{\exp^w}$$

- $s$  (Grammar Scale Factor) como  $w$  (Word Insertion Penalty) son establecidos en forma empírica.
- Usualmente las probabilidades del modelo de lenguaje se expresan en logaritmos (naturales), por lo que la transformación anterior queda:

$$\ln(P'(q, c, q')) = s \ln(P(q, c, q')) - w$$

# GSF y WIP (1)

- Inconsistencia de los modelos: “scores” acústicos son mayores que uno ( $\log > 0$ ). Gaussianas con varianza pequeña.
- Conseguir un balance adecuado de los órdenes de magnitud entre  $P_{\text{HMM}}(\mathbf{x}|\mathbf{w})$  y  $P_{\text{ML}}(\mathbf{w})$ .
- Las transiciones del modelo de lenguaje son afectadas por la siguiente transformación:

$$P'(q, c, q') = \frac{[P(q, c, q')]^s}{\exp^w}$$

- $s$  (Grammar Scale Factor) como  $w$  (Word Insertion Penalty) son establecidos en forma empírica.
- Usualmente las probabilidades del modelo de lenguaje se expresan en logaritmos (naturales), por lo que la transformación anterior queda:

$$\ln(P'(q, c, q')) = s \ln(P(q, c, q')) - w$$

# GSF y WIP (2)

- En la formulación estadística general del problema del reconocimiento manuscrito:

$$\hat{\mathbf{w}} = \arg \max_{\mathbf{w}} P_{\text{HMM}}(\mathbf{x} | \mathbf{w}) \cdot P_{\text{ML}}(\mathbf{w})$$

$P_{\text{ML}}(\mathbf{w})$ , asumiendo la utilización de  $n$ -gramas para modelar una frase de  $L$  palabras, puede ser expresada como:

$$\begin{aligned} P_{\text{ML}}(\mathbf{w}) &\approx \prod_{i=1}^L \left[ s \cdot P(w_i | w_{i-n+1}^{i-1}) - w \right] \\ &= s \cdot \prod_{i=1}^L P(w_i | w_{i-n+1}^{i-1}) - L \cdot w \end{aligned}$$

$s$  afecta al término de probabilidades de transición del modelo, mientras que  $w$  tiene efecto solo sobre el número de transiciones que se realizan.

# Temario

1

## Uso de Modelos de Lenguaje en RES

- Decodificación (Reconocimiento)
- Diferentes Niveles de Percepción
- Autómata Estocástico de Estados Finitos (AEEF)
- Modelado a Nivel Morfológico
- Modelado a Nivel Léxico
- Modelado a Nivel Sintáctico
- Modelo Integrado
- Modelos de Lenguaje de N-gramas
- Marco Teórico: Formulación del Problema de RES
- Mixtura de Gaussianas, GSF y WIP
- Métricas de Evaluación**

2

## Aplicaciones

- Aplicación RNCN
- Aplicación RMFE
- Prototipo PAITYT

# Métricas de Evaluación

Métricas utilizadas para evaluar las prestaciones de los sistemas:

**SER** : *Sentence Error Rate* se calcula contabilizando el porcentaje de no coincidencias entre las hipótesis y sus transcripciones de referencia.

**WER** : *Word Error Rate* contabiliza el número de palabras que difieren entre la hipótesis y la referencia.

Respecto al WER:

- El número de palabras de la hipótesis y la referencia pueden diferir. Se calcula programación dinámica (alineamiento entre hipt. y ref.).
- En el alineamiento de las dos frases se pueden dar cuatro casos.
  - a) **acuerdo**: La referencia y la hipótesis alineada coinciden.
  - b) **sustitución**: La referencia es alineada a una palabra diferente.
  - c) **inserción**: La hipótesis no ha podido ser alineada con la referencia.
  - d) **borrado**: Una referencia que no aparece en la hipótesis.

$$\text{WER} = \frac{n_b + n_s + n_i}{n_b + n_s + n_a}$$

# Métricas de Evaluación

Métricas utilizadas para evaluar las prestaciones de los sistemas:

**SER** : *Sentence Error Rate* se calcula contabilizando el porcentaje de no coincidencias entre las hipótesis y sus transcripciones de referencia.

**WER** : *Word Error Rate* contabiliza el número de palabras que difieren entre la hipótesis y la referencia.

Respecto al WER:

- El número de palabras de la hipótesis y la referencia pueden diferir. Se calcula programación dinámica (alineamiento entre hipt. y ref.).
- En el alineamiento de las dos frases se pueden dar cuatro casos.
  - a) **acuerdo**: La referencia y la hipótesis alineada coinciden.
  - b) **sustitución**: La referencia es alineada a una palabra diferente.
  - c) **inserción**: La hipótesis no ha podido ser alineada con la referencia.
  - d) **borrado**: Una referencia que no aparece en la hipótesis.

$$\text{WER} = \frac{n_b + n_s + n_i}{n_b + n_s + n_a}$$

# Métricas de Evaluación

Métricas utilizadas para evaluar las prestaciones de los sistemas:

**SER** : *Sentence Error Rate* se calcula contabilizando el porcentaje de no coincidencias entre las hipótesis y sus transcripciones de referencia.

**WER** : *Word Error Rate* contabiliza el número de palabras que difieren entre la hipótesis y la referencia.

Respecto al WER:

- El número de palabras de la hipótesis y la referencia pueden diferir. Se calcula programación dinámica (alineamiento entre hipt. y ref.).
- En el alineamiento de las dos frases se pueden dar cuatro casos.
  - a) **acuerdo**: La referencia y la hipótesis alineada coinciden.
  - b) **sustitución**: La referencia es alineada a una palabra diferente.
  - c) **inserción**: La hipótesis no ha podido ser alineada con la referencia.
  - d) **borrado**: Una referencia que no aparece en la hipótesis.

$$\text{WER} = \frac{n_b + n_s + n_i}{n_b + n_s + n_a}$$

# Métricas de Evaluación

Métricas utilizadas para evaluar las prestaciones de los sistemas:

**SER** : *Sentence Error Rate* se calcula contabilizando el porcentaje de no coincidencias entre las hipótesis y sus transcripciones de referencia.

**WER** : *Word Error Rate* contabiliza el número de palabras que difieren entre la hipótesis y la referencia.

Respecto al WER:

- El número de palabras de la hipótesis y la referencia pueden diferir. Se calcula programación dinámica (alineamiento entre hipt. y ref.).
- En el alineamiento de las dos frases se pueden dar cuatro casos.
  - a) **acuerdo**: La referencia y la hipótesis alineada coinciden.
  - b) **sustitución**: La referencia es alineada a una palabra diferente.
  - c) **inserción**: La hipótesis no ha podido ser alineada con la referencia.
  - d) **borrado**: Una referencia que no aparece en la hipótesis.

$$\text{WER} = \frac{n_b + n_s + n_i}{n_b + n_s + n_a}$$

# Aplicaciones

Consideramos tres aplicaciones:

- 1 Reconocimiento de nombres de cantidades numéricas en español (RNCN).
- 2 Reconocimiento de Respuestas Manuscritas extraídas de Formularios de Encuestas (RMFE).
- 3 Prototipo para Alineamiento de Imagen de Texto y Transcripción (PAITYT).

# Aplicaciones

Consideramos tres aplicaciones:

- ① Reconocimiento de nombres de cantidades numéricas en español (RNCN).
- ② Reconocimiento de Respuestas Manuscritas extraídas de Formularios de Encuestas (RMFE).
- ③ Prototipo para Alineamiento de Imagen de Texto y Transcripción (PAITYT).

# Aplicaciones

Consideramos tres aplicaciones:

- ① Reconocimiento de nombres de cantidades numéricas en español (RNCN).
- ② Reconocimiento de Respuestas Manuscritas extraídas de Formularios de Encuestas (RMFE).
- ③ Prototipo para Alineamiento de Imagen de Texto y Transcripción (PAITYT).

# Temario

1

## Uso de Modelos de Lenguaje en RES

- Decodificación (Reconocimiento)
- Diferentes Niveles de Percepción
- Autómata Estocástico de Estados Finitos (AEEF)
- Modelado a Nivel Morfológico
- Modelado a Nivel Léxico
- Modelado a Nivel Sintáctico
- Modelo Integrado
- Modelos de Lenguaje de N-gramas
- Marco Teórico: Formulación del Problema de RES
- Mixtura de Gaussianas, GSF y WIP
- Métricas de Evaluación

2

## Aplicaciones

- Aplicación RNCN
- Aplicación RMFE
- Prototipo PAITYT

# Objetivo y Adquisición

## Aplicación:

- Destinada para su uso en bancos.
- Que involucra el procesamiento y reconocimiento de nombres de cantidades numéricas escritas por clientes en los cheques.

## Adquisición de imágenes de texto manuscrito:

- Se generaron 35 hojas impresas con un listado diferente de 17 frases (nombres de cantidades).
- 35 escritores diferentes realizaron la transcripción a mano de cada hoja impresa.
- Las transcripciones se realizaron a mano, utilizando siempre el mismo bolígrafo (color negro), sobre hojas provistas con 17 renglones.
- Se transcribieron en total unas 595 ( $35 \times 17$ ) frases.
- Las hojas fueron finalmente digitalizadas utilizando un escáner “Hewlett Packard HP ScanJet 4C” a 300dpi.

# Objetivo y Adquisición

## Aplicación:

- Destinada para su uso en bancos.
- Que involucra el procesamiento y reconocimiento de nombres de cantidades numéricas escritas por clientes en los cheques.

## Adquisición de imágenes de texto manuscrito:

- Se generaron 35 hojas impresas con un listado diferente de 17 frases (nombres de cantidades).
- 35 escritores diferentes realizaron la transcripción a mano de cada hoja impresa.
- Las transcripciones se realizaron a mano, utilizando siempre el mismo bolígrafo (color negro), sobre hojas provistas con 17 renglones.
- Se transcribieron en total unas 595 ( $35 \times 17$ ) frases.
- Las hojas fueron finalmente digitalizadas utilizando un escáner “Hewlett Packard HP ScanJet 4C” a 300dpi.

# Ej. de Formulario de Recopilación de Muestras

Escritor : 19 Frases : 307 - 323

=====

307 seiscientos noventa  
308 un millon veintiseis mil veintiseis  
309 un millon dieciseis mil veintiocho  
310 treinta y ocho millones  
311 veintitres mil veintiseis  
312 setenta mil quinientos  
313 cincuenta mil diecinueve millones  
314 quinientos millones doce  
315 mil once millones cuarenta  
316 cuatro mil once millones veintiocho mil veinticuatro  
317 un millon diecinueve mil cincuenta  
318 un millon noventa mil quince  
319 mil cuatrocientos millones trescientos mil cuarenta  
320 mil veinticuatro millones dieciseis  
321 novecientos millones veintidos mil  
322 mil seiscientos diecisiete millones  
323 doce millones novecientos veinte mil

seiscientos noventa

un millon veintiseis mil veintiseis

un millon dieciseis mil veintiocho

treinta y ocho millones

veintitres mil veintiseis

setenta mil quinientos

cincuenta mil diecinueve millones

quinientos millones doce

mil once millones cuarenta

cuatro mil once millones veintiocho mil veinticuatro

un millon diecinueve mil cincuenta

un millon noventa mil quince

mil cuatrocientos millones trescientos mil cuarenta

mil veinticuatro millones dieciseis

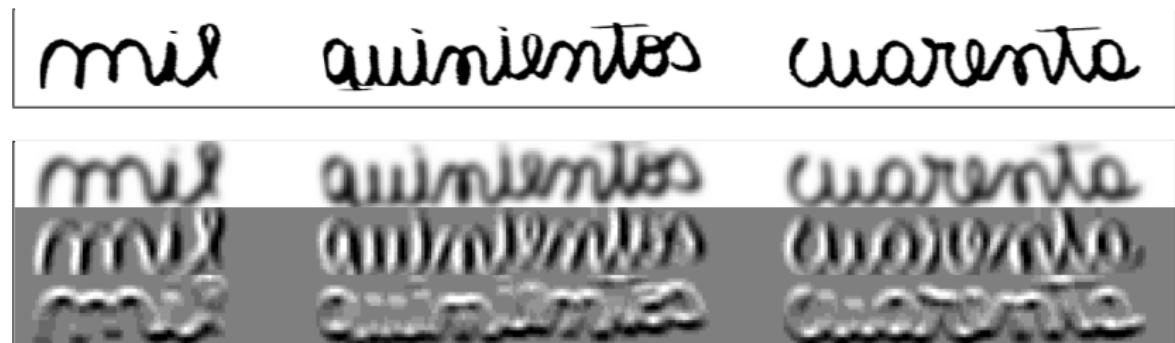
novecientos millones veintidos mil

mil seiscientos diecisiete millones

doce millones novecientos veinte mil

# Extracción de Características

En el proceso de extracción de características:



Cada vector de características se construye a partir de cada columna de celdas, por el apilamiento de las 3 tipos de características computadas.

# Modelos de Léxico y Lenguaje (1)

- Los 19 modelos HMMs (incluido el símbolo @ para representar el espacio en blanco entre palabras) son:

```
@ a c d e h i l m n o q r s t u v y z
```

- Vocabulario de 52 palabras.
- Modelo de Lenguaje representado como AEEF:
  - Acepta todos los textos de cantidades numéricas escritas en español comprendidas entre el rango  $[0, 10^{12} - 1]$ .
  - Construido a mano.
  - Topología: 32 nodos y 660 arcos de transiciones equiprobables en donde se emiten las palabras del vocabulario.
  - Este AEEF “genera” frases entre  $[0, 10^{12} - 1]$  con una distribución exponencial de frecuencias de longitud (frases cortas más probables que frases largas).

# Modelos de Léxico y Lenguaje (1)

- Los 19 modelos HMMs (incluido el símbolo @ para representar el espacio en blanco entre palabras) son:

@ a c d e h i l m n o q r s t u v y z

- Vocabulario de 52 palabras.
- Modelo de Lenguaje representado como AEEF:
  - Acepta todos los textos de cantidades numéricas escritas en español comprendidas entre el rango  $[0, 10^{12} - 1]$ .
  - Construido a mano.
  - Topología: 32 nodos y 660 arcos de transiciones equiprobables en donde se emiten las palabras del vocabulario.
  - Este AEEF “genera” frases entre  $[0, 10^{12} - 1]$  con una distribución exponencial de frecuencias de longitud (frases cortas más probables que frases largas).

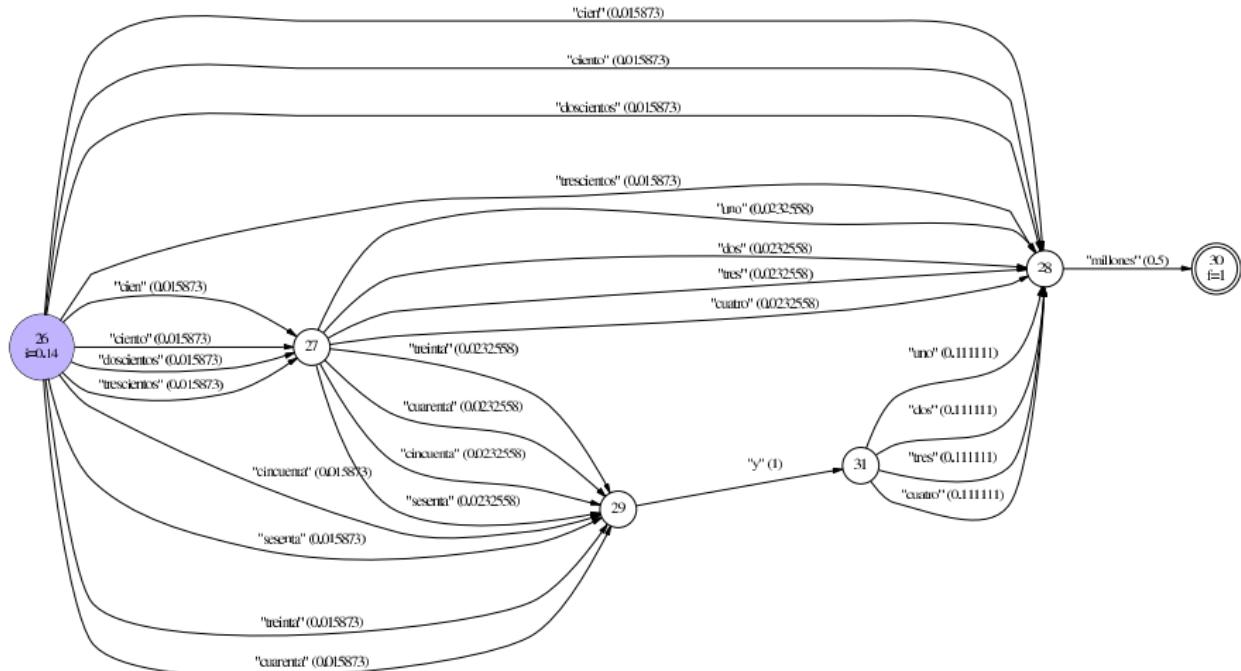
# Modelos de Léxico y Lenguaje (1)

- Los 19 modelos HMMs (incluido el símbolo @ para representar el espacio en blanco entre palabras) son:

@ a c d e h i l m n o q r s t u v y z

- Vocabulario de 52 palabras.
- Modelo de Lenguaje representado como AEEF:
  - Acepta todos los textos de cantidades numéricas escritas en español comprendidas entre el rango  $[0, 10^{12} - 1]$ .
  - Construido a mano.
  - Topología: 32 nodos y 660 arcos de transiciones equiprobables en donde se emiten las palabras del vocabulario.
  - Este AEEF “genera” frases entre  $[0, 10^{12} - 1]$  con una distribución exponencial de frecuencias de longitud (frases cortas más probables que frases largas).

# Modelos de Léxico y Lenguaje (2)



# Experimentación: Parametrización de los Modelos

- 19 clases de caracteres con 1D-HMMs izq-der: 6 estados y 10 transiciones (2 por estado).
- Densidades de probabilidad de emisión de cada estado: mixturas de Gaussianas con covarianzas diagonales.  $NG \in \{1, 2, 4, 8, 16, 32, 64\}$ .
- 52 modelos de palabras (autómatas) que contemplan todas sus posibles variaciones.
- M.L.: gramática (autómata) de frases de cantidades numéricas entre 1 y  $10^{12} - 1$ . Topología: 32 estados y 660 transic. **equiprobables**.
- *Baum-Welch "embedded training"* en el entrenamiento de los 19 modelos individuales de caracteres.
- Entrenamiento y reconocimiento con la herramienta HTK.

# Experimentación: Parametrización de los Modelos

- 19 clases de caracteres con 1D-HMMs izq-der: 6 estados y 10 transiciones (2 por estado).
- Densidades de probabilidad de emisión de cada estado: mixturas de Gaussianas con covarianzas diagonales.  $NG \in \{1, 2, 4, 8, 16, 32, 64\}$ .
- 52 modelos de palabras (autómatas) que contemplan todas sus posibles variaciones.
- M.L.: gramática (autómata) de frases de cantidades numéricas entre 1 y  $10^{12} - 1$ . Topología: 32 estados y 660 transic. **equiprobables**.
- *Baum-Welch "embedded training"* en el entrenamiento de los 19 modelos individuales de caracteres.
- Entrenamiento y reconocimiento con la herramienta HTK.

# Experimentación: Parametrización de los Modelos

- 19 clases de caracteres con 1D-HMMs izq-der: 6 estados y 10 transiciones (2 por estado).
- Densidades de probabilidad de emisión de cada estado: mixturas de Gaussianas con covarianzas diagonales.  $NG \in \{1, 2, 4, 8, 16, 32, 64\}$ .
- 52 modelos de palabras (autómatas) que contemplan todas sus posibles variaciones.
- M.L.: gramática (autómata) de frases de cantidades numéricas entre 1 y  $10^{12} - 1$ . Topología: 32 estados y 660 transic. **equiprobables**.
- *Baum-Welch "embedded training"* en el entrenamiento de los 19 modelos individuales de caracteres.
- Entrenamiento y reconocimiento con la herramienta HTK.

# Experimentación: Parametrización de los Modelos

- 19 clases de caracteres con 1D-HMMs izq-der: 6 estados y 10 transiciones (2 por estado).
- Densidades de probabilidad de emisión de cada estado: mixturas de Gaussianas con covarianzas diagonales.  $NG \in \{1, 2, 4, 8, 16, 32, 64\}$ .
- 52 modelos de palabras (autómatas) que contemplan todas sus posibles variaciones.
- M.L.: gramática (autómata) de frases de cantidades numéricas entre 1 y  $10^{12} - 1$ . Topología: 32 estados y 660 transic. **equiprobables**.
- *Baum-Welch "embedded training"* en el entrenamiento de los 19 modelos individuales de caracteres.
- Entrenamiento y reconocimiento con la herramienta HTK.

# Experimentación: Parametrización de los Modelos

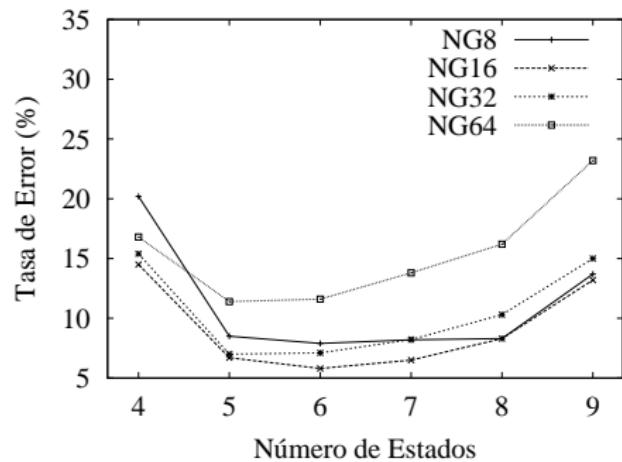
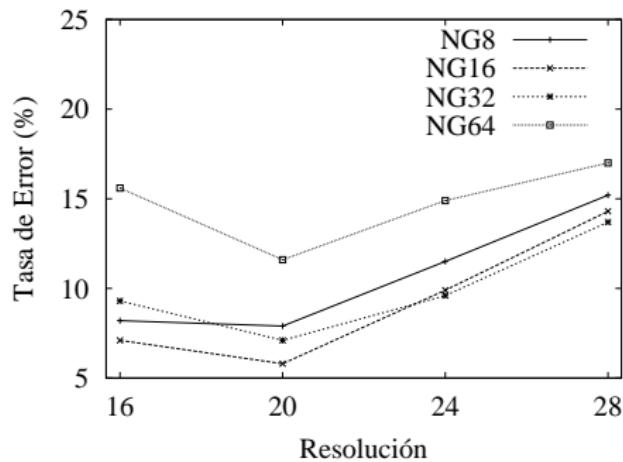
- 19 clases de caracteres con 1D-HMMs izq-der: 6 estados y 10 transiciones (2 por estado).
- Densidades de probabilidad de emisión de cada estado: mixturas de Gaussianas con covarianzas diagonales.  $NG \in \{1, 2, 4, 8, 16, 32, 64\}$ .
- 52 modelos de palabras (autómatas) que contemplan todas sus posibles variaciones.
- M.L.: gramática (autómata) de frases de cantidades numéricas entre 1 y  $10^{12} - 1$ . Topología: 32 estados y 660 transic. **equiprobables**.
- *Baum-Welch “embedded training”* en el entrenamiento de los 19 modelos individuales de caracteres.
- Entrenamiento y reconocimiento con la herramienta HTK.

# Experimentación: Parametrización de los Modelos

- 19 clases de caracteres con 1D-HMMs izq-der: 6 estados y 10 transiciones (2 por estado).
- Densidades de probabilidad de emisión de cada estado: mixturas de Gaussianas con covarianzas diagonales.  $NG \in \{1, 2, 4, 8, 16, 32, 64\}$ .
- 52 modelos de palabras (autómatas) que contemplan todas sus posibles variaciones.
- M.L.: gramática (autómata) de frases de cantidades numéricas entre 1 y  $10^{12} - 1$ . Topología: 32 estados y 660 transic. **equiprobables**.
- *Baum-Welch “embedded training”* en el entrenamiento de los 19 modelos individuales de caracteres.
- Entrenamiento y reconocimiento con la herramienta HTK.

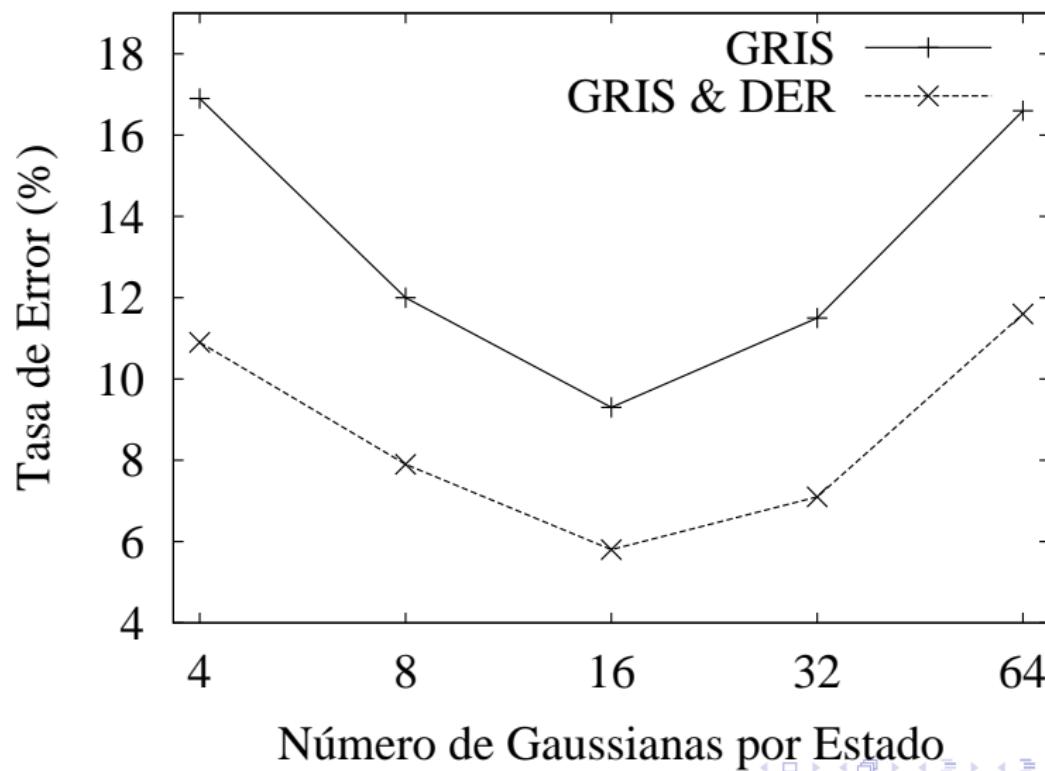
# Resultados: Resolución y Número de Estados

Viterbi “beam-search” se utilizó en el proceso de reconocimiento sobre el modelo integrado global.



Resol.	# Gauss.	# Estados	T.E. (%)
1/20	16	6	5.8

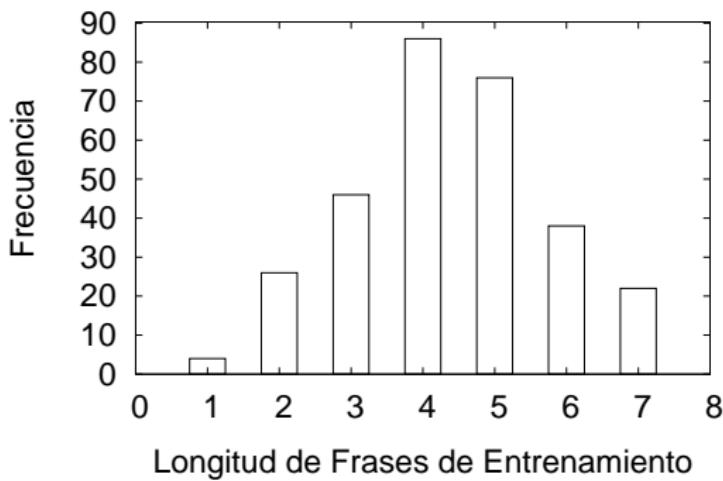
# Resultados: Extracción de Características



# Resultados: Cambios en la topología del AEEF del ML

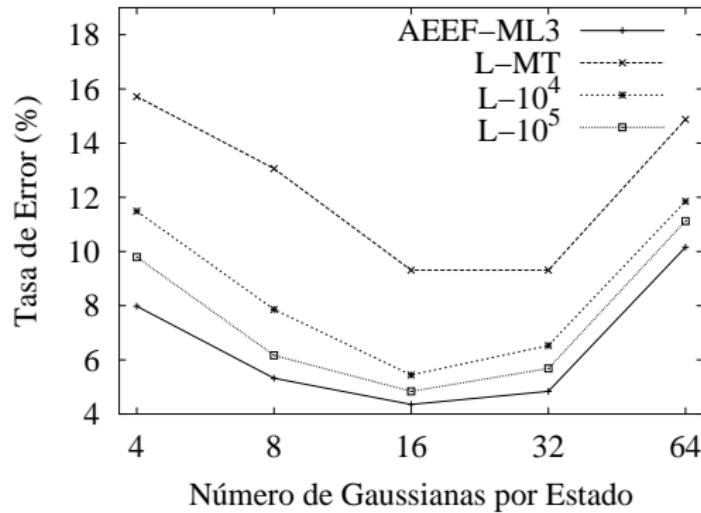
**ML0:** transiciones equiprobables, modelado equiprobable de espacios en blanco entre palabras y al final de las frases.

**ML3:** “fuerza” la ocurrencia de espacios en blanco entre palabras. Probabilidades de transición en función de la distribución de longitudes de las frases de entrenamiento.



Topol. AEEF	WER (%)
ML0	5.8
ML3	4.4

# Resultados: Utilizando $n$ -gramas como ML



**L-MT:** Lista de las 298 transcripciones de frases de entrenamiento.

**L-10<sup>4</sup>:** Lista de 10000 nombres de cantidades numéricas entre  $[1, 10^{12})$  generadas aleatoriamente con la misma distribución de frecuencias de longitud de L-MT.

**L-10<sup>5</sup>:** Lista de 100000 nombres de cantidades numéricas  $[1, 10^{12})$  generadas aleatoriamente con la misma distribución de frecuencias de longitud de L-MT.

# Temario

1

## Uso de Modelos de Lenguaje en RES

- Decodificación (Reconocimiento)
- Diferentes Niveles de Percepción
- Autómata Estocástico de Estados Finitos (AEEF)
- Modelado a Nivel Morfológico
- Modelado a Nivel Léxico
- Modelado a Nivel Sintáctico
- Modelo Integrado
- Modelos de Lenguaje de N-gramas
- Marco Teórico: Formulación del Problema de RES
- Mixtura de Gaussianas, GSF y WIP
- Métricas de Evaluación

2

## Aplicaciones

- Aplicación RNCN
- **Aplicación RMFE**
- Prototipo PAITYT

# Objetivo y Características

## Objetivo

*La clasificación de párrafos manuscritos y espontáneos extraídos de encuestas, realizadas por un compañía de telecomunicaciones.*

## Características de la Aplicación

- *Clasificación manual de respuestas manuscritas extraídas de los formularios.*
- *Lectura rápida de las respuestas; idea esencial para proceder a su clasificación en un número reducido de clases predefinidas.*

Implementación de un sistema que realice esta tarea de clasificación tan rápido como sea posible y con un mínimo de intervención humana.

# Objetivo y Características

## Objetivo

*La clasificación de párrafos manuscritos y espontáneos extraídos de encuestas, realizadas por un compañía de telecomunicaciones.*

## Características de la Aplicación

- *Clasificación manual de respuestas manuscritas extraídas de los formularios.*
- *Lectura rápida de las respuestas; idea esencial para proceder a su clasificación en un número reducido de clases predefinidas.*

Implementación de un sistema que realice esta tarea de clasificación tan rápido como sea posible y con un mínimo de intervención humana.

# Ejemplo de Formulario

**1** ¿A través de qué medios le informa habitualmente Telefónica Movistar del servicio telefónico que le ofrece? (marque con una X tantos como considere)

Publicidad en medios de comunicación  Mensajes cortos  
 Información recibida por correo  Información recibida con la factura  
 Otras (conteste en mayúsculas)

**2** ¿Se considera bien informado sobre los servicios y novedades de Telefónica Movistar?

Sí  No

**3** ¿Recuerda haber recibido en su domicilio algún tipo de información sobre el servicio de Telefónica Movistar junto con la factura?

Sí  No

Si ha contestado "Sí", ¿podría indicarnos que hace normalmente con esta información?

La leo con atención  
 La miro sin prestar mucha atención  
 La leo sin leerla

Si ha marcado "La leo sin leerla", ¿podría señalar el motivo?

No tengo tiempo  No me interesa  
 No leo nunca la publicidad  Otras (conteste en mayúsculas)

**4** ¿Recuerda haber recibido en su domicilio algún tipo de comunicación sobre el servicio de Telefónica Movistar, distinta a los encartes enviados junto con la factura?

Sí  No

Si ha contestado "Sí", ¿podría indicarnos que hace normalmente con estas comunicaciones?

Las leo con atención  
 Las miro sin prestar mucha atención  
 Las leo sin leerlas

Si ha marcado "Las leo sin leerlas", ¿podría señalar el motivo?

No tengo tiempo  No me interesa  
 No leo nunca la publicidad  Otras (conteste en mayúsculas)

**5** Si ha marcado "Las leo sin leerlas", ¿podría señalar el motivo?

No tengo tiempo  No me interesa  
 No leo nunca la publicidad  Otras (conteste en mayúsculas)

**6** ¿Como valora la frecuencia con que recibe esta información?

Excesiva  Bastante  Adecuada  
 Escasa  Muy escasa

**7** ¿Le gustaría recibir información de algún tema en especial?

Sí  No (conteste en mayúsculas)

**8** ¿Cómo valora la frecuencia con que recibe esta información?

Excesiva  Bastante  Adecuada  
 Escasa  Muy escasa

**9** ¿Cuál es su grado de satisfacción general con el servicio de teléfono móvil prestado actualmente. Para responder, utilice una escala de 1 a 10, en la que 1 significa "Nada satisfecho" y 10 "Muy satisfecho"

1  2  3  4  5  6  7  8  9  10

Nada satisfecho Muy satisfecho

**10** ¿Qué aspectos del servicio considera que deberían mejorar para que aumentara su satisfacción con el mismo? (conteste en mayúsculas)

**ESPECIFICAR MAS CLARO EL TEMA TARIFAS**

**11** ¿Señale su grado de satisfacción general con el servicio de teléfono móvil prestado actualmente. Para responder, utilice una escala de 1 a 10, en la que 1 significa "Nada satisfecho" y 10 "Muy satisfecho"

1  2  3  4  5  6  7  8  9  10

Nada satisfecho Muy satisfecho

**12** Si tiene que recomendar a un amigo o conocido una empresa de telefonía móvil, recomendaría Telefónica Movistar? Utilice la escala de 1 a 10 en la que 1 significa "Nunca la recomendaré" y 10 "Siempre la recomendaré"

1  2  3  4  5  6  7  8  9  10

Nunca la recomendaré Siempre la recomendaré

**13** Por último, si desea realizar algún comentario o sugerencia sobre el servicio que presta Telefónica Movistar o las comunicaciones que le envía, hágallo a continuación (conteste en mayúsculas)

**QUE FUERA MAS ASEQUIBLE ECONOMICAMENTE Y QUE HICIERA UNA PUBLICIDAD MAS SENCILLA**

En cumplimiento de la Ley Orgánica 15/1999 de 13 de diciembre de Protección de Datos de Carácter Personal, conforme al art. 5 apartado b) del derecho de información en la recogida de datos. Telefónica Movistar S.A.U. le informa de que sus datos se destinan a la realización de encuestas y estudios de opinión que tienen el carácter de investigación científica y se obtendrán sobre 1000 partes del programa de partes de Telefónica Movistar. La información proporcionada será incluida en ficheros informatizados de la filial de Telefónica Móviles España S.A.U. con domicilio en Plaza de la Independencia nº 6, 28001 Madrid, que será la destinataria de la información facilitada con la finalidad de ofrecerle informaciones publicitarias u ofertas personalizadas o no de productos y servicios que puedan ser de su interés. Los datos se conservarán durante el plazo de 5 años. Puede ejercer sus derechos de acceso, rectificación, cancelación y oposición en la dirección: Ctra. de Alcalá, 100, 28001 Madrid. Ref. DATOS Apartado de Correos 053 28000 Madrid. Asimismo dispone de la posibilidad de ejercer los derechos de acceso, rectificación, cancelación y oposición previstos en la Ley Salvo manifestación expresa dirigida a la dirección indicada se entenderá que el cliente no tiene objeción en que sus datos sean utilizados por las empresas del Grupo Telefónica exclusivamente para dirigirle ofertas de los servicios que puedan ser de su interés.



# Pregunta 8 y su Esquema de Clasificación

**8** - Por último, si desea realizar algún comentario o sugerencia sobre el servicio que presta <nombre de compañía> o las comunicaciones que le envía, hágalo a continuación (conteste en mayúsculas).

## Esquema Propuesto de Clasificación de Respuestas

Etp.	Descripción
01	Precios mas baratos,descuentos,eliminar cuotas, mas ofertas en precios, etc. Eliminar el establecimiento de llamada. Reducir tarifas. Promociones.
02	Mejoras geográficas en cobertura, más antenas, etc.
03	Terminales: facilitar cambios, baterías, más sencillos de manejo, etc.
04	Mejorar la atención e información al cliente.
07	Estoy satisfecho.
09	Insatisfacción con el servicio o las promociones.
10	Servicios: mas y nuevos servicios técnicos en general (o cambio, modificación o ampliación de los existentes) no relativos al precio. Ampliar horarios. Internet.
12	Solicita una respuesta. Pide acción por parte de <nombre de compañía>.

- 01 Precios mas baratos,descuentos,eliminar cuotas, mas ofertas en precios, etc. Eliminar el establecimiento de llamada. Reducir tarifas. Promociones.
- 02 Mejoras geográficas en cobertura, más antenas, etc.
- 03 Terminales: facilitar cambios, baterías, más sencillos de manejo, etc.
- 04 Mejorar la atención e información al cliente.
- 07 Estoy satisfecho.
- 09 Insatisfacción con el servicio o las promociones.
- 10 Servicios: mas y nuevos servicios técnicos en general (o cambio, modificación o ampliación de los existentes) no relativos al precio. Ampliar horarios. Internet.
- 12 Solicitud de respuesta. Pide acción por parte de <nombre de compañía>.

# Dificultades de la Tarea

La escritura manuscrita espontánea: amplia variedad de estilos y ninguna de restricción explícita o formal respecto al vocabulario.

## Causas:

- Los párrafos manuscritos realizados por un grupo heterogéneo de personas.
- Ningún tipo de especificación fue impuesta respecto al tipo de bolígrafo, estilo de escritura y vocabulario.

## Consecuencias:

- La combinación de estilos de escritura y ruido, resulta en la aparición de muestras parcial o totalmente ilegibles.
- El vocabulario de la tarea llega a ser muy grande.

A pesar de la dificultad de la tarea, la misma es compensada en cierta medida por la simplicidad del esquema de clasificación definido.

# Dificultades de la Tarea

La escritura manuscrita espontánea: amplia variedad de estilos y ninguna de restricción explícita o formal respecto al vocabulario.

## Causas:

- *Los párrafos manuscritos realizados por un grupo heterogéneo de personas.*
- *Ningún tipo de especificación fue impuesta respecto al tipo de bolígrafo, estilo de escritura y vocabulario.*

## Consecuencias:

- *La combinación de estilos de escritura y ruido, resulta en la aparición de muestras parcial o totalmente ilegibles.*
- *El vocabulario de la tarea llega a ser muy grande.*

A pesar de la dificultad de la tarea, la misma es compensada en cierta medida por la simplicidad del esquema de clasificación definido.

# Dificultades de la Tarea

La escritura manuscrita espontánea: amplia variedad de estilos y ninguna de restricción explícita o formal respecto al vocabulario.

## Causas:

- Los párrafos manuscritos realizados por un grupo heterogéneo de personas.
- Ningún tipo de especificación fue impuesta respecto al tipo de bolígrafo, estilo de escritura y vocabulario.

## Consecuencias:

- La combinación de estilos de escritura y ruido, resulta en la aparición de muestras parcial o totalmente ilegibles.
- El vocabulario de la tarea llega a ser muy grande.

A pesar de la dificultad de la tarea, la misma es compensada en cierta medida por la simplicidad del esquema de clasificación definido.

# Ejemplos de Muestras con Inconvenientes

<p><b>ALTURA DE LA LETRA</b></p> <p>MÁS BARATO ANQUE SOLO SEA A MI Por lo pronto</p>	<p><b>IDIOMA</b></p> <p>CAP COMENTARI MIGRATORIA DISPOSAR DE 1 MÓBIL</p>
<p><b>SEPARACIÓN DE PALABRAS</b></p> <p>TRABAJAN DO MUY LENTI Los TELEFONOS malfuncionamiento desservicios</p>	<p><b>SEGMENTACIÓN EN LÍNEAS</b></p> <p>DAR LA OPORTUNIDAD A QUE UNA PERSONA QUE LLEVA TANTO TIEMPO CONFIAndo EN TECNICA OPREBRIERSELO OTRO TI- AD</p>
<p><b>TACHADOS</b></p> <p>Por el numero <del>██████</del> que estoy <del>██████</del> ANESES LA <del>██████</del> REBASAR EL NÚMERO</p>	<p><b>RENGLONES MARCADOS</b></p> <p>Maya trabajad para cambiar di teléfono cortaron el nuevo m-c cuando no se acerc a rotura por antiguedad del apart. (Me refiero al precio)</p>
<p><b>MAYÚSCULAS Y MINÚSCULAS</b></p> <p>ESToy SATISfecho UTILIZAMOS EL SERVICIO</p>	<p><b>ABREVIATURAS</b></p> <p>TELEF. TFNO. CREO q. EL SERVICIO</p>

# Ejemplos de Muestras con Inconvenientes

<p><b>ESTILOS</b></p> <p>por mejorar el servicio.</p> <p>DEBERÍAN DAR MÁS RAPIDA BETOR DULCE LUMOSAS</p> <p>el establecimiento de lluvia</p> <p>ME GUSTA PIAZAGÜE. PL</p> <p>El concierto es que espero malfuncionamiento de servicios</p>	<p><b>ORTOGRAFÍA</b></p> <p>CAMBIAR EXCESIVAS HUTILIZAR FALTURA HESTO i VAJAR</p>
<p><b>GROSOR DEL TRAZO</b></p> <p>INFORMACIÓN DE <b>CUANDO LLUVIAS</b></p> <p>mejor en mi deseo de tener</p>	<p><b>OTROS</b></p> <p>GRACIAS <del>PA</del> TU BIENSE REPETIDAS X CORTE SIEMPRE SERVICIOS</p>

# Preproceso y Extracción de Características

## Preproceso

Ahora se suma la tarea de unir las líneas de cada párrafo manuscrito en una única línea, para su posterior extracción de características.

CONSIDERO TODO CORRECTO .	→	CONSIDERO TODO CORRECTO .
DEBERIA MEJORAR LA COBERTURA.	→	DEBERIA MEJORAR LA COBERTURA.

## Extracción de Características

- Adecuada con HMMs unidimensionales: secuencias de vectores de características de dimensión fija (60 componentes).
- Nivel de gris normalizado y sus respectivas dos derivadas.

# Modelado del Sistema

Nuevamente, para el modelado de cada uno de los niveles de percepción, se utilizan los modelos de estados finitos.

1 HMMs para modelado de los caracteres.

- Topología izquierda-derecha.
- 6 estados.
- Mixtura de 16 Gaussianas por estado.

2 AEEFs para el modelado de cada una de las palabras del léxico.

3 N-gramas para el modelo de lenguaje y los clasificadores de texto.

- Se utilizaron uni-gramas y bi-gramas.
- Suavizado por *back-off*.
- Método de descuento de Witten-Bell.

# Modelado del Sistema

Nuevamente, para el modelado de cada uno de los niveles de percepción, se utilizan los modelos de estados finitos.

## 1 HMMs para modelado de los caracteres.

- Topología izquierda-derecha.
- 6 estados.
- Mixtura de 16 Gaussianas por estado.

## 2 AEEFs para el modelado de cada una de las palabras del léxico.

## 3 N-gramas para el modelo de lenguaje y los clasificadores de texto.

- Se utilizaron uni-gramas y bi-gramas.
- Suavizado por *back-off*.
- Método de descuento de Witten-Bell.

# Modelado del Sistema

Nuevamente, para el modelado de cada uno de los niveles de percepción, se utilizan los modelos de estados finitos.

## 1 HMMs para modelado de los caracteres.

- Topología izquierda-derecha.
- 6 estados.
- Mixtura de 16 Gaussianas por estado.

## 2 AEEFs para el modelado de cada una de las palabras del léxico.

## 3 N-gramas para el modelo de lenguaje y los clasificadores de texto.

- Se utilizaron uni-gramas y bi-gramas.
- Suavizado por *back-off*.
- Método de descuento de Witten-Bell.

# Modelado del Sistema

Nuevamente, para el modelado de cada uno de los niveles de percepción, se utilizan los modelos de estados finitos.

## 1 HMMs para modelado de los caracteres.

- Topología izquierda-derecha.
- 6 estados.
- Mixtura de 16 Gaussianas por estado.

## 2 AEEFs para el modelado de cada una de las palabras del léxico.

## 3 N-gramas para el modelo de lenguaje y los clasificadores de texto.

- Se utilizaron uni-gramas y bi-gramas.
- Suavizado por *back-off*.
- Método de descuento de Witten-Bell.

# Marco Estadist. del Esquema Desacoplado: $\mathbf{x} \rightarrow \mathbf{c} \rightarrow I$

Sea  $\mathbf{x}$  una secuencia de v.c.,  $\mathbf{c}$  una secuencia de palabras y  $I$  un identificativo de la categoría.

$$\hat{I} = \arg \max_I P(I | \vec{x}) = \arg \max_I \sum_{\mathbf{c}} P(\mathbf{c}, I | \mathbf{x})$$

$$= \arg \max_I \sum_{\mathbf{c}} p(\mathbf{x} | \mathbf{c}, I) P(\mathbf{c}, I) = \arg \max_I \sum_{\mathbf{c}} p(\mathbf{x} | \mathbf{c}) P(\mathbf{c}, I)$$

$$(\hat{I}, \hat{\mathbf{c}}) = \arg \max_{I, \mathbf{c}} p(\vec{x} | \mathbf{c}) P(I | \mathbf{c}) P(\mathbf{c})$$

$$\hat{\mathbf{c}} \approx \arg \max_{\mathbf{c}} p_{\text{HMM}}(\mathbf{x} | \mathbf{c}) P_{\text{ML}}(\mathbf{c})$$

$$\hat{I} \approx \arg \max_I P(I | \hat{\mathbf{c}}) = \arg \max_I P_{\text{CL}}(\hat{\mathbf{c}} | I) P(I)$$

# Marco Estadist. del Esquema Desacoplado: $\mathbf{x} \rightarrow \mathbf{c} \rightarrow I$

Sea  $\mathbf{x}$  una secuencia de v.c.,  $\mathbf{c}$  una secuencia de palabras y  $I$  un identificativo de la categoría.

$$\hat{I} = \arg \max_I P(I | \vec{x}) = \arg \max_I \sum_{\mathbf{c}} P(\mathbf{c}, I | \mathbf{x})$$

$$= \arg \max_I \sum_{\mathbf{c}} p(\mathbf{x} | \mathbf{c}, I) P(\mathbf{c}, I) = \arg \max_I \sum_{\mathbf{c}} p(\mathbf{x} | \mathbf{c}) P(\mathbf{c}, I)$$

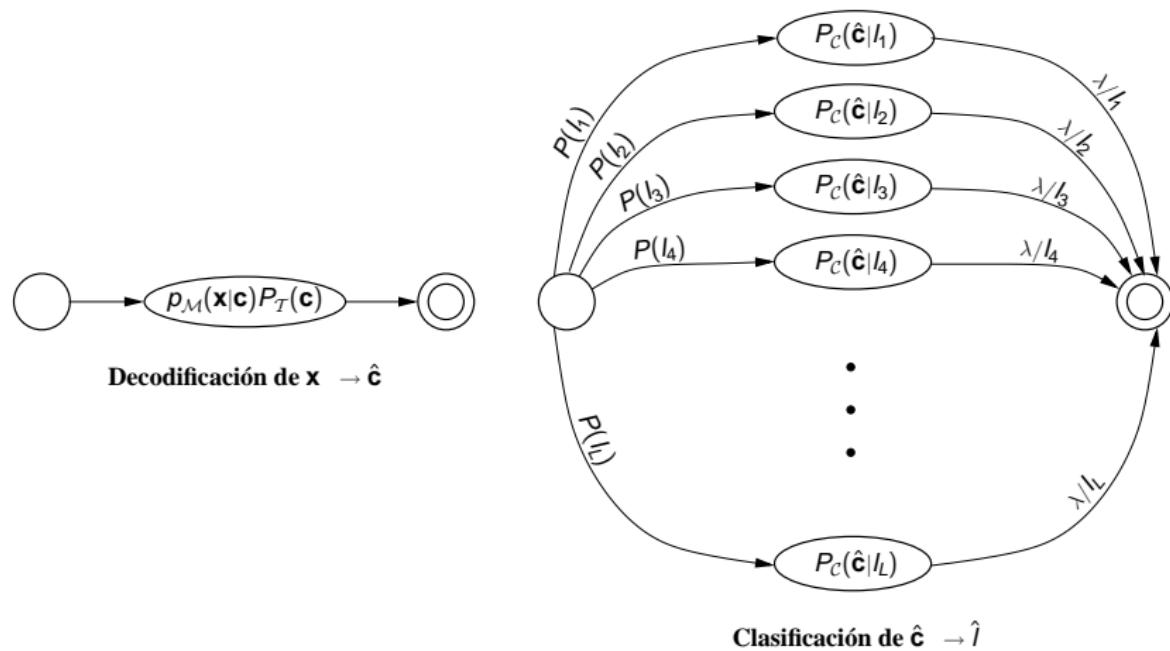
$$(\hat{I}, \hat{\mathbf{c}}) = \arg \max_{I, \mathbf{c}} p(\vec{x} | \mathbf{c}) P(I | \mathbf{c}) P(\mathbf{c})$$

$$\hat{\mathbf{c}} \approx \arg \max_c p_{\text{HMM}}(\mathbf{x} | \mathbf{c}) P_{\text{ML}}(\mathbf{c})$$

$$\hat{I} \approx \arg \max_I P(I | \hat{\mathbf{c}}) = \arg \max_I P_{\text{CL}}(\hat{\mathbf{c}} | I) P(I)$$

# Esquema de Reconoc. y Clasif. Desacoplado

**Esquema Serie:** en la que primero se realiza el reconocimiento de cada muestra manuscrita, y luego se procede a su clasificación.



# Marco Estadístico del Esquema Integrado

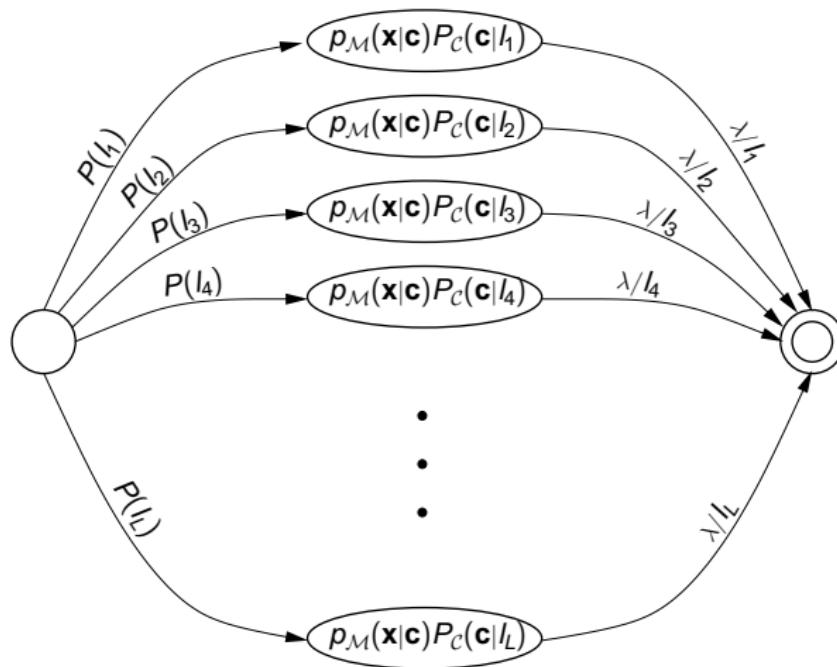
Sea  $\mathbf{x}$  una secuencia de v.c.,  $\mathbf{c}$  una secuencia de palabras y  $I$  un identificativo de la categoría.

$$\begin{aligned}
 \hat{I} &= \arg \max_I P(I | \vec{\mathbf{x}}) = \arg \max_I \sum_{\mathbf{c}} P(\mathbf{c}, I | \mathbf{x}) \\
 &= \arg \max_I \sum_{\mathbf{c}} p(\mathbf{x} | \mathbf{c}, I) P(\mathbf{c}, I) = \arg \max_I \sum_{\mathbf{c}} p(\mathbf{x} | \mathbf{c}) P(\mathbf{c}, I) \\
 &= \arg \max_I \sum_{\mathbf{c}} p(\mathbf{x} | \mathbf{c}) P(\mathbf{c} | I) P(I) = \arg \max_I P(I) \sum_{\mathbf{c}} p(\mathbf{x} | \mathbf{c}) P(\mathbf{c} | I) \\
 \hat{I} &\approx \arg \max_I P(I) \max_{\mathbf{c}} p(\mathbf{x} | \mathbf{c}) P(\mathbf{c} | I)
 \end{aligned}$$

$$\hat{I} \approx \arg \max_I P(I) \max_{\mathbf{c}} p_{\text{HMM}}(\mathbf{x} | \mathbf{c}) P_{\text{ML}}(\mathbf{c} | I)$$

# Esquema de Reconocimiento y Clasificación Integrado

**Esquema Integrado:** tanto el reconocimiento como la clasificación se realizan en forma simultánea.



# Corpus de Experimentación

Número de:	Corpus	Entrenamiento	Test	Léxico
párrafos	913	676	237	—
palabras	16371	12287	4084	3308
caracteres	86199	64666	21533	80

Etiq.	# Parraf.	# Palabras	Voc	Voc <sub>frec&gt;1</sub>	Voc <sub>frec&gt;2</sub>
01	239	4097	847	329	192
02	87	1415	419	127	75
03	134	3135	599	230	159
04	198	3470	771	260	158
07	111	1189	290	91	60
09	54	1071	324	98	48
10	55	995	343	85	55
12	36	891	302	81	45

# Resultados Experimentales

## Esquema Desacoplado

	WER( %)	ER( %)	
		1-gram(C)	2-gram(C)
1-gram(R)	34.3	51.1	—
2-gram(R)	32.5	—	57.0

## Esquema Integrado

	WER( %)	ER( %)	
		1-gram(C)	2-gram(C)
1-gram(R)	34.4	50.2	n/a
2-gram(R)	33.0	n/a	59.1

# Varios Ejemplos de Reconocimiento

I	<b>MAYOR RESPONSABILIDAD EN LA RESOLUCIÓN DE PROBLEMAS</b>					
T	MAYOR	RESPONSABILIDAD	EN	LA	RESOLUCIÓN	DE
R	MAYOR	RESPONSABILIDAD	EN	LA	RESOLUCIÓN	DE NO LE MAS

I	<b>SIEMPRE LLEGAN TARDE LAS OFERTAS Y REGALOS</b>					
T	SIEMPRE	LLEGAN	TARDE	LAS	OFERTAS	Y REGALOS
R	SIEMPRE	LLEGAN	TARDE	LAS ,	OFERTAS .	TRES AÑOS

I	<b>MEJORAR EL PROGRAMA DE PUNTOS</b>					
T	MEJORAR	EL	PROGRAMA	DE	PUNTOS	
R	ME DURAR .	EL	PROGRAMA	DE .	PUNTOS	

I	<b>ESTOY BASTANTE SATISFECHO</b>					
T	ESTOY	BASTANTE	SATISFECHO			
R	ESTO	MAS TANTO .	SATISFECHO			

# Ejemplos de párrafos perfectamente reconocidas

Párrafo Clasificado Erróneamente (E:07 R:03)

I-T	<i>ESTOY CONTENTO PERO CREO QUE FALTA ALGO MAS DE APOYO A LOS QUE TRABAJAMOS CON ELLOS .</i>	ESTOY CONTENTO PERO CREO QUE FALTA ALGO MAS DE APOYO A LOS QUE TRABAJAMOS CON ELLOS .
R	ESTOY CONTENTO PERO CREO QUE FALTA ALGO MAS DE APOYO A LOS QUE TRABAJAMOS CON ELLOS	

Párrafo Clasificado Correctamente (E:03 R:03)

I-T	<i>MAYOR FACILIDAD PARA CAMBIAR DE TELÉFONO CONSERVANDO EL NÚMERO ANTIGUO.</i>	MAYOR FACILIDAD PARA CAMBIAR DE TELÉFONO CONSERVANDO EL NÚMERO ANTIGUO.
R	MAYOR FACILIDAD PARA CAMBIAR DE TELÉFONO CONSERVANDO EL NUMERO ANTIGUO	

# Ejemplos de párrafos reconocidos con algunos errores

## Párrafo Clasificado Erróneamente (E:03 R:09)

I-T	CONSERVACION DEL N° DEL TELEFONO GRATUITO Y MANTENERLO AUNQUE SE HA DE TARJETA.	CONSERVACION DEL N° DEL TELEFONO GRATUITO Y MANTENERLO AUNQUE SE HA DE TARJETA.
R	CONSERVACIÓN DEL <u>NI</u> DEL TELÉFONO GRATUITO Y MANTENERLO AUNQUE SE <u>YA</u> DE TARJETA	

## Párrafo Clasificado Correctamente (E:03 R:03)

I-T	LO PUESTO EN EL PUNTO 6, PUESTO QUE LLEVO 4 AÑOS CON EL MISMO TLEFNO Y ME SUPONE MUCHO DINERO CON VOSOTROS CAMBIARLE, LO QUE HACE QUE ESTE MIRANDO OTRAS OFERTAS.	LO PUESTO EN EL PUNTO 6, PUESTO QUE LLEVO 4 AÑOS CON EL MISMO TLEFNO Y ME SUPONE MUCHO DINERO CON VOSOTROS CAMBIARLE, LO QUE HACE QUE ESTE MIRANDO OTRAS OFERTAS.
R	LO PUESTO EN EL <b>PUNTOS</b> PUESTO QUE LLEVO <u>LA</u> AÑOS CON EL MISMO TELÉFONO <u>MI</u> ME SUPONE MUCHO DINERO CON VOSOTROS CAMBIARLE LO QUE HACE QUE ESTE MIRANDO OTRAS OFERTAS	

# Ejemplos de párrafos reconocidos con más errores

## Párrafo Clasificado Correctamente (E:01 R:01)

I-T	<u>DEBERÍAN IMITAR A LA COMPETENCIA BAJANDO LOS PRECIOS Y FACILITANDO MÁS INFORMACIÓN GRATIS EN LAS TARJETAS ACTIVA MEJORAR LA COBERTURA</u>	DEBERÍAN IMITAR A LA COMPETENCIA BAJANDO LOS PRECIOS Y FACILITANDO MÁS INFORMACIÓN GRATIS EN LAS TARJETAS ACTIVA MEJORAR LA COBERTURA
R	DEBERÍAN IMITAR A LA COMPETENCIA <b>DANDO NI</b> PRECIOS Y FACILITANDO MAS INFORMACIÓN GRATIS EN LAS <b>TARIFAS</b> ACTIVA <b>ME BORRAR</b> LA COBERTURA	

## Párrafo Clasificado Erróneamente (E:04 R:01)

I-T	<u>INFORMACION MAS PUNTUAL Y DETALLADA POR PARTE DE LOS DISTRIBUIDORES.</u>	INFORMACION MAS PUNTUAL Y DETALLADA POR PARTE DE LOS DISTRIBUIDORES.
R	<u>O A EN TEMA A DÚO NIÑOS PRIMAN MI HAY M O AMENA O A POR A A E TE O ESO OTRAS TERMINAR O</u>	

# Temario

1

## Uso de Modelos de Lenguaje en RES

- Decodificación (Reconocimiento)
- Diferentes Niveles de Percepción
- Autómata Estocástico de Estados Finitos (AEEF)
- Modelado a Nivel Morfológico
- Modelado a Nivel Léxico
- Modelado a Nivel Sintáctico
- Modelo Integrado
- Modelos de Lenguaje de N-gramas
- Marco Teórico: Formulación del Problema de RES
- Mixtura de Gaussianas, GSF y WIP
- Métricas de Evaluación

2

## Aplicaciones

- Aplicación RNCN
- Aplicación RMFE
- **Prototipo PAITYT**

# Necesidad y Aplicación

Causa:

- Las librerías digitales vienen publicando en internet una gran cantidad de material bibliográfico, que incluyen numerosos documentos manuscritos antiguos.
- Muchos de estos documentos están acompañados de su respectiva transcripción (en formato ASCII, PDF, DOC, etc.).

Se ha incrementado el desarrollo de metodologías para alinear las imágenes de palabras con las palabras correspondientes en su transcripción.

Esto puede ayudar a:

- Paleógrafos expertos a localizar una imagen de texto mientras leen su transcripción.
- Ser de utilidad a la gente en general, cuando leen unos de estos documentos y arriban a partes del mismo que son más complejas o están dañadas.

# Necesidad y Aplicación

Causa:

- Las librerías digitales vienen publicando en internet una gran cantidad de material bibliográfico, que incluyen numerosos documentos manuscritos antiguos.
- Muchos de estos documentos están acompañados de su respectiva transcripción (en formato ASCII, PDF, DOC, etc.).

Se ha incrementado el desarrollo de metodologías para alinear las imágenes de palabras con las palabras correspondientes en su transcripción.

Esto puede ayudar a:

- Paleógrafos expertos a localizar una imagen de texto mientras leen su transcripción.
- Ser de utilidad a la gente en general, cuando leen unos de estos documentos y arriban a partes del mismo que son más complejas o están dañadas.

# Necesidad y Aplicación

Causa:

- Las librerías digitales vienen publicando en internet una gran cantidad de material bibliográfico, que incluyen numerosos documentos manuscritos antiguos.
- Muchos de estos documentos están acompañados de su respectiva transcripción (en formato ASCII, PDF, DOC, etc.).

Se ha incrementado el desarrollo de metodologías para alinear las imágenes de palabras con las palabras correspondientes en su transcripción.

Esto puede ayudar a:

- Paleógrafos expertos a localizar una imagen de texto mientras leen su transcripción.
- Ser de utilidad a la gente en general, cuando leen unos de estos documentos y arriban a partes del mismo que son más complejas o están dañadas.

# Necesidad y Aplicación

Causa:

- Las librerías digitales vienen publicando en internet una gran cantidad de material bibliográfico, que incluyen numerosos documentos manuscritos antiguos.
- Muchos de estos documentos están acompañados de su respectiva transcripción (en formato ASCII, PDF, DOC, etc.).

Se ha incrementado el desarrollo de metodologías para alinear las imágenes de palabras con las palabras correspondientes en su transcripción.

Esto puede ayudar a:

- Paleógrafos expertos a localizar una imagen de texto mientras leen su transcripción.
- Ser de utilidad a la gente en general, cuando leen unos de estos documentos y arriban a partes del mismo que son más complejas o están dañadas.

# Necesidad y Aplicación

Causa:

- Las librerías digitales vienen publicando en internet una gran cantidad de material bibliográfico, que incluyen numerosos documentos manuscritos antiguos.
- Muchos de estos documentos están acompañados de su respectiva transcripción (en formato ASCII, PDF, DOC, etc.).

Se ha incrementado el desarrollo de metodologías para alinear las imágenes de palabras con las palabras correspondientes en su transcripción.

Esto puede ayudar a:

- Paleógrafos expertos a localizar una imagen de texto mientras leen su transcripción.
- Ser de utilidad a la gente en general, cuando leen unos de estos documentos y arriban a partes del mismo que son más complejas o están dañadas.

# Interfaz del Prototipo de Alineamientos

Prototype to align Text-Images and Transcriptions

ORIGINAL IMAGE

Noticia histórica  
de las fiestas con que Valencia  
celebró el siglo sexto de  
la venida a esta Capital  
de la milagrosa imagen del Salvador

por

D. Vicente Boix  
cronista de la misma Ciudad

1853.

TRANSCRIPTION

NOTICIA HISTÓRICA  
DE LAS FIESTAS CON QUE VALENCIA  
CELEBRÓ EL SIGLO SEXTO DE  
LA VENIDA Á ESTA CAPITAL  
DE LA MILAGROSA IMAGEN DEL SALVADOR

POR

D. VICENTE BOIX,  
CRONISTA DE LA MISMA CIUDAD

1853

Pos x: 0000 Lin I: 0710 Col I: 0389  
Pos y: 0000 Lin E: 0628 Col E: 0588

Pos Ins: 0163 Wrd I: 0160  
Wrd Idx: 0015 Wrd E: 0166

SA GO  
◀ 001/051 ▶

GENERALITAT  
VALENCIANA

RES Off-Line

# Alineación basada en Viterbi

- Muchas de las técnicas que aparecen en la literatura, se basan en una segmentación explícita de las palabras que aparecen en la imagen de la página.
- La que emplea este prototipo se basa en el algoritmo de Viterbi para reconocimiento de texto manuscrito con HMMs.
- No requiere de una segmentación explícita de las palabras.
- El alineamiento en sí, es un subproducto del proceso propio de reconocimiento, donde cada segmento corresponde a una palabra reconocida.
- En la literatura esta técnica de segmentación es conocida como “reconocimiento forzado”.
- La segmentación de palabras es obtenida a nivel de línea.

# Alineación basada en Viterbi

- Muchas de las técnicas que aparecen en la literatura, se basan en una segmentación explícita de las palabras que aparecen en la imagen de la página.
- La que emplea este prototipo se basa en el algoritmo de Viterbi para reconocimiento de texto manuscrito con HMMs.
- No requiere de una segmentación explícita de las palabras.
- El alineamiento en sí, es un subproducto del proceso propio de reconocimiento, donde cada segmento corresponde a una palabra reconocida.
- En la literatura esta técnica de segmentación es conocida como “reconocimiento forzado”.
- La segmentación de palabras es obtenida a nivel de línea.

# Alineación basada en Viterbi

- Muchas de las técnicas que aparecen en la literatura, se basan en una segmentación explícita de las palabras que aparecen en la imagen de la página.
- La que emplea este prototipo se basa en el algoritmo de Viterbi para reconocimiento de texto manuscrito con HMMs.
- No requiere de una segmentación explícita de las palabras.
- El alineamiento en sí, es un subproducto del proceso propio de reconocimiento, donde cada segmento corresponde a una palabra reconocida.
- En la literatura esta técnica de segmentación es conocida como “reconocimiento forzado”.
- La segmentación de palabras es obtenida a nivel de línea.

# Alineación basada en Viterbi

- Muchas de las técnicas que aparecen en la literatura, se basan en una segmentación explícita de las palabras que aparecen en la imagen de la página.
- La que emplea este prototipo se basa en el algoritmo de Viterbi para reconocimiento de texto manuscrito con HMMs.
- No requiere de una segmentación explícita de las palabras.
- El alineamiento en sí, es un subproducto del proceso propio de reconocimiento, donde cada segmento corresponde a una palabra reconocida.
- En la literatura esta técnica de segmentación es conocida como “reconocimiento forzado”.
- La segmentación de palabras es obtenida a nivel de línea.

# Alineación basada en Viterbi

- Muchas de las técnicas que aparecen en la literatura, se basan en una segmentación explícita de las palabras que aparecen en la imagen de la página.
- La que emplea este prototipo se basa en el algoritmo de Viterbi para reconocimiento de texto manuscrito con HMMs.
- No requiere de una segmentación explícita de las palabras.
- El alineamiento en sí, es un subproducto del proceso propio de reconocimiento, donde cada segmento corresponde a una palabra reconocida.
- En la literatura esta técnica de segmentación es conocida como “*reconocimiento forzado*”.
- La segmentación de palabras es obtenida a nivel de línea.

# Alineación basada en Viterbi

- Muchas de las técnicas que aparecen en la literatura, se basan en una segmentación explícita de las palabras que aparecen en la imagen de la página.
- La que emplea este prototipo se basa en el algoritmo de Viterbi para reconocimiento de texto manuscrito con HMMs.
- No requiere de una segmentación explícita de las palabras.
- El alineamiento en sí, es un subproducto del proceso propio de reconocimiento, donde cada segmento corresponde a una palabra reconocida.
- En la literatura esta técnica de segmentación es conocida como “*reconocimiento forzado*”.
- La segmentación de palabras es obtenida a nivel de línea.

# Reconocimiento Forzado

Sea

$$\hat{\mathbf{w}} = \langle w_1, w_2, \dots, w_n \rangle$$

$$\mathbf{x} = \langle x_1, x_2, \dots, x_p \rangle$$

$$\mathbf{b} = \langle b_0, b_1, \dots, b_n \rangle$$

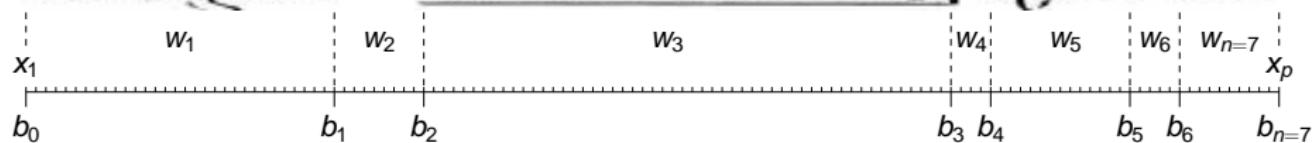
secuencia de palabras conocidas

secuencia de vectores de características

secuencia de marcas de segmentación

*antiguos ciudadanos, que en*

*antiguos ciudadanos, que en*



# Formulación Matemática

Formulación estadística del problema del reconocimiento de escritura manuscrita:

$$\hat{\mathbf{w}} = \arg \max_{\mathbf{w}} Pr(\mathbf{x}|\mathbf{w}) \cdot Pr(\mathbf{w})$$

$$\hat{\mathbf{w}} = \arg \max_{\mathbf{w}} \sum_{\mathbf{b}} Pr(\mathbf{x}, \mathbf{b}|\mathbf{w}) \cdot Pr(\mathbf{w})$$

$$\hat{\mathbf{w}} \approx \arg \max_{\mathbf{w}} \max_{\mathbf{b}} Pr(\mathbf{x}, \mathbf{b}|\mathbf{w}) \cdot Pr(\mathbf{w})$$

En este caso, la transcripción es conocida de antemano. Sea  $\tilde{\mathbf{w}}$  una transcripción dada, entonces  $Pr(\tilde{\mathbf{w}}) = 1$ :

$$\hat{\mathbf{b}} = \arg \max_{\mathbf{b}} Pr(\mathbf{x}, \mathbf{b}|\tilde{\mathbf{w}})$$

$$= Pr(x_{b_0}^{b_1}, x_{b_1}^{b_2}, \dots, x_{b_{n-1}}^{b_n} | \tilde{\mathbf{w}})$$

$$\approx \arg \max_{\mathbf{b}} Pr(x_{b_0}^{b_1} | \tilde{w}_1) \dots Pr(x_{b_{n-1}}^{b_n} | \tilde{w}_n)$$

# Formulación Matemática

Formulación estadística del problema del reconocimiento de escritura manuscrita:

$$\hat{\mathbf{w}} = \arg \max_{\mathbf{w}} Pr(\mathbf{x}|\mathbf{w}) \cdot Pr(\mathbf{w})$$

$$\hat{\mathbf{w}} = \arg \max_{\mathbf{w}} \sum_{\mathbf{b}} Pr(\mathbf{x}, \mathbf{b}|\mathbf{w}) \cdot Pr(\mathbf{w})$$

$$\hat{\mathbf{w}} \approx \arg \max_{\mathbf{w}} \max_{\mathbf{b}} Pr(\mathbf{x}, \mathbf{b}|\mathbf{w}) \cdot Pr(\mathbf{w})$$

En este caso, la transcripción es conocida de antemano. Sea  $\tilde{\mathbf{w}}$  una transcripción dada, entonces  $Pr(\tilde{\mathbf{w}}) = 1$ :

$$\hat{\mathbf{b}} = \arg \max_{\mathbf{b}} Pr(\mathbf{x}, \mathbf{b}|\tilde{\mathbf{w}})$$

$$= Pr(x_{b_0}^{b_1}, x_{b_1}^{b_2}, \dots, x_{b_{n-1}}^{b_n} | \tilde{\mathbf{w}})$$

$$\approx \arg \max_{\mathbf{b}} Pr(x_{b_0}^{b_1} | \tilde{w}_1) \dots Pr(x_{b_{n-1}}^{b_n} | \tilde{w}_n)$$

# Formulación Matemática

Formulación estadística del problema del reconocimiento de escritura manuscrita:

$$\hat{\mathbf{w}} = \arg \max_{\mathbf{w}} Pr(\mathbf{x}|\mathbf{w}) \cdot Pr(\mathbf{w})$$

$$\hat{\mathbf{w}} = \arg \max_{\mathbf{w}} \sum_{\mathbf{b}} Pr(\mathbf{x}, \mathbf{b}|\mathbf{w}) \cdot Pr(\mathbf{w})$$

$$\hat{\mathbf{w}} \approx \arg \max_{\mathbf{w}} \max_{\mathbf{b}} Pr(\mathbf{x}, \mathbf{b}|\mathbf{w}) \cdot Pr(\mathbf{w})$$

En este caso, la transcripción es conocida de antemano. Sea  $\tilde{\mathbf{w}}$  una transcripción dada, entonces  $Pr(\tilde{\mathbf{w}}) = 1$ :

$$\hat{\mathbf{b}} = \arg \max_{\mathbf{b}} Pr(\mathbf{x}, \mathbf{b}|\tilde{\mathbf{w}})$$

$$= Pr(x_{b_0}^{b_1}, x_{b_1}^{b_2}, \dots, x_{b_{n-1}}^{b_n} | \tilde{\mathbf{w}})$$

$$\approx \arg \max_{\mathbf{b}} Pr(x_{b_0}^{b_1} | \tilde{w}_1) \dots Pr(x_{b_{n-1}}^{b_n} | \tilde{w}_n)$$

# Visión General del Prototipo de Alineamientos

La implementación del Prototipo de Alineamientos involucró 4 etapas diferentes:

- ① Preproceso de Imágenes.
- ② Extracción de Características.
- ③ Entrenamiento de HMMs.
- ④ Generación de Mapas de Alineamiento.

El procesamiento de imágenes comprende:

- Corrección del “Skew”
- Filtrado de Ruido y Eliminación del Fondo.
- Detección y Extracción de Imágenes de Línea de Texto Manuscrito.
- Corrección de “Slant”.
- Normalización de Tamaño.

# Visión General del Prototipo de Alineamientos

La implementación del Prototipo de Alineamientos involucró 4 etapas diferentes:

- ① Preproceso de Imágenes.
- ② Extracción de Características.
- ③ Entrenamiento de HMMs.
- ④ Generación de Mapas de Alineamiento.

El procesamiento de imágenes comprende:

- Corrección del “Skew”
- Filtrado de Ruido y Eliminación del Fondo.
- Detección y Extracción de Imágenes de Línea de Texto Manuscrito.
- Corrección de “Slant”.
- Normalización de Tamaño.

# Modelos HMM de Caracteres

- 78 modelos (caracteres mayúsculas y minúsculas).
- **HMMs** con topología izquierda-derecha de seis estados, fueron empleados para modelar cada carácter.
- Número de Gaussianas en la mixtura de cada estado:
- Gaussianas utilizan matrices de covarianza diagonales.
- El entrenamiento de los HMMs es llevado a cabo utilizando el algoritmo *forward-backward* o *Baum-Welch re-estimation*.

# Métricas para Evaluación de Alineamientos (1)

Para evaluar la calidad de los alineamientos, se emplean:

- **MEAN-STD:**  $\mu$  y  $\sigma$  de las diferencias absolutas entre las marcas de referencia y las propuestas por el prototipo de alineamientos, están dadas por:

$$\mu = \frac{\sum_{i=1}^{n-1} d_i}{n - 1} \quad \sigma = \sqrt{\frac{\sum_{i=1}^{n-1} (d_i - \mu)^2}{n - 1}}$$

donde  $d_i = |r_i - b_i|$  y  $\mathbf{r} = \langle r_0, r_1, \dots, r_n \rangle$  es una sec. de marcas de ref.

- **AER:** Tasa de Error de Alineamientos, definida como:

$$\text{AER}(\%) = \frac{100}{N} \sum_{j: w_j \neq s} e_j \quad \text{where} \quad e_j = \begin{cases} 0 & b_{j-1} < m_j < b_j \\ 1 & \text{otherwise} \end{cases}$$

donde  $s$  son las palabras en blanco,  $N < n$  es el número de palabras reales, y  $m_j = (r_{j-1} + r_j)/2$ .

# Métricas para Evaluación de Alineamientos (1)

Para evaluar la calidad de los alineamientos, se emplean:

- **MEAN-STD:**  $\mu$  y  $\sigma$  de las diferencias absolutas entre las marcas de referencia y las propuestas por el prototipo de alineamientos, están dadas por:

$$\mu = \frac{\sum_{i=1}^{n-1} d_i}{n - 1} \quad \sigma = \sqrt{\frac{\sum_{i=1}^{n-1} (d_i - \mu)^2}{n - 1}}$$

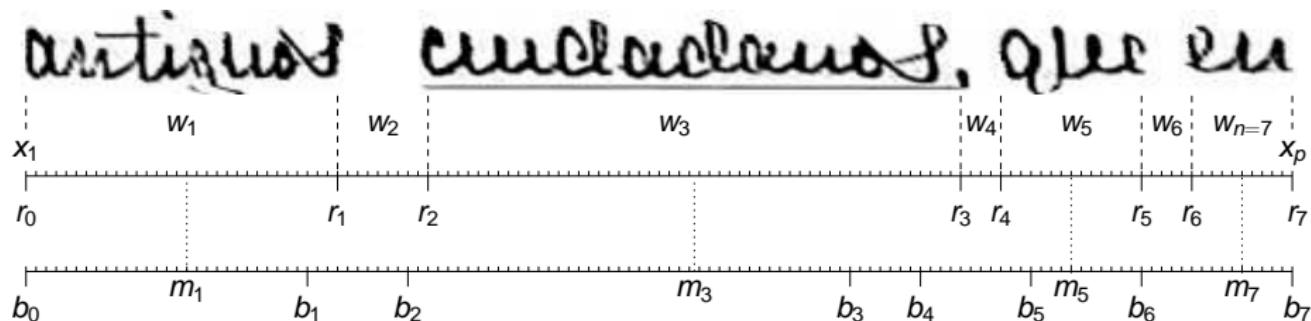
donde  $d_i = |r_i - b_i|$  y  $\mathbf{r} = \langle r_0, r_1, \dots, r_n \rangle$  es una sec. de marcas de ref.

- **AER:** Tasa de Error de Alineamientos, definida como:

$$\text{AER}(\%) = \frac{100}{N} \sum_{j: w_j \neq s} e_j \quad \text{where} \quad e_j = \begin{cases} 0 & b_{j-1} < m_j < b_j \\ 1 & \text{otherwise} \end{cases}$$

donde  $s$  son las palabras en blanco,  $N < n$  es el número de palabras reales, y  $m_j = (r_{j-1} + r_j)/2$ .

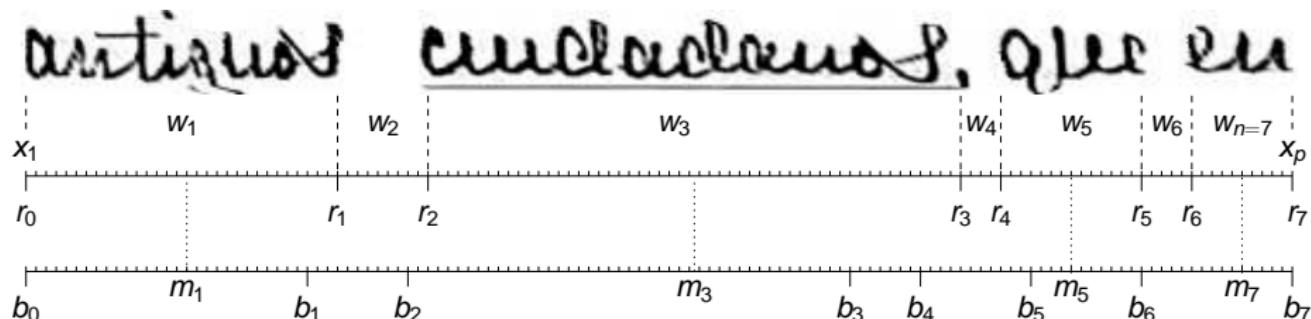
# Métricas para Evaluación de Alineamientos (2)



En este caso ilustrado, el AER sería 25 %.

- Un buen alineamiento tendría un  $\mu$  cercano a 0 y un  $\sigma$  pequeño.
- MEAN-STD nos da una idea acerca de la precisión de las marcas de alineamiento, computadas automáticamente por el prototipo.
- AER mide la tasa de no coincidencia entre las imágenes de palabras y sus transcripciones ASCII asignadas, excluyendo del cómputo las palabras en blanco.

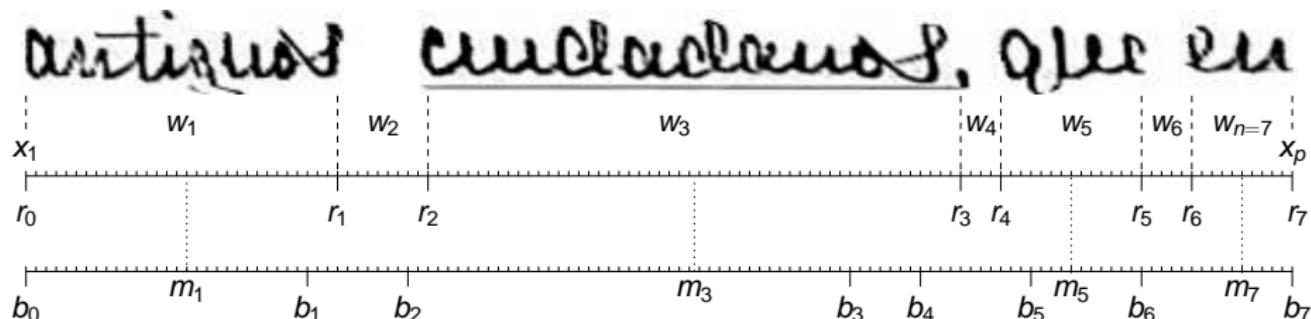
# Métricas para Evaluación de Alineamientos (2)



En este caso ilustrado, el AER sería 25 %.

- Un buen alineamiento tendría un  $\mu$  cercano a 0 y un  $\sigma$  pequeño.
- MEAN-STD nos da una idea acerca de la precisión de las marcas de alineamiento, computadas automáticamente por el prototipo.
- AER mide la tasa de no coincidencia entre las imágenes de palabras y sus transcripciones ASCII asignadas, excluyendo del cómputo las palabras en blanco.

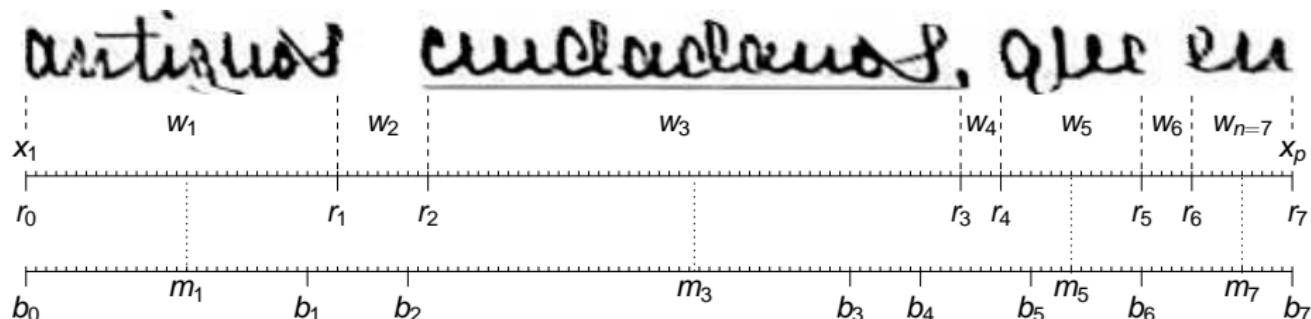
# Métricas para Evaluación de Alineamientos (2)



En este caso ilustrado, el AER sería 25 %.

- Un buen alineamiento tendría un  $\mu$  cercano a 0 y un  $\sigma$  pequeño.
- MEAN-STD nos da una idea acerca de la precisión de las marcas de alineamiento, computadas automáticamente por el prototipo.
- AER mide la tasa de no coincidencia entre las imágenes de palabras y sus transcripciones ASCII asignadas, excluyendo del cómputo las palabras en blanco.

# Métricas para Evaluación de Alineamientos (2)



En este caso ilustrado, el AER sería 25 %.

- Un buen alineamiento tendría un  $\mu$  cercano a 0 y un  $\sigma$  pequeño.
- MEAN-STD nos da una idea acerca de la precisión de las marcas de alineamiento, computadas automáticamente por el prototipo.
- AER mide la tasa de no coincidencia entre las imágenes de palabras y sus transcripciones ASCII asignadas, excluyendo del cómputo las palabras en blanco.

# Ejemplos de Imágenes de Páginas del corpus CS



# Partición del Corpus CS

- El corpus CS está compuesto por 53 imágenes de páginas de texto manuscrito, escaneadas a 300dpi.
- Texto manuscrito por un solo escritor.
- Las transcripciones de cada página están disponibles (PDF originalmente).
- Para testear la calidad de los alineamientos computados, 12 páginas fueron elegidas aleatoriamente y segmentadas manualmente en palabras para ser utilizadas como referencias.

Numero de:	Referencias	Total	Léxico
paginas	12	53	—
lineas de texto	312	1.172	—
palabras	2.955	10.911	3.408
caracteres	16.893	62.159	78

# Partición del Corpus CS

- El corpus CS está compuesto por 53 imágenes de páginas de texto manuscrito, escaneadas a 300dpi.
- Texto manuscrito por un solo escritor.
- Las transcripciones de cada página están disponibles (PDF originalmente).
- Para testear la calidad de los alineamientos computados, 12 páginas fueron elegidas aleatoriamente y segmentadas manualmente en palabras para ser utilizadas como referencias.

Numero de:	Referencias	Total	Léxico
paginas	12	53	—
lineas de texto	312	1.172	—
palabras	2.955	10.911	3.408
caracteres	16.893	62.159	78

# Partición del Corpus CS

- El corpus CS está compuesto por 53 imágenes de páginas de texto manuscrito, escaneadas a 300dpi.
- Texto manuscrito por un solo escritor.
- Las transcripciones de cada página están disponibles (PDF originalmente).
- Para testear la calidad de los alineamientos computados, 12 páginas fueron elegidas aleatoriamente y segmentadas manualmente en palabras para ser utilizadas como referencias.

Numero de:	Referencias	Total	Léxico
paginas	12	53	—
lineas de texto	312	1.172	—
palabras	2.955	10.911	3.408
caracteres	16.893	62.159	78

# Partición del Corpus CS

- El corpus CS está compuesto por 53 imágenes de páginas de texto manuscrito, escaneadas a 300dpi.
- Texto manuscrito por un solo escritor.
- Las transcripciones de cada página están disponibles (PDF originalmente).
- Para testear la calidad de los alineamientos computados, 12 páginas fueron elegidas aleatoriamente y segmentadas manualmente en palabras para ser utilizadas como referencias.

Numero de:	Referencias	Total	Léxico
paginas	12	53	–
lineas de texto	312	1.172	–
palabras	2.955	10.911	3.408
caracteres	16.893	62.159	78

# Resultados Experimentales

- Dos esquemas de modelado diferentes de HMMs, fueron empleados:

- El primer esquema modela cada uno de los 78 clases de caracteres utilizando diferentes HMMs por clase.
- El segundo esquema utiliza 2 HMMs: uno para modelar el carácter de espacio en blanco, y otro para modelar los restantes 77 caracteres conjuntamente.

La topología de los HMMs es idéntica en ambos esquema:  
izquierda-derecha de 6 estados y 64 Gaussianas en la mixtura de cada estado.

- **RESULTADOS:**

	78-HMMs	2-HMMs
AER (%)	7.20	25.98
$\mu$ (mm)	1.15	2.95
$\sigma$ (mm)	3.90	6.56

# Resultados Experimentales

- Dos esquemas de modelado diferentes de HMMs, fueron empleados:
  - El primer esquema modela cada uno de los 78 clases de caracteres utilizando diferentes HMMs por clase.
  - El segundo esquema utiliza 2 HMMs: uno para modelar el carácter de espacio en blanco, y otro para modelar los restantes 77 caracteres conjuntamente.

La topología de los HMMs es idéntica en ambos esquema: izquierda-derecha de 6 estados y 64 Gaussianas en la mixtura de cada estado.

- **RESULTADOS:**

	78-HMMs	2-HMMs
AER (%)	7.20	25.98
$\mu$ (mm)	1.15	2.95
$\sigma$ (mm)	3.90	6.56

# Resultados Experimentales

- Dos esquemas de modelado diferentes de HMMs, fueron empleados:
  - El primer esquema modela cada uno de los 78 clases de caracteres utilizando diferentes HMMs por clase.
  - El segundo esquema utiliza 2 HMMs: uno para modelar el carácter de espacio en blanco, y otro para modelar los restantes 77 caracteres conjuntamente.

La topología de los HMMs es idéntica en ambos esquema:  
izquierda-derecha de 6 estados y 64 Gaussianas en la mixtura de cada estado.

## • RESULTADOS:

	78-HMMs	2-HMMs
AER (%)	7.20	25.98
$\mu$ (mm)	1.15	2.95
$\sigma$ (mm)	3.90	6.56

# Resultados Experimentales

- Dos esquemas de modelado diferentes de HMMs, fueron empleados:
  - El primer esquema modela cada uno de los 78 clases de caracteres utilizando diferentes HMMs por clase.
  - El segundo esquema utiliza 2 HMMs: uno para modelar el carácter de espacio en blanco, y otro para modelar los restantes 77 caracteres conjuntamente.

La topología de los HMMs es idéntica en ambos esquema: izquierda-derecha de 6 estados y 64 Gaussianas en la mixtura de cada estado.

## • RESULTADOS:

	78-HMMs	2-HMMs
AER (%)	7.20	25.98
$\mu$ (mm)	1.15	2.95
$\sigma$ (mm)	3.90	6.56

# Resultados Experimentales

- Dos esquemas de modelado diferentes de HMMs, fueron empleados:
  - El primer esquema modela cada uno de los 78 clases de caracteres utilizando diferentes HMMs por clase.
  - El segundo esquema utiliza 2 HMMs: uno para modelar el carácter de espacio en blanco, y otro para modelar los restantes 77 caracteres conjuntamente.

La topología de los HMMs es idéntica en ambos esquema:  
izquierda-derecha de 6 estados y 64 Gaussianas en la mixtura de cada estado.

- **RESULTADOS:**

	78-HMMs	2-HMMs
AER (%)	7.20	25.98
$\mu$ (mm)	1.15	2.95
$\sigma$ (mm)	3.90	6.56

# Ejemplo de Marcas Computadas de Alineamiento

