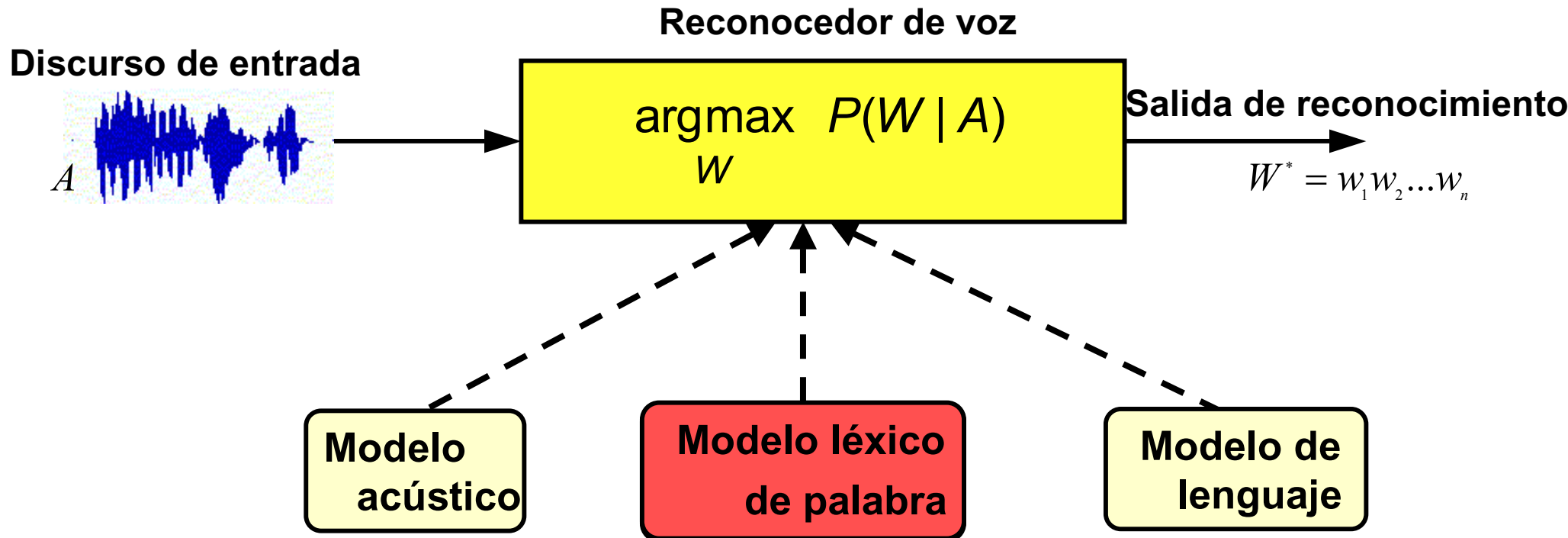


Modelando nuevas palabras

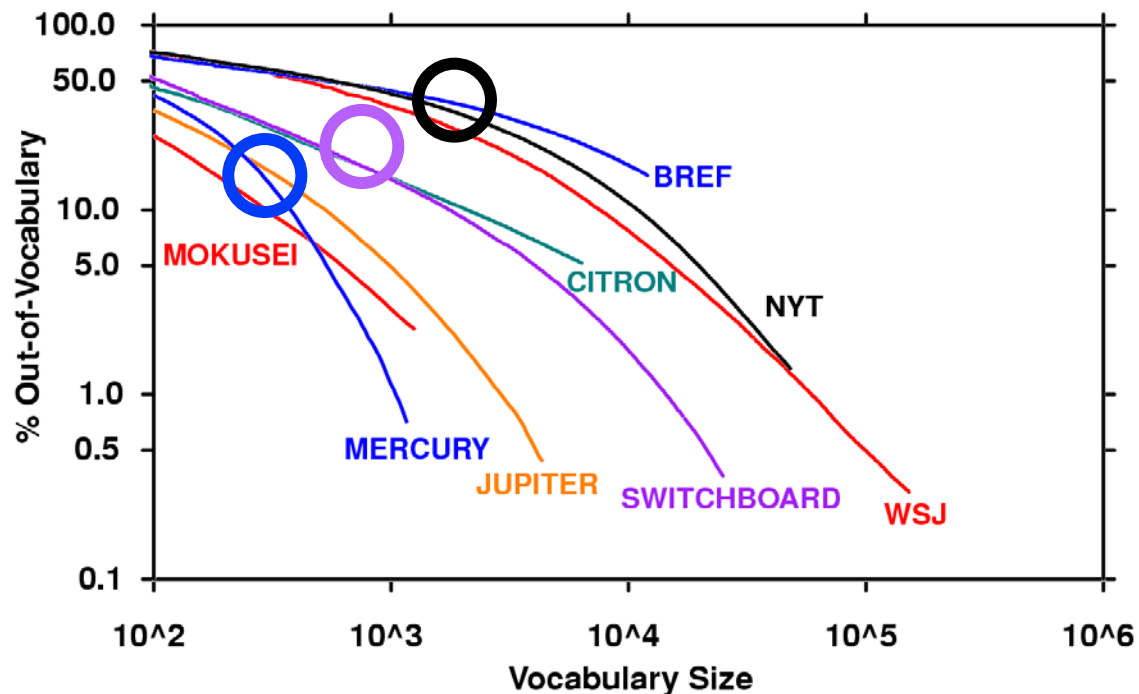
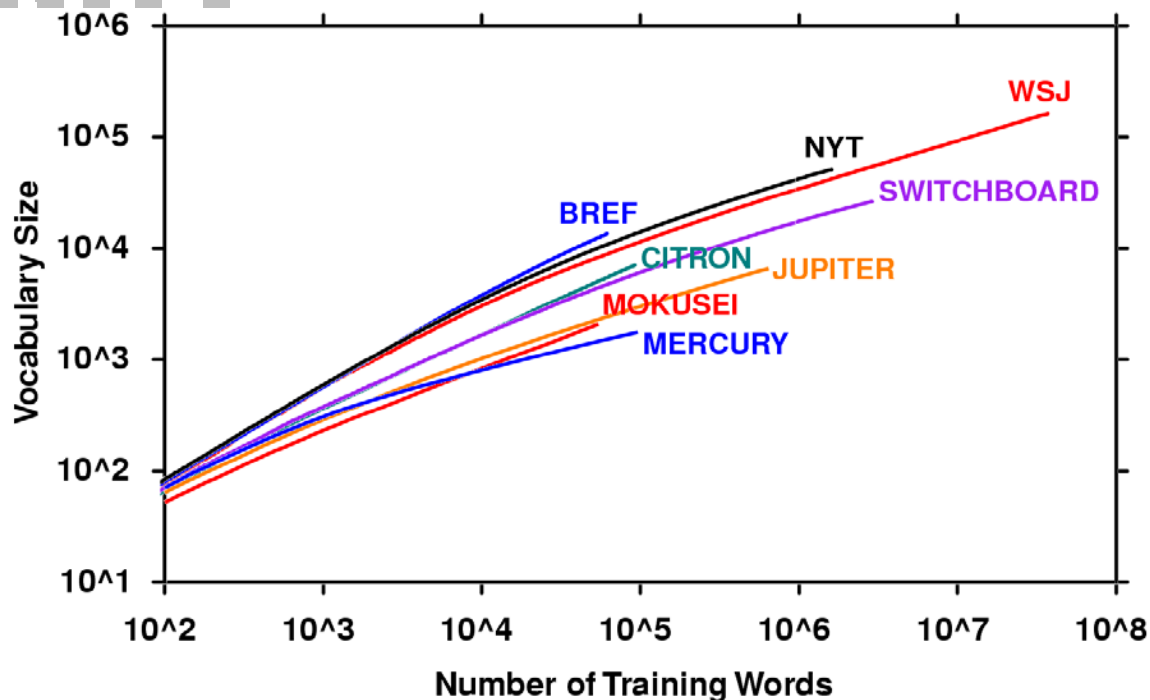
- **Introducción**
- **Modelando palabras no presentes en el vocabulario (OOV)**
 - **Formulación probabilística**
 - **Métodos independientes del dominio**
 - **Aprendizaje de unidades de subpalabra OOV**
 - **Modelos de OOV multiclase**

¿Qué es una nueva palabra?



- **Casi todos los reconocedores de voz buscan un léxico finito**
 - Una palabra no contenida por el léxico es referida como palabra no presente en el vocabulario
 - Las palabras no presentes en el vocabulario (OOV) son inevitables a la vez que problemáticas

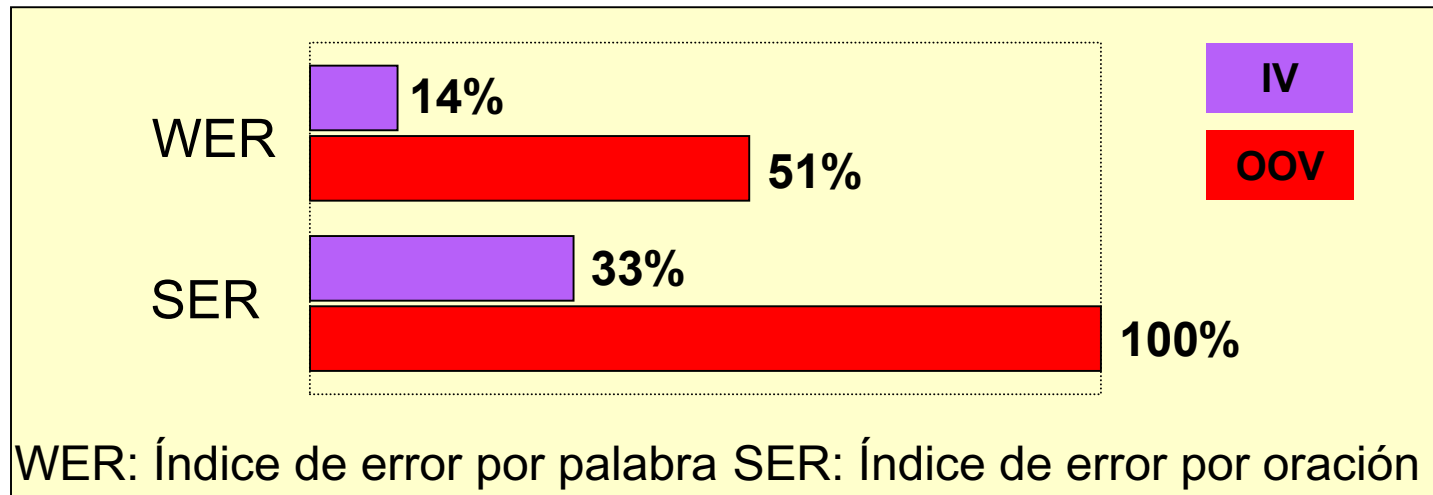
Las nuevas palabras son inevitables



- **Análisis de discurso múltiple y corpus de texto**
 - Tamaño del vocabulario frente a cantidad de datos de entrenamiento
 - Índice no presente en el vocabulario frente a tamaño del vocabulario
- **El crecimiento del vocabulario parece ilimitado**
 - Nuevas palabras van apareciendo constantemente
 - El crecimiento parece ser independiente del lenguaje
- **Índice de error para palabras no presentes en el vocabulario, una función de tipo de datos**
 - **Discurso entre humano y máquina**
 - **Discurso de humano a humano**
 - **Texto de un periódico**

Las nuevas palabras causan errores

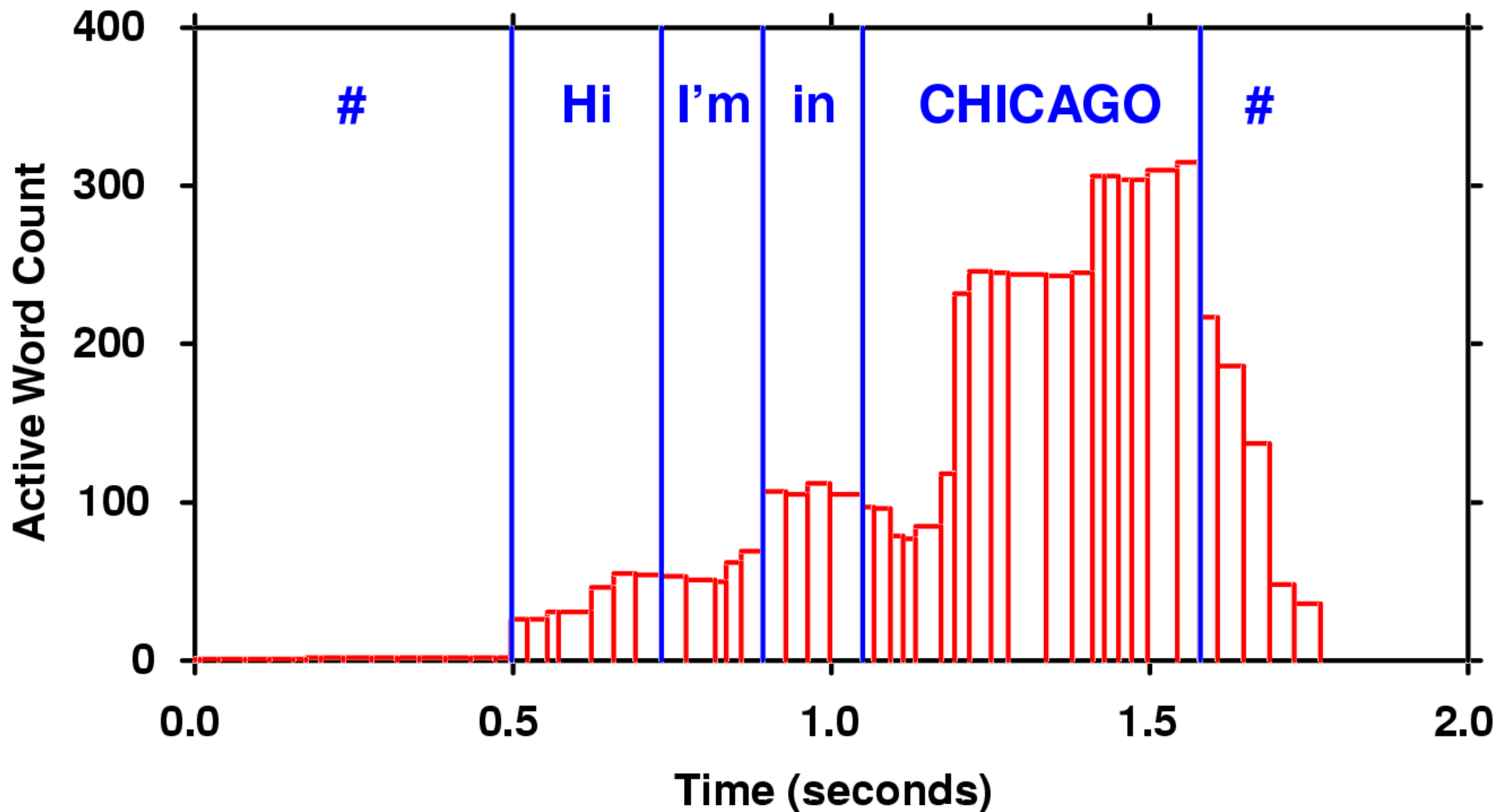
- Las palabras no presentes en el vocabulario (OOV) presentan un índice de error más alto a nivel de palabra y oración, que las palabras presentes en el vocabulario (IV)



- Las palabras OOV normalmente generan múltiples errores, ej., “Symphony”
Ref.: “Members of Charleston Symphony Orchestra are being treated...”
Hip.: “Members of Charleston simple your stroke are being treated...”

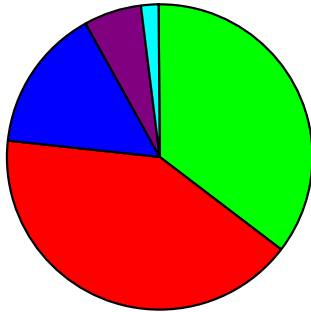
Las palabras nuevas bloquean a los reconocedores

- La computación de la búsqueda aumenta la presencia cercana de nuevas palabras



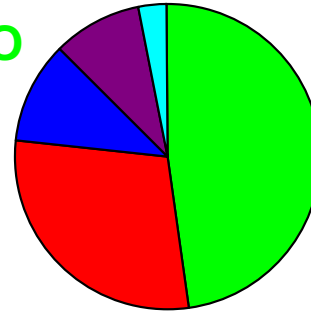
Las nuevas palabras son importantes

- Las nuevas palabras son normalmente palabras de contenido importantes



Weather (tiempo)

NOMBRE PROPIO
SUSTANTIVO
VERBO
ADJETIVO
ADVERBIO



Broadcast News(parte meteorológico)

- Es más probable que las palabras de contenido sean reutilizadas (ej., *persistent*)

- **Cuatro desafíos con nuevas palabras:**
 - 1) **Detectar** la presencia de la palabra
 - 2) Determinar su **lugar** dentro del enunciado
 - 3) Reconocer la **secuencia fonética** subyacente
 - 4) Identificar la **ortografía** de la palabra
- **Aplicaciones para nuevos modelos de palabras:**
 - Mejorar el reconocimiento, detectar errores de reconocimiento
 - Manejar palabras parciales
 - Procesar estrategias de diálogo
 - Incorporar dinámicamente nuevas palabras en el vocabulario

- **Aumentar el tamaño del vocabulario**
- **Emplear una puntuación de seguridad para detectar palabras OOV**
- **Utilizar unidades de subpalabra en la primera de un sistema de dos fases**
- **Incorporar un modelo de palabras desconocido en un reconocedor de voz**
 - **Una extensión de un modelo rellenedor o basurero para las no palabras**

Incorporación de un modelo OOV en ASR (Bazzi, 2002)

- Espacio de búsqueda híbrida: una unión de espacios de búsqueda IV y OOV

1) Comenzar con una red léxica estándar

2) Construir una red de subpalabra separada

3) Añadir una red de subpalabra a una red de palabras como una nueva palabra,

W_{OOV}

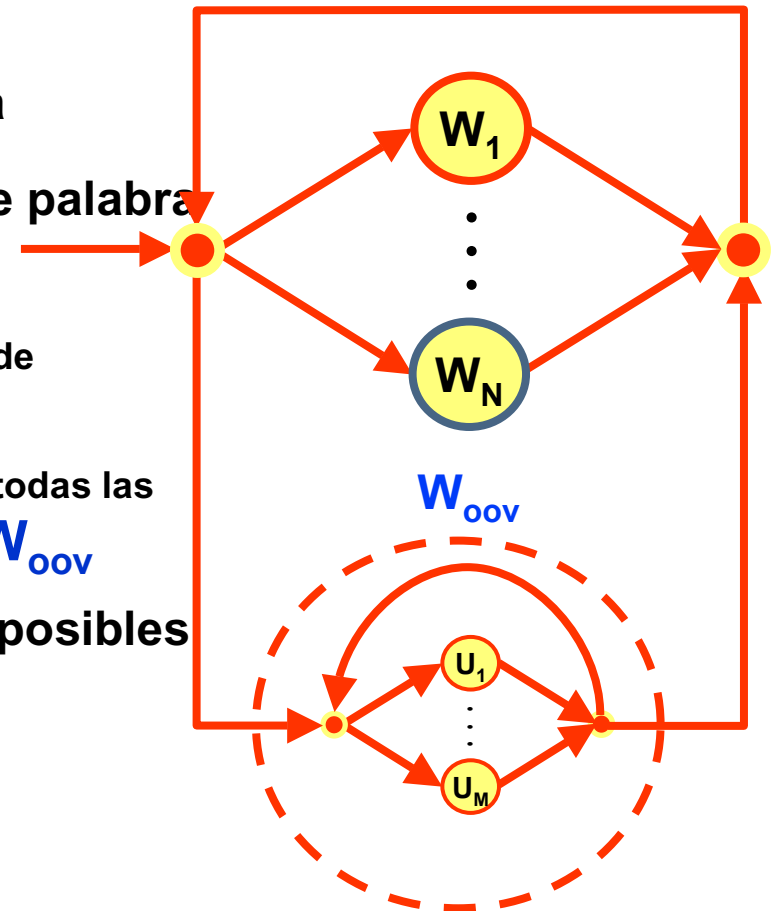
- El coste C_{OOV} , se añade para controlar el índice de detección OOV

- Durante el entrenamiento del modelo de lenguaje, todas las palabras OOV van asociadas a la etiqueta W_{OOV}

- Una variedad de unidades de subpalabra son posibles (ej., fonos, sílabas ...)

- Una variedad de restricciones topológicas

- Restricciones fonético-acústicas
- Restricciones de duración
- ...



El modelo de probabilidad de OOV

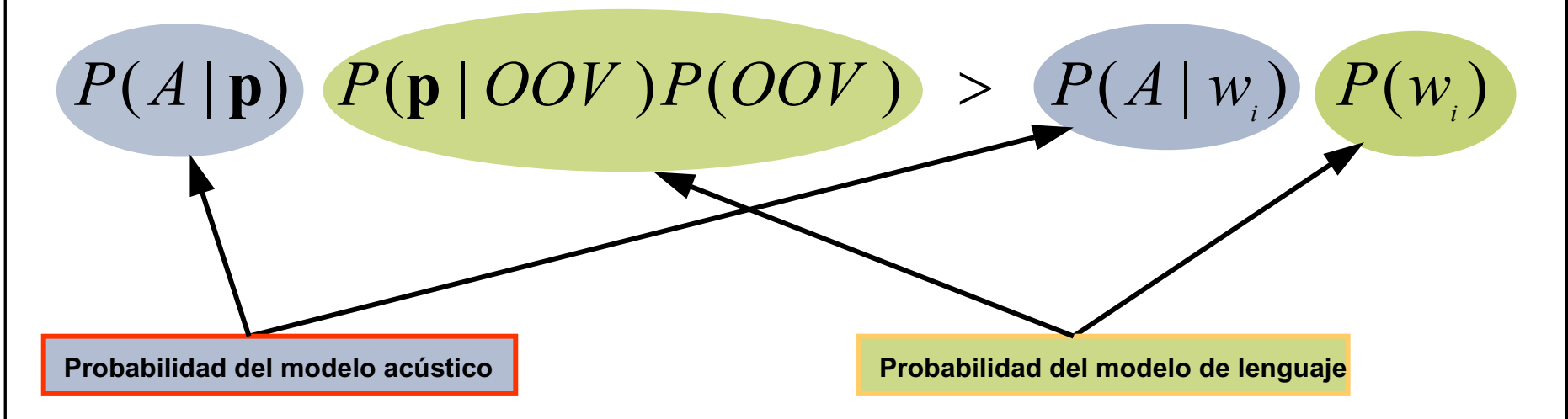
- El modelo de probabilidad estándar:

$$W^* = \arg \max_W P(A | W) P(W)$$

- Modelos acústicos: igual para palabras IV y OOV
- Modelos de lenguaje: se utiliza una clase n -grama para palabras OOV

Una palabra OOV es planteada como hipótesis si:

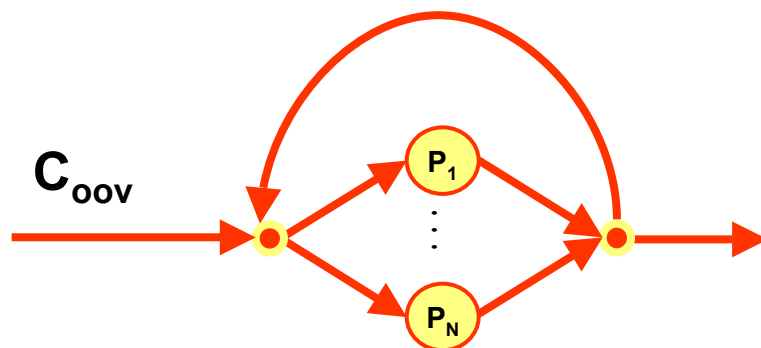
$\exists \mathbf{p} = p_1 p_2 \dots p_N$, tal que $\forall w_i \in L$



- **Comparado con modelos de relleno**
 - **Mismos modelos acústicos para palabras IV y OOV**
 - * Las estimaciones de probabilidad son comparables
 - **Modelo de lenguaje de subpalabra**
 - * Estimado para el propósito de reconocimiento de palabras OOV
 - **Modelo de lenguaje a nivel de palabra prediciendo la palabra OOV**
 - **Uso de grandes unidades de subpalabra**
 - **Todo lo mencionado dentro de un marco sencillo**
- **Lo mejor de ambos mundos: los rellenos y las dos fases**
 - **Temprana utilización del conocimiento léxico (rellenos)**
 - **Modelado subléxico detallado (dos fases)**

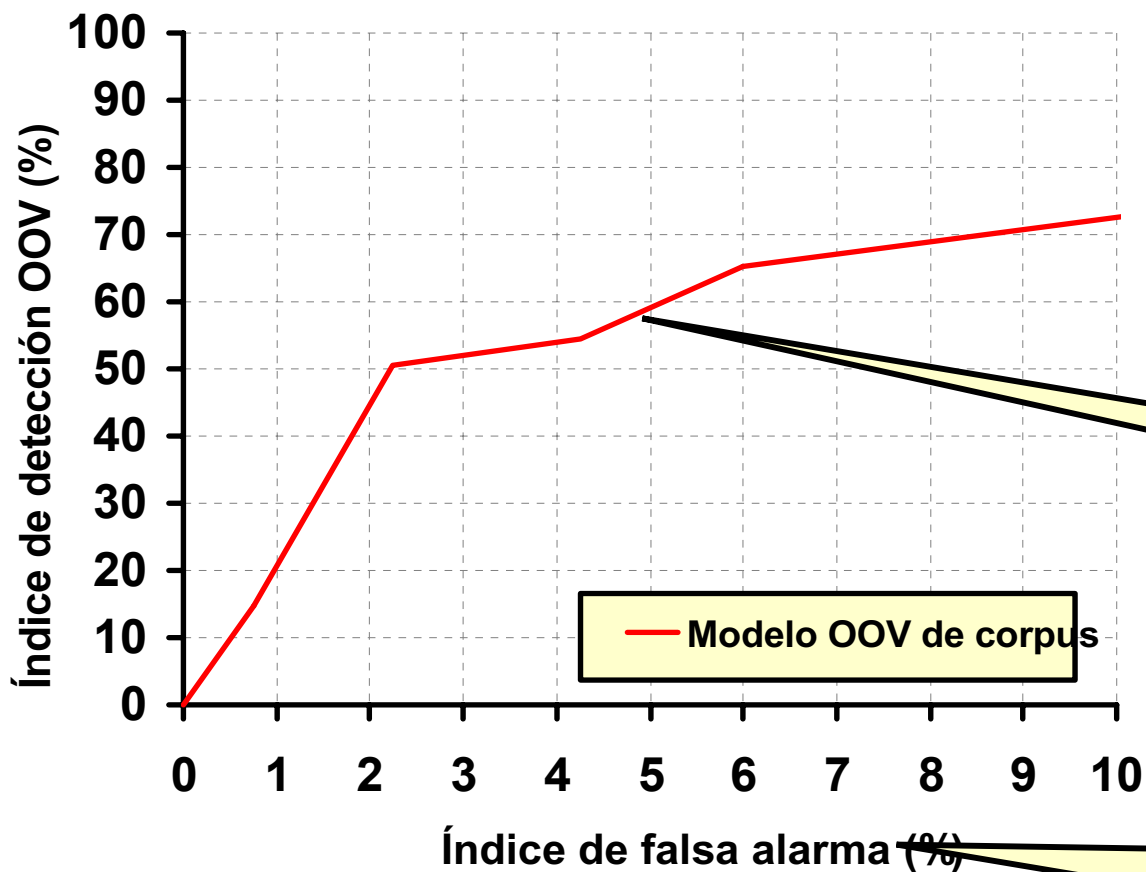
Un modelo de OOV basado en corpus

- El modelo de OOV basado en corpus usa una configuración de reconocimiento de fonos típica
 - Se permite cualquier secuencia de fonos de cualquier longitud
 - Durante el reconocimiento, las secuencias de fonos están limitadas por una n -grama de fonos
 - La n -grama de fonos es calculada desde el mismo **corpus de entrenamiento** utilizado para entrenar al reconocedor de palabras



- **Los experimentos emplean el reconocedor del sistema de información meteorológica JUPITER**
 - **Reconocedor SUMMIT basado en segmentos**
 - **Modelos difonos dependientes del contexto**
 - **88,755 enunciados de datos de entrenamiento**
 - **2,009 palabras en el vocabulario del reconocedor**
 - **Índice de OOV: 2.2% (15.5% a nivel de enunciado)**
 - **El modelo de OOV utiliza un bigrama de fonos**
- **Los experimentos emplean 2,029 enunciados de prueba de llamadas a JUPITER**
 - **1,715 enunciados sólo con palabras IV**
 - **314 enunciados contienen palabras OOV**

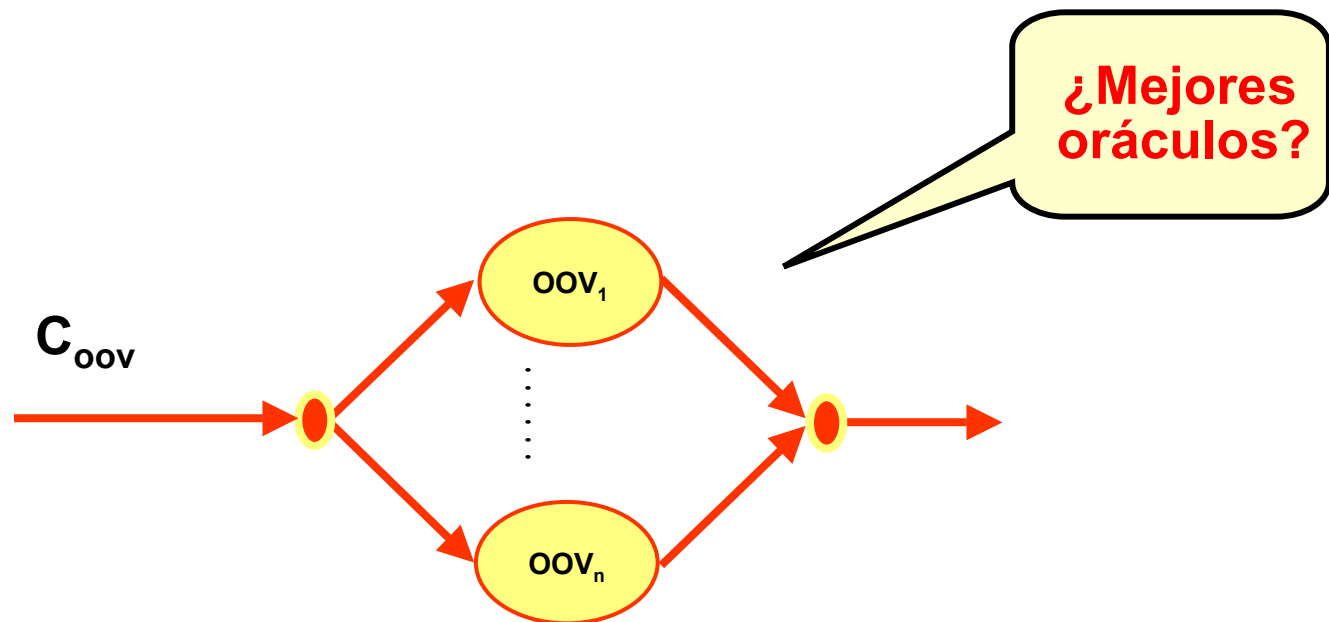
Resultados de la detección del modelo OOV de corpus



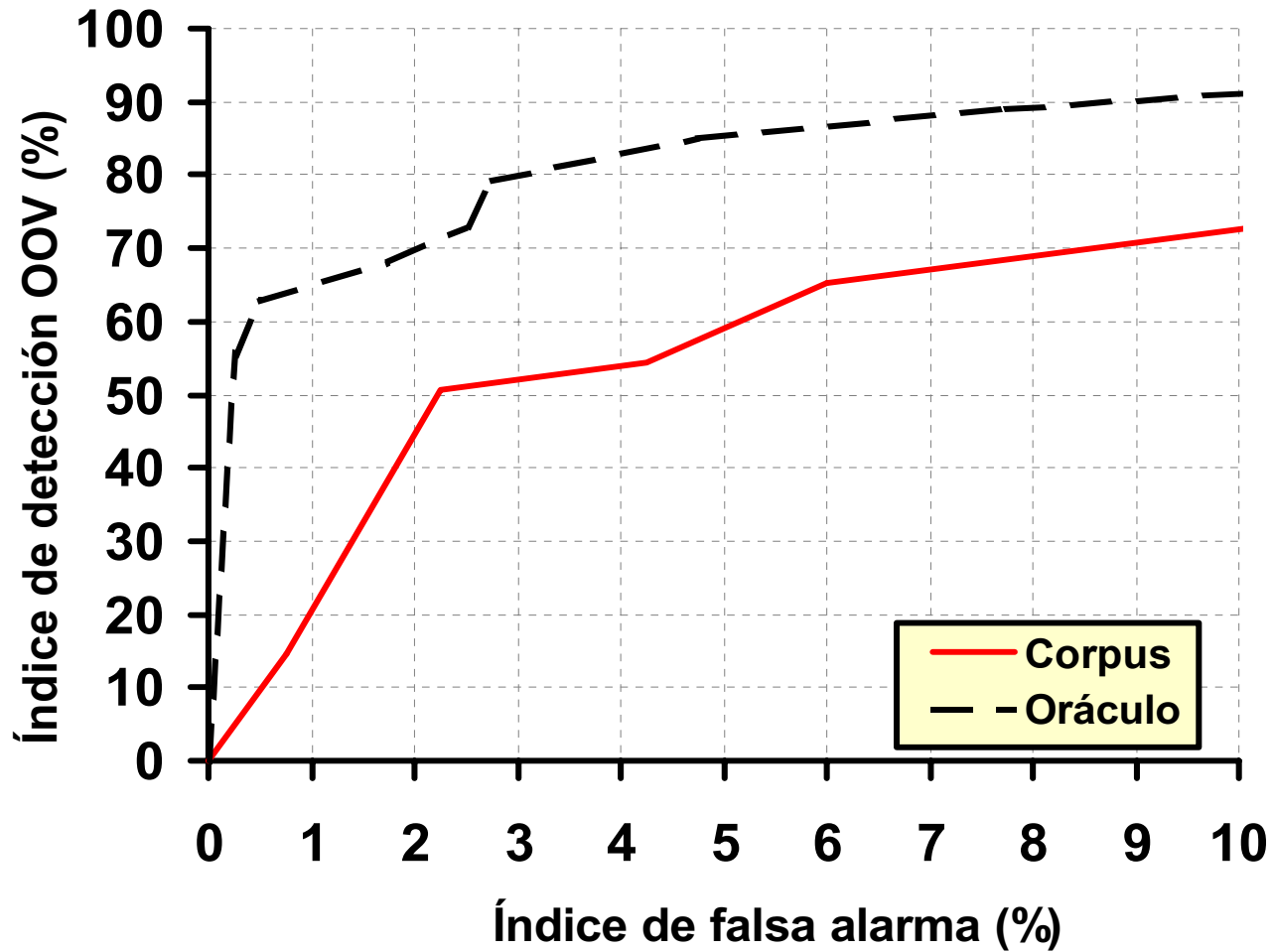
- La mitad de las palabras OOV detectadas presentan un 2% de falsa alarma
- En un índice de detección del 70%, la falsa alarma es del 8.5%

EL modelo de OOV oráculo

- **Objetivo:** cuantificar el mejor rendimiento posible con el marco propuesto
- **Enfoque:** construir un modelo de OOV que permita sólo las secuencias de fonos de palabras OOV del conjunto de pruebas
- La configuración oráculo no es equivalente a la adición de palabras OOV al vocabulario

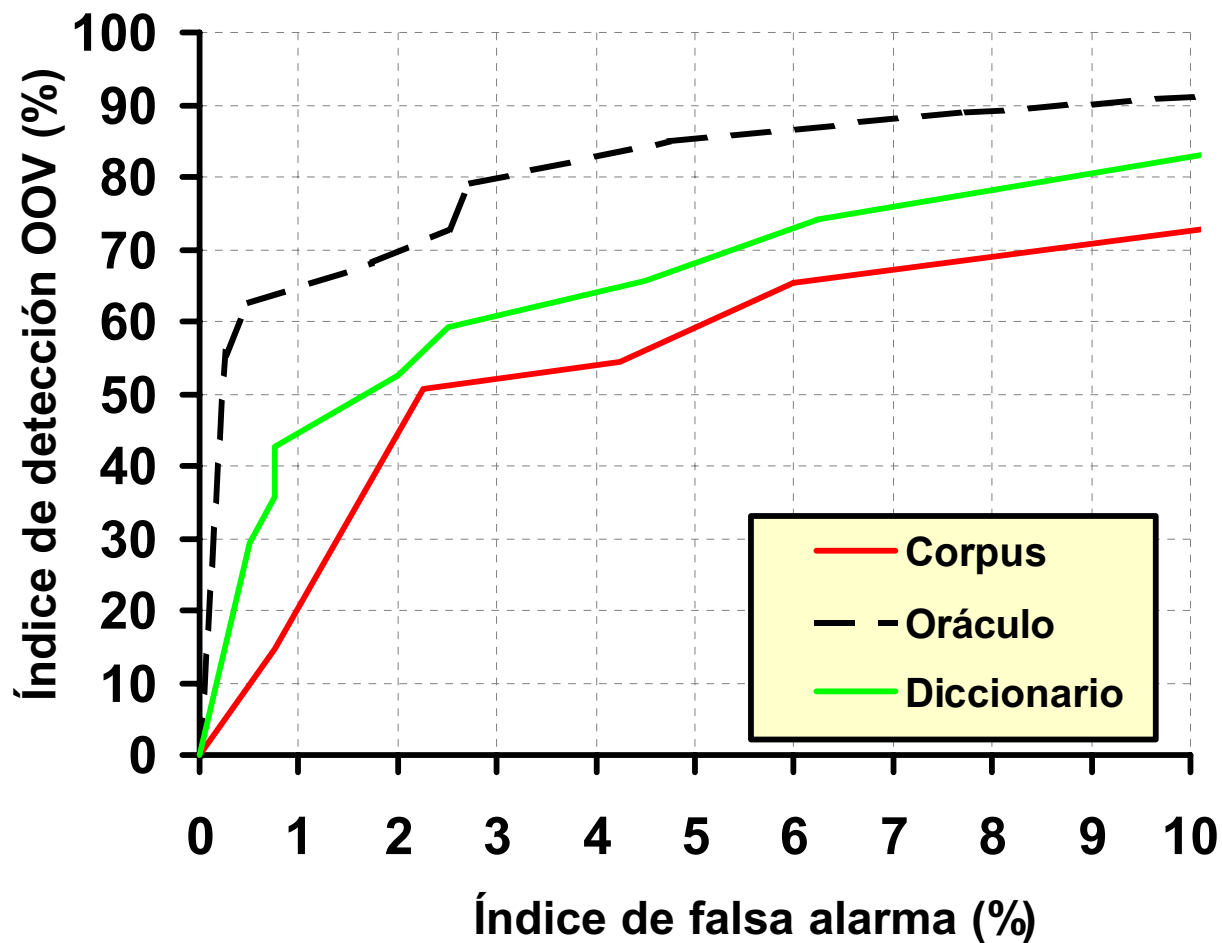


Detección de resultados del modelo de OOV oráculo



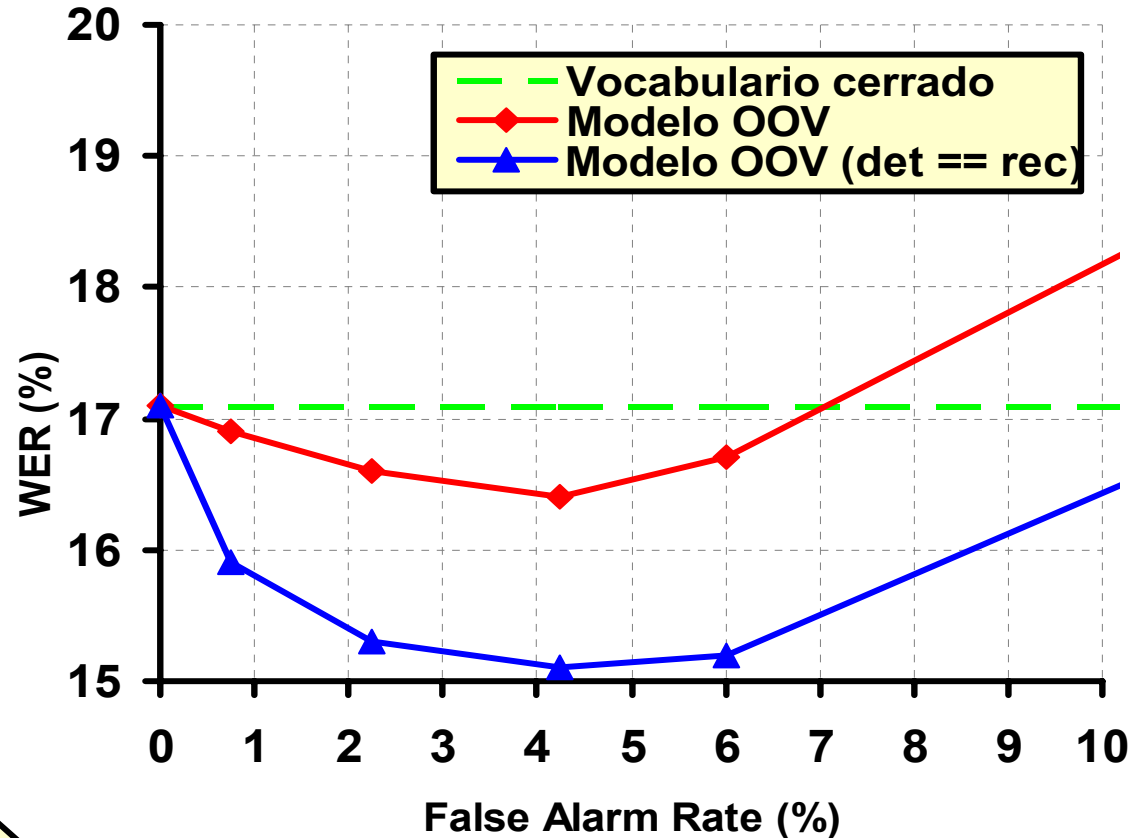
Considerable espacio para una mejora

- **Inconvenientes del modelo corpus**
 - Favorece a las palabras más frecuentes dado que está entrenado con transcripciones fonéticas de enunciados completos
 - Dedicar una porción de la masa de probabilidad n -grama para las secuencias entre palabras
 - El modelo de OOV dependiente del dominio podría no generalizar
- **Un modelo OOV diccionario se construye desde un diccionario de palabras genérico en vez de a partir de un corpus de enunciados**
 - Elimina la dependencia del dominio y el margen de error de las palabras frecuentes
- **En los experimentos se emplea el diccionario LDC PRONLEX**
 - **90.694 palabras con un total de 99.202 pronunciaciones**



En un índice de detección del 70%, el índice de falsa alarma queda reducido de 8.5% a 5.3%

Impacto en el índice de error por palabra

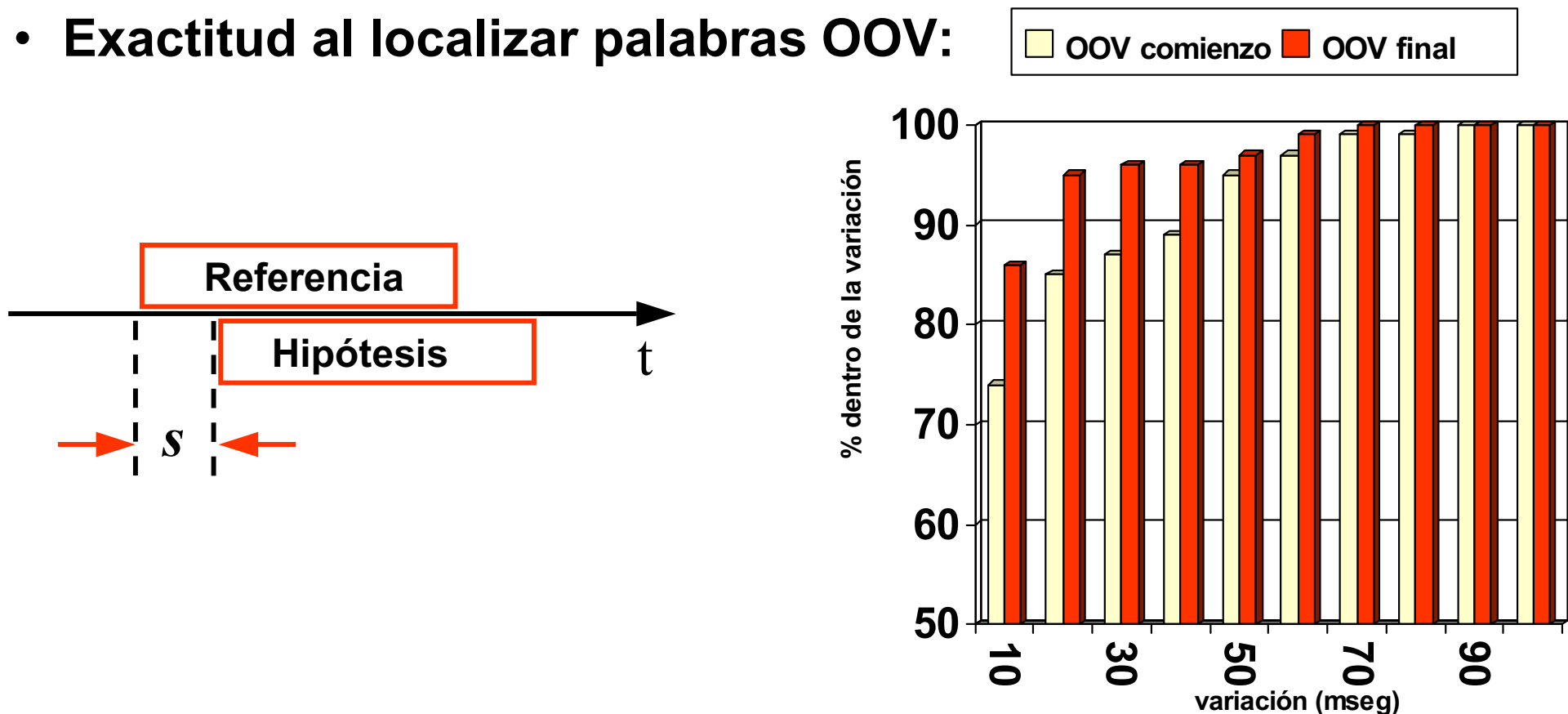


¿Qué ocurre con los datos de prueba IV?

- La WER en un conjunto de pruebas completos es reducida de un 17.1% a un 16.4%
- La WER puede reducirse de un 17.1% a un 15.1% con un mecanismo de identificación

Otras medidas de rendimiento

- Exactitud al localizar palabras OOV:



- Tasa de error fonético de OOV (PER):

PER	Sustituciones	Inserciones	Eliminaciones
37.8%	18.9%	6.0%	12.9%

¿Sirve?
bien?

- **Objetivo:** incorporar restricciones estructurales adicionales para reducir la falsa hipótesis de las palabras OOV
- **Idea:** restringir el reconocimiento de la red OOV para unidades multi-fono específicas

¿Cómo obtenemos el conjunto de unidades multifono?

- **Un enfoque orientado a datos:** medir la estadística de coaparición de fonos (ej., información mutua) dentro de un amplio diccionario para proponer incrementalmente nuevas unidades multifono

Aprendizaje de unidades multifono

- Un algoritmo iterativo ascendente
 - Comienza con fonos individuales
 - Combina iterativamente pares de unidades para formar unidades más largas
- El criterio para combinar pares de unidades se basa en la información mutua de pesos (MI_w) de un par:

$$MI_w(u_1, u_2) = p(u_1, u_2) \log \frac{p(u_1, u_2)}{p(u_1)p(u_2)}$$

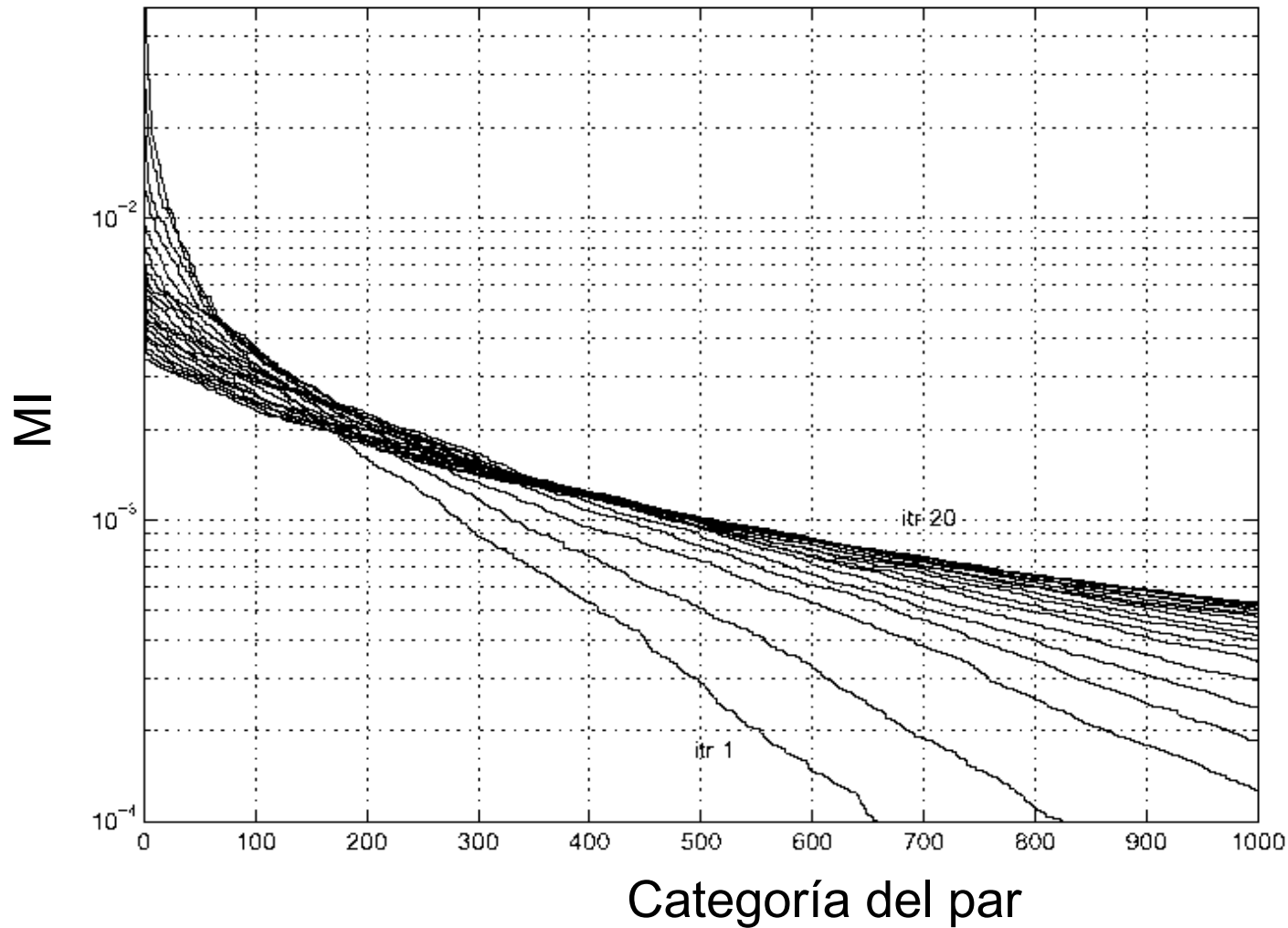
- En cada iteración, se combinan los pares n con el MI_w más alto
- El número de unidades multifono derivado depende del número de iteraciones
- Un subproducto es un análisis completo de todas las palabras del vocabulario en función de las unidades aprendidas

Resultados MMI

- El conjunto inicial de unidades es el conjunto del fono (62 fonos)
- El tamaño del inventario de la unidad final es de 1.977 unidades (después de 200 iteraciones y 10 combinaciones por iteración)
- La perplejidad del modelo OOV desciende desde 14'0 para el conjunto del fono inicial hasta 7'1 para el conjunto multifono derivado
- El 67% de las unidades derivadas son sílabas inglesas legales
- La longitud media de una unidad derivada es de 3'2 fonos
- Ejemplos:

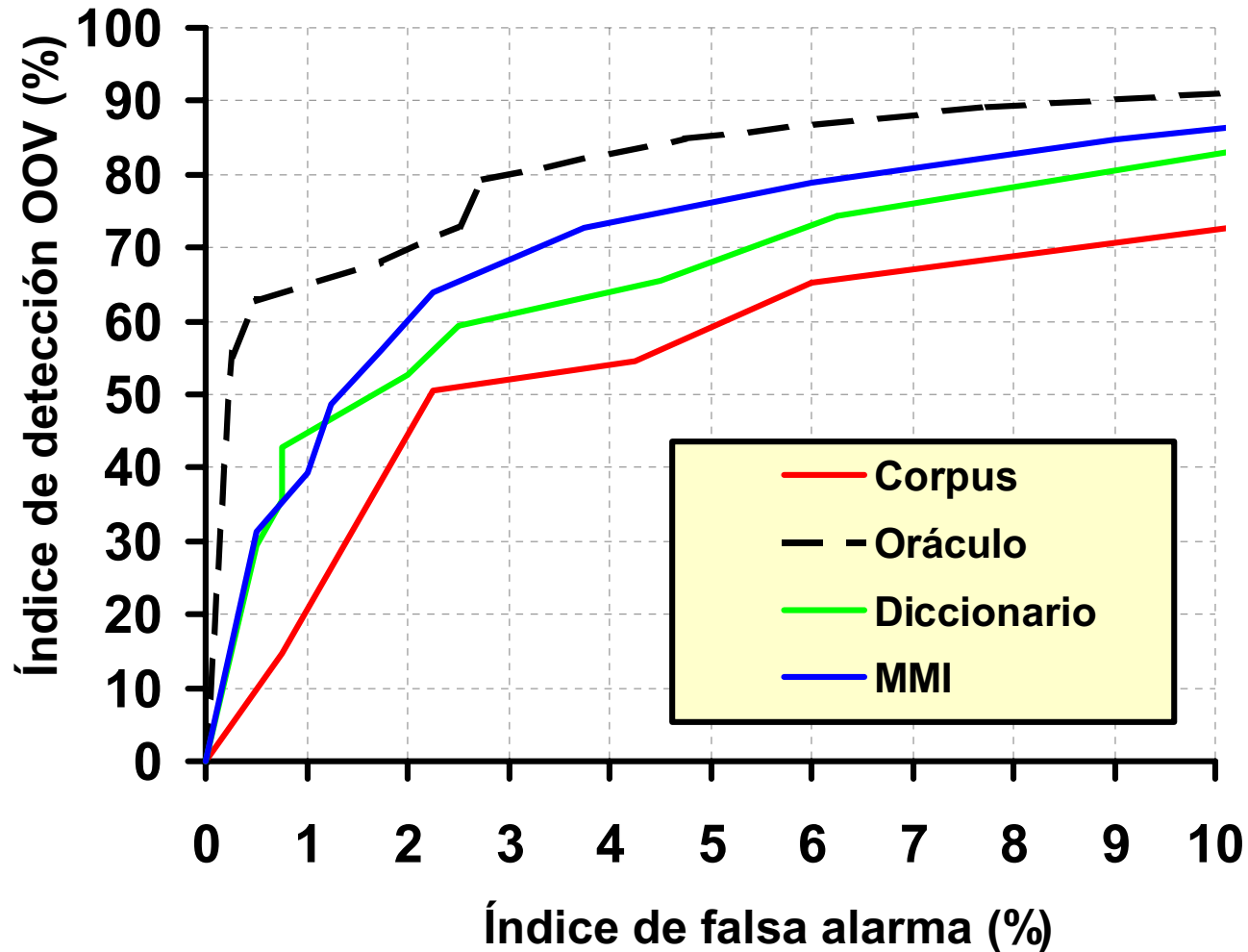
Palabra	Pronunciación
whisperers	(w_ ih) (s) (p_ ax_ r) (axr_ z)
yugoslavian	(y_ uw) (g_ ow) (s_ l_ aa) (v_ iy) (ax_ n)
shortage	(sh_ ao_ r) (tf_ ax) (jh)

Comportamiento del agrupamiento MMI



Niveles MI externos para pares de clasificación superior; después de varias iteraciones (puede ser útil como criterio de parada)

Resultados de detección del modelo de OOV MMI



- En un índice de detección del 70%, el índice de falsa alarma es reducido a un 3'2%
- El índice de error fonético es reducido de un 37'8% a un 31'2%

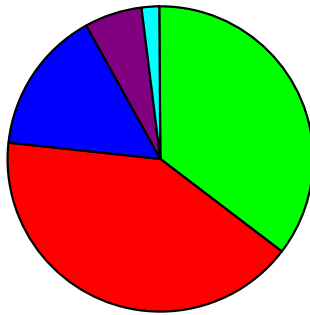
Figura de mérito de la detección de OOV

- La figura de mérito (FOM) mide el área bajo el primer 10% y el 100% total de la curva ROC
- La FOM aleatoria muestra el rendimiento para un modelo OOV aleatoriamente adivinatorio (ROC es la diagonal $y=x$)

Modelo de OOV	100% FOM	10% FOM
Corpus	0.89	0.54
Diccionario	0.93	0.64
MMI	0.95	0.70
Oráculo	0.97	0.80
Aleatorio	0.50	0.10

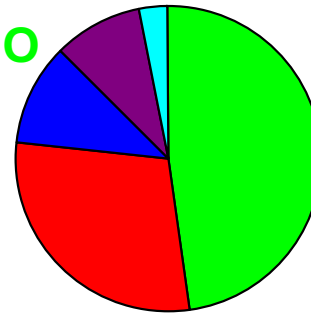
Un modelo de OOV multiclase

- **Motivación:** modelado superior de clases de palabras desconocidas
 - En el nivel fonético: estructura fonotáctica semejante
 - En el nivel del modelo de lenguaje: patrones de uso lingüístico similar



Weather (tiempo)

NOMBRE PROPIO
SUSTANTIVO
VERBO
ADJETIVO
ADVERBIO

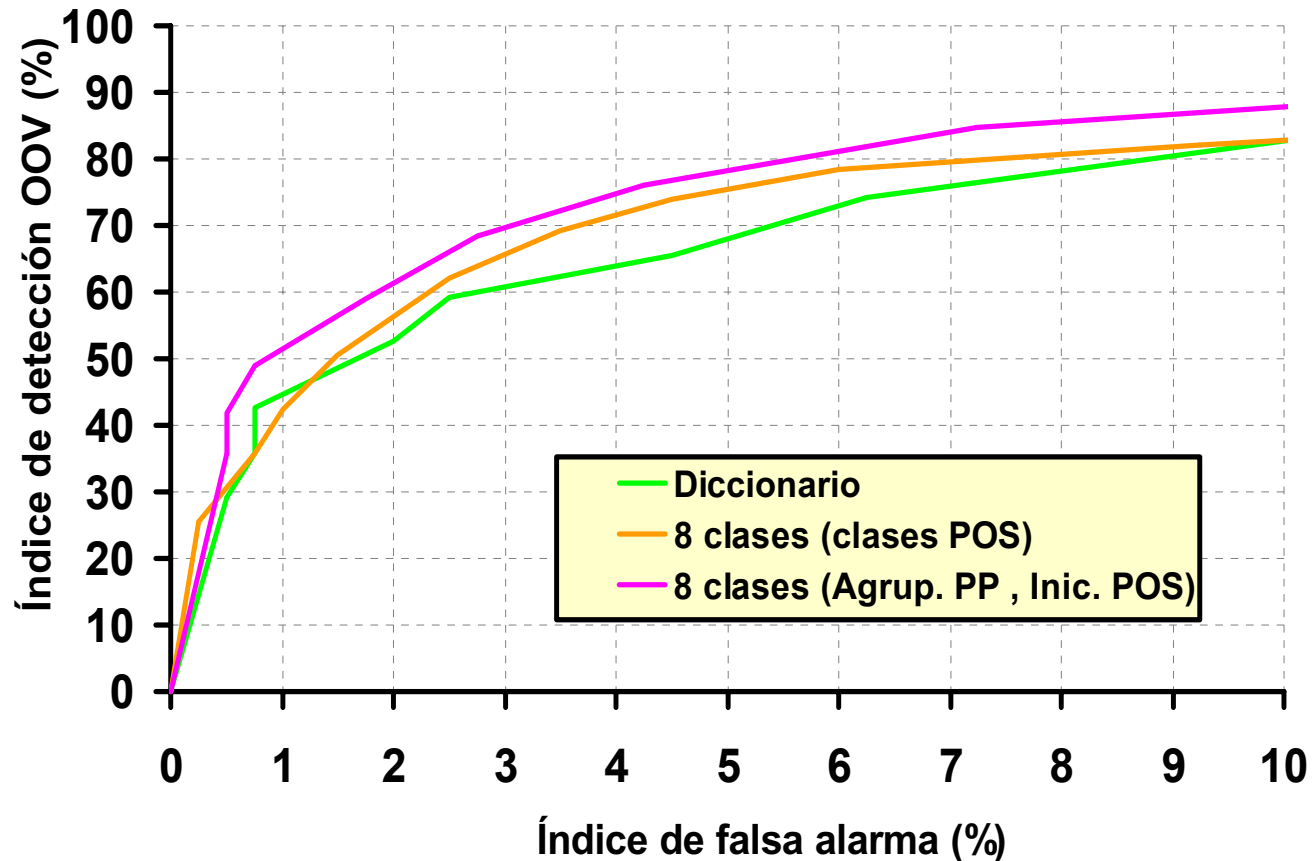


Broadcast News (parte meteorológico)

- **Enfoque:** extender el marco de OOV para modelar categorías múltiples de palabras desconocidas
 - Una colección de redes de OOV en paralelo con la red IV
 - La Gramática G_N a nivel de palabras predice múltiples clases de OOV

Experimentos multiclase

- **Trabajos de clase en función de etiquetas de partes del discurso**
 - **Derivado de un diccionario etiquetado de palabras (LDC COMLEX)**
 - **Modelo de lenguaje a nivel de palabras entrenado con 8 clases POS**
 - **Múltiples modelos de lenguajes de subpalabra para las distintas clases POS**
- **Trabajos de clase basados en el agrupamiento de perplejidad**
 - **Crear un modelo de lenguaje de bigrama de fono a partir de grupos iniciales**
 - **Usar el agrupamiento K-medias para cambiar palabras de un grupo a otro**
 - **En cada iteración, cada palabra es trasladada al grupo de menor perplejidad (mayor probabilidad)**



- El método multiclase supera al modelo OOV de diccionario
- El modelo POS consigue el 81% de exactitud en la identificación de la clase
- El agrupamiento de perplejidad funciona mejor que las clases POS

Condición/FOM	G_1 n -grama	G_8 n -grama
1 red OOV	0.64	0.65
8 redes OOV	0.68	0.68

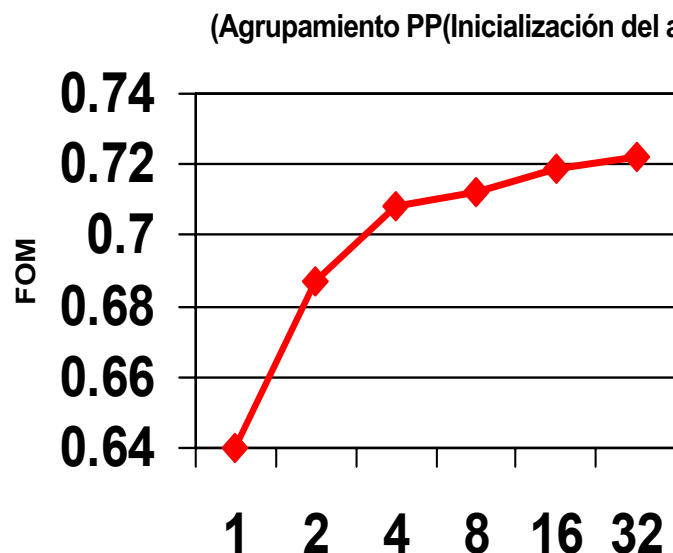
- **La mayor parte del beneficio proviene de redes múltiples OOV**
 - La fonotáctica es más importante que las restricciones del modelo de lenguaje
- **El comportamiento puede ser distinto para otros dominios**

Derivando multiclases mediante agrupamiento

- El agrupamiento puede utilizarse para proponer multiclases iniciales
 - Agrupamiento ascendente para inicializar el trabajo de clase de palabras
 - Distancia métrica basada en la similitud del bigrama de fono
 - Una medida de similitud media se utiliza para combinar grupos:

$$d_{avg}(X_m, X_n) = \frac{1}{C_m C_n} \sum_{w_i \in X_m} \sum_{w_j \in X_n} d(w_i, w_j)$$

- Un número arbitrario de clases pueden ser agrupadas
- Las clases se pueden suavizar con el agrupamiento de perplejidad



Modelo	Clases	10% FOM
Diccionario	1	0.64
Clases POS	8	0.68
PPClus (AggClus Init)	8	0.71
PPClus (POS Init)	8	0.72

Otras áreas de investigación relacionadas

- **Medición del impacto en el reconocimiento de OOV para la comprensión**
- **Mejorar la exactitud fonética de OOV**
- **Extender el enfoque para modelar enunciados no pertenecientes al dominio**
- **Desarrollo de puntuaciones de seguridad específicas de OOV**
 - **Mejorar la calidad de detección**
- **Modelar otros tipos de sonidos no pertenecientes al dominio (ej., ruido)**

A. Asadi, “Automatic detection and modeling of new words in a large vocabulary continuous speech recognition system,” tesis doctoral, Northeastern University, 1991.

I. Bazzi, “Modelling out-of-vocabulary words for robust speech recognition,” tesis doctoral, MIT, 2002.

G. Chung, “Towards multi-domain speech understanding with flexible and dynamic vocabulary,” tesis doctoral, MIT, 2001.

L. Hetherington, “The problem of new, out-of-vocabulary words in spoken language systems,” tesis doctoral, MIT, 1994.