

Medidas de Confianza

Rubén San Segundo Hernández

(5 Febrero 2004)



Introducción



RECONOCIMIENTO

COMPRENSIÓN
DE LENGUAJE

Medidas de Confianza

- Análisis de medidas de confianza al nivel de palabra, concepto y frase en el sistema CU Communicator
- Aplicación de medidas de confianza para la mejora de la tasa de reconocimiento

Diagrama de un **Servidor Vocal Interactivo (SVI)**

Medidas de Confianza (I)

Tipos de parámetros para Estima de Confianza:

- ▣ **Parámetros de Decodificación** (Chase,1997a; Wessel,1999; Moreau, 1999;Jouvet,1999a; Gunawardana,1999; Kamppari,2000; Dasmahapatra,2000; Moreau,2000; Macías-Guarasa,2000b; Wessel,2001; Charlet,2001)
- ▣ **Parámetros exclusivos del Modelo de Lenguaje** (Uhrik,1997; San-Segundo,2000b)
- ▣ **Parámetros de Comprensión** (Bouwman, 1999; Souvignier,2000; Hazen, 2000a; Hazen, 2000b; San-Segundo, 2001a; Zhang, 2001)
- ▣ **Parámetros del Gestor de Diálogo** (Carpenter et al, 2001)

Medidas de Confianza (II)

Resolución de las Medidas de Confianza:

☞ **Nivel de Palabra** (Chase,1997a; Wessel,1999; Moreau, 1999; Jouvét,1999a; Gunawardana,1999; Kamppari,2000; Dasmahapatra,2000; Moreau,2000; Macías-Guarasa,2000b; Wessel,2001; Charlet,2001)

☞ **Nivel de Concepto Semántico** (Bouwman, 1999; Souvignier, 2000; Hazen, 2000a; Hazen, 2000b; San-Segundo, 2001a; Zhang, 2001)

☞ **Nivel de Frase** (Uhrík y Ward, 1997; Pao, 1998; San-Segundo et al, 2000b, San-Segundo, 2001a)

☞ **Nivel de Interacción** (Carpenter et al, 2001)

Análisis de medidas de confianza

Sistemas analizados

CU Communicator

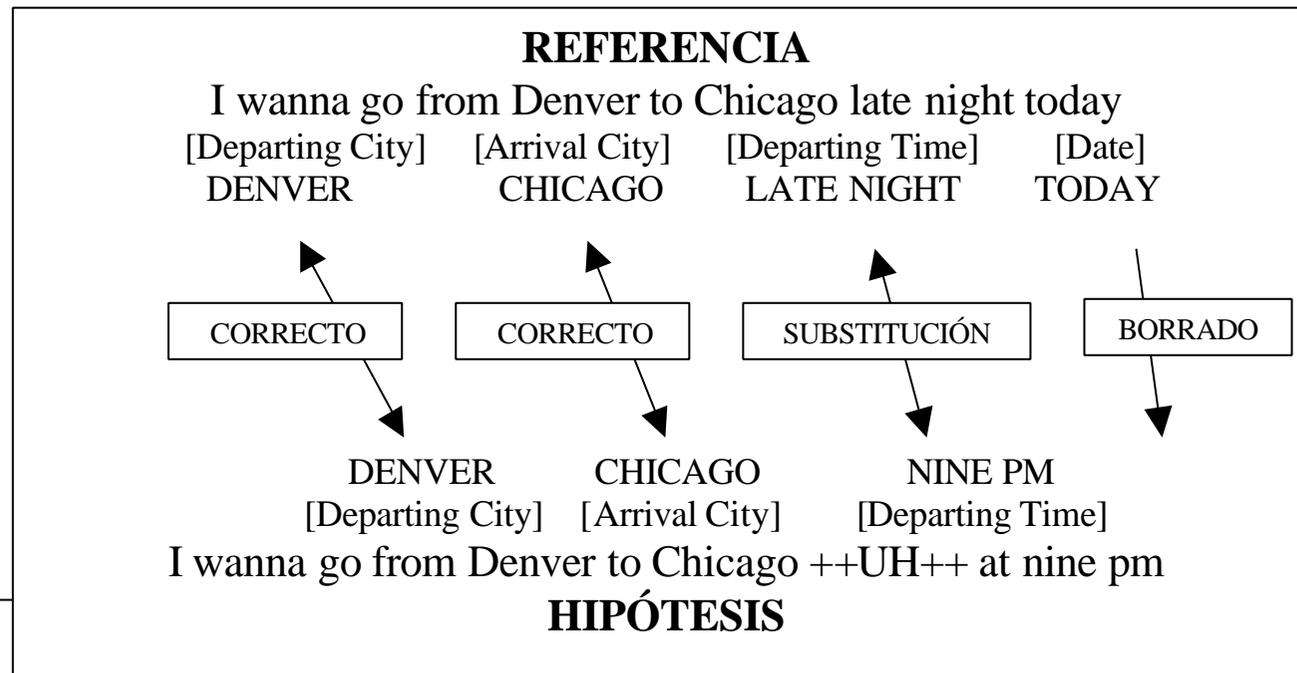
- Detección de errores al nivel de palabra, concepto y frase
- Parámetros acústicos, lingüísticos y semánticos
- Aplicación de medidas de confianza para la recuperación de errores al nivel de palabra y de concepto

Etiquetación de Ejemplos (1)

Necesidad de etiquetar los ejemplos como “A Rechazar” (errores o OOV) o “A Aceptar” (aciertos)

Nivel de Palabra: comparando hipótesis y referencia (ins, borr, subs y aciertos). Para los borrados no se estima confianza.

Nivel de Concepto: comparando las secuencias de conceptos de la hipótesis y de la frase de referencia.



Etiquetación de Ejemplos (II)

Nivel de Frase: podríamos considerar varios criterios:

1.- Detección de frases **Fuera del Dominio**. Etiquetación manual.

2.- (Pao, 1998): **Análisis Semántico de las N mejores hipótesis:** Rechazamos la frase si el análisis semántico de alguna de las N mejores hipótesis no coincide con la frase de referencia

3.- (Hazen, 2000b): **Tasa de reconocimiento:** Rechazamos la frase si alguna de las 3 primeras hipótesis no coincide con la frase de referencia y la tasa de la primera hipótesis es menor de 70%.

4.- (San-Segundo, 2001a): **Análisis Semántico:** Rechazamos la frase si todos los conceptos semánticos son erróneos en comparación con la frase de referencia

Nivel de Palabra: parámetros (I)

Del Proceso de Decodificación

- **Verosimilitud normalizada:** logaritmo de la verosimilitud acumulada a lo largo de la palabra bajo análisis.
- **Homogeneidad de la palabra en las 100 mejores hipótesis:** porcentaje de veces que una misma palabra aparece en posición análoga (mismo segmento de voz) en las 100 mejores hipótesis.
- **Densidad del grafo palabras:** número de enlaces en el lattice desde cualquier palabra a la palabra considerada.
- **Perplejidad de fonemas:** número medio de modelos de alófono activos (sobreviven al Beam Search) a lo largo de las tramas en las que permanece activa la palabra bajo estudio.

Nivel de Palabra: parámetros (II)

Del Modelo del Lenguaje

- Comportamiento Back-off del modelo de lenguaje

Valor de Confianza	Comportamiento
1.0	$P(W_j)$ como sucesión trigram: $P(W_j, W_{j-1}, W_{j-2})$
0.8	$P(W_j)$ como sucesión bigram-bigram: $P(W_j, W_{j-1})$ y $P(W_{j-1}, W_{j-2})$
0.6	$P(W_j)$ como sucesión bigram: $P(W_j, W_{j-1})$
0.4	$P(W_j)$ como sucesión unigram-bigram: $P(W_j)$ y $P(W_{j-1}, W_{j-2})$
0.3	$P(W_j)$ como sucesión unigram-unigram: $P(W_j)$ y $P(W_{j-1})$
0.2	$P(W_j)$ como unigram: $P(W_j)$
0.1	Palabra desconocida. Nunca se da en la salida del reconocedor.

- Probabilidad de la palabra en la secuencia $P(W_j)$, obtenida del modelo de lenguaje

Nivel de Concepto: parámetros (I)

Proceso de Decodificación y del Modelo del Lenguaje

- **Confianza Media de las palabras pertenecientes a la Regla utilizada en la obtención del Concepto.**
- **Confianza Media de las palabras pertenecientes al Valor del concepto.**

REGLA: prepos. (“to”) + nom. ciudad (“Denver”)

VALOR: Ciudad Destino (“Denver”)

Analizador semántico

- **Número de palabras contenidas en la Regla utilizada para la obtención del Concepto.**
- **Número de palabras que forman el Valor del Concepto.**

Nivel de Concepto: parámetros (II)

- **Homogeneidad del concepto** en las 100 mejores hipótesis. Se analizan semánticamente todas las hipótesis
- **Homogeneidad del concepto y su valor** en las 100 mejores hipótesis. Repeticiones del concepto con el mismo VALOR.

Modelo de Lenguaje Conceptual

- **Comportamiento Back-Off** del modelo de lenguaje conceptual.
- **Probabilidad del concepto** en la secuencia $P(C_j)$, obtenida del modelo de lenguaje.

El modelo de lenguaje conceptual se obtiene con las secuencias de conceptos obtenidas de las frases referencia (el valor del concepto no se considera)

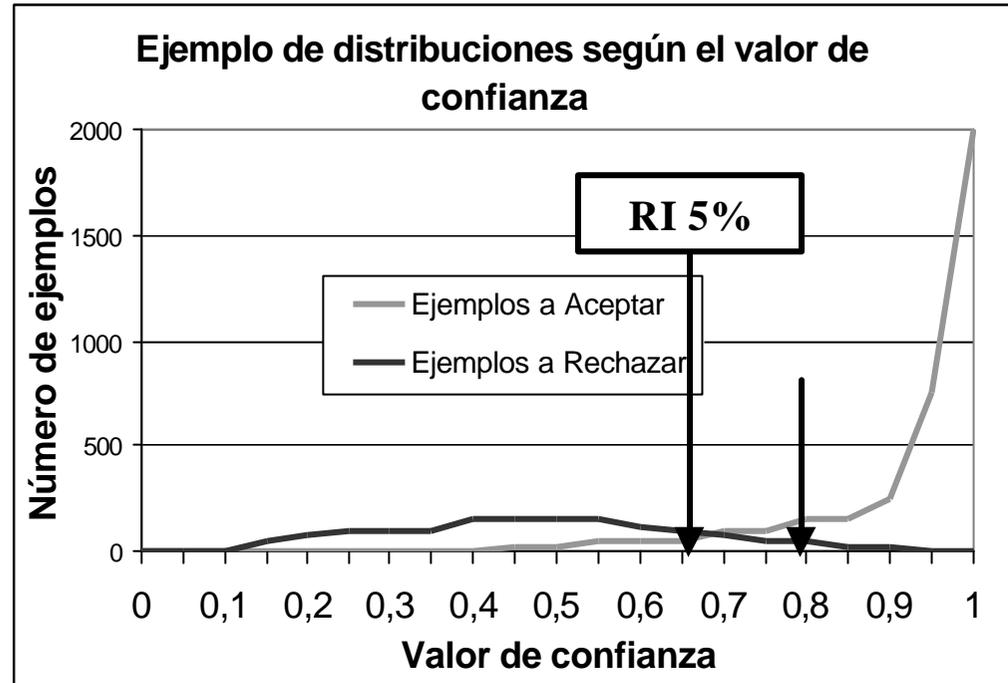
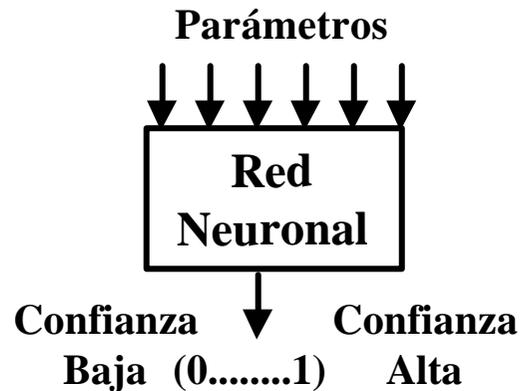
Nivel de Frase: parámetros

Medidas al nivel de palabra o concepto extendidas a toda la frase

- **Confianza media de las Palabras que forman la frase**
- **Confianza media de los Conceptos que forman la frase**
- **Porcentaje de palabras que pertenecen a la tarea:** palabras que forman parte de algún concepto o de alguna regla que se aplica para obtener un concepto.
- **Relación entre el número de conceptos y el número de palabras de la frase.**
- **Porcentaje de hipótesis (de entre las 100 mejores) que permite obtener algún concepto.**

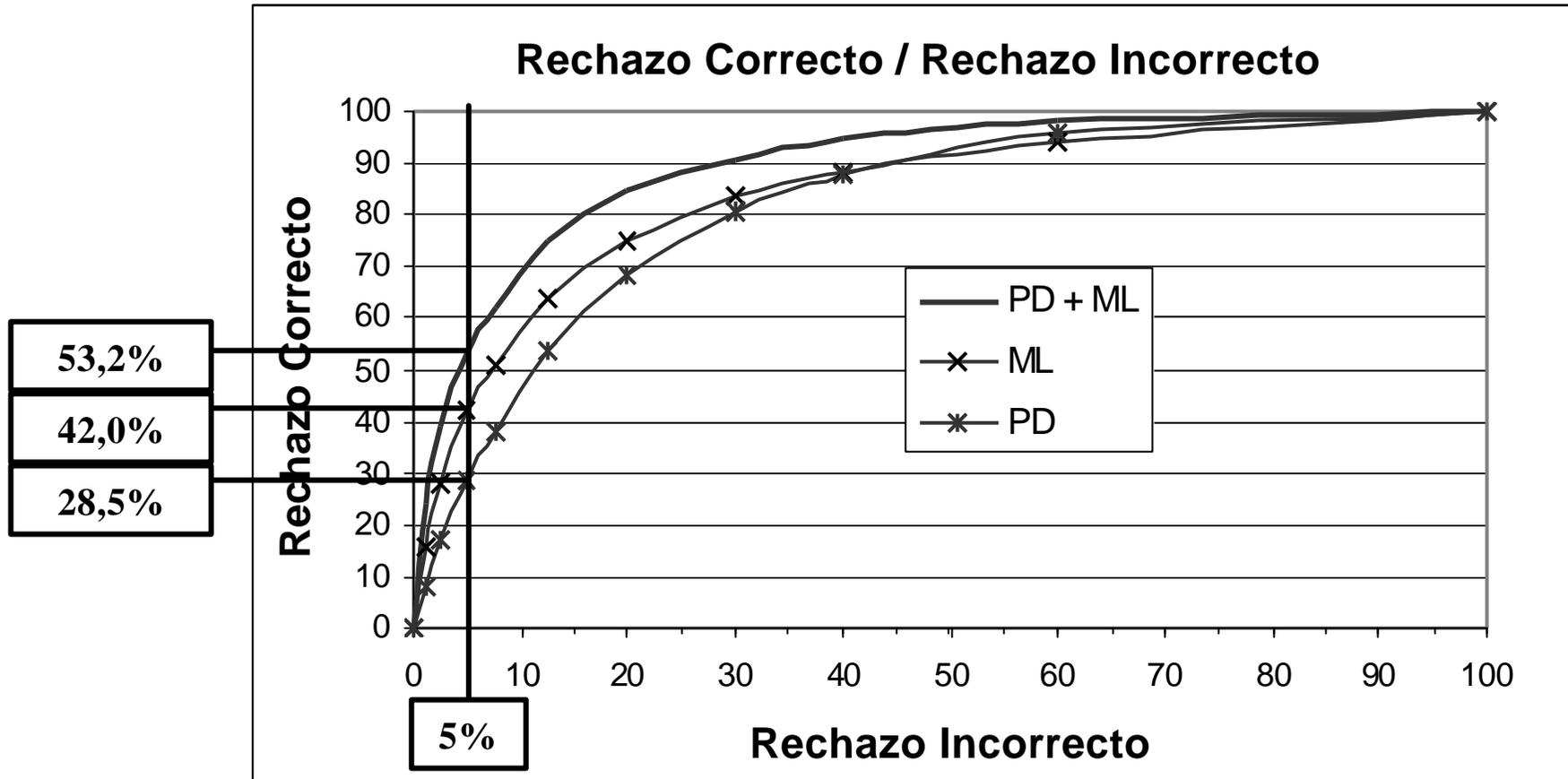
La combinación de los parámetros a todos los niveles se ha realizado mediante una Red Neuronal (Perceptrón Multi-Capa)

Evaluación



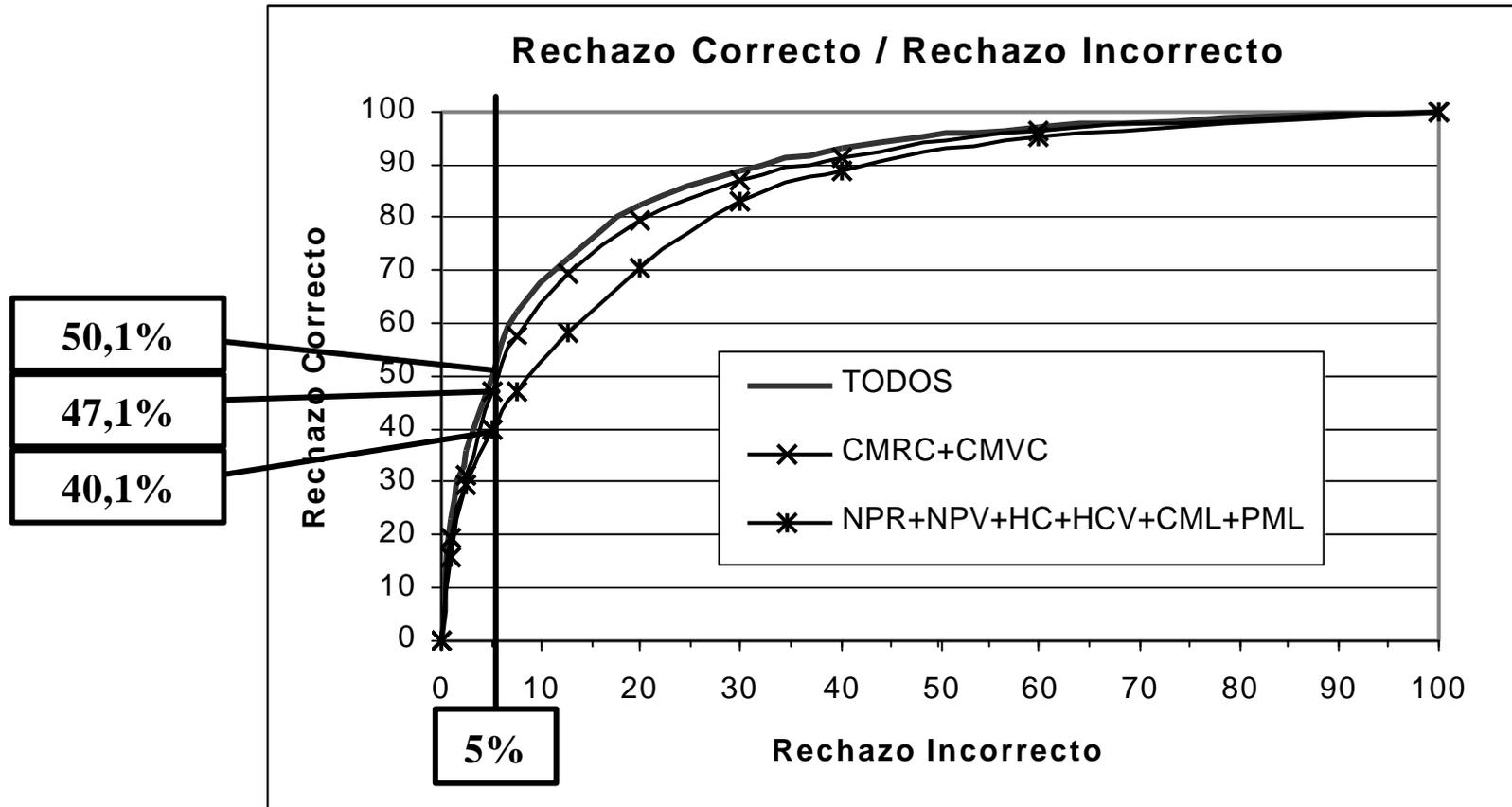
- **Rechazo Correcto (RC):** % de casos de error detectados
- **Rechazo Incorrecto (RI):** % de casos de acierto rechazados por error
- **Error de Clasificación (EC):** % de casos mal clasificados
- **Error de Referencia (ER):** distribución inicial de ejemplos

Resultados al nivel de Palabra



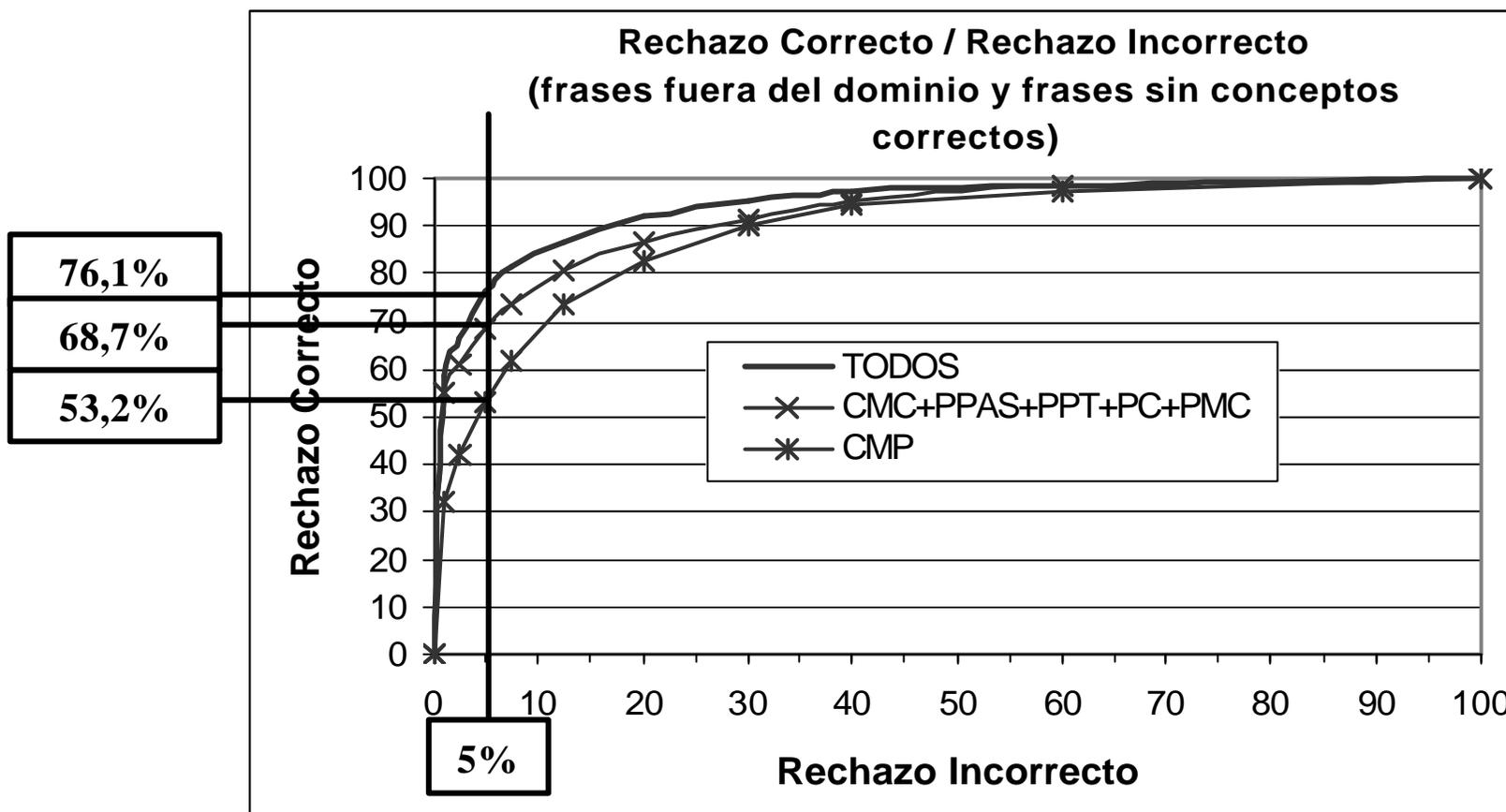
- **PD:** Parámetros del Proceso de Decodificación
- **ML:** Parámetros del Modelo de Lenguaje

Resultados al nivel de Concepto



- **CMRC y CMVC:** Confianza Media en la Regla y el Valor del Concepto.
- **NPR, NPV, HC, HCV, CML y PML:** Parámetros exclusivos del analizador semántico.

Resultados al nivel de Frase



- **CMC, PPAS, PPT, PC y PMC:** Parámetros exclusivos del analizador semántico.
- **CMP:** Confianza Media de las Palabras de la frase.

Aplicación: recuperación de errores

Combinación de varias hipótesis de reconocimiento mediante

Medidas de Confianza:

- **FLCR (Flat List Confidence Rescoring):** selección de la mejor hipótesis según la mayor confianza media por palabra.
- **WGCR (Word Graph Confidence Rescoring):** construcción de un grafo de palabras y selección del camino en el grafo con mayor confianza media. **IDEA:** generar una nueva hipótesis con las partes de mayor confianza de cada hipótesis.

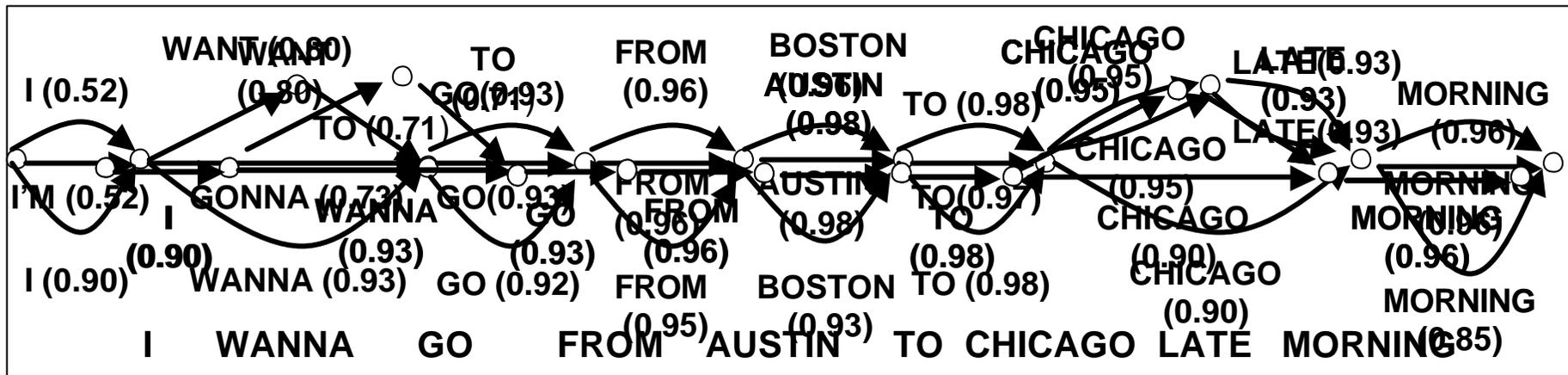
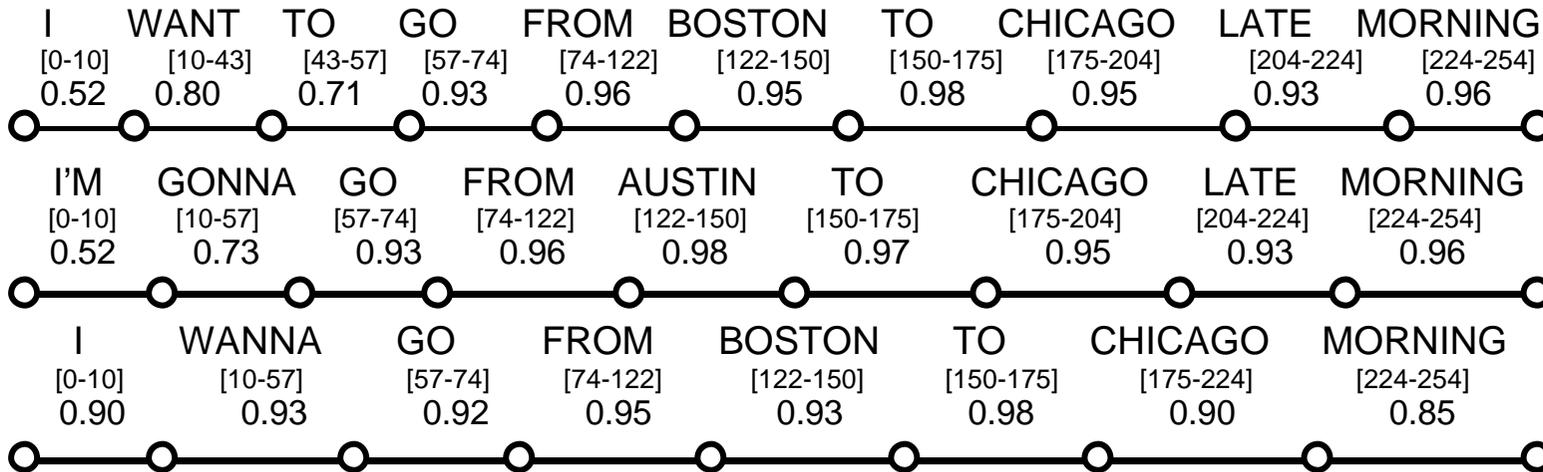
MÉTODOS:

↓ **Alineamiento Temporal:** se definen nodos de unión entre palabras de diferentes hipótesis cuyos inicios o finales temporales sean iguales. Método más preciso

↓ **Alineamiento de Texto:** se alinean las hipótesis con un algoritmo de programación dinámica.

Ejemplo de construcción de un grafo

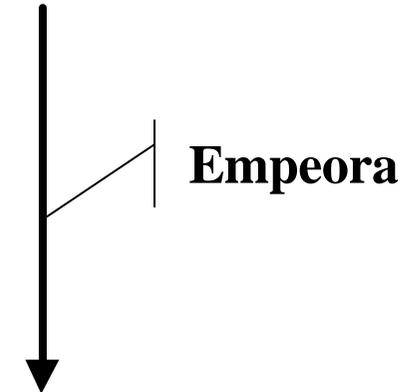
REFERENCIA: I WANNA GO FROM AUSTIN TO CHICAGO LATE MORNING
 HIPÓTESIS



Resultados de Combinación de Hipótesis

N Hipótesis del mismo Reconocedor

Número de Hipótesis	FLCR	WGCR
1	27,2%	27,2%
2	29,2%	29,2%
4	30,8%	31,2%
8	32,2%	32,6%
16	33,5%	33,6%



1 hipótesis de varios Reconocedores (H/M)

Método Utilizado	Tasa de Error
Referencia	27,2%
Verosimilitud Acumulada	26,2%
Confianza: FLCR	24,2%
Confianza: WGCR	23,4%

N Hipótesis de varios Reconocedores (H/M)

Número de Hipótesis	FLCR	WGCR
1	24,2%	23,4%
2	23,3%	22,9%
4	29,4%	26,4%
8	30,5%	28,0%
16	31,4%	28,7%