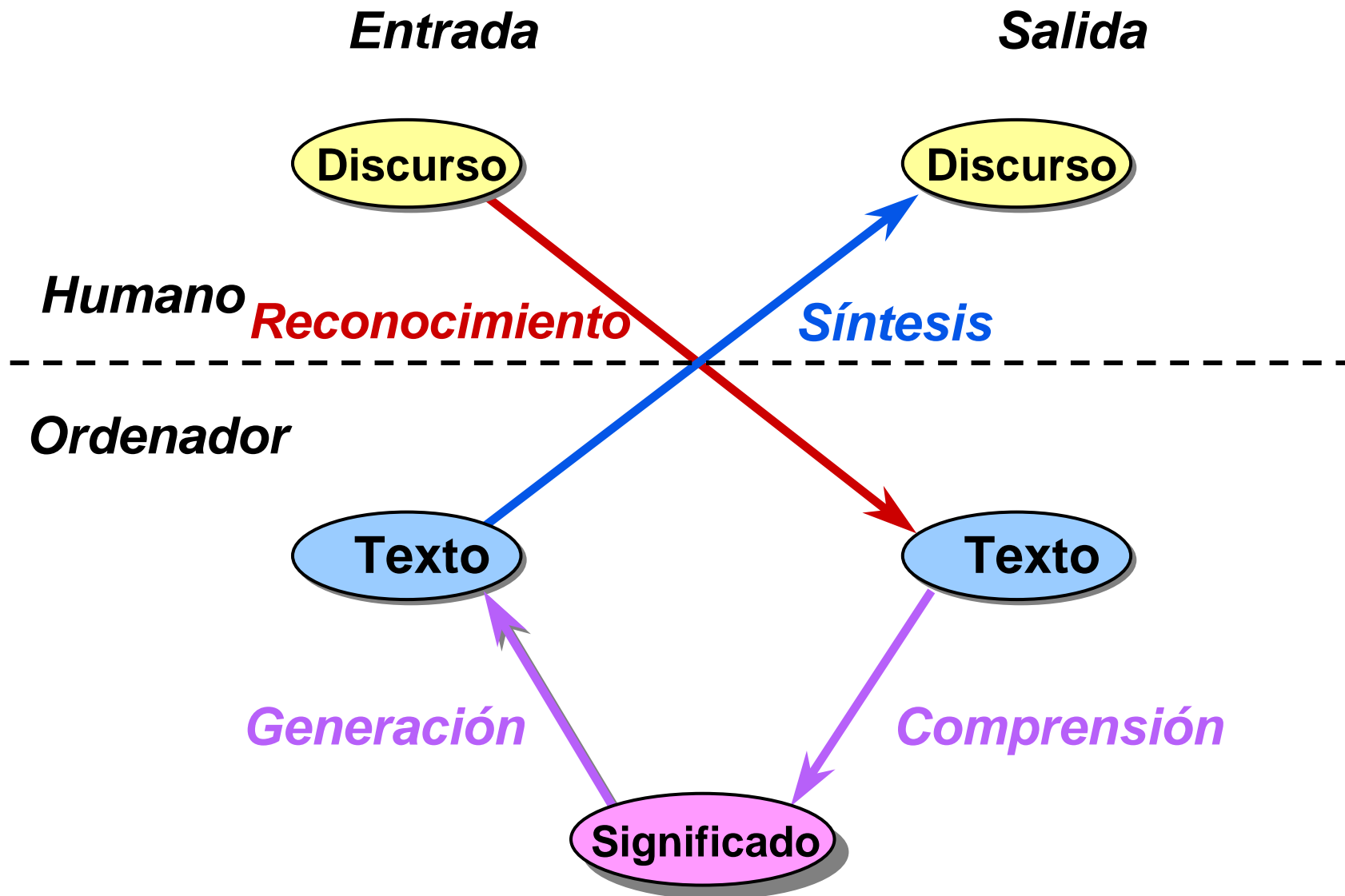


- **Clases: Jim Glass y conferenciantes invitados**
- **Introducción al RAH**
 - Definición del problema
 - Ejemplos vanguardistas
- **Visión global del curso**
 - Esquema de las clases
 - Trabajos
 - Proyecto trimestral
 - Calificación



Ventajas del lenguaje hablado

- Natural:** No requiere un entrenamiento especial
- Flexible:** No requiere la presencia de otros sentidos
- Eficiente:** Presenta una alta velocidad de datos
- Económico:** Puede transmitirse/recibirse sin coste alguno

Las interfaces de voz son ideales para la gestión y el acceso a la información cuando:

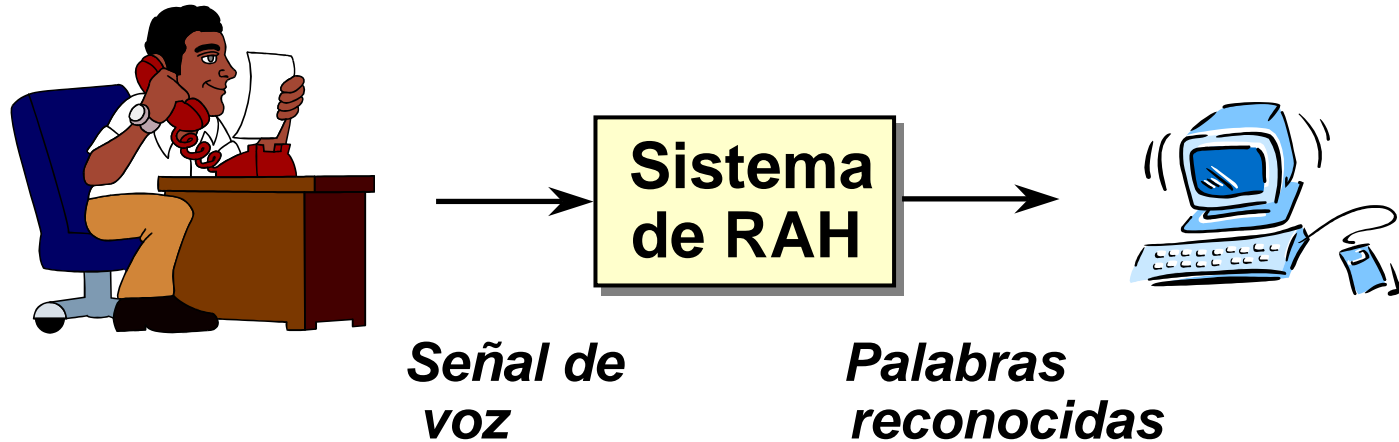
- El espacio para la información es amplio y complejo;
- Los usuarios son incapaces, técnicamente hablando;
- Los teléfonos es el único medio disponible.



Distintas fuentes de restricción para la comunicación a través del lenguaje hablado

Acústica:	tracto vocal humano
Fonética:	let us pray lettuce spray
Fonológica:	gas shortage fish sandwich
Fonotáctica:	blit vnuk
Sintáctica:	I am flying to Chicago tomorrow tomorrow I flying Chicago am to
Semántica:	Is the baby crying Is the bay bee crying
Contextual:	It is easy to recognize speech It is easy to wreck a nice beach

Reconocimiento automático del habla



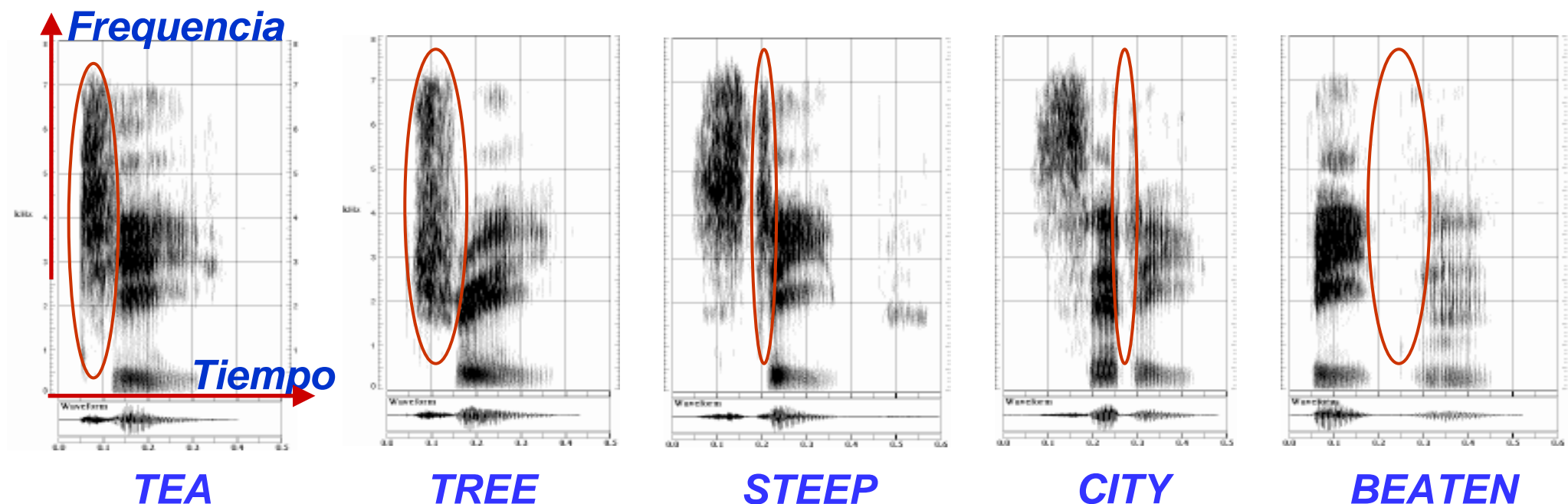
- Un sistema de RAH convierte la señal de voz en palabras.
- Las palabras reconocidas pueden ser:
 - La salida final.
 - Los datos de entrada para el procesamiento del lenguaje natural.

- **Principalmente datos de entrada (sólo reconocimiento)**
 - Control y orden simple
 - Entrada de datos simples (por teléfono)
 - Dictado
- **Conversación interactiva (se requiere comprensión)**
 - Cabinas de información
 - Procesamiento transaccional
 - Agentes inteligentes

- **Co-articulación**
- **Independencia con respecto al hablante**
 - Variaciones dialectales
 - Hablantes no nativos
- **Habla espontánea**
 - Disfluencia
 - Palabras no presentes en el vocabulario
- **Modelado del lenguaje**
- **Robustez en cuanto al ruido**

Ejemplo de variación fonológica

- La realización acústica de un fonema depende en gran medida del contexto en el que éste aparece



Ejemplos contrastivos de discurso leído y discurso espontáneo (Entorno de navegación)

Pausas rellenas y no rellenas:	leído, espontáneo
Alargamiento de palabras:	leído, espontáneo
Inicios falsos:	leído, espontáneo



En ocasiones, los datos reales establecerán
los requisitos tecnológicos (Entorno de nombres de ciudades)

Tecnología requerida

Extracción de palabra clave simple

Extracción de palabra clave compleja

Comprensión del discurso

Ejemplo

Um, Braintree

Eh yes, Avis rent-a-car in
Boston

Hello, please Brighton,
uh, can I have the number
of Earthscape, in, uh, on
Nonantum Street

Woburn, uh, Somerville.

I'm sorry



Parámetros que caracterizan las capacidades de los sistemas de RAH

Parámetros	Ámbito
Modo de habla:	Palabra aislada a discurso ininterrumpido
Estilo de habla:	Discurso leído a discurso espontáneo
Inscripción:	Dependencia del hablante a independencia del hablante
Vocabulario:	Pequeño (<20 palabras) a extenso (>50,000 palabras)
Modelo de lenguaje:	De estado finito a sensible al contexto
Perplejidad:	Baja (<10) a alta(>200)
SNR _(proporción señal ruido)	Alta (>30dB) a baja (<10dB)
Transductor:	Micrófono cancelador de ruido a teléfono móvil

Tendencias en RAH*: antes y ahora

	anterior a la mitad de los 70	mitad años 70 - mitad años 80	posterior a mitad de los 80
Unidades de reconocimiento	unidades de palabra y de subpalabra	unidades de subpalabra	unidades de subpalabra
Enfoques de modelado	heurística y ad hoc	correspondencia de plantillas	matemático y formal
	basado en reglas y declarativo	determinístico y dirigido por datos	probabilístico y dirigido por datos
Representación del conocimiento	heterogéneo y complejo	homogéneo y simple	homogéneo y simple
Adquisición del conocimiento	ingeniería del conocimiento profundo	insertado en estructura simple	aprendizaje automático

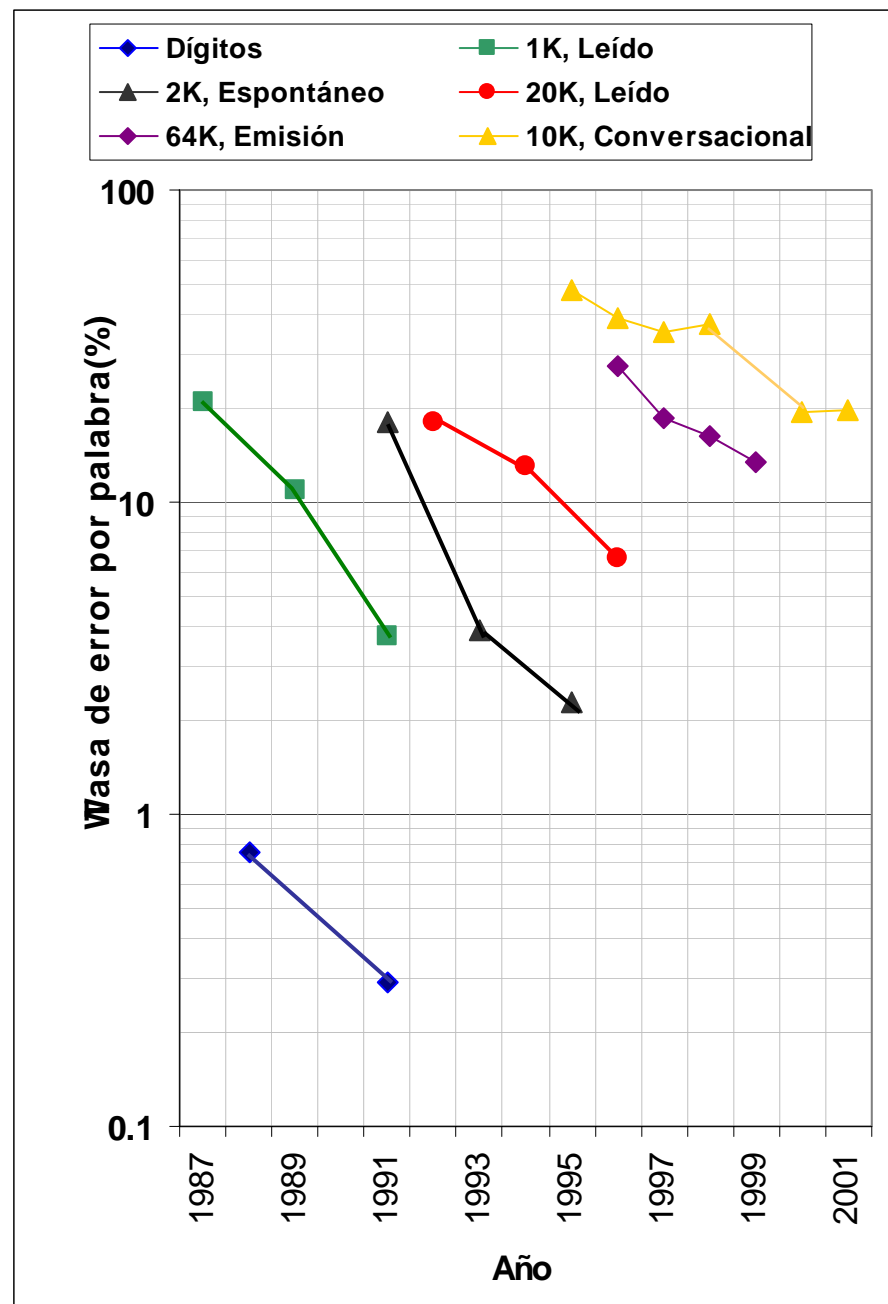
* Existen por supuesto, muchas excepciones.

MIT Reconocimiento del habla: ¿Dónde estamos ahora?

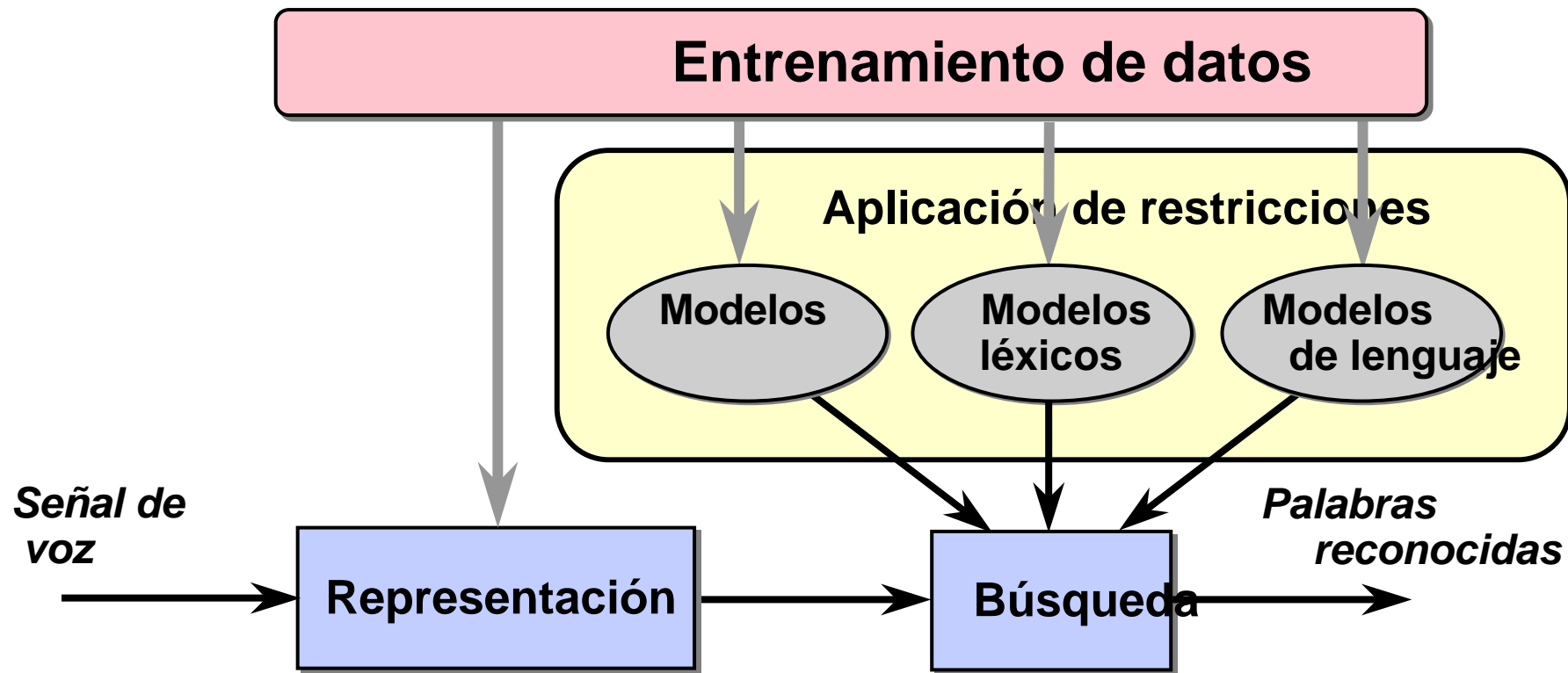
- **Alto rendimiento, ya es posible el reconocimiento de voz con independencia del hablante**
 - **Extenso vocabulario (para hablantes de colaboración en entornos propicios)**
 - **Vocabulario moderado (para el discurso espontáneo por teléfono)**
- **Ya hay disponibles sistemas de reconocimiento comerciales**
 - **Dictado (ej., Dragon, IBM, L&H, Philips) Scansoft**
 - **Transacciones telefónicas (ej., AT&T, Nuance, Philips, SpeechWorks, TellMe, etc.)**
- **Cuando se combinan con las aplicaciones, la tecnología puede incrementar el rendimiento del trabajo real.**

Ejemplos del funcionamiento del RAH

- Independencia del hablante, el RAH del discurso ininterrumpido es ahora posible
- Reconocimiento digital por teléfono con una tasa de error por palabra del 0.3%
- Tasa de error reducida a la mitad cada 2 años para módicas tareas de vocabulario
- El error en el discurso espontáneo es doble o triplemente más frecuente que en el discurso leído
- El discurso conversacional en un entorno acústico pobre y con múltiples hablantes, constituye aún un desafío
- Decenas de horas de entrenamiento de datos para probarlos en un entorno distinto
- Los modelos estadísticos que utilizan el entrenamiento automático consiguen importantes progresos



- **El modelado estadístico y los enfoques dirigidos por datos han demostrado ser potentes**
- **La infraestructura de investigación es crucial:**
 - **Enormes cantidades de datos lingüísticos**
 - **Metodologías de evaluación**
- **La disponibilidad y asequibilidad del poder computacional condujo a una reducción de los ciclos de desarrollo tecnológico y a los sistemas en tiempo real**
- **El paradigma dirigido por rendimiento acelera el desarrollo tecnológico**
- **La colaboración interdisciplinaria genera un aumento de las aptitudes (ej., la comprensión del lenguaje hablado)**



- El reconocimiento del habla es un problema de:
 - Cómo **representar** la señal
 - Cómo **modelar** las restricciones
 - Cómo **buscar** la respuesta más óptima

Demostración: Dictado continuo

- **ViaVoice de IBM se ejecuta en un ThinkPad**
- **Entrenado para una oficina tranquila (rendimiento en clase no óptimo)**

- **Desarrollado por SpeechWorks International (existen otros)**
- **Información de gastos de envío para Fedex (1-800-GO-FEDEX)**
 - **Proporciona información sobre:**
 - * Tipos de paquetes
 - * Códigos zip de origen y destino
 - * Peso, tamaño, valor
 - * Tipo de servicio
 - **Controla todo el flujo de llamadas de información de EEUU**
- **Sistema de corretaje automatizado para el comercio E*_(electrónico)**
 - **Respalda las cotizaciones y el comercio**
 - * Utiliza símbolos o nombres
 - * Para valores, opciones y fondos mutuos
 - **Los usuarios pueden "meterse" en cualquier momento**
 - **Utilización a nivel nacional por más de 450,000 clientes**

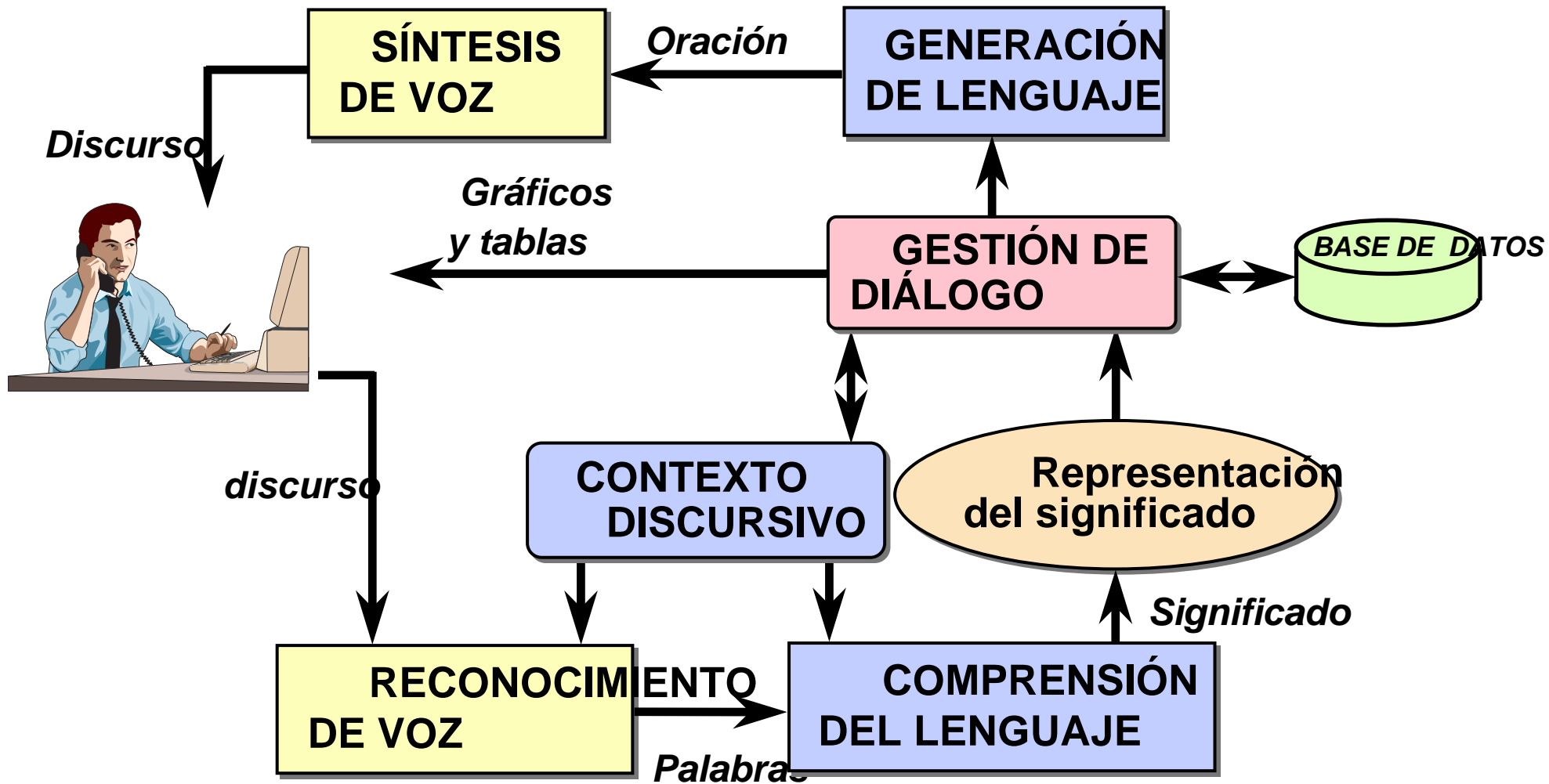


Interfaces conversacionales: La próxima generación

- Nos permite **conversar** con máquinas (del mismo modo que lo hacemos entre nosotros) para crear, acceder y controlar información y resolver problemas
- Aumenta la tecnología del reconocimiento del habla con tecnología de lenguaje natural para **comprender** la entrada verbal
- Puede entablar un **diálogo** con un usuario durante la interacción
- Utiliza el lenguaje natural para **dar** la respuesta adecuada
- Es lo que Hollywood y cualquier “futurista” preve



Arquitectura de un sistema conversacional



- **Sistema de información meteorológica Jupiter**
 - Acceso a través del teléfono
 - 500 ciudades a nivel mundial
 - Información meteorológica sobre el campo desde la Web varias veces al día

Jupiter

Una interfaz conversacional para información meteorológica en línea por teléfono.

1-888-573-8255

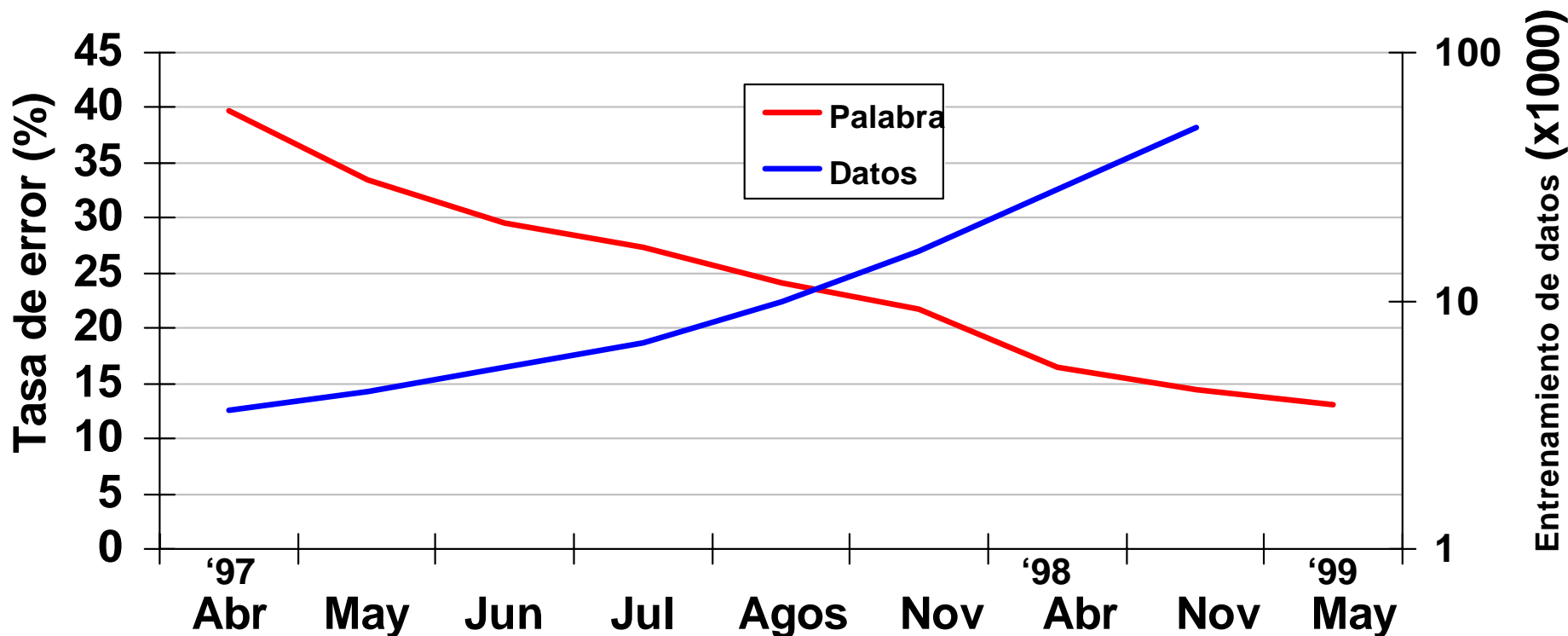
(fuera de EEUU: 1-617-258-0300)

<http://www.sls.lcs.mit.edu/jupiter>

Grupo de sistemas de lenguaje hablado,
MIT Laboratory for Computer Science



Datos (reales) mejoran el funcionamiento (Entorno meteorológico)



- Las evaluaciones longitudinales muestran las mejoras.
- La recogida de datos reales mejora el funcionamiento:
 - Permite una complejidad cada vez mayor y una mayor robustez para modelos acústicos y de lenguaje.
 - Mejor opción que las condiciones bajo las que se realizaban las grabaciones en laboratorios.
- Participación continua por parte de los usuarios.



Pero estamos muy lejos de conseguirlo

Corpus	Tipo de discurso	Tamaño del léxico	Tasa de error por palabra (%)	Tasa de error humano (%)
Cadenas de dígitos (teléfono)	espontáneo	10	0.3	0.009
Gestión de recursos	leído	1000	3.6	0.1
ATIS (Sistema de información aérea)	espontáneo	2000	2	--
Periódico Wall Street	leído	64000	6.6	1
Noticias de la radio	mixto	64000	13.5	--
Centralita (teléfono)	conversación	10000	19.3	4
Llamadas particulares (teléfono)	conversación	10000	30	--

Esquema del curso

Información paralingüística
Comprensión del discurso
Interfaces multimodales

Modelado
Fonético-
Acústico

Reconocimiento
de patrones

Transductores de
estado finito

Modelado del
lenguaje

Teoría acústica de la
producción del habla

RAH
robusto

Modelos
acústicos

Modelos
léxicos

Modelos
de lenguaje

Adaptación

*Señal de
VOZ*

Representación

Búsqueda

Palabras reconocidas

Propiedades de
los sonidos del discurso

Representación de
la señal

Algoritmos
de búsqueda

Cuantización vectorial
y agrupamiento

Modelos ocultos
de Markov

Modelos
gráficos

Modelos
segmentales

Logística del curso

- **Clases:** Dos sesiones por semana, 90 min. por sesión
- **Prácticas:** Toda la semana en horario lectivo

Calificación

- **9 trabajos** **45%**
- **2 pruebas** **30%**
- **Proyecto trimestral (sobre 4 semanas)** **25%**

- **Habr  9 trabajos semanales**
 - Problemas que ampl an el material de clase
 - Trabajos pr cticos para reforzar el material de clase
 - Los trabajos se entregar n todos los mi rcoles de la siguiente semana
- **El trabajo pr ctico se realizar  en el laboratorio inform tico**
- **Es obligatorio registrarse para las pr cticas (Web del curso)**
- **Se facilitar n las soluciones**

Proyecto trimestral

- **Investigar un aspecto comparado en un experimento de RAH**
- **Le daremos distintos reconocedores y entornos para que elija, y le ayudaremos a escoger un tema:**
- **Podrá elegir:**
 - **Condición de evaluación:** ej., clasificación fonética, reconocimiento de la palabra)
 - **Base de datos** (ej., TIMIT, RM, Jupiter, Aurora, ...)
 - **Reconocedor** (ej., Sphinx, Summit, GMTK, ...)
 - **Aspecto contrastivo** (ej., representación de la señal, modelo acústico, modelo de lenguaje)
- **Requisitos:**
 - **Propuesta**
 - **Experimentos (el peso del trabajo)**
 - **Redacción**
 - **Presentación el último día de clase**

- Huang, Acero y Hon, *Spoken Language Processing*, Prentice-Hall, 2001.
- Jelinek, *Statistical Methods for Speech Recognition*, MIT Press, 1997.
- Rabiner y Juang, *Fundamentals of Speech Recognition*, Prentice-Hall, 1983.
- Duda, Hart y Stork, *Pattern Classification*, Wiley & Sons, 2001.
- Stevens, *Acoustic Phonetics*, MIT Press, 1998.