



# **Nuevos Paradigmas de Interacción**

## **Tutorial de Sistemas de Diálogo**

**Prof. Ramón López-Cózar Delgado**  
**Dpto. Lenguajes y Sistemas Informáticos**  
**ETS Informática y Telecomunicación**  
**Universidad de Granada**  
**<http://lsi.ugr.es/rlopezc>**



# Puntos a tratar

- 
- Introducción a sistemas de diálogo
  - Sistemas de diálogo hablado
  - Sistemas de diálogo multimodal
  - Sistemas de diálogo multilingüe
  - Recomendaciones de implementación

# Introducción a sistemas de diálogo

- Sistema de diálogo
  - Sistema informático capaz de comunicarse con las personas mediante lenguaje natural y/o otras modalidades de interacción para realizar tareas concretas
  - La comunicación se realiza simulando el diálogo entre personas
- Proyectos precursores
  - **DARPA SLS** (*Spoken Language Systems*) en Estados Unidos
  - **SUNDIAL** (*Speech UNDerstanding and DIALog*) en Europa

# Introducción a sistemas de diálogo

- **DARPA SLS**
  - sólo idioma Inglés
  - dominios restringidos (principalmente, viajes en avión)
  - centrado en interfaz de entrada (reconocimiento + comprensión)
  - competiciones periódicas de sistemas desarrollados

# Introducción a sistemas de diálogo

- **SUNDIAL**

- varios idiomas (Inglés, Francés, Alemán e Italiano)
- dominios de aplicación: viajes en avión y en tren
- centrado en interfaz de entrada, control del diálogo e interfaz de salida
- favorece la cooperación en lugar de la competición (integración de módulos creados en diferentes laboratorios)

# Introducción a sistemas de diálogo

- **Clasificación de sistemas de diálogo**
  - **Sistemas diálogo unimodal:** sólo usan un canal de comunicación
    - Habla (lenguaje natural mediante voz)
    - Texto (lenguaje natural escrito)
  - **Sistemas diálogo multimodal:** usan varios canales de comunicación
    - **Entrada:** habla, movimiento labios, gestos usuario, etc.
    - **Salida:** habla, imágenes, gráficos, etc.
  - **Sistemas diálogo multilingüe:** permiten interacción mediante varios idiomas (p. e. Castellano e Inglés)

# Puntos a tratar

- Introducción a sistemas de diálogo
- ➔ ● Sistemas de diálogo hablado
- Sistemas de diálogo multimodal
- Sistemas de diálogo multilingüe
- Recomendaciones de implementación

# Sistemas de diálogo hablado

- Principales áreas (o dominios) de aplicación:
  - Información y reserva de viajes (p.e., Air Travel Information Service, ATIS)
  - Información meteorológica
  - Información bancaria
  - Venta productos
  - Información de directorios
  - etc.



# Sistemas de diálogo hablado

- **Ventajas**
  - ahorro económico para empresas
  - uso del teléfono: comodidad y flexibilidad
  - comunicación natural y variada
  - usuarios gozan de mayor grado de expresión
  - no limitados a responder mediante determinados comandos
- **Algunos retos actuales**
  - mayor robustez
  - diálogo más natural e “inteligente”

# Sistemas de diálogo hablado

- Estrategias de interacción
  - dirigida por sistema (system-directed)
  - dirigida por usuario (user-directed)
  - mixta (mixed-initiative)

# Sistemas de diálogo hablado

- Estrategias de confirmación
  - Funcionamiento de interfaz de entrada no es perfecto
  - Técnicas para confirmar datos
    - Confirmación directa (o explícita)
    - Confirmación indirecta (o implícita)

# Sistemas de diálogo hablado

- Estrategias de confirmación
  - Confirmación directa (o explícita)
    - El sistema dato mediante turno específico del diálogo dedicado a ello
    - Se generan turnos adicionales para obtener datos (diálogos menos eficientes)
    - Facilidad de corrección para usuario

# Sistemas de diálogo hablado

- Estrategias de confirmación

- Confirmación directa (o explícita)

S1: ¿A qué ciudad deseas viajar?

U1: A Madrid

→ S2: ¿Has dicho que deseas viajar a Madrid?

U2: Sí

S3: De acuerdo, destino Madrid, ¿Desde que ciudad deseas salir?

U3: Desde Granada

→ S4: ¿Has dicho que deseas salir desde Granada?

U4: Sí

S5: De acuerdo, salida desde Granada, ¿Qué día de la semana deseas viajar?

U5: El lunes

# Sistemas de diálogo hablado

- Estrategias de confirmación
  - Confirmación indirecta (o implícita)
    - El sistema confirma dato usando turnos empleados para solicitar otros datos
    - Diálogos más eficientes
    - Los usuarios pueden tener poco claro qué hacer

# Sistemas de diálogo hablado

- Estrategias de confirmación

- Confirmación indirecta (implícita)

S1: ¿A qué ciudad deseas viajar?

U1: A Madrid

→ S2: De acuerdo, destino Valladolid. ¿Desde qué ciudad deseas salir?

U2: Te has equivocado, he dicho Madrid, no Valladolid

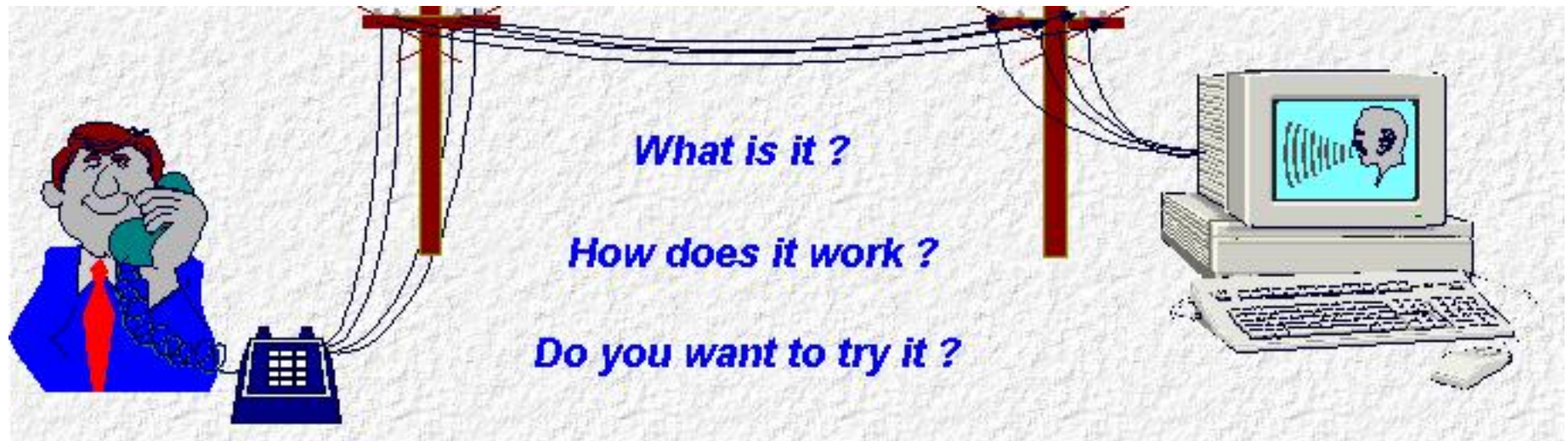
S3: De acuerdo, destino Madrid. ¿Desde qué ciudad deseas salir?

U3: Desde Granada

→ S4: De acuerdo, salida desde Granada, ¿Qué día de la semana deseas viajar?

U4: El lunes

# Sistemas de diálogo hablado



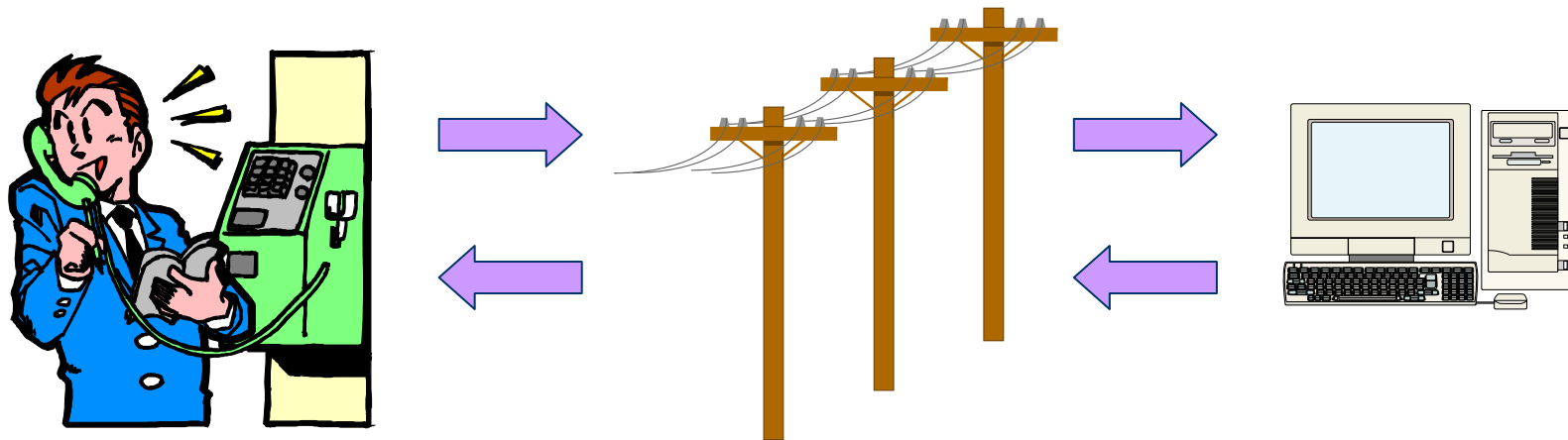


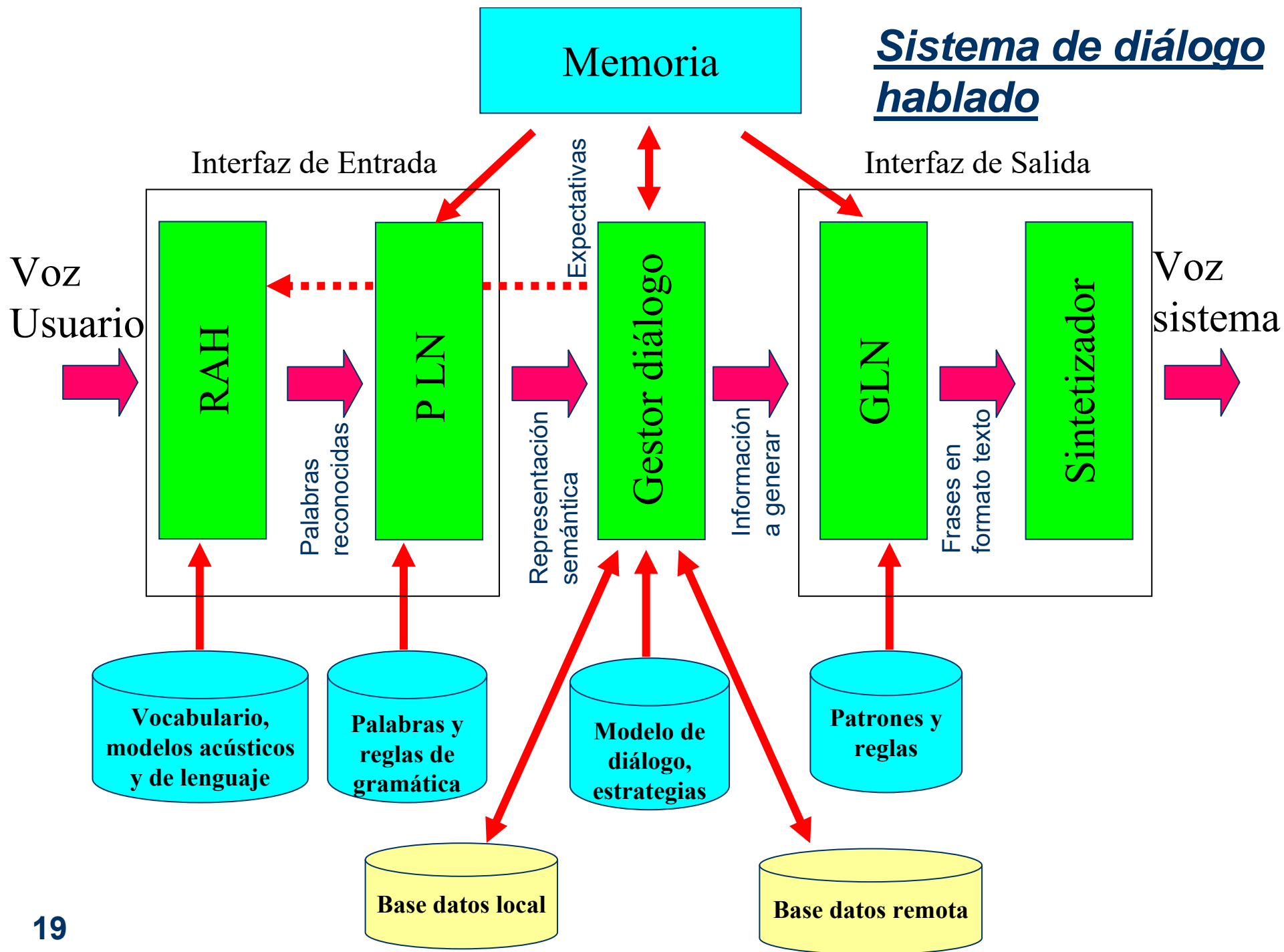
# Sistemas de diálogo hablado

- **Tecnologías utilizadas en la implementación**
  - Reconocimiento Automático del Habla (RAH)
  - Procesamiento de Lenguaje Natural (PLN)
  - Gestión diálogo
  - Generación de Lenguaje Natural (GLN)
  - Síntesis del habla



# Sistemas de diálogo hablado





# Sistemas de diálogo hablado

- **Interfaz de entrada**
  - **Reconocedor del habla (Speech Recogniser)**
    - proporciona secuencia de palabras reconocidas
    - una o varias hipótesis de reconocimiento
    - valores de confianza
  - **Procesador del lenguaje natural (Semantic Analyser)**
    - obtiene significado de frases de usuario (generalmente mediante **frames**)

# Sistemas de diálogo hablado

- **Gestor del diálogo (Dialogue Manager)**
  - Decide qué respuesta debe generar el sistema
  - Solicita datos adicionales, confirmaciones o aclaraciones
  - Genera expectativas para mejorar el proceso de RAH
- **Memoria (Memory Module)**
  - Almacena datos de interacciones de usuarios
  - Proporciona historia del diálogo al gestor del diálogo
- **Bases de datos (Database)**
  - Contienen información para usuarios

# Sistemas de diálogo hablado

- **Interfaz de salida**
  - **Generador de lenguaje natural (Natural Language Generator o Response Generator)**
    - Genera respuestas en formato de texto
  - **Sintetizador del habla (Speech Synthesizer)**
    - Concatenación de palabras. Poco flexible, gran inteligibilidad
    - Concatenación de unidades menores que las palabras (p. e. difonemas). Gran flexibilidad, menos inteligibilidad

# Sistemas de diálogo hablado

- **Problemas relacionados con el RAH**
  - **Coarticulación:** fonemas se ven afectados por fonemas vecinos
  - **Variabilidad:** diferencias de pronunciación, distintos acentos, etc.
  - **Tamaño del diccionario:** al crecer, aumenta probabilidad de error
  - **Ruidos e interferencias:** provocan errores de reconocimiento
  - **Palabras OOV (*Out-of-vocabulary*):** no incluidas en diccionario
  - **Cross-talk:** terceras personas hablando cerca del micrófono
  - **Otros factores:** estado anímico del locutor, resfriados, etc.

# Sistemas de diálogo hablado

- **Problemas relacionados con el RAH**
  - Como consecuencia de los problemas aparecen tres tipos de errores:
    - **Inserciones:** en frase reconocida hay más palabras que en frase pronunciada
    - **Borrados:** en frase reconocida hay menos palabras que en frase pronunciada
    - **Sustituciones:** en frase reconocida algunas palabras son cambiadas



# Sistemas de diálogo hablado

- **Problemas relacionados con el PLN**
  - **Elipsis** (omisión de palabras)
    - Ej. “*uno de lomo con queso*” (¿a qué se refiere “uno”?)
  - **Anáfora**
    - Ej. “*lo quiero grande*” (¿a qué se refiere “lo”?)
  - **Ambigüedad** (léxica y estructural)
    - Ej. “*vale*” (¿qué quiere decir “vale”?)
    - Ej. “*Mi hermana vio a un niño jugando en el parque con un telescopio*” (¿cuál es el significado?)
  - **Derivados de errores de RAH**
    - inserciones, borrados y sustituciones pueden provocar errores en análisis lingüístico de las frases

# Sistemas de diálogo hablado

- **Problemas relacionados con la gestión del diálogo**
  - **Gestionar la interacción**
    - ¿Quién tiene la iniciativa en la conversación?
  - **Gestionar errores de reconocimiento y/o comprensión**
    - ¿Qué hacer cuando se han producido errores?
  - **Gestionar las confirmaciones**
    - ¿Qué datos proporcionados por el usuario se han de confirmar?
    - ¿Cómo confirmar dichos datos?
  - **Intentar minimizar aparición de errores**
    - ¿Qué expectativas generar para evitar errores en proceso de RAH?

# Sistemas de diálogo hablado

- **Relacionados con la GLN**
  - **Generar frases sintáctica y semánticamente correctas**
    - ¿Quieres *uno* bocadillo de lomo y *una* coca-cola?
    - ¿Has dicho que quieres una cerveza *de limón*?
  - **Elegir las palabras y expresiones adecuadas para un contexto dado**
    - **Anáfora:** ¿Has dicho que *la* quieres de tamaño grande?
    - **Elipsis:** El bocadillo cantábrico cuesta 2 euros y el (...) de lomo cuesta 2 euros con 20 céntimos

# Sistemas de diálogo hablado

- Relacionados con la síntesis de habla
  - Lograr gran **inteligibilidad**
    - Las frases deben ser fácilmente entendibles
  - Lograr gran **naturalidad**
    - La voz sintética debe tener una gran semejanza con la voz real
  - Elección de unidades lingüísticas a usar
    - **Fonemas**: menor naturalidad, más complejo, más flexible
    - **Difonemas**: menor naturalidad, más complejo, más flexible
    - **Trifonemas**: menor naturalidad, más complejo, más flexible
    - **Palabras**: gran naturalidad, simple, poco flexible
    - **Frases**: gran naturalidad, simple, muy poco flexible

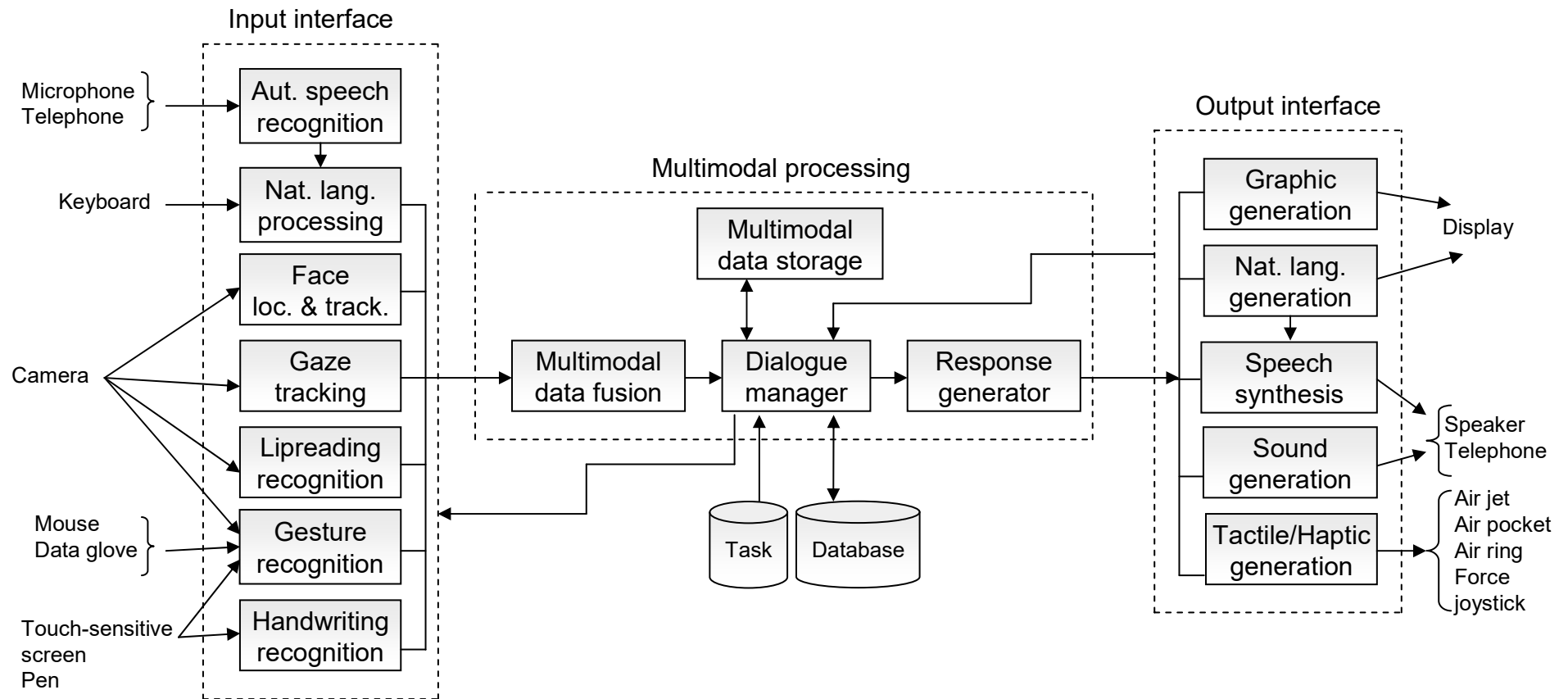
# Sistemas de diálogo hablado

- Relacionados con el tiempo de respuesta
  - Los usuarios **no aceptan** sistemas lentos, especialmente en aplicaciones telefónicas
- Relacionados con expectativas de usuarios
  - Los usuarios no suelen ser conscientes de limitaciones técnicas

# Puntos a tratar

- Introducción a sistemas de diálogo
- Sistemas de diálogo hablado
- ➔ ● Sistemas de diálogo multimodal
- Sistemas de diálogo multilingüe
- Recomendaciones de implementación

# Sistema de diálogo multimodal



# Sistemas de diálogo multimodal

- **Interfaz de entrada**
  - Modalidades adicionales de entrada
    - Gestos del usuario
    - Movimiento de labios de usuario
    - Dirección de mirada de usuario
    - Expresiones faciales de usuario
    - Texto escrito por usuario
    - Etc.



# Sistemas de diálogo multimodal

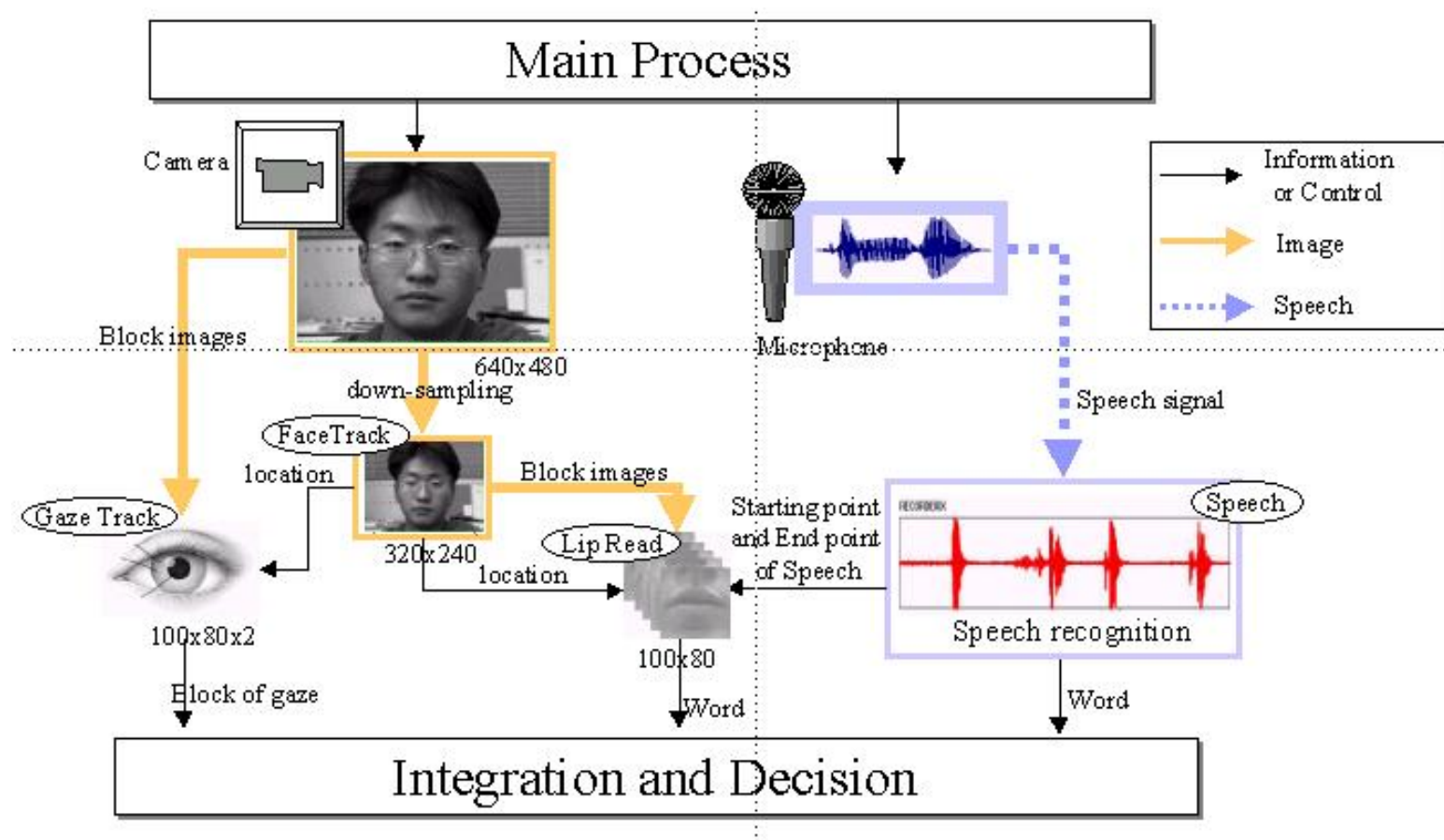
- **Procesamiento multimodal**
  - Procesar información multimodal de entrada
    - Sincronización
    - Fusión de información
  - Realizar gestión del diálogo
    - Estrategias de interacción y confirmación
  - Almacenar datos multimodales
  - Generar información multimodal para salida
    - Determinar qué modalidades de salida usar
    - Sincronización de modalidades de salida

# Estructura modular de un sistema de diálogo multimodal

- **Interfaz de salida**
  - Modalidades adicionales de salida
    - Gráficos
    - Agente animado (cara o cuerpo), movimientos sincronizados
    - Otros sonidos
    - Etc.

# Sistemas de diálogo multimodal

- Fusión de información multimodal



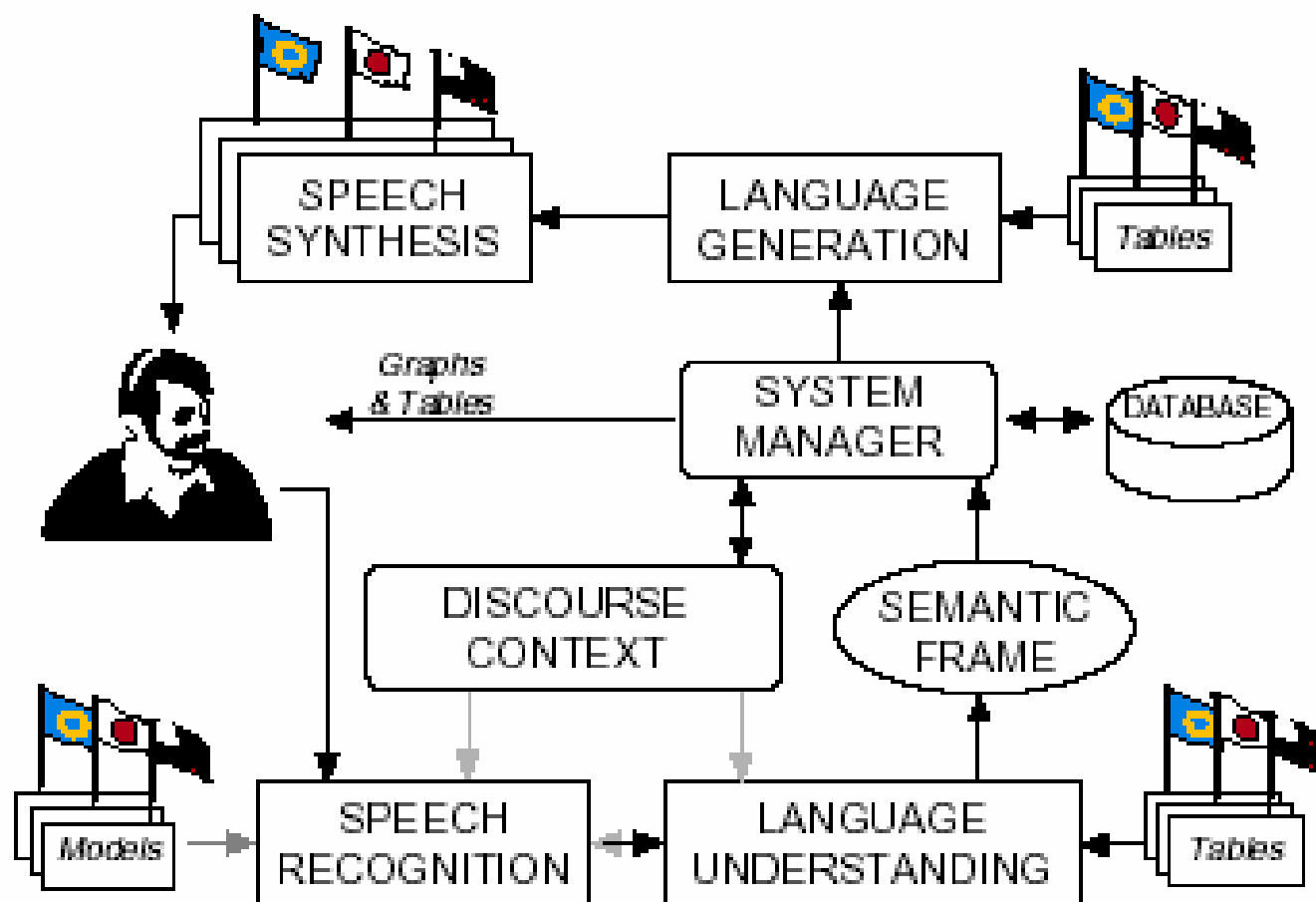
# Puntos a tratar

- Introducción a sistemas de diálogo
- Sistemas de diálogo hablado
- Sistemas de diálogo multimodal
- ➔ ● Sistemas de diálogo multilingüe
- Recomendaciones de implementación

# Sistemas de diálogo multilingüe

- Estrategias de implementación
  - Elección idioma al inicio del diálogo
  - Detección automática del idioma
- Ideas fundamentales
  - Extracción de **misma representación semántica** independientemente del idioma utilizado (*interlingua*)
  - Crear módulos que sean **independientes del idioma**
    - Información dependiente de idioma almacenada externamente
    - Estrategias de interacción no cambian
    - Usar sistema TTS multilingüe (Ej. Festival)

# Sistemas de diálogo multilingüe



# Puntos a tratar

- Introducción a sistemas de diálogo
- Sistemas de diálogo hablado
- Sistemas de diálogo multimodal
- Sistemas de diálogo multilingüe
- ➔ ● Recomendaciones de implementación

# Recomendaciones de implementación

- Usar indicadores de comprensión del sistema (*feedback*)
  - Usuarios deben ser conscientes del estado de conversación
- Incluir indicadores “sistema ocupado”
  - Evitar que usuarios creen que sistema está funcionando mal
  - si respuesta del sistema se retrasa mucho, usar música de fondo, por ejemplo
- Si hace falta, usar indicadores de “posesión del turno”
  - Interfaz gráfica: icono de un micrófono, por ejemplo
  - Interfaz telefónica: tonos



# Recomendaciones de implementación

- Limitar posibles entradas del usuario para evitar errores
  - “Disculpe, ¿Ha dicho Sevilla o Melilla?”
- Usar confirmaciones
  - “¿Ha dicho que quiere ir a Sevilla?”
- Solicitar ayuda al usuario
  - “Por favor, ¿puede deletrear la ciudad de destino?”

# Recomendaciones de implementación

- En sistemas multimodales, usar las diversas modalidades de entrada disponibles. Por ejemplo:
  - Usar lápiz de PDA para escribir palabra que sistema no reconoce
  - Usar cámaras para detectar cuándo usuario está hablando
- Estrategia de prompts acertada
  - Generar mensajes breves y concisos
  - Permitir interrupciones de usuarios (barge-in)

# Referencias

- **Libro**
  - Ramón López-Cózar, Masahiro Araki, 2005. Spoken, Multilingual and Multimodal Dialogue Systems. Development and Assessment. Wiley

