

Capítulo1. *Interfaces Conversacionales e Iniciativa Mixta*

1.1 Introducción

El reconocimiento de voz tiene una larga historia, y se distingue por ser uno de los problemas difíciles dentro de la ciencia de la computación y de la inteligencia artificial entre otros.

En el desarrollo de tareas de reconocimiento, y en el caso particular del habla, se crea un desafío en un dominio de conocimiento, en el cual se incluyen muchos de los requerimientos del comportamiento inteligente, ejemplo de esto es: operar en tiempo real, explotación de un amplio campo de conocimiento y de información, tolerancia al error, manejo de situaciones desconocidas e inesperadas, uso de símbolos y abstracciones, comunicación en Lenguaje Natural y aprendizaje del entorno.

En la actualidad para que los sistemas de cómputo, puedan ser adoptados por la mayoría de los usuarios, es necesario hacerlos más accesibles, más atractivos y fáciles de usar. Dicho en otras palabras los sistemas al poseer la capacidad de aprender las preferencias, intereses y hábitos de los usuarios, podrán responder de forma más similar a sus deseos, interactuando y hasta prediciendo sus necesidades de información. Además, estos sistemas también deberán comprender ideas incompletas o ambiguas. Este tipo de ideas suelen presentarse en las primeras etapas que son concebidas. En el caso contrario, dar mayor agilidad y precisión a transacciones entre el o los usuarios con los sistemas, en específico para usuarios experimentados en el uso de estos sistemas, o para aquellos con ideas concretas, sobre lo que quieren lograr a través de esta interacción.

Una forma de satisfacer las necesidades de información, o como un medio para hallar una solución viable a los problemas de los usuarios, es la de crear interfaces con las que se pueda interactuar con los sistemas de cómputo de una forma más natural, como los humanos, en su forma más utilizada lo logran; haciendo uso del lenguaje hablado.

1.2 Punto de partida para las Interfaces Conversacionales

Las interfaces conversacionales son una tecnología relativamente nueva. Teniendo su primera aparición a finales de la década de los 80's, surgen como resultado de los esfuerzos financiados por gobiernos en ambos lados del Atlántico, por una parte Estados Unidos lanza su programa en Sistemas de Lenguaje Hablado a través de la organización **DARPA** (**D**efense **A**dvanced **R**esearch **P**rojects **A**gency), y por la otra en Europa el programa Esprit **SUNDIAL** (**S**peech **U**nderstanding and **DIAL**og). Los dos programas se desarrollaron en paralelo en cada continente, y generalmente involucraban el acceso a una base de datos para la planeación de viajes. Es importante mencionar que el programa europeo fue un trabajo de fin multilingüe (siendo precisos, para el Alemán, Francés, Inglés e Italiano). Estos sistemas a la fecha han alcanzado capacidades de operación en tiempo real, y de instalación en cualquier estación de trabajo o PC, sin agregar hardware adicional.[Zue 1999]

1.2.1 Interfaces Conversacionales

El termino “interfaz conversacional” puede tener significados diferentes para diferentes personas, sin embargo por lo general implica un sistema interactivo el cual opera en un dominio restringido. Asimismo muchas interfaces de voz pueden ser consideradas conversacionales sin embargo, “Un aspecto que diferencia a este tipo de interfaces, es el grado en el cual los sistemas toman un rol activo en la conversación.”[Glass 1999]

Estos sistemas permiten a los usuarios interactuar con un equipo de computo para recobrar información, conducir transacciones, o para desarrollar otras tareas en la solución de problemas. La arquitectura de este tipo de interfaces varia significativamente, de acuerdo a las capacidades ofrecidas por la misma, desde aquellas que sólo obedecen a comandos específicos, hasta aquellas que permiten un diálogo continuo y fluido.

1.2.2 Arquitectura general de las Interfaces Conversacionales

En la figura 1.1 se muestra una arquitectura típica de una interfaz conversacional, la cual, como mínimo se compone de los siguientes elementos:

- ?? Un *Reconocedor Automático de Voz* (ASR, Automatic Speech Recognizer), para desarrollar la conversión del habla a texto.
- ?? Módulo de *Entendimiento de Lenguaje Natural* “Parser” (NLU, Natural Language Understanding), el cual se encarga de obtener una representación significativa de lo obtenido por el Reconocedor.
- ?? *Manejador de Diálogo* (o controlador de diálogo), el cual se encarga de controlar la interacción con el usuario.
- ?? Un módulo *Generador de Lenguaje*, este va a formar oraciones, para presentar o pedir información, y que estas puedan ser entendidas por el usuario.
- ?? Un mecanismo para transmitir la información obtenida al usuario, por lo general se utiliza un *Sintetizador de Voz* (TTS, Text to Speech System), pero también pueden incluirse pantallas o cualquier otro medio gráfico de visualización.

En sistemas más complejos además de agregar los módulos antes mencionados también incluyen un mecanismo para el manejo de fenómenos del Diálogo (disfluencias, interrupciones, confirmaciones, clarificaciones, elipsis co-referencias etc.), agregado al módulo de Manejo del Diálogo.

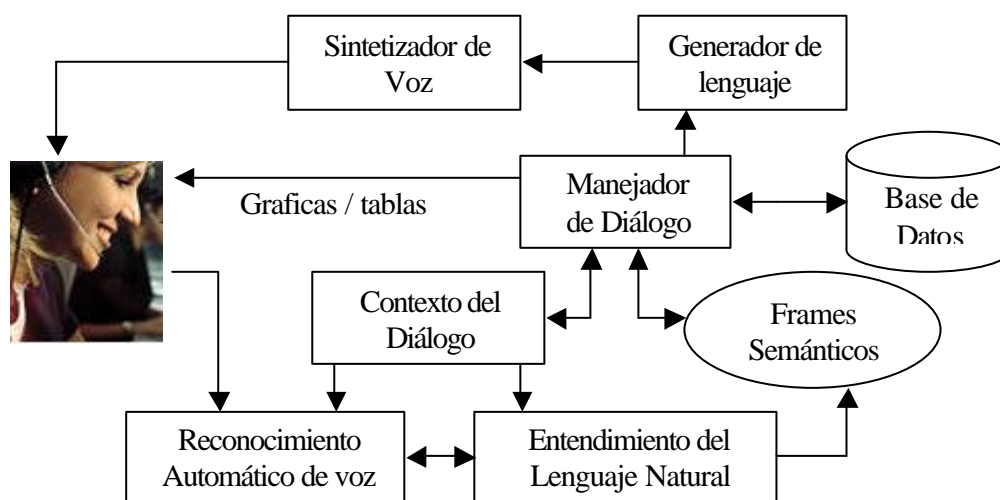


Figura 1.1. Diagrama a Bloques de una Interfaz Conversacional Típica. [Glass 1999]

1.2.3 El proceso incluido en las Interfaces Conversacionales

La manera en que opera una interfaz conversacional se puede describir brevemente de la siguiente manera: La entrada de información hecha a través de la voz es en primer término procesada por el componente de Reconocimiento de Voz Automático (ASR), después el módulo de Entendimiento de Lenguaje Natural (NLU), trabaja en coordinación con el reconocedor obteniendo una representación significativa, de lo que el primero produce (generalmente una cadena de texto), analizando y verificando que sea una entrada válida para el sistema, esta representación significativa puede usarse para recuperar información en forma de texto, tablas o gráficas. Si la información obtenida por el reconocedor es insuficiente o ambigua, el sistema puede hacer preguntas al usuario formuladas por el módulo Generador de Lenguaje, y coordinado por el manejador del Diálogo, para resolver las ambigüedades presentadas en la entrada. El mecanismo para transmitir la información obtenida es generalmente un Sintetizador de Voz (TTS) el cual produce respuestas habladas de la información requerida o bien transmite las preguntas para resolver las ambigüedades.

A lo largo del proceso, la información del diálogo se mantiene y retroalimenta a los elementos de reconocimiento y entendimiento del lenguaje para que las oraciones puedan ser entendidas propiamente de acuerdo al contexto en que aparecen.

1.3 En busca de una comunicación confiable

Cuando un usuario interactúa con un sistema de cómputo de capacidades determinadas a través de la voz, la confiabilidad que otorgue el sistema es esencial. Hasta el momento la mayoría de las aplicaciones que se han desarrollado son frágiles, por llamarlo de alguna manera; si se logran transacciones exitosas con su uso, estas se deben en su mayoría a la cooperación del usuario y a su notable capacidad de adaptarse a la conversación en situaciones de ruido extremo o a un canal de comunicación deficiente.

1.3.1 Mejorando la interacción con las Interfaces Conversacionales

Para mejorar la interacción con las interfaces conversacionales, se debe incorporar características tales como: dar mayor flexibilidad en su uso, una mayor capacidad de adaptación en situaciones de ruido extremo o a canales de comunicación deficientes. De lograrse esto las interfaces conversacionales podrían ser más robustas y confiables, agregando un mejor desempeño.

Una interfaz conversacional además deberá poseer la capacidad para responder a preguntas, que puedan ser contenidas en diferentes categorías. Situaciones como esta, se presentan por ejemplo en la frase: “ ¿Si viajo el miércoles, el boleto me cuesta menos?”, esta frase es el prelude de una transacción por realizarse y que establece las bases para una nueva pregunta como la siguiente: “¿Bueno, que tal el viernes?”. Otro ejemplo puede ser todas aquellas frases que representan un acto indirecto por ejemplo “¿Cuál es el precio actual por viajar en su línea?”.

Agregando la combinación de algunos procesos, como lo es el de clarificación y el uso de las características propias de un proceso de conversación, se puede analizar un diálogo para resolver ambigüedades sobre lo que alguna de las partes involucradas en este proceso acaba de decir. Con esto también se contribuye al logro de un mutuo entendimiento, basado en intervenciones dentro de un contexto de diálogo o conversación.

1.3.2 Acerca del propósito de las Interfaces Conversacionales.

Cuando hablamos del uso de una interfaz conversacional no sólo nos referimos al entendimiento de una y otra parte, y ciertamente desde el punto de vista del usuario, él además desea que el sistema desarrolle una actividad o servicio determinado. Esto es muy difícil e importante, ya que de lograrse lo anterior, se obtendría una base de gran alcance para poder obtener una interacción conversacional refinada.

La respuesta a preguntas sobre las capacidades que ofrece una interfaz conversacional, nos permiten darnos cuenta, que no son la solución para todo, pero que pueden lograr un buen desempeño en determinado dominio de problemas. Además si esta misma interfaz, tiene la cualidad de que al modificar ciertos parámetros de ella, esta puede ser aplicada a un dominio diferente, la hace más versátil.

1.3.3 Progreso reciente en las Interfaces Conversacionales

En la pasada década se ha presentado un incremento en la actividad investigadora, en el área de las interfaces conversacionales. A partir del lanzamiento de los programas de desarrollo de este tipo de sistemas, tanto en EU a través de la organización DARPA y en Europa el Esprit SUNDIAL, se ha venido enfocando el trabajo investigador para dar un énfasis a las interfaces basadas en el diálogo, asimismo por parte del proyecto de DARPA ha tratado de estandarizar las arquitecturas de estas interfaces, con el fin de obtener un enfoque compartido en los componentes de sus prototipos [Sennef 1998]. En Europa desde su inicio, el programa tuvo un fin multilingüe. Este programa promueve la cooperación de diferentes países, obteniendo de cada uno de ellos, componentes que puedan ser integrados en un solo sistema que responda a diferentes lenguajes (Alemán, Francés, Italiano, etc).

Por la parte independiente, existen diversos proyectos; podemos mencionar algunos: El sistema administrador de oficina desarrollado por el CMU (Carnegie Melon University) enfocado a proveer el desarrollo de aplicaciones de voz para una oficina del futuro [Rudnicky 1991]. En los laboratorios de la AT&T se desarrolla el sistema "AutoRes", este sistema permite a los usuarios la renta de autos vía telefónica, El sistema WAXHOLM permite obtener horarios e información turística sobre los servicios disponibles en el Archipiélago de Estocolmo. En EUA en la *Universidad de Colorado en el CSLR (Center for Spoken Language Research)* se ha desarrollado el sistema *CU Communicator* el cual permite, también por vía telefónica elaborar un plan de viaje, permitiendo hacer reservaciones para vuelos, habitaciones de hotel y renta de autos en distintas ciudades de E.U. De este sistema se presentara su composición y funcionamiento con un mayor detalle, en el siguiente capítulo de esta tesis.

Por último un área de investigación concerniente a las interfaces conversacionales que atrae mucho la atención al conjunto de investigadores, son los *Agentes Conversacionales*, los cuales son típicamente interactivos con el usuario, además estos mismos agentes pueden trabajar independientemente (por ejemplo al ordenarle: “Avísame cuando falte hora y media para la salida de X vuelo, en Y aeropuerto). Estos sistemas son candidatos potenciales para incorporar a ellos entradas y salidas multimodales de información. Agregado a esto se busca también tener avances en la obtención de estrategias para el manejo del diálogo [Pérez 2001].

1.3.4 Interfaces Conversacionales orientadas al diálogo

En el desarrollo de interfaces conversacionales existen diversos enfoques. Una estrategia para el desarrollo de este tipo de interfaces, son las orientadas al diálogo, en donde se hace uso de Técnicas de Iniciativa Mixta, y donde también se reafirma el paradigma conversacional [Glass 1999].

Aquí el sistema es más flexible al manejar restricciones para el usuario, permitiendo procesar entradas lingüísticas más complejas. Estos sistemas usualmente intentan negociar con el usuario las restricciones que debe cumplir para efectuar determinada tarea, esto se logra permitiendo que el usuario, tenga una mayor flexibilidad en el uso del sistema.

Sin embargo este tipo de sistemas pueden tener altos índices de error, y pueden mostrarse confusos para usuarios que no están familiarizados con esta tecnología. Para reducir este tipo de situaciones algunos sistemas toman un enfoque híbrido, presentando de nuevo un diálogo restringido, cuando el sistema detecta que el usuario está teniendo problemas para expresarse o para entender la forma de operar del sistema [Sennef 1992].

1.4 Introducción a la Teoría de la Iniciativa Mixta

En una situación en la que se desea resolver un problema, la información y las habilidades necesarias para esta tarea, son a menudo distribuidas entre los colaboradores. Como un resultado, la dirección y el control de la interacción cambia entre los participantes. Si los sistemas computacionales están colaborando efectivamente con los usuarios para resolver problemas, estos también tendrán en algún momento la habilidad de tomar y abandonar el control del proceso para la solución del problema y la comunicación involucrada en el mismo.

La teoría y los mecanismos que son la razón de este comportamiento, son los tópicos de esta cuestión en especial, y son estos mismos los que permiten modelar la interacción adaptativa del usuario a modelos computacionales, para lograr una interacción que pueda ser incluida como de Iniciativa Mixta.

1.4.1 El objetivo de la investigación en Iniciativa Mixta

Como esta descrito en [Burstein & Mc.Dermont 1996] “El objetivo total de la investigación en la planeación con Iniciativa Mixta es el de explorar síntesis productivas de fuerzas complementarias en los humanos y las máquinas para construir planes efectivos, más rápidamente y con una gran confiabilidad”.

Así mismo ellos también hacen una clara distinción sobre la motivación de esta investigación. “Nuestro más grande interés en los sistemas de planeación con Iniciativa Mixta crece sobre algunas observaciones de las fortalezas y debilidades de los humanos, de los sistemas de planeación automática, y de cómo estos son usados. Los humanos son... mejores en la planeación de tareas y las máquinas son mejores en la búsqueda sistemática de los posibles espacios para estos planes”.

1.4.2 Términos involucrados en la teoría de la Iniciativa Mixta

Como un punto de partida en el estudio de esta teoría, es necesario definir los siguientes términos:

Iniciativa: ‘s. Habilidad para iniciar cosas, capacidad de emprender, primer paso, idea que inicia una acción, poder o derecho de comenzar, inventar’.

Iniciar: ‘v.t. Originar, comenzar, poner en marcha, introducir’.

Control: ‘s. Capacidad de dirigir o moderar, autorregulación, significado de moderar o regular algo’.

Flujo: ‘v.i. Deslizamiento a lo largo de una corriente, moverse como un líquido, raudal, moverse diluida y sostenidamente’.

Diálogo: ‘s. Coloquio, conversación entre dos o más personas, debate’.

Plan: ‘s. Intento, proyecto que se tiene de realizar una cosa, programa detallado de una obra, acción, conjunto de disposiciones tomadas para llevarla a cabo’.

Tarea: ‘s. Obra, trabajo, trabajo que debe hacerse en tiempo limitado, afán, penalidad por un trabajo continuo’.

Tópico: ‘s. Tema, asunto, cuestión, tema de conversación en general’.

Meta: ‘s. fin que se desea alcanzar’.

Interacción: ‘v.i. Ejercer una acción recíproca’.

Estas definiciones fueron tomadas de [Oxford 1984].

1.4.3 Definiciones propuestas de Iniciativa

1ª DEFINICIÓN: “Iniciativa es el control sobre el flujo de una conversación”

[Haller et al 1999]

Como resultado de esta definición se desprenden tres razones esenciales por las cuales se puede tomar la iniciativa, las cuales son:

- ?? Cuando el escucha sabe o intenta inducir el significado de lo que el locutor esta tratando de decir, antes de que este termine.
- ?? Cuando el escucha tiene otra meta más importante para él, la cual cree que necesita que sea satisfecha antes de la actual.
- ?? Cuando el escucha no esta muy interesado en el plan de solución actual.

2ª DEFINICIÓN: “Iniciativa es el ejercicio del poder o la habilidad, de un participante en un diálogo, para sugerir (o desarrollar) un plan (o tarea) que contribuye a resolver un problema presente” [Haller et al 1999].

En esta definición se propone la existencia de diferentes niveles de iniciativa entre los participantes, que dicen tener la iniciativa. Aquí implicando a un contexto de diálogo colaborativo, la iniciativa es tomada como el proceso para proveer soluciones, los participantes que son los más aptos en portar soluciones son los que pueden tomar la iniciativa. Esta definición argumenta también que para ciertos dominios la iniciativa no necesita modelarse explícitamente, sin embargo no abandona la importancia de seguir el transcurso del intercambio de la iniciativa, con el fin de optimizar el proceso de solución.

3^a DEFINICIÓN: “ La Iniciativa en una conversación ocurre en el momento en que una persona obtiene el control de la conversación, al hacer una intervención, que se encuentra dentro del dominio de la meta, que los participantes intentan alcanzar” [Haller et al 1999].

En esta definición se aprecia un interés de reconocer más fácilmente a las discusiones interactivas, y de poder obtener un modelo basado en turnos además de que este adquiere un carácter mixto en la participación. También esta definición se diferencia de las anteriores en que toma a la Iniciativa no como un control en el diálogo, sino como una ocurrencia discreta. Esto permite a los diseñadores de sistemas, determinar la profundidad de una meta por alcanzar, y de escoger el nivel de complejidad en la interacción que se desea modelar.

4^a DEFINICIÓN “ Es cuando los sistemas con Adaptación Interactiva, son capaces de interactuar de forma adecuada y efectiva con un humano” [Allen 1994].

La definición involucra tanto a un agente humano como a uno computacional, que pueden tomar la iniciativa en varios puntos del diálogo. Esta definición implica que esta teoría puede ser aplicable a los sistemas de planeación colaborativa, donde la Iniciativa sigue siendo el primer paso de un proceso, y un proceso es un conjunto de turnos, en el cual se persigue una meta. El enfoque de esta definición está orientado al diálogo y al control del mismo diálogo, y es esta misma definición la que nos da un acercamiento mayor para la obtención de una definición universal sobre Iniciativa Mixta. Con la obtención de los modelos de participación interactiva, estos pueden intentar ser plasmados a modelos computacionales.

1.4.4 Comparación entre las definiciones propuestas sobre Iniciativa

La primera definición se enfoca en el desarrollo del control en una conversación, sin embargo simples afirmaciones no siempre implican un cambio en la iniciativa. De las demás definiciones la cuarta es la más cercana a la primera, porque esta también se enfoca a modelar la estructura de un diálogo. Esta definición tiene características adicionales que permiten identificar grados de iniciativa y una forma de agrupar los turnos de intervención en el proceso de diálogo. La cuarta definición considera un proceso dirigido a la meta en un conjunto de intervenciones, y la primera definición permite a un participante obtener la iniciativa cuando una sub-meta es propuesta.

La segunda y tercera definiciones se enfocan en las metas descritas en una conversación, y no desarrollan un modelado del diálogo. Las dos definiciones tratan de diferente forma a las intervenciones hechas por los participantes que dialogan en busca de una solución. La segunda definición trata a la iniciativa en un diálogo, como situaciones donde no únicamente un participante tiene la iniciativa. La tercera definición intenta tratar los casos donde un participante esta ofreciendo información, dirigiendo el plan de solución, y actualmente contribuye también con una nueva dirección para el mismo plan, y esta dado como una situación en donde la iniciativa ha cambiado hacia este participante.

1.4.5 Definición de Iniciativa Mixta

Cuando existe una interacción, si un participante es un sistema y tiene la iniciativa, se dice que es una interacción dirigida por el sistema, en el caso de que el otro participante sea un usuario de este sistema y él tiene la iniciativa, entonces se dice que la interacción está dirigida por el usuario. Cuando ambos pueden tomar o abandonar la iniciativa, en cualquier estado de la interacción se dice que existe *Iniciativa Mixta*.

Definimos Iniciativa Mixta como: “*Un modo de interacción particular que se presenta en un diálogo entre uno o más participantes, el cual está orientado a lograr una meta, y en donde para el caso particular de los sistemas, el o los usuarios junto con ellos participan interactivamente para resolver un problema, usando un paradigma conversacional.*”

1.4.6 ¿Quién decide el grado de Iniciativa en una interacción?

En los turnos de una conversación, pueden encontrarse distintos grados de Iniciativa, si estos contienen intervenciones en las cuales puede haber una transferencia en el control del diálogo, las propiedades que muestran grados de iniciativa son:

- ?? Un turno que resulta con un cambio de tópico tiene fuerte iniciativa, debido a que este no solo comienza un nuevo proceso, sino un nuevo tópico (secuencia de procesos) también.
 - ?? Los turnos en los cuales se da información que no es solicitada, también tienen fuerte iniciativa, porque esta información a menudo comienza un nuevo proceso.
 - ?? Los turnos en los cuales el locutor pide un consejo posee fuerte iniciativa, pero no tanta como en los turnos en los que son dadas ordenes.
 - ?? Los turnos en los cuales el locutor interviene con una mayor intención o seguridad, posee fuerte iniciativa más que en aquellas que hace deliberadamente.
-

Hasta aquí, podemos decir que hemos dado el primer paso en el intento de ganar un mayor entendimiento de lo que significa la Iniciativa Mixta, y de su potencial mostrado en el uso de esta, para el diseño de sistemas conversacionales que incluyen técnicas de este tipo. En el siguiente capítulo mostraremos el análisis a detalle de la interfaz conversacional desarrollada por la Universidad de Colorado el “CU Communicator”, el cual a su vez tiene integrado un analizador semántico robusto que es conocido como el sistema “Phoenix”, y que de su forma original ha sufrido modificaciones para lograr un grado mayor de integración al módulo de reconocimiento.