

Procesamiento del Lenguaje Natural

Tema 10. Agentes conversacionales

Índice

Esquema

Ideas clave

- 10.1. Introducción y objetivos
- 10.2. Definición de agente conversacional
- 10.3. Características de las conversaciones entre humanos
- 10.4. Tipos de agentes conversacionales
- 10.5. Estructura de los agentes conversacionales
- 10.6. Diseño de chatbots
- 10.7. Referencias bibliográficas

A fondo

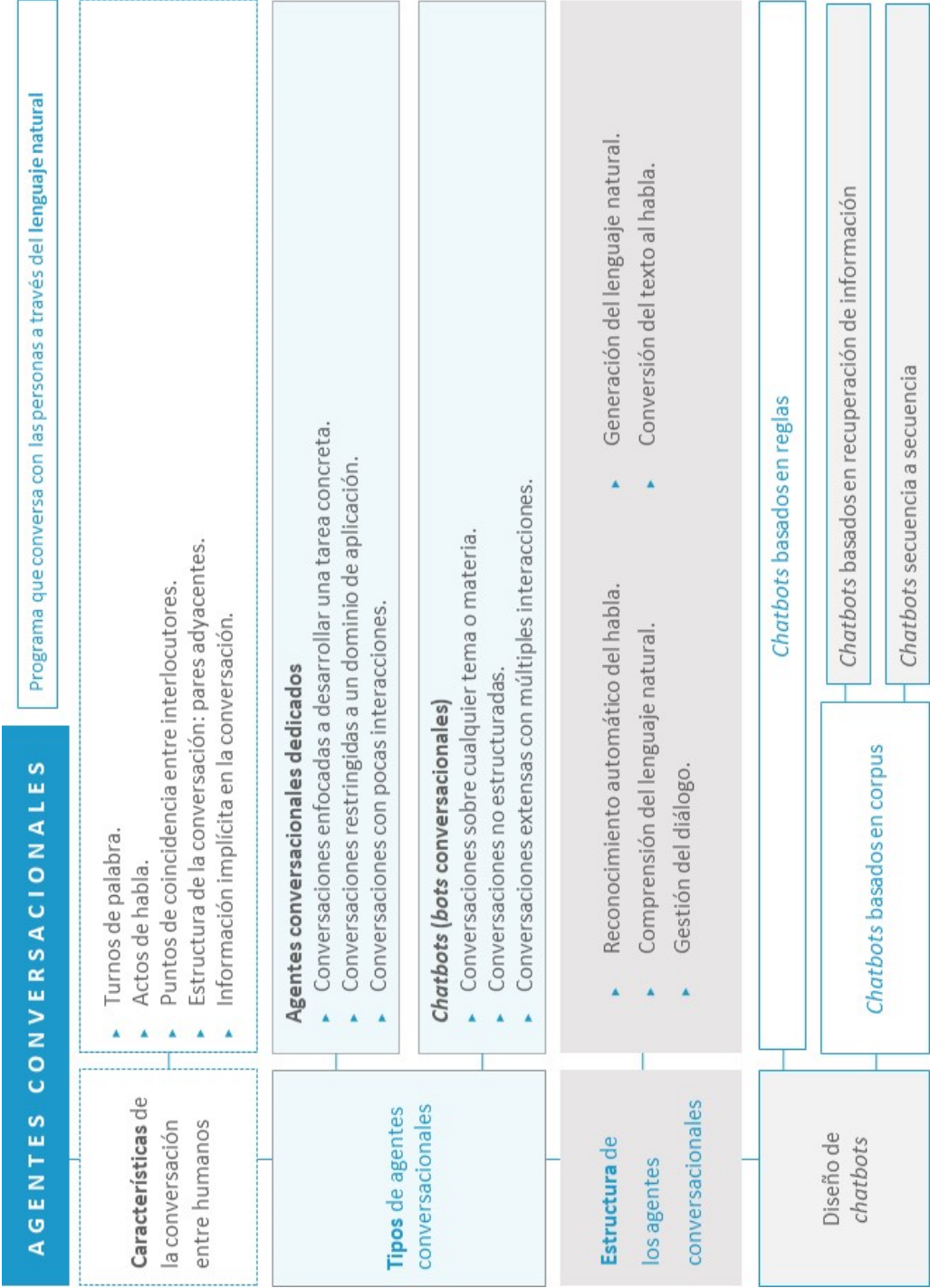
Selección de las transcripciones de los finalistas del Premio Lobner 2017

Como superar el test de Turing (ganar el Premio Loebner)

Google Assistant realizando una llamada durante el Google I/O '18

Discurso y diálogo

Test



10.1. Introducción y objetivos

A continuación, se definirá lo que es un **agente conversacional**. Analizamos las características más relevantes de las conversaciones entre humanos que son las que determinan las características de los agentes conversacionales. Además, se describen los **agentes conversacionales dedicados y los *chatbots***, dos tipos de agentes conversacionales que se diferencian por el tipo de conversación que son capaces de mantener y la funcionalidad que pretenden implementar. También, presentamos la estructura básica de un agente conversacional y se describe la utilidad de cada uno de los módulos que la componen. Por último, veremos los **diferentes enfoques de cómo diseñar un *chatbot***.

Objetivos

- ▶ Conocer lo que es un agente conversacional.
- ▶ Conocer las características más relevantes de las conversaciones entre humanos que son las que determinan las características de los agentes conversacionales.
- ▶ Identificar los diferentes tipos de agentes conversacionales dependiendo del tipo de conversación que son capaces de mantener y la funcionalidad que pretenden implementar.
- ▶ Representar gráficamente la estructura básica de los agentes conversacionales y describir la utilidad de cada uno de los módulos que la componen.
- ▶ Conocer los diferentes enfoques utilizados para diseñar un *chatbot*.

10.2. Definición de agente conversacional

La Real Academia Española (RAE, s. f.) en su diccionario de la lengua española define la palabra diálogo:

«Del lat. dialógus, y este del gr. διλογος diálogos. 1. m. Plática entre dos o más personas, que alternativamente manifiestan sus ideas o afectos». Este mismo diccionario define la palabra plática: «De práctica. 1. f. conversación (|| acción de hablar)».

Los agentes conversacionales, también llamados sistemas de diálogo, son programas que conversan con las personas a través del lenguaje natural.

Los agentes conversacionales pueden interactuar con el humano, ya sea a través de la voz (hablando con el usuario), de texto, en el caso de que la conversación se lleve a cabo a través de un chat de texto, o utilizando ambas modalidades a la vez. Los **agentes conversacionales multimodales integran** el habla con **otras modalidades** como los gráficos o las imágenes.

Algunos agentes conversacionales modernos están empezando a incorporar las emociones, por ejemplo: sorpresa, enfado, alegría, tristeza o estrés, en la conversión de texto en habla para dar más realismo a la respuesta del agente conversacional. De hecho, las emociones también pueden formar parte del proceso de comprensión del lenguaje y la posterior generación del diálogo más adecuado a las emociones detectadas.

Las emociones se pueden detectar del propio contenido del mensaje, pero también se podrían extraer del análisis de los gestos y expresiones faciales de la persona que interactúa con el agente conversacional.

En el caso más sencillo de agente conversacional, este es capaz de contestar a las preguntas del usuario o de realizar una acción de control a partir de un comando de voz. Este es el caso de los asistentes virtuales que tienen una única interacción con la persona. Sin embargo, también existen agentes conversacionales más complejos capaces de continuar con la interacción con el usuario y, por ejemplo, contestar a una segunda pregunta relacionada con la primera. Sería el caso de agentes conversacionales que ya incluyen diálogo y que se han aplicado en asistentes para planificar y gestionar viajes, como en el caso de GUS (Bobrow et al., 1977).

Los agentes conversacionales incluso pueden usarse en dominios mucho más complejos como los agentes conversacionales pedagógicos que actúan como **tutor del estudiante**, como sería el caso de ITSPOKE (Forbes-Riley y Litman, 2011). Por último, los agentes conversacionales se pueden usar simplemente como **divertimiento para charlar con ellos**, pero sin que la idea sea realizar una tarea concreta, como en Cleverbot.

10.3. Características de las conversaciones entre humanos

Las principales características de los agentes conversacionales se deben a las características intrínsecas que tiene una conversación entre humanos. Una conversación es una acción que se realiza de forma compartida entre varias personas y que implica un alto grado de complejidad, ante esto, las conversaciones entre una persona y una máquina intentan **emular los comportamientos que se dan en las conversaciones entre humanos**.

Turnos de palabra

Una conversación se compone de **turnos de palabra** que se dan de **forma consecutiva** y que alternan las opiniones de los interlocutores. En el caso de que dos personas mantengan una conversación, el interlocutor A declara algo, el interlocutor B le responde, luego el interlocutor A vuelve a expresarse, y así sucesivamente.

Las personas somos capaces de identificar fácilmente quién es la siguiente persona a la que le toca expresarse y el momento exacto en el que debe intervenir. Por lo tanto, durante una conversación hablada, los humanos somos capaces de esperar a que nuestro interlocutor acabe de hablar antes de responder, identificar que nuestro interlocutor ya ha acabado su frase y saber que somos nosotros los que debemos responder, determinar el momento preciso en el que debemos responder (por ejemplo, un instante después de que acabe de hablar nuestro interlocutor), o interpretar que un silencio muy prolongado de nuestro interlocutor puede significar que le incomoda la pregunta que le hemos realizado y que no la va a contestar.

En las conversaciones entre humanos, los **turnos de palabra están claros** y se asignan de **forma intuitiva**, sin embargo, para los agentes conversacionales llegar a manejar correctamente los turnos de palabra y gestionar eficientemente los tiempos

es una tarea complicada. La gestión de los turnos de palabra en los agentes conversacionales se estudia en el campo del **análisis conversacional** (Levinson, 1983) y se basa en la idea de que el intercambio de turnos se puede modelar con una serie de reglas. Estas reglas, que indican cómo manejar los turnos de palabra, se evalúan en los momentos en que la estructura del lenguaje permite intercambiar el interlocutor, momentos llamados en inglés *transition-relevance places* (TRP) (Sacks, Schegloff y Jefferson, 1974).

Actos de habla

Se ha comentado anteriormente que una conversación se compone de **una secuencia de turnos de palabra**. Además, se puede añadir que cada uno de los turnos incluye una o varias expresiones o declaraciones. La teoría de los actos de habla sugiere que **cada expresión en una conversación** es un **tipo de acción** que **realiza el interlocutor** (Austin, 1962). Esta teoría fue desarrollada por el filósofo británico John L. Austin antes de morir en 1960 y presentada en su obra póstuma, en la cual se acuñó el término acto de habla.

Un acto de habla es cada uno de los tipos de acciones que involucran el uso del lenguaje natural.

En una expresión tal como:

- ▶ A mi nuevo gato lo he llamado Felix.

La acción de llamar a mi gato Felix tiene la consecuencia de que el gato tenga el nombre de Felix. En este caso queda claro que la acción que realiza el interlocutor y que se recoge en la expresión ha permitido cambiar el estado del mundo haciendo que el gato que no tenía nombre tenga ahora uno. Sin embargo, el trabajo de Austin sugiere que todas las expresiones tienen asociado algún acto de habla, aunque este no sea tan evidente como el descrito anteriormente.

Los actos de habla definidos por Austin se han ido extendiendo con el paso de los años y hoy en día podemos hablar de actos que cambian el estado del mundo, sugieren al interlocutor que haga algo, le dan directrices para que haga algo, prometen o planifican que el interlocutor va a hacer algo en el futuro o expresan el estado psicológico o la actitud del interlocutor sobre una acción.

En una conversación los humanos somos capaces de hacer preguntas, dar órdenes o informar sobre algo sin darnos cuenta de que nuestras expresiones contienen diferentes tipos de los actos de habla definidos por Austin. Sin embargo, para los agentes conversacionales no es trivial saber interpretar en cada momento en qué acto de habla se encuentran y tomar las decisiones correctas para responder de la forma más adecuada.

No se puede afirmar que todos los agentes conversacionales sean capaces de tratar con múltiples de actos de habla. De hecho, solo algunos agentes conversacionales modernos implementan funcionalidades más avanzadas y son capaces de lanzar una pregunta, dar una orden, hacer una propuesta, rechazar una sugerencia o estar de acuerdo con una expresión presentada por su interlocutor.

Puntos de coincidencia entre interlocutores

Como hemos visto, cada turno de palabra y, más en concreto, cada expresión del turno se puede ver como una acción de un interlocutor. Sin embargo, un diálogo no se compone de una serie de actos independientes sin relación alguna. Sino que, a diferencia de en un monólogo, **una conversación es un acto colectivo** realizado entre ambos interlocutores. Debido a esta acción conjunta que es una conversación, se deben establecer constantemente unos puntos de coincidencia entre los interlocutores (Stalnaker, 1978).

Los puntos de coincidencia son el conjunto de fundamentos en los que creen los interlocutores que participan en una conversación.

Es necesario asegurar que se han establecido unos puntos de coincidencia entre interlocutores para que la conversación pueda desarrollarse con éxito. Esto significa que el receptor del mensaje debe dejar claro que ha entendido el significado y la intención de la expresión lanzada por el emisor, y también el caso contrario, indicar que no ha entendido al emisor del mensaje (Clark, 1996).

En las conversaciones entre humanos **se utilizan diferentes técnicas** para **garantizar** que se están estableciendo los **puntos de coincidencia entre interlocutores** (Clark y Schaefer, 1989). En una conversación, el receptor del mensaje puede asentir o utilizar palabras tipo «vale» o «genial» u onomatopeyas tipo «ajá» para mostrar que está entendiendo y, además, está de acuerdo con su interlocutor.

Otras opciones que garantizan los puntos de coincidencia serían que el receptor del mensaje reproduzca textualmente, reformule o parafrasee lo que le ha transmitido su interlocutor o que le ayude a acabar de construir su argumentación. Además, el receptor del mensaje puede mostrar interés por la conversación y entonces se interpreta directamente que está conforme con lo que le comenta su interlocutor. Por último, el receptor puede simplemente pasar a proponer su siguiente contribución al diálogo, por lo que se supone que está de acuerdo con lo que se ha comentado hasta ese momento.

Es imprescindible que los agentes conversacionales imiten todos los comportamientos que se acaban de describir para poder asegurar que se establecen puntos de coincidencia entre el agente y su interlocutor humano. De hecho, se ha demostrado que, si el agente no garantiza que ha entendido al humano, va a crear **confusión a la persona** y esta no se va a sentir cómoda utilizando el agente

conversacional (Yankelovich, Levow y Marx, 1995).

Ejemplo ilustrativo 1

Posibles conversaciones de un asistente virtual que ayuda a un cliente de una operadora de telefonía a configurar su perfil de usuario, consultar su consumo, etc.

Conversación A:

Asistente: ¿Desea revisar algo más de su perfil de usuario?

Cliente: No.

Asistente: ¿Qué desea hacer ahora?

Conversación B:

Asistente: ¿Desea revisar algo más de su perfil de usuario?

Cliente: No.

Asistente: Entonces ¿qué desea hacer ahora?

En el ejemplo de la conversación A, el cliente no puede asegurar que el asistente virtual haya entendido que le ha dicho que no porque este agente conversacional no le ha corroborado la recepción de la información, ya sea de forma explícita o implícita. Entonces, se puede afirmar que en la conversación A no se ha establecido un punto de coincidencia entre la persona y el agente conversacional.

Sin embargo, en la conversación B, por el mero hecho de que el asistente virtual haya utilizado el adverbio «entonces», se sobreentiende que el agente conversacional ha entendido que el cliente ha respondido que no y le está lanzando una nueva pregunta. En este caso, sí se ha establecido el

punto de coincidencia entre la persona y el agente conversacional, tan necesario para el correcto funcionamiento del agente.

Estructura de la conversación

Se ha comentado en los apartados anteriores que la conversación se compone de una serie de turnos de palabra relacionados entre sí y en los que los interlocutores exponen diferentes expresiones. También, se ha explicado que la conversación es un acto colectivo en el que requiere que ambos interlocutores lleguen a establecer una serie de puntos de coincidencia. Para asegurar los fundamentos de la conversación y garantizar que el receptor del mensaje ha entendido al emisor, se han presentado algunas técnicas en las que se veía una clara correlación entre los actos de un turno de palabra y los del siguiente turno. De hecho, aunque las conversaciones tienen una estructura libre, se pueden **identificar algunos patrones bien definidos** en algunas partes de los diálogos entre humanos.

En numerosas ocasiones se va a dar el caso de que un turno de palabra esté relacionado con el siguiente. Por ejemplo, cuando se lanza una pregunta en un turno, se condiciona claramente a que en el siguiente turno de palabra se dé una respuesta. Es por esta razón que se puede hablar de que en las conversaciones se da un patrón de **pares adyacentes** (Schegloff, 1968).

Los pares adyacentes son estructuras que se dan en una conversación del tipo pregunta-respuesta (pregunta seguida de una respuesta), propuesta-aceptación o propuesta-rechazo, petición-concesión, saludo-saludo.

Según el patrón de los pares adyacentes, en una conversación en la que se lance una propuesta (primera parte del par), acto seguido va a darse una aceptación o un rechazo (segunda parte del par). Sin embargo, se puede dar el caso de que, en un

diálogo, la segunda parte del par no aparezca directamente justo después de la primera parte, sino que haya algunas expresiones intermedias. Esta situación se podría dar si el interlocutor necesita hacer una pregunta aclaratoria para recabar información antes de poder aceptar o rechazar la propuesta que le había lanzado el otro interlocutor. En este caso se daría un **subdiálogo** entre la propuesta (primera parte del par) y la aceptación o rechazo de la misma (segunda parte del par).

Aparte de la estructura de la conversación marcada por los pares adyacentes, se pueden encontrar otros patrones en **cómo se organiza una conversación**. Por ejemplo, en una llamada de teléfono siempre habla primero la persona que responde al teléfono, luego lo hace la persona que llama indicándole quién es; en los pasos siguientes, muy posiblemente, se intercambien saludos y luego, la persona que llama indica el motivo de la llamada. Este patrón da comienzo a una llamada telefónica y es solo un ejemplo concreto, pero podríamos tener también un patrón para el final de la llamada. Además, en otros tipos de conversaciones podríamos encontrar otros patrones que describan la estructura de la conversación.

Los agentes conversacionales se van a beneficiar de los patrones en la estructura de la conversación para facilitar la generación de diálogos de forma automática. Por ejemplo, saber que la persona ha lanzado una pregunta y que, por lo tanto, el agente debe proporcionar una respuesta será una información muy útil a la hora de crear un diálogo. Además, si se sabe que se está desarrollando un agente que responde a llamadas telefónicas, se podrá fijar el patrón que dé comienzo o que finalice la conversación.

Información implícita en la conversación

En los apartados anteriores se han tratado las características de la conversación relacionadas con lo que se podría llamar la «infraestructura» de la conversación:

- La conversación es una actividad que realizan de forma conjunta varios interlocutores.

- ▶ Estos generan **turnos de palabra** en los que se exponen una serie de contribuciones que requieren la aceptación por parte del interlocutor y que cumplen con unos patrones predeterminados.

Sin embargo, hasta este punto no se ha comentado nada acerca de la información que se comunica entre los interlocutores durante el diálogo.

En una conversación, el significado de una contribución no se limita al significado de las palabras entendidas de forma individual. Es más, este significado tampoco se limita al significado que pueda tener una expresión concreta.

Esto se debe a que la interpretación de una expresión se basa en algo más que el significado literal de las oraciones que la componen.

En la interpretación de una expresión intervienen el conocimiento que comparten los interlocutores, el contexto en el que se desarrolla la conversación o la intención del emisor del mensaje. Es por eso por lo que se puede decir que en una conversación **hay información** que tiene un **significado literal o explícito**, pero también hay otra información que el receptor del mensaje es capaz de inferir, por lo tanto, **información implícita**.

Las implicaturas o informaciones implícitas son los significados que no aparecen de forma explícita en la conversación, pero que el receptor del mensaje infiere.

La **inferencia** juega un papel crucial en una conversación, el emisor de un mensaje puede hacer llegar al receptor más información de la que le transmite de forma explícita en las expresiones que componen su mensaje. Por ejemplo, si para hacer una reserva de un vuelo, un agente de viajes le pregunta a su cliente que qué día quiere viajar y este le responde que tiene que llegar para una reunión el 15 de mayo,

entonces, el agente de viajes va a interpretar que tiene que reservar el vuelo cuya llegada se dé el día 14 de mayo. En este caso se ve cómo claramente el cliente no ha respondido al agente de viajes con la información que este le había preguntado, sino que le ha proporcionado otra información que de forma implícita responde a su pregunta. De hecho, en este caso el cliente ha decidido de forma intencionada proporcionarle al agente una información explícita que le permita a través de un proceso de inferencia obtener la información implícita requerida.

Los agentes conversacionales también deben ser capaces de extraer la información implícita que les hace llegar un humano cuando mantienen un diálogo.

Para ello, los agentes implementan mecanismos de inferencia que les permiten obtener los significados adicionales de un mensaje a partir del significado literal de las expresiones que lo componen.

Además, los agentes conversacionales para poder funcionar correctamente utilizando las informaciones implícitas deben ser capaces de determinar la intención de su interlocutor. El agente conversacional debe interpretar que el humano no se ha equivocado porque no haya respondido explícitamente la pregunta que le había formulado, sino que lo ha hecho a propósito porque la intención de la persona era transmitir al agente una información implícita.

La inferencia de información implícita en la conversación es una de las tareas más complicadas a las que se enfrenta un agente conversacional a la hora de comportarse como una persona y de reproducir las características que tiene una conversación entre humanos.

10.4. Tipos de agentes conversacionales

Existen dos tipos de agentes conversacionales: los **agentes conversacionales dedicados a una tarea concreta y los *chatbots***. Esta clasificación se hace con base en el **tipo de conversación** que son capaces de mantener los agentes conversacionales y **la funcionalidad u objetivo final** que pretenden implementar.

Agentes conversacionales dedicados

Los agentes conversacionales dedicados a una tarea concreta están **diseñados para desarrollar una tarea en particular** y están restringidos a un dominio de aplicación. Esta única tarea que puede realizar este tipo de agentes sería, por ejemplo, obtener algún tipo de información del usuario o ayudarlo a completar una acción sencilla.

Los agentes conversacionales dedicados están configurados para mantener conversaciones cortas con la persona, en general, conversaciones que involucren desde una interacción a unas pocas de interacciones (en el rango de media docena).

Además, cabe resaltar que las conversaciones mantenidas por estos agentes son **sencillas** y no se realizan de una manera natural como cuando conversan dos humanos.

Los asistentes personales que encontramos en los teléfonos móviles o en algunos dispositivos inteligentes para el control domótico (Google Assistant, Siri de Apple, Cortana de Windows o Alexa de Amazon) son agentes conversacionales que forman parte de la categoría de agentes conversacionales dedicados a una tarea concreta. Estos pueden realizar tareas simples como, por ejemplo, dar instrucciones sobre un viaje, controlar electrodomésticos, encontrar restaurantes, ayudar a hacer llamadas telefónicas o enviar mensajes de texto.

Muchas empresas implementan agentes conversacionales dedicados en sus sitios web para ayudar a los clientes a responder a sus preguntas y dudas para resolver los problemas que pudieran tener con los productos o servicios contratados. Además, los agentes conversacionales juegan un papel importante como **interfaz hombre-máquina en algunos robots**.

Chatbots

Los *chatbots*, llamados en español *bots* conversacionales, son aquellos agentes conversacionales que permiten entablar conversaciones extensas, es decir, con múltiples interacciones entre la persona y el agente conversacional.

Permiten mantener conversaciones no estructuradas, una característica fundamental de las conversaciones entre personas, y sobre cualquier tema o materia a diferencia de agentes conversacionales dedicados a una tarea concreta.

Por tanto, los *chatbots* ofrecen una conversación realista entre una persona y el agente conversacional utilizando el lenguaje natural.

Ejemplo ilustrativo 2

Mitsuku: el *chatbot* ganador del Premio Loebner 2017

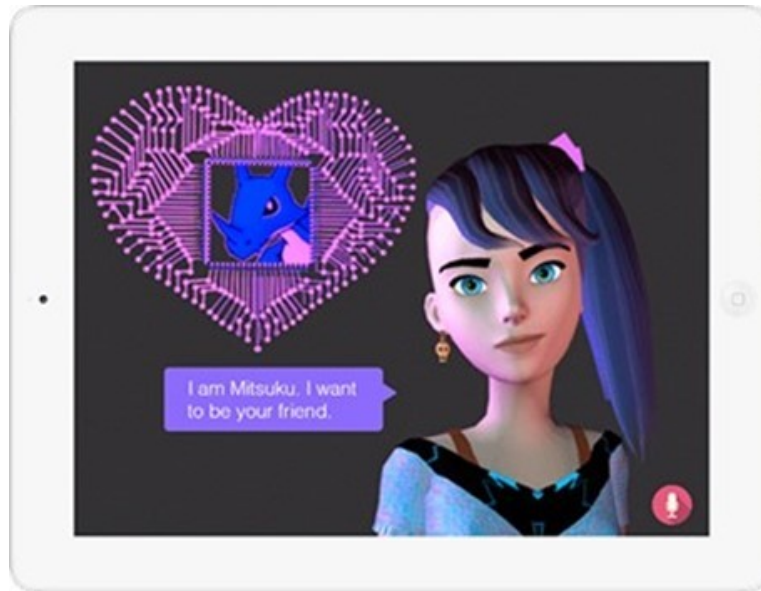


Figura 1. Avatar del chatbot Mitsuku. Fuente: Pandorabots, 2017.

Mitsuku está considerado como el mejor *chatbot* en la actualidad, ya que ha sido el ganador del Premio Loebner en las ediciones 2017, 2016 y 2013.

El Premio Loebner es una competición anual que premia a la aplicación de inteligencia artificial más inteligente del mundo. Para determinar el ganador de la competición lo que se hace es aplicar el test de Turing. A través de este test, un juez humano tiene que decidir si detrás de la pantalla a la que realiza preguntas está un programa de ordenador o un ser humano. Mitsuku consiguió superar el test de Turing gracias a que respondió de forma meditada y a un ritmo razonable y, además, contestó las preguntas de forma convincente.

Puedes observar el funcionamiento de Mitsuku, en el siguiente enlace:

<https://www.pandorabots.com/mitsuku/>

Aunque en la prensa y la industria se utilice la palabra *chatbot* para nombrar cualquier agente conversacional, en este curso solo se va a utilizar este término para designar al grupo de agentes conversacionales que pueden mantener una conversación prolongada y realista sobre cualquier tema, tal como se hace en la comunidad del procesamiento del lenguaje natural. Por lo tanto, Mitsuku es considerado un *chatbot*, pero no Google Assistant, aunque sea capaz de desarrollar llamadas de forma fluida según las últimas funcionalidades presentadas en el congreso de desarrolladores Google I/O que tuvo lugar en mayo de 2018.

10.5. Estructura de los agentes conversacionales

Los agentes conversacionales, aunque puedan utilizar diferentes modalidades para interactuar con el usuario e implementen **diferentes métodos para gestionar el diálogo**, siguen una estructura básica común.

La Figura 2 muestra la arquitectura de un agente conversacional basado en voz:

- ▶ Un módulo de entrada para el reconocimiento automático de la voz.
- ▶ Un módulo de comprensión del lenguaje natural.
- ▶ Un módulo de gestión del diálogo.
- ▶ Un módulo de generación del lenguaje natural.
- ▶ Un módulo de una salida de conversión de texto al habla.

La arquitectura para un agente conversacional basado en texto sería equivalente, aunque el módulo de entrada implementaría el reconocimiento automático de texto y el módulo de salida renderizaría el texto a un formato de presentación gráfica adecuado.

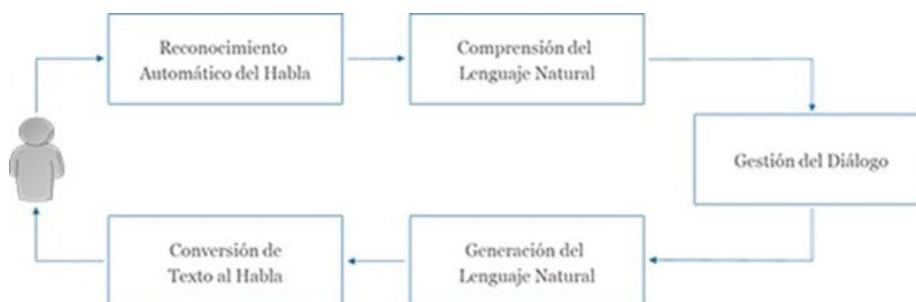


Figura 2. Arquitectura de un agente conversacional. Fuente: elaboración propia.

Reconocimiento automático del habla

El módulo de entrada tiene como función el **reconocimiento automático del habla**, es decir, **extraer las palabras** a partir de la **señal de audio** que se ha grabado con un micrófono y que recoge la voz del usuario. Aquí se utilizan tecnologías estándar para el modelado acústico-fonético, sin embargo, otros aspectos se pueden optimizar para hacer que su implementación sea más adecuada a los agentes conversacionales. Los métodos utilizados para el reconocimiento del habla no son objeto de estudio en este curso, aunque en los siguientes párrafos se describen brevemente algunas **estrategias** que permiten optimizar estos métodos.

Si el agente conversacional está dedicado a una tarea concreta y, por tanto, no se espera que vaya a responder a cualquier pregunta, solo aquellas que tienen que ver con la aplicación concreta, se puede hacer que solo sea capaz de transcribir frases en ese ámbito. Con este fin, los agentes conversacionales pueden utilizar modelos basados en gramáticas de estados finitos que especifican todas las respuestas posibles que el sistema entiende.

Otra opción para optimizar los modelos de lenguaje es que los agentes conversacionales utilicen un modelo de lenguaje N-gram en el cual las probabilidades en el diálogo de estados se adaptan para incorporar las restricciones concretas relacionadas con el contexto del diálogo.

Comprensión del lenguaje natural

El módulo de comprensión del lenguaje natural realiza las tareas básicas de procesamiento del lenguaje natural para extraer la semántica de las frases que recibe del módulo de reconocimiento automático del habla. Para representar la semántica se pueden utilizar diferentes alternativas y en el proceso de extracción se usan algunas de las técnicas de procesamiento del lenguaje natural que se han explicado en los capítulos anteriores.

En el caso de los agentes conversacionales dedicados, lo más común es utilizar **frames** y pares atributo-valor para representar la semántica de las frases. Entonces, estos agentes conversacionales basados en *frames* utilizan normalmente una ontología para la representación formal de la semántica de la conversación. Esta ontología define los diferentes *frames*, es decir, las posibles plantillas para una frase, una colección de los huecos (*slots*) en cada plantilla y los posibles valores para rellenar estos huecos.

Los pares atributo-valor restringen la tipología semántica de los valores que pueden ir asociados a cada hueco en la plantilla.

El módulo de comprensión del lenguaje natural de un agente conversacional basado en *frames* extrae la semántica de la petición lanzada por el usuario identificando diferentes elementos. De hecho, si el agente conversacional puede realizar tareas en múltiples dominios, **primero identifica el dominio al que hace referencia el usuario.**

Por ejemplo, detecta si el usuario está hablando de reservar vuelos, programando la alarma de un reloj o gestionando las citas de su calendario. Evidentemente, en los agentes dedicados que trabajan en un único dominio no sería necesario realizar este paso. Sin embargo, la mayoría de los agentes conversacionales modernos tipo Siri de Apple, Alexa de Amazon o Google Assistant realizan varias tareas en algunos dominios predefinidos y necesitan implementar esta funcionalidad para clasificar la petición del usuario de acuerdo con el dominio al que hace referencia.

Una vez se ha identificado el dominio en el que se clasifica la petición lanzada por el usuario, lo siguiente es determinar la **intención del usuario**. Es decir, se tiene que identificar cuál es el objetivo del usuario y qué tarea pretende que realice el agente conversacional. Por ejemplo, en el dominio de la búsqueda de vuelos, la intención del usuario podría ser mostrar todos los vuelos entre dos ciudades para una fecha en

concreto. En el otro ejemplo, el agente conversacional que gestiona la alarma de un reloj, las tareas que podría realizar serían activar la alarma a una hora concreta o desactivarla dependiendo de la intención del usuario.

Finalmente, cuando el agente conversacional ya ha detectado el dominio y la intención del usuario, a partir de la petición lanzada por el usuario también debe identificar los huecos (slots) en la plantilla que define la tarea que tiene que realizar el agente conversacional y obtener los valores necesarios para rellenar esos diferentes huecos de la plantilla.

El análisis semántico realizado por el módulo de comprensión del lenguaje natural requiere no solo extraer y representar el significado de la oración, incluyendo el dominio, intención y valores de los huecos del *frame*, sino también identificar otros aspectos propios de un diálogo como son **los puntos de coincidencia**. Entonces, en el proceso de extracción de la semántica de la petición lanzada por el usuario se deben identificar también los **turnos de confirmación, negación o cortesía**. Por ejemplo, si el módulo de comprensión del lenguaje natural recibe la frase:

- Sí, me gustaría saber el precio del vuelo de Madrid a Logroño.

Este debe identificar que la palabra «Sí» al inicio de la frase se refiere a una afirmación y es independiente de la tarea que debe realizar el agente conversacional, que es la consulta del precio de un vuelo concreto.

Ejemplo ilustrativo 3

Representación semántica utilizando *frames* de una interacción con un agente conversacional dedicado a la reserva de vuelos.

El agente conversacional basado en *frames* define con una ontología los pares atributo-valor, es decir los huecos (*slots*) en la plantilla y los posibles valores para rellenar estos huecos. La Tabla 1 muestra algunos ejemplos de pares atributo-valor definidos para el agente conversacional que reserva vuelos:

Hueco (slot)	Valor (type)
ORIGIN_CITY	city
DESTINATION_CITY	city
DEPARTURE_DATE	date
ARRIVAL_DATE	date
DEPARTURE_TIME	time
ARRIVAL_TIME	time

Tabla 1. Pares atributo-valor definidos para el agente conversacional que reserva vuelos.

Fuente: elaboración propia.

Se observa en la Tabla 1 que el hueco llamado ORIGIN_CITY y el llamado DESTINATION_CITY deben tener un valor que sea una ciudad (*city*). Por tanto, si aparecen esos huecos en la plantilla, solo se pueden rellenar con nombres de ciudades.

Además, la ontología que define los pares atributo-valor para los huecos y sus valores puede proporcionar una estructura jerárquica. Así, el valor *date* para los huecos DEPARTURE_DATE y ARRIVAL_DATE de la Tabla 1 puede ser a su vez el siguiente *frame* con valores del tipo entero (*INTEGER*), nominal (*NAME*) o miembros de un conjunto de nombres (*MEMBER*):

DATE

MONTH NAME

DAY (BOUNDED-INTEGER 1 31)

YEAR INTEGER

WEEKDAY (MEMBER (SUNDAY MONDAY TUESDAY WEDNESDAY THURSDAY FRIDAY SATURDAY))

Cuando en un turno de palabra el módulo de comprensión del lenguaje natural recibe la petición en inglés «*Show me flights from Boston to San Francisco on Tuesday morning*», lo primero que hace es identificar el dominio al que hace referencia el usuario: el de los viajes en avión (*AIR-TRAVEL*). Además, este módulo identifica que la intención del usuario es ver los vuelos disponibles y que, por lo tanto, la tarea que debe realizar es mostrar los posibles vuelos (*SHOW-FLIGHTS*).

El módulo de comprensión del lenguaje natural es también capaz de identificar varios huecos y obtener sus valores a partir de la petición lanzada por el usuario. Este módulo extrae la información que el origen del vuelo es Boston y el destino es San Francisco, y determina que la fecha del vuelo es el martes (*tuesday*) por la mañana (*morning*).

Entonces, el módulo de comprensión del lenguaje natural basado en *frames* genera la siguiente representación semántica de la petición en inglés «*Show me flights from Boston to San Francisco on Tuesday morning*»:

DOMAIN: AIR-TRAVEL

INTENT: SHOW-FLIGHTS

ORIGIN_CITY: Boston

DEPARTURE_DATE: Tuesday

DEPARTURE_TIME: morning

DESTINATION_CITY: San Francisco

El módulo de comprensión del lenguaje natural de un agente conversacional basado en *frames* extrae la semántica de la petición lanzada por el usuario aplicando algunas de las técnicas de procesamiento del lenguaje natural que se han explicado en los capítulos anteriores. Por ejemplo, algunos agentes conversacionales como el Core Language Engine utilizan gramáticas de unificación con la semántica adjunta para identificar el significado de la frase y, a partir de esta, se extraen las posibles palabras que pueden completar los huecos en el *frame*.

Muchos de los agentes conversacionales comerciales en la actualidad y también el clásico agente conversacional para la reserva de viajes GUS (Bobrow et al., 1977) utilizan analizadores semánticos basados en **gramáticas semánticas**. Estas pertenecen al grupo de las gramáticas libres de contexto y tienen como característica que en la parte izquierda de las reglas que componen la gramática aparecen los nombres de los posibles huecos en las plantillas (*frames*).

Ventajas

La ventaja de las soluciones basadas en gramáticas semánticas es que el análisis semántico se puede implementar utilizando cualquier algoritmo estándar que permita realizar el análisis basado en una gramática libre de contexto o basado en programación dinámica como, por ejemplo, CKY.

El resultado del analizador es la cadena de caracteres de entrada etiquetada jerárquicamente con nodos semánticos. Muchas soluciones para la comprensión del lenguaje natural utilizan gramáticas semánticas porque algunos nodos obtenidos del análisis se corresponden directamente con los huecos en un *frame*.

Ejemplo ilustrativo 4

Gramática semántica utilizada por el módulo de comprensión del lenguaje natural para la identificación de la intención y la obtención de los huecos y sus valores a partir de una petición lanzada por el usuario.

Para obtener la representación semántica de la petición en inglés «*Show me flights from Boston to San Francisco on Tuesday morning*» presentada en el ejemplo ilustrativo 3, es necesario que el módulo de comprensión del lenguaje natural basado en *frames* trabaje con una gramática semántica.

A continuación, se presenta un fragmento de una gramática semántica que permite realizar el análisis semántico de la petición del usuario que quiere ver distintos vuelos:

SHOW → show me | i want | can i see | ...

DEPART_TIME_RANGE → (after | around | before) HOUR |

morning | afternoon | evening

HOUR → one | two | three | four ... | twelve (AMPM)

FLIGHTS → (a) flight | flights

AMPM → am | pm

ORIGIN → from CITY

DESTINATION → to CITY

CITY → Boston | San Francisco | Denver | Washington

Entonces, utilizando cualquier analizador para gramáticas libres de contexto y que permita generar un árbol a partir de la gramática anterior, se obtendría la estructura jerárquica de las etiquetas semánticas resultantes de analizar la petición del usuario. Considerando la gramática presentada justo arriba y la petición «*Show me flights from Boston to San Francisco on Tuesday morning*», la figura 3 presenta el resultado del análisis semántico utilizando la gramática semántica.

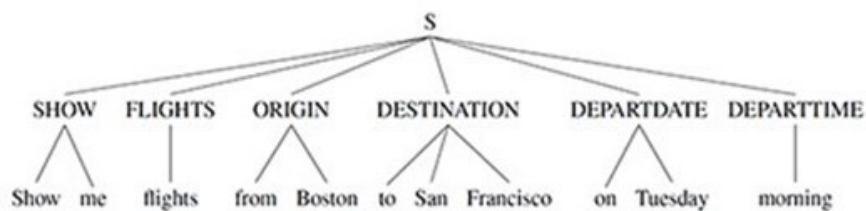


Figura 3. Resultado del análisis semántico utilizando la gramática semántica que permite crear un árbol donde los nodos son los posibles huecos de los *frames*. Fuente: Jurafsky y Martin, 2009.

Por ejemplo, en la Figura 3 se observa que del nodo ORIGIN en el árbol se puede obtener el valor a asignar al hueco ORIGIN_CITY en el *frame* que representa la petición del Ejemplo ilustrativo 3. Además, la intención de la petición se puede obtener de la combinación de los nodos *SHOW* y *FLIGHTS*.

Problemas

Aunque muchas soluciones para la comprensión del lenguaje natural utilizan gramáticas semánticas debido a sus ventajas y su alta precisión para un dominio restringido, las gramáticas **semánticas tienen varios problemas**.

- ▶ El primero es que no funcionan cuando se dan ambigüedades. Para solucionarlo, se pueden añadir probabilidades en la gramática semántica y utilizar un modelo basado en las gramáticas libres de contexto probabilísticos.
- ▶ Otro problema de las gramáticas semánticas es que estas se tienen que generar de forma manual, lo que es lento y costoso. Como alternativa existe la probabilidad de utilizar aprendizaje automático supervisado para aprender las reglas, suponiendo que se disponga de un conjunto de entrenamiento donde cada oración está etiquetada con la semántica correcta, se puede entrenar un clasificador para que asigne a cada oración su dominio y sus intenciones, y generar un modelo que mapee de la frase a los valores de los diferentes huecos en el *frame*.

Una alternativa a las gramáticas semánticas que se podría implementar en el módulo de comprensión del lenguaje natural sería un modelo **HMM semántico** (Pieraccini y Levin, 1991); este método probabilístico **evita** que se tengan que **codificar las gramáticas de forma manual**.

En este modelo HMM semántico los estados ocultos son las etiquetas de los huecos, mientras que las palabras observadas serían los valores con las que se podrían rellenar los huecos.

Se acaba de describir en detalle el funcionamiento del módulo de comprensión del lenguaje natural para agentes conversacionales basados en *frames*. Los agentes conversacionales comerciales actuales utilizan esa arquitectura, sin embargo, en el ámbito de la investigación se están desarrollando agentes más modernos que implementan una arquitectura basada en el diálogo o, mejor dicho, en las características del diálogo.

Este nuevo tipo de agentes conversacionales **basados en el diálogo** permiten, además de hacer preguntas, realizar otro tipo de acciones como **informar de un hecho, sugerir algo o dar órdenes**. Por lo tanto, estos agentes necesitan identificar los diferentes actos de habla que se dan en la conversación con el usuario. De hecho, la identificación de los actos de habla permite saber si el usuario simplemente está proporcionando una información al agente conversacional, si está afirmando o confirmando algo, haciendo una pregunta, etc.

El módulo de comprensión del lenguaje natural en un agente conversacional basado en el diálogo debe identificar los actos de habla, además de obtener los valores de los huecos del *frame* para representar el significado de la interacción.

Primero, se identifican los actos de habla aplicando técnicas de aprendizaje supervisado. Se entrena un clasificador a partir de un corpus etiquetado donde cada expresión se ha anotado a mano con su acto de habla. Para entrenar, se utiliza una gran variedad de características: características léxicas (bigramas; palabras como «no», que indican negación; o «muéstreme», que indica una petición; la longitud de la expresión o la puntuación) o características semánticas (contexto del diálogo o el acto de habla de la interacción previa).

Luego se obtienen los valores de los huecos, tal como se hacía en el caso de los agentes conversacionales basados en *frames*. En estos módulos se tienen en cuenta otros aspectos propios del diálogo como son los puntos de coincidencia.

Entonces, en el proceso de extracción de la semántica de una interacción con el agente conversacional se identifican también los turnos de confirmación, negación o cortesía.

Como hemos ejemplificado antes, si el módulo de comprensión del lenguaje natural recibe la frase «Sí, me gustaría saber el precio del vuelo de Madrid a Logroño», este debe identificar que la palabra «Sí» al inicio de la frase se refiere a una afirmación y que es independiente de la petición de la consulta del precio del vuelo.

Gestión del diálogo

El módulo de gestión del diálogo, también llamado **controlador del dialogo**, es el elemento principal de los agentes conversacionales y se encarga de **identificar la acción** que debe realizar el agente conversacional en el **siguiente turno de palabra** y **cómo se debe continuar la conversación**. Para ello, este módulo analiza la representación semántica de la petición que ha efectuado el usuario del agente conversacional o de la frase que ha pronunciado la persona.

Así, analiza la representación obtenida por el módulo de comprensión del lenguaje natural y proporciona a su salida la acción a realizar o el concepto que se quiere transmitir al usuario del sistema conversacional en el siguiente turno de palabra y como respuesta a la petición realizada por el usuario.

La gestión del diálogo requiere **mantener el estado y el flujo de la conversación** y en algunos casos pedir más información al usuario para poder completar la información necesaria y dar una respuesta. Por ejemplo, si el usuario ha realizado la petición

- «Me gustaría saber el precio del vuelo de Madrid a Logroño.

El agente conversacional solo sabrá a qué vuelo se refiere el usuario si mantiene el estado de la conversación y en alguna interacción anterior este le ha dado más detalles sobre el vuelo en cuestión, por ejemplo, el día en el que quiere viajar. Si ese no fuera el caso, el agente conversacional no puede responder directamente al usuario con el precio del vuelo y necesitaría realizarle una serie de preguntas para obtener la información que le falta sobre la fecha concreta del vuelo antes de poder ejecutar la búsqueda de los precios y contestar a la petición del usuario mostrándole los precios de los vuelos.

De esta forma, cualquier agente conversacional que necesite asegurar que puede responder a las peticiones del usuario, aunque estas sean incompletas, y que no solamente conteste a las peticiones para las cuales el usuario proporcione toda la información en una única petición, debe implementar un módulo de gestión del diálogo que mantenga el estado de la conversación y que permita múltiples interacciones con el usuario para obtener la información no disponible. De hecho, si el agente conversacional solo respondiera a las peticiones completas y no mantuviera ninguna interacción con el usuario, se hablaría de un simple sistema de preguntas y respuestas y no de un agente conversacional.

Existen **diferentes estrategias** para la **gestión del diálogo**. Por ejemplo, la mayoría de los agentes conversacionales dedicados que utilizan *frames* tienen un módulo de diálogo basado en una máquina de estados finitos que define las acciones necesarias para llevar a cabo una tarea concreta:

- ▶ La máquina de estados finitos que rige un agente conversacional dedicado basado en *frames* gestiona el diálogo y modela cómo este debe desarrollarse.
- ▶ Los estados de la máquina de estados finitos se corresponden con las preguntas a realizar al usuario y que son necesarias para obtener los valores de los diferentes huecos en la plantilla.
- ▶ Los arcos de la máquina de estados finitos se corresponden a las acciones que deben realizarse en función de lo que el usuario responda.

Entonces, la máquina de estados finitos controla completamente la conversación con el usuario. El agente conversacional le hace una serie de preguntas al usuario, ignorando (o malinterpretando) cualquier contestación que no sea una respuesta directa a la pregunta y luego pasa a la siguiente pregunta. De hecho, la arquitectura de control para un agente conversacional basado en *frames* también utiliza plantillas para generar el contenido de las nuevas oraciones con las que se debe continuar la conversación. Por lo que va a ser necesario rellenar los huecos en las plantillas con los valores relevantes que se han obtenido de las interacciones previas con el usuario.

Ejemplo ilustrativo 5

Gestión del diálogo en un agente conversacional dedicado basado en *frames* utilizando una máquina de estados finitos.

La máquina de estados finitos que rige un agente conversacional dedicado a la reserva de vuelos (similar al que se ha presentado en el ejemplo ilustrativo 3 y 4) se muestra en la Figura 4. Este agente conversacional

basado en *frames* gestiona el diálogo y modela cómo se debe desarrollar con base en esta máquina de estados finitos. De hecho, se observan las diferentes preguntas que se van a realizar para obtener la información necesaria para rellenar los diferentes huecos del *frame* antes de ejecutar la tarea de reservar el vuelo.

Por ejemplo, en uno de los turnos de palabra el agente conversacional pregunta:

«Do you want to go from <FROM> to <TO> on <DATE>? ».

Se observa que esta pregunta es una plantilla y que tiene tres huecos *FROM*, *TO* y *DATE* que deben ser rellenados a partir de las respuestas obtenidas de las preguntas realizadas en turnos anteriores.

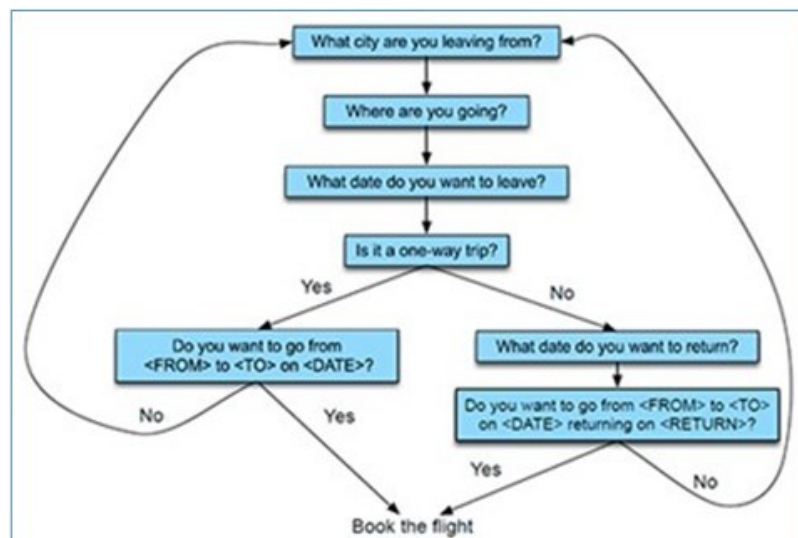


Figura 4. Máquina de estados finitos para modelar la gestión del diálogo de un agente conversacional basado en *frames* para reservar vuelos. Fuente: Jurafsky y Martin, 2009.

La Tabla 2 muestra la asociación entre los huecos y las preguntas a partir de las cuales se va a obtener el valor para rellenar esos huecos.

Hueco (slot)	Pregunta
FROM	What city are you leaving from?
TO	Where are you going?
DATE	What date do you want to leave?
RETURN	What date do you want to return?

Tabla 2. Asociación entre los huecos del *frame* y las preguntas que necesita realizar el agente conversacional para determinar el valor de esos huecos. Fuente: elaboración propia.

Una estrategia de gestión del diálogo basada en una máquina de estados finitos en la cual el agente conversacional toma totalmente la iniciativa tiene la ventaja de que el sistema siempre sabe a qué pregunta responde el usuario. Esto significa que el sistema puede preparar el módulo de reconocimiento automático del habla con un modelo de lenguaje adaptado a las respuestas para esta pregunta. Esto también facilita el funcionamiento del módulo de comprensión del lenguaje natural que puede obtener la semántica de una forma más fácil. No obstante, una arquitectura basada en una máquina de estados finitos tan simplista solo se puede aplicar a tareas muy sencillas y no es suficiente para los agentes conversacionales que necesitan algo más de flexibilidad en el diálogo.

Google Assistant, Siri de Apple o Alexa de Amazon utilizan una estrategia de gestión del diálogo más compleja y que está basada en la arquitectura del agente conversacional GUS (Bobrow et al., 1977). Esta estrategia permite al usuario tomar cierta iniciativa y ayuda al agente conversacional a tratar los *frames* de una forma mucho más flexible. El agente conversacional hace preguntas al usuario, rellenando cualquier hueco en la plantilla al que haga referencia la información que el usuario le transmite. Entonces, en cada interacción, el agente conversacional puede rellenar varios huecos a la vez, incluso si la respuesta del usuario no se corresponde con la

pregunta realizada. El sistema simplemente salta las preguntas asociadas con los huecos que ya han sido rellenados previamente.

Los diferentes huecos del *frame* se rellenan sin seguir una secuencia concreta y en función de la información proporcionada por el usuario.

Una vez que el agente conversacional tiene la suficiente información, ya puede realizar la acción, por ejemplo, **consultar la base de datos de vuelos**, y finalmente devolver el resultado al usuario.

Los agentes conversacionales basados en *frames* que siguen una estrategia de control flexible como, por ejemplo, muchos de los agentes conversacionales actuales, necesitan cambiar el control del diálogo y saltar entre *frames* dependiendo de la información proporcionada por el usuario. Debido a esta necesidad de cambiar dinámicamente el control, la **arquitectura GUS** para la gestión del diálogo se implementa como un **sistema de reglas de producción y la información** proporcionada por el usuario **puede activar varias reglas de producción**, el resultado de lanzarlas permite rellenar múltiples huecos en diferentes *frames*.

A diferencia de los agentes conversacionales dedicados basados en *frames*, los agentes conversacionales basados en el diálogo **aplican estrategias** mucho más **avanzadas para la gestión del diálogo**. Por ejemplo, para la gestión del diálogo, algunos de estos agentes conversacionales modernos, que todavía están en el ámbito de la investigación, aplican una estrategia basada en el contexto local y otros una estrategia basada en procesos de decisión de Markov y aprendizaje por refuerzo.

Los agentes conversacionales que se basan en el diálogo e implementan en el módulo de gestión del diálogo una estrategia basada en el contexto local lo que hacen es predecir, para un turno de palabra, cuál es la acción para realizar según el estado del diálogo, es decir, con base en las acciones que han tomado el agente y el

usuario en los turnos de palabra anteriores.

Para simplificar, se puede considerar que el estado del diálogo principalmente solo va a depender de la última acción que ha realizado el agente conversacional, la última acción que ha realizado el usuario y del estado actual del *frame*, es decir, del conjunto de los huecos rellenos y los valores que estos toman.

A partir de un corpus de conversaciones lo suficientemente grande se puede estimar la probabilidad de que el agente conversacional tome una acción dadas las acciones que han tomado el agente conversacional y el usuario en el turno de palabra anterior y el estado actual del *frame*.

Entonces, el módulo de gestión del diálogo del agente conversacional va a seleccionar la acción que maximice la probabilidad para ese turno de palabra.

Esta estrategia de gestión del diálogo que decide qué acción debe realizar el agente conversacional tiene un problema: solo se fundamenta en el pasado del diálogo, ignorando completamente si la acción que toma el agente conversacional es probable que conduzca al resultado esperado o no (por ejemplo, a completar correctamente la reserva de un vuelo). De hecho, en el momento de planificar la acción aún falta mucho para poder saber si el resultado será exitoso. Es por eso por lo que hoy en día se están estudiando **nuevas estrategias para la gestión del diálogo** asentadas en procesos de decisión de Markov y aprendizaje por refuerzo.

Generación del lenguaje natural

El módulo de generación del lenguaje natural tiene como **objetivo elegir los conceptos que se quieren expresar al usuario y, además, planificar cómo**

expresarlos en palabras. Entonces, la función de este módulo se puede separar en dos etapas: **qué decir y cómo decirlo.**

Primera etapa

En la primera etapa del módulo de generación del lenguaje se realiza una tarea de **planificación del contenido**, es decir, se decide **qué contenido** se debe **expresar al usuario en cada turno de palabra**. Por ejemplo, si el usuario ha hecho una pregunta, se presenta la respuesta a esa pregunta o, si ha hecho una propuesta, el agente conversacional puede aceptarla o rechazarla.

Así, la tarea de planificación del contenido está basada en el patrón de los pares adyacentes que define la estructura de las conversaciones y tiene en consideración los actos de habla asociados a cada una de las expresiones en la conversación.

Muchas veces la funcionalidad de planificación del contenido se integra con el módulo de gestión de diálogo y no aparece como un componente separado en el módulo de generación del lenguaje. Este sería el caso de los agentes conversacionales basados en *frames* que se han presentado en los ejemplos ilustrativos 3, 4 y 5. Estos agentes gestionan el diálogo utilizando una máquina de estados finitos de forma que, para cada estado, se obtiene directamente la expresión que el sistema le transmite al usuario en ese turno de palabra. De forma similar, los agentes conversacionales basados en el diálogo que aplican una estrategia según el contexto local también planifican la acción a realizar para un turno de palabra en el módulo de gestión de diálogo.

Segunda etapa

En la segunda etapa del módulo de generación del lenguaje se escogen **las estructuras sintácticas y las palabras** que se necesitan para expresar el concepto

que se quiere transmitir al usuario. Existen diferentes formas para realizar esta tarea dependiendo de la estrategia utilizada para la gestión del diálogo en el módulo de gestión de diálogo o la implementación concreta de primera etapa del módulo de generación del lenguaje.

En el caso de un agente conversacional dedicado basado en *frames*, el módulo de gestión de diálogo ya ha determinado la pregunta que el agente conversacional debe efectuar al usuario en el siguiente turno de palabra. De hecho, esa pregunta que se transmitirá al usuario puede estar completa o tener algunos huecos que se deben rellenar con los valores obtenidos por el agente conversacional en las interacciones anteriores. Por lo tanto, el módulo de generación del lenguaje natural no tiene que realizar ninguna acción complicada, solo identificar los huecos en la frase y rellenarlos con los valores correctos para ese turno de palabra.

En el caso de un agente conversacional que se basa en el diálogo e implementa una **estrategia de gestión basada en el contexto local**, el módulo de gestión de diálogo ha determinado el acto de habla del siguiente turno de palabra, es decir, la acción que el agente conversacional debe realizar en el siguiente turno de palabra. Por lo tanto, el módulo de generación del lenguaje natural conoce el acto de habla del siguiente turno de palabra y el estado actual del *frame*, es decir, el conjunto de los huecos y los valores que estos toman. Entonces, el módulo de generación del lenguaje natural debe generar el mensaje que el agente conversacional va a transmitir al usuario en el siguiente turno de palabra.

Para generar el mensaje asociado a un acto de habla se utiliza un proceso con dos pasos:

- ▶ Primero, se genera una frase deslexicalizada (una frase con huecos).
- ▶ Luego, se rellenan los huecos con los valores adecuados que se obtienen del estado actual del *frame*.

Se sigue esta estrategia de dos pasos porque **generar una frase deslexicalizada es más fácil** que generar la frase final. Para obtener la frase deslexicalizada lo más normal es utilizar un N-grama entrenado a partir de las frases de un corpus que están etiquetadas con el acto de habla que se quiere expresar en el siguiente turno de palabra. Sin embargo, algún trabajo reciente, en lugar de utilizar N-gramas, utiliza modelos neuronales que aprenden cómo obtener la frase resultante a partir del siguiente turno de palabra y del estado actual del *frame* (Wen et al., 2015).

Conversión del texto al habla

El módulo de salida tiene como función **la conversión del texto** al habla sintetizando en formato de voz el mensaje a transmitir al usuario en el siguiente turno de palabra: Este módulo de conversión del texto al habla obtiene del módulo de generación del lenguaje natural el mensaje en formato de texto y sintetiza la onda de sonido para este mensaje. Los métodos utilizados para la conversión del texto al habla no son objeto de estudio en este curso, aunque describiremos brevemente algunas ideas de cómo se realiza la síntesis de la voz.

El módulo de conversión del texto al habla asigna anotaciones sobre el acento y la entonación de cada una de las palabras que componen el mensaje (Reiter y Dale, 2000) y luego las utiliza para sintetizar una onda de sonido. Los métodos de síntesis que más se han utilizado hasta la fecha son los basados en concatenación. En estos métodos se graban unas muestras del habla y se guardan en una base de datos, después se combinan para crear nuevas palabras y frases. Sin embargo, hay otros métodos más modernos fundamentados en hacer las transiciones más suaves o en la articulación.

10.6. Diseño de chatbots

Los *chatbots* son aquellos agentes conversacionales que **permiten reproducir conversaciones extensas y no estructuradas** que se han utilizado en el ámbito del entretenimiento o para propósitos prácticos como probar teorías de la práctica psicológica, se dividen en **dos clases** en función de su implementación: **sistemas basados en reglas y sistemas basados en corpus**.

Chatbots basados en reglas

Entre estos se incluye uno de los primeros agentes conversacionales, **ELIZA** (Weizenbaum, 1966). ELIZA fue diseñado para simular a un psiquiatra cuyo método involucra hacer reflexionar al paciente devolviéndole sus propias declaraciones. Por ejemplo, si un paciente le dice al psiquiatra: «Fui a dar un paseo en bote», el psiquiatra le va a responder: «Hábleme de los botes». En este caso no se asumirá que el psiquiatra no sabe qué es un bote, se asume que la respuesta tiene un objetivo puramente conversacional.

Entonces, un *chatbot* basado en reglas como ELIZA implementa este tipo de conversación poco común en la cual el *chatbot* no conoce casi nada del mundo real y solo se dedica a reformular las palabras del usuario.

Un *chatbot* basado en reglas como ELIZA **utiliza reglas patrón→transformación**. Cada regla describe la transformación que se debe aplicar a la declaración hecha por el usuario para obtener la respuesta que le va a proporcionar el *chatbot*, siempre y cuando la declaración del usuario cumpla un cierto patrón. Cada regla está asociada a una **palabra clave** que puede ocurrir en la frase, además, a las palabras clave se les asigna un rango, las palabras clave más específicas tienen un rango más alto y las más generales tienen un rango más bajo.

En cada turno de palabra, el *chatbot* responde con la transformación del patrón asociado a la palabra clave que tiene un rango más alto de todas las palabras contenidas en la declaración hecha por el usuario. En el caso de que no hubiera ninguna palabra en la declaración del usuario que coincidiera con las palabras claves asociadas a las reglas, el *chatbot* daría una respuesta evasiva como «esto es muy interesante» o «por favor, sigue». También puede utilizar algún truco para intentar evitar estas respuestas evasivas, por ejemplo, cuando le llega una frase en la que la palabra clave más probable es el determinante «mi», se aplica una transformación con un patrón similar a «Antes has dicho que tu ...» y el resultado se guarda en memoria para utilizarlo en el próximo turno de palabra en el que no coincida ninguna palabra clave. La Figura 5 muestra la lógica del algoritmo implementado por ELIZA.

```

function ELIZA GENERATOR(user sentence) returns response

Find the word w in sentence that has the highest keyword rank
if w exists
    Choose the highest ranked rule r for w that matches sentence
    response ← Apply the transform in r to sentence
    if w = 'my'
        future ← Apply a transformation from the 'memory' rule list to sentence
        Push future onto memory stack
    else (no keyword applies)
        either
            response ← Apply the transform for the NONE keyword to sentence
        or
            response ← Pop the top response from the memory stack
return(response)
  
```

Figura 5. Lógica del algoritmo implementado por ELIZA, *chatbot* basado en reglas. Fuente: Jurafsky y Martin, 2009.

El modelo de reglas patrón→transformación de ELIZA aún se utiliza hoy en día. Por ejemplo, un *chatbot* moderno como ALICE se basa en una versión actualizada del *chatbot* basado en reglas ELIZA.

Ejemplo ilustrativo 5

ELIZA, uno de los primeros agentes conversacionales, un *chatbot* basado en reglas que simula a un psiquiatra cuyo método involucra hacer reflexionar al paciente contestándole con sus propias declaraciones.

A continuación, se presenta un fragmento de la conversación más famosa de ELIZA recogida en el artículo de Weizenbaum (1966). Las frases en minúscula son las que pronunció el paciente y las frases en mayúscula las contestaciones generadas por parte del *chatbot* que simulaba el psiquiatra:

Men are all alike.

IN WHAT WAY?

They're always bugging us about something or other.

CAN YOU THINK OF A SPECIFIC EXAMPLE?

Well my boyfriend made me come here.

YOUR BOYFRIEND MADE YOU COME HERE

He says I'm depressed much of the time.

I AM SORRY TO HEAR YOU ARE DEPRESSED

...

WHO ELSE IN YOUR FAMILY TAKES CARE OF YOU?

My father

YOUR FATHER

You are like my father in some ways.

WHAT RESEMBLANCE DO YOU SEE?

You are not very aggressive but I think you don't want me to notice that.

WHAT MAKES YOU THINK I AM NOT AGGRESSIVE?

You don't argue with me.

WHY DO YOU THINK I DON'T ARGUE WITH YOU?

You are afraid of me.

DOES IT PLEASE YOU TO BELIEVE I'M AFRAID OF YOU?

My father is afraid of everybody.

WHAT ELSE COMES TO MIND WHEN YOU THINK OF YOUR FATHER?

Bullies.

DOES THAT HAVE ANYTHING TO DO WITH THE FACT THAT YOUR BOYFRIEND MADE YOU COME?

Algunas de las reglas patrón→transformación que implementa ELIZA son las siguientes:

(0 YOU 0 ME) -> (WHAT MAKES YOU THINK I 3 YOU)

(I *) -> (You say you 2)

En la primera regla, cuando se da el patrón 0 YOU 0 ME, es decir, algo seguido de la palabra YOU, otro algo y luego la palabra ME, entonces se transforma en la frase en WHAT MAKES YOU THINK I 3 YOU, donde el número tres debe substituirse por el valor del segundo cero en el patrón. Por lo tanto, a partir de la primera regla, la oración pronunciada por el usuario «You hate me» iría seguida de la oración obtenida en la

transformación «WHAT MAKES YOU THINK I HATE YOU».

Si el usuario pronuncia la siguiente frase: «I know everybody laughed at me», ELIZA le contestaría «YOU SAY YOU KNOW EVERYBODY LAUGHED AT YOU» después de aplicar la transformación recogida en la segunda regla.

Chatbots basados en corpus

Los *chatbots* basados en corpus **aplican técnicas de minería de datos** a grandes conjuntos de conversaciones entre humanos **con el fin de extraer las posibles respuestas del *chatbot* al usuario** y para no tener que generar reglas a mano.

Los principales corpus que permiten aprender de las conversaciones están basados en conversaciones en plataformas de chat, en Twitter o en diálogos de películas (Serban et al., 2017). De hecho, a veces estos *chatbots* también pueden minar conversaciones entre un humano y una máquina e incluso extraer oraciones de un texto que no sea un diálogo.

Al igual que los basados en reglas, la mayoría de *chatbots* basados en corpus **tienden a utilizar poca información del contexto de la conversación** para generar una respuesta y suelen generar la respuesta basándose solamente en el turno de palabra inmediatamente anterior.

Entonces, los *chatbots* basados en corpus tienen cierta similitud con los sistemas de búsqueda de respuestas (*Question Answering*) que permiten la recuperación de información basada en el lenguaje natural.

Hay dos tipos de *chatbots* basados en corpus:

- ▶ Basados en recuperación de información.
- ▶ Los *chatbots* secuencia a secuencia.

Por ejemplo, Cleverbot, que permite que un usuario pueda hablar con él por puro divertimento, pertenece al grupo de los *chatbots* basados en recuperación de información.

Chatbots basados en recuperación de información

Este tipo de *chatbots* funcionan utilizando un método que les permite obtener una respuesta proporcionada por un humano en una conversación previa y utilizarla para responder al turno de palabra actual.

Así, usan una respuesta que aparece en el corpus y cuyo turno anterior es lo más parecido posible a la interacción que justo acaba de efectuar el usuario con el *chatbot* para el turno de respuesta de esta interacción. La idea es que se debe buscar un turno que se parezca lo más posible al turno del usuario y devolver la respuesta humana a ese turno (Jafarpour, Burges y Ritter, 2009; Leuski y Traum, 2011).

Otra opción sería que el *chatbot* basado en recuperación de información, en lugar de devolver la respuesta al turno más similar, devolviera directamente el turno más similar. La idea en este caso es **buscar las coincidencias entre la interacción del usuario** y los turnos en el corpus, ya que una buena respuesta a menudo compartirá palabras o semántica con el turno anterior.

Aunque devolver la respuesta al turno más similar parece ser un algoritmo más intuitivo, en la práctica parece que devolver el turno más similar funciona mejor (Ritter, Cherry y Dolan, 2011; Wang et al. 2013).

Para medir la similitud entre turnos se puede utilizar cualquier medida de similitud, por ejemplo, la similitud del coseno computada sobre las palabras (usando una medida de *term frequency–inverse document frequency*). De hecho, se puede extender el algoritmo para que incluya más características, además de las palabras en el turno actual, por ejemplo, las palabras en turnos anteriores o información sobre el usuario.

Debido a la importancia que tiene el corpus para el correcto funcionamiento de los *chatbots* basados en recuperación de información, las respuestas que va generando el usuario durante su interacción con el *chatbot* también se pueden usar para seguir entrenando este.

Chatbots secuencia a secuencia

Estos *chatbots* están basados en el paradigma de la traducción automática y aplican técnicas de aprendizaje automático supervisado. Una forma alternativa de utilizar un corpus para generar un diálogo es pensar en la generación de respuestas como la tarea de transducción desde el turno del usuario hasta el siguiente turno del sistema, es decir, un *chatbot* que implemente esta idea sería básicamente una versión de ELIZA, pero basado en aprendizaje automático.

En la primera implementación de esta idea se utilizaron técnicas de traducción automática basada en frases (Ritter et al., 2011) para traducir un turno de usuario en una respuesta del sistema. Sin embargo, rápidamente quedó claro que la tarea de generación de respuestas era demasiado diferente de la traducción automática. En esta, las palabras o frases en las oraciones de origen y de destino se alinean entre sí. No obstante, en una conversación, una declaración del usuario puede no compartir palabras o frases con una respuesta coherente. Entonces, aproximadamente de forma simultánea se empezaron a modelar los transductores para la generación de respuestas utilizando modelos secuencia a secuencia (seq2seq) (Shang, Lu y Li, 2015; Vinyals y Le, 2015; Sordani et al. 2015).

El **modelo seq2seq** tiene una serie de problemas, por ejemplo, es incapaz de modelar el contexto anterior a la conversación y genera respuestas que no tienen por qué ser coherentes entre los diferentes turnos de palabra. Entonces, se han realizado algunas modificaciones al modelo seq2seq básico para que pueda funcionar en *chatbots* secuencia a secuencia. Por ejemplo, se ha introducido información sobre los turnos de palabra previos utilizando un modelo jerárquico (Lowe et al., 2017) y se han implementado métodos de aprendizaje por refuerzo para aprender a escoger las respuestas que hacen que la conversación sea más natural (Li et al., 2017).

10.7. Referencias bibliográficas

- Austin, J. L. (1962). *How to Do Things with Words*. Harvard University Press.
- Bobrow, D. G., Kaplan, R. M., Kay, M., Norman, D. A., Thompson, H. y Winograd, T. (1977). GUS, A frame driven dialog system. *Artificial Intelligence*, 8, 155–173.
- Clark, H. H. (1996). Cambridge University Press.
- Clark, H. H. y Schaefer, E. F. (1989). Contributing to discourse. *Cognitive Science*, 13, 259-294.
- Forbes-Riley, K. y Litman, D. J. (2011). Benefits and challenges of real-time uncertainty detection and adaptation in a spoken dialogue computer tutor. *Speech Communication*, 53(9), 1115–1136.
- Jafarpour, S., Burges, C. y Ritter, A. (2009). Filter, rank, and transfer the knowledge: Learning to chat. <https://aritter.github.io/chat.pdf>
- Jurafsky, D. y Martin, J. H. (2009). *Speech and Language Processing: An Introduction to Natural Language Processing, Speech Recognition and Computational Linguistics*. Prentice-Hall.
- Leuski, A. y Traum, D. (2011). NPCEditor: Creating virtual human dialogue using information retrieval techniques. *AI Magazine*, 32(2), 42-56.
- Levinson, S. C. (1983). Conversational Analysis. En *Pragmatics*. Cambridge University Press.
- Li, J., Monroe, W., Shi, T., Ritter, A. y Jurafsky, D. (2017). Adversarial learning for neural dialogue generation. En *EMNLP* (pp. 2157-2169). Association for Computational Linguistics.

- Lowe, R. T., Pow, N., Serban, I. V., Charlin, L., Liu, C. W. y Pineau, J. (2017). Training end-to-end dialogue systems with the ubuntu dialogue corpus. *Dialogue & Discourse*, 8(1), 31–65.
- Pieraccini, R. y Levin, E. (1992). Stochastic representation of semantic structure for speech understanding. *Speech Communication*, 11(2-3), 283-288.
- RAE. (s. f.). Diálogo. En Diccionario de la lengua española (actualización de la 23ª ed.). <https://dle.rae.es/?id=DetWqMJ>
- Reiter, E. y Dale, R. (2000). *Building Natural Language Generation Systems*. Cambridge University Press.
- Ritter, A., Cherry, C. y Dolan, B. (2011). Data-driven response generation in social media. In EMNLP-11, pp. 583-593.
- Sacks, H., Schegloff, E. A. y Jefferson, G. (1974). A simplest systematics for the organization of turn-taking for conversation. *Language*, 50(4), 696–735.
- Schegloff, E. A. (1968). *Sequencing in conversational openings*. *American Anthropologist*, 70, 1075–1095.
- Serban, I. V., Lowe, R. T., Charlin, L. y Pineau, J. (2017). A survey of available corpora for building data-driven dialogue systems. arXiv:1512.05742.
- Shang, L., Lu, Z. y Li, H. (2015). Neural responding machine for short-text conversation. En *Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing* (pp. 1577-1586). Beijing: Association for Computational Linguistic.

Sordoni, A., Galley, M., Auli, M., Brockett, C., Ji, Y., Mitchell, M., Nie, J. Y., Gao, J. y Dolan, B. (2015). A neural network approach to context-sensitive generation of conversational responses.

<http://rali.iro.umontreal.ca/rali/sites/default/files/publis/1506.06714.pdf>

Stalnaker, R. C. (1978). Assertion. En P. Cole (Ed.), *Pragmatics: Syntax and Semantics* (vol. 9, pp. 315-332). Academic Press.

Vinyals, O. y Le, Q. (2015). A neural conversational model. En *Proceedings of the International Conference on Machine Learning*. <https://arxiv.org/pdf/1506.05869.pdf>

Wang, H., Lu, Z., Li, H. y Chen, E. (2013). A dataset for research on short-text conversations. In *Proceedings of the 2013 Conference on Empirical Methods in Natural Language Processing* (pp. 935-945). Association for Computational Linguistics.

Weizenbaum, J. (1966). ELIZA—A computer program for the study of natural language communication between man and machine. *Communications of the ACM*, 9(1), 36-45.

Wen, T. H., Gasic, M., Kim, D., Mrksic, N., Su, P. H., Vandyke, D. y Young, S. J. (2015). Stochastic language generation in dialogue using recurrent neural networks with convolutional sentence reranking. En *SIGDIAL 2015 Meeting on Discourse and Dialogue* (pp. 275-284). Association for Computational Linguistics.

Yankelovich, N., Levow, G. A. y Marx, M. (1995). Designing SpeechActs: issues in speech user interfaces. En *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (CHI '95) (pp. 369-376). ACM Press/Addison-Wesley Publishing Co.

Selección de las transcripciones de los finalistas del Premio Lobner 2017

Martin, A. O. (2017). AISB Loebner Prize 2017 Finalist Selection Transcripts.
http://www.aomartin.co.uk/uploads/loebner_2017_finalist_selection_transcripts.pdf

El documento presenta una selección de las transcripciones de las conversaciones llevadas a cabo por el juez y los diferentes chatbots durante los test de Turing que permitieron decidir el ganador del Premio Lobner 2017.

Como superar el test de Turing (ganar el Premio Loebner)

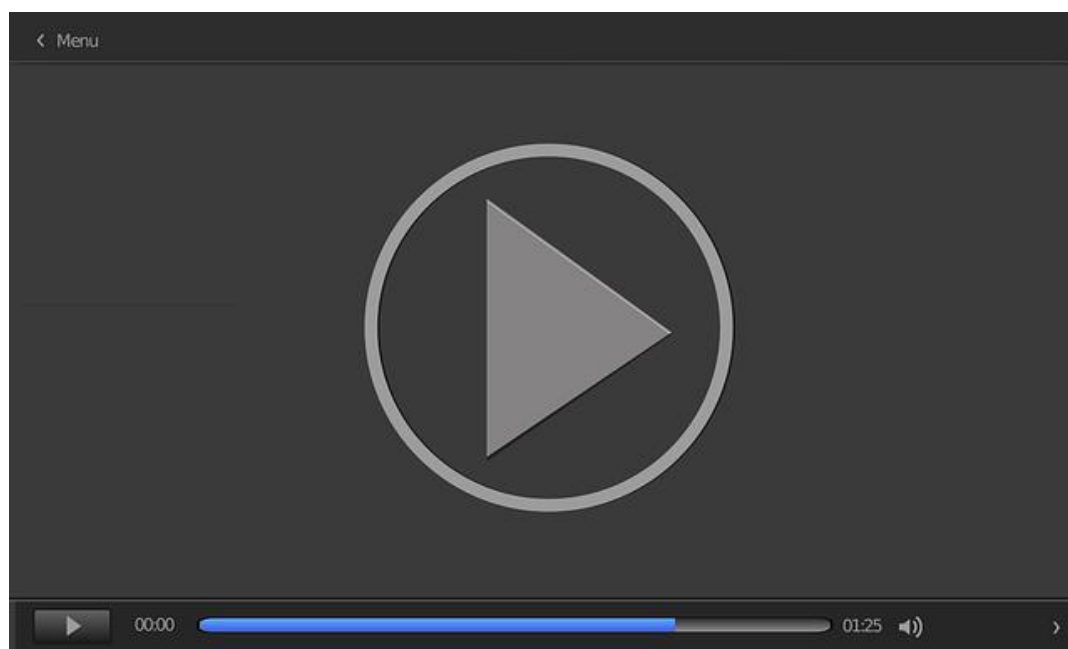
Moloney, C. (24 de septiembre de 2017). How to win a Turing Test (the Loebner prize) . <https://chatbotsmagazine.com/how-to-win-a-turing-test-the-loebner-prize-3ac2752250f1>

El artículo recoge la experiencia de uno de los jueces que formó parte del jurado que falló el Premio Loebner 2017. Además, en el artículo se expone una idea muy interesante de este juez: el hecho de que se deberían programar los chatbots para que cometieran errores como pasa en el caso de los humanos y así engañar a los jueces.

Google Assistant realizando una llamada durante el Google I/O '18

Presentación de apertura. (8 de mayo de 2018). Youtube.
<https://youtu.be/ogfYd705cRs>

El vídeo (desde el minuto 35:00 al minuto 40:24) muestra la presentación de la nueva funcionalidad del Google Assistant que permite realizar llamadas de forma automática. La demostración de esta funcionalidad la realizó el CEO de Google, Sundar Pichai, durante el congreso de desarrolladores Google I/O que tuvo lugar en mayo de 2018.



Presentación de apertura (Google I/O '18)

Accede al vídeo:

<https://www.youtube.com/embed/ogfYd705cRs>

Discurso y diálogo

Sanchís, E. (2016). Discurso y diálogo. En Á. L. Gonzalo (Coords.), *Tecnologías del lenguaje en España: comunicación inteligente entre personas y máquinas*. Barcelona: Fundación Telefónica-Ariel.
https://www.fundaciontelefonica.com/artes_cultura/publicaciones-listado/pagina-item-publicaciones/itempubli/565/

Este capítulo, que forma parte del estudio impulsado por la Fundación Telefónica y el Instituto para la Promoción de las Tecnologías de la Lengua, introduce las complejidades lingüísticas, psicológicas y cognitivas de los procesos de discurso y diálogo, y concluye que:

«Uno de los retos más importantes actualmente, además de aumentar la robustez en reconocimiento y comprensión, es la posibilidad de diseñar sistemas que aprendan de las interacciones con usuarios reales y se adapten automáticamente a nuevas situaciones. Debe tenerse en cuenta que la usabilidad de estos sistemas sólo se popularizará cuando sus prestaciones sean muy altas, ya que de otra forma el posible usuario se frustra y escoge otro tipo de interacción».

1. Indica las afirmaciones correctas sobre los agentes conversacionales:
 - A. Son programas que conversan con las personas a través del lenguaje natural.
 - B. Interactúan con el humano o a través de la voz o a través de texto, pero nunca a través de los dos.
 - C. Son capaces de continuar la interacción con el usuario contestando a preguntas dependientes de las anteriores.
 - D. También se llaman sistemas de diálogo.

2. Indica las afirmaciones correctas sobre las características de las conversaciones entre humanos que influyen en los agentes conversacionales:
 - A. Todos los agentes conversacionales son capaces de tratar con múltiples de actos de habla, es decir, que son capaces de lanzar una pregunta, dar una orden, hacer una propuesta, rechazar una sugerencia o estar de acuerdo con una expresión presentada por su interlocutor.
 - B. La gestión de los turnos de palabra en agentes conversacionales se puede modelar con una serie de reglas que se evalúan en los momentos en que la estructura del lenguaje permite intercambiar el interlocutor.
 - C. Para establecer puntos de coincidencia con el interlocutor, el agente conversacional debe dejar claro explícitamente que ha entendido el significado y la intención de la expresión lanzada por el humano.
 - D. En las conversaciones se da un patrón de pares adyacentes, por ejemplo, si en un turno de palabra se lanza una pregunta (primera parte del par), en el siguiente turno se espera que se dé una respuesta (segunda parte del par). Sin embargo, puede darse el caso que la segunda parte del par no aparezca directamente justo después de la primera.

3. Indica las afirmaciones correctas sobre los diferentes tipos agentes conversacionales:

- A. Los agentes conversacionales dedicados permiten entablar conversaciones extensas con múltiples interacciones.
- B. Los *chatbots* son aquellos agentes conversacionales que permiten entablar conversaciones sobre cualquier tema o materia.
- C. Los agentes conversacionales disponibles en las páginas web que proporcionan ayuda a los clientes para resolver dudas son agentes conversacionales dedicados.
- D. Cortana de Windows es un *chatbot*.

4. Indica las afirmaciones correctas sobre la estructura de los agentes conversacionales:

- A. Los agentes conversacionales, independientemente de las modalidades que utilicen para interactuar con el usuario, siguen una estructura básica común.
- B. La mayoría de los agentes conversacionales implementan el mismo método para gestionar el diálogo.
- C. Los agentes conversacionales más sencillos pueden integrar las funcionalidades de comprensión del lenguaje natural, de gestión del diálogo y de generación del lenguaje natural en un único módulo.
- D. Un agente conversacional basado en voz se compone de un módulo de entrada para el reconocimiento automático de la voz, un módulo de comprensión del lenguaje natural, un módulo de gestión del diálogo, un módulo de generación del lenguaje natural y un módulo de una salida de conversión de texto al habla.

5. Indica las afirmaciones correctas sobre el módulo de comprensión del lenguaje natural en un agente conversacional:

- A. Es el módulo de entrada de la arquitectura del agente conversacional.
- B. Para representar la semántica de las frases es muy común que utilice *frames* y pares atributo-valor.
- C. Aparte de determinar los huecos y sus posibles valores expresados en la frase, tiene que identificar los aspectos propios de un diálogo como, por ejemplo, los puntos de coincidencia o los actos del habla.
- D. Tiene como objetivo extraer la semántica de las frases en un turno de palabra.

6. Indica las afirmaciones correctas sobre el módulo de gestión del diálogo en un agente conversacional:

- A. Es el elemento principal del agente conversacional.
- B. Analiza la representación semántica de la frase extraída por el módulo de comprensión del lenguaje natural.
- C. Es imprescindible mantener el estado del diálogo y el flujo de la conversación para la gestión del diálogo.
- D. Se puede implementar como una máquina de estados finitos o como un proceso de decisión de Márkov.

7. Indica las afirmaciones correctas sobre el módulo de generación del lenguaje natural en un agente conversacional:

- A. Tiene como objetivo elegir los conceptos que se quieren expresar al usuario y además planificar cómo expresar estos conceptos en palabras.
- B. Realiza una tarea de planificación del contenido totalmente independiente de la tarea realizada por el módulo de gestión de diálogo.
- C. Escoge las estructuras sintácticas y las palabras que se necesitan para expresar el concepto.
- D. Es el módulo de salida de la arquitectura del agente conversacional.

8. Indica las afirmaciones correctas sobre los tipos de *chatbot* basados en corpus:

- A. Los *chatbot* basados en la recuperación de información usan como turno de respuesta a una consulta una respuesta que aparezca en el corpus y cuyo turno anterior sea lo más parecido posible a la consulta.
- B. Los *chatbot* basados en el aprendizaje supervisado de la transducción de secuencias se implementan utilizando transductores de autómatas finitos.
- C. Los *chatbot* secuencia a secuencia se basan en la idea de que generar una respuesta es una tarea de transducción del turno de usuario al turno del sistema, es decir, traducir el turno de usuario en una respuesta del sistema.
- D. Los *chatbot* secuencia a secuencia utilizan métodos de aprendizaje por refuerzo para aprender a escoger las respuestas que hacen que la conversación sea más natural.

9. Indica las afirmaciones correctas sobre los *chatbots* basados en corpus:
- A. Aplican técnicas de minería de datos a un conjunto de conversaciones entre humanos para extraer las posibles respuestas del *chatbot* al usuario.
 - B. Las respuestas del *chatbot* se pueden extraer de un texto que no sea un diálogo.
 - C. Utilizan poca información del contexto de la conversación para generar una respuesta.
 - D. Suelen generar la respuesta basándose solamente en el turno de palabra inmediatamente anterior.
10. Indica las afirmaciones correctas sobre los *chatbots* basados en reglas:
- A. Las reglas conforman una gramática libre de contexto y están anotadas semánticamente.
 - B. Cada regla está asociada a una palabra clave en la frase.
 - C. Implementan un tipo de conversación poco común en la cual el *chatbot* no conoce casi nada del mundo real.
 - D. ELIZA, uno de los primeros agentes conversacionales, implementa este tipo de estrategia.