

Telefonica

Whitepaper Diseño Conversacional

Marta Pérez García
Sarita Saffon López
Ana Molina Alarcón



Introducción

“No longer the sovereign property of humans, Speech has become an ability we share with machines” - Sarah Borruso

Como en cada nuevo capítulo desde que la tecnología se puso al servicio de los humanos, el diseño de productos digitales está experimentando un nuevo cambio de paradigma que impactará indudablemente en la manera en la que diseñamos los servicios: la conversación como modelo de interacción.

Hace varias décadas, estaríamos enfocándonos únicamente en productos físicos, en qué materiales tienen, cómo es su ergonomía física, cómo lo usan las personas, etc. Sin embargo, desde hace un par de décadas, con el cambio de paradigma de los productos físicos a los digitales, ese escenario también viró, y se comenzaron a tener en cuenta la manera en la que las personas interactuamos con la tecnología, qué buscamos, cómo interactuamos con las interfaces, y cómo nos vamos adaptando cada vez a esa vida digital en el terreno personal y profesional. Aparte del cambio antes descrito, ahora, y desde hace ya varios años, nos encontramos con un contexto aún más nuevo; las interfaces visuales han dado paso a nuevos modelos de interacción, como es el caso de las interfaces de voz. Modelos que se popularizan con la llegada de los asistentes virtuales presentes en la mayoría de los dispositivos inteligentes. Estos asistentes, que durante los últimos tres años han perseguido simplificar la relación entre los humanos y la tecnología, incorporan este nuevo mecanismo de interacción permitiéndonos mediante conversaciones naturales y sencillas: gestionar mejor nuestro día a día, obtener la información que buscamos sin tocar la pantalla, o realizar compras de forma sencilla. Un prometedor inicio en el que mucho está por llegar.

Estas nuevas interfaces, que incorporan la conversación, representan nuevos retos en el mundo del diseño y de la experiencia de usuario, en el mundo de los negocios y en el modelo de relación entre empresas y clientes. Este nuevo canal o punto de contacto con los clientes requiere de una gran profundidad de desarrollo de experiencia, para no solo dar respuestas correctas, sino para generar relaciones de calidad basadas en la confianza entre las personas y las empresas.

Pero antes de adentrarnos en todo eso, empecemos por lo básico: **¿Qué es una conversación?**

El diccionario de la Real Academia define una conversación como la acción de hablar familiarmente una o varias personas con otra u otras. El diccionario de Cambridge, a su vez, establece que para que una conversación se produzca, se deben expresar ideas y sentimientos, responder preguntas e intercambiar información entre las varias personas involucradas. Ambas definiciones recalcan el hecho de que en una conversación participa más de una persona, sin embargo, cuando nos referimos a sistemas digitales como los asistentes virtuales (AVs), una conversación puede ocurrir implicando a una sola persona interactuando con una máquina. De hecho, cuando consideramos la disciplina emergente de Diseño Conversacional, el propio término no ha sido ampliamente explorado y documentado. Actualmente el término está definido como el campo que estudia cómo se pueden crear óptimas conversaciones entre un agente artificial y una persona. En este artículo vamos a explorar más en profundidad el estado actual en el que se encuentra este campo y la forma en la que lo estamos abordando en Aura, el asistente virtual de Movistar.

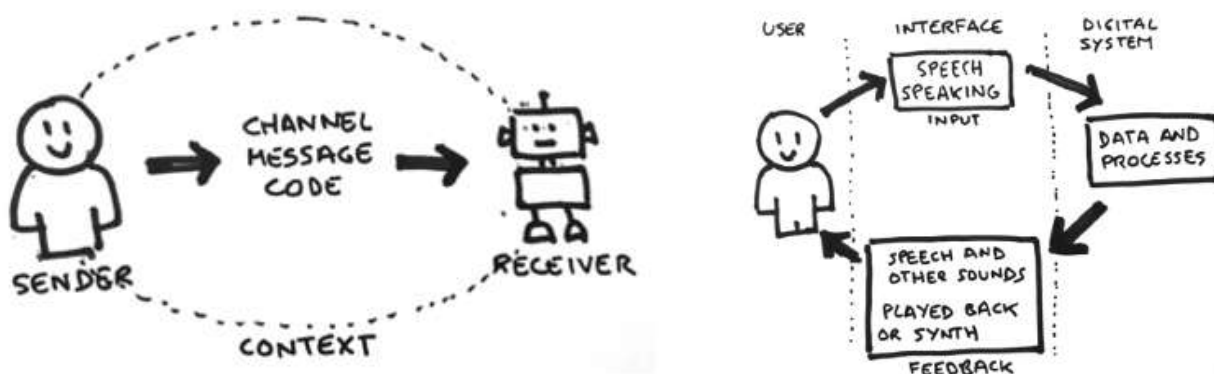


Ilustración: Cómo se articula una conversación entre una persona y una máquina (Autoría: Cristina Santa Cecilia).

¿Cómo están las organizaciones gestionando estas conversaciones entre humanos y máquinas?

Aunque las conversaciones entre humanos y máquinas ya son un hecho habitual dentro de las empresas y se aplica en la estrategia de experiencia de cliente para generar una nueva relación y un nuevo punto de contacto con sus clientes, varias aún se enfrentan al problema de comprender realmente cómo resolver un diálogo, ya que aun teniendo en mente algunas frases tipo que podrían usarse para solicitar una tarea, la manera en que las diferentes personas adaptan su manera de hablar, los giros que puede dar una misma conversación, no están, ni pueden estar, totalmente definidos.

Una posible razón por la que las empresas se encuentran en esta encrucijada enfrentándose al tema de las conversaciones entre las personas y las máquinas es que, si bien el lenguaje ha penetrado el ámbito del diseño, conocido por ayudar a enfrentarse a temas complejos, el enfoque que se le ha dado es de priorizar el intercambio de datos, el desarrollo del diálogo como tal y del flujo de la conversación, más que diseñar el resultado final. Aunque lo primero es importante, no se está aprovechando al máximo lo que esta disciplina puede ofrecer, que es concentrarse en lo que va más allá de la interacción literal, y ver la experiencia de usuario como un todo que debe ser diseñado en su totalidad.

Es decir, las empresas tecnológicas tienen más claras las funcionalidades de las que quieren disponer en sus asistentes virtuales que la propia articulación de las conversaciones que los usuarios van a tener con ellos, por lo que existe una disyuntiva entre lo que la empresa puede hacer a nivel tecnológico y la definición de la forma en la que se desarrollará la experiencia. Es más, a menudo se trabaja más en resolver mediante el asistente cosas que ya se resuelven de forma eficiente por otros caminos, antes que en ofrecer un valor añadido con las experiencias conversacionales.

¿Podemos diseñar una conversación con las herramientas que tenemos hoy?

Hoy en día la tecnología está cambiando estos paradigmas, permitiéndonos diseñar y desarrollar conversaciones con lenguaje natural entre personas y sistemas digitales. Existen nuevas tecnologías como pueden ser ASR (Reconocimiento del habla automático) o NLP (Procesado del lenguaje natural), que hacen que estas interacciones entre los dos sujetos (hombre y máquina) parezcan realmente una conversación.

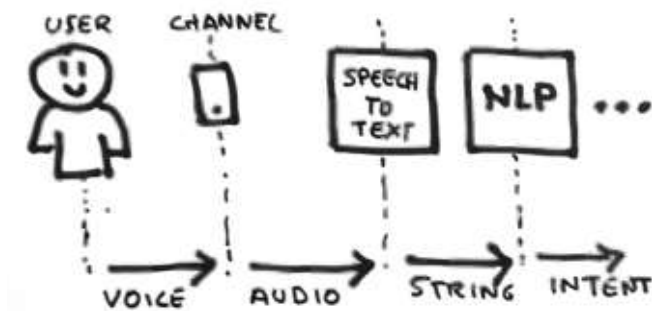


Ilustración: Procesado del lenguaje natural y el reconocimiento automático del diálogo (Autoría: Cristina Santa Cecilia)

Sin embargo, en este tipo de procesos, como podemos observar en la ilustración, existen una gran variedad de pasos, lo que conlleva muchas oportunidades para que algo pueda no funcionar correctamente en el proceso. Es por ello por lo que, aun teniendo este tipo de tecnologías, debemos estar preparados para lo inesperado, puesto que nuestras conversaciones no responden a un flujo lineal. Interactuamos y cooperamos entre nosotros para que la comunicación llegue a buen término porque compartimos un contexto común y porque estamos entrenados para ello. Emular esa capacidad humana de forma natural es un reto enorme, mientras la tecnología resuelve ese gran reto, diseñar y gestionar los errores podría ser la solución. De hecho, considerar los errores en las conversaciones es algo que debemos de tener muy en cuenta en el diseño conversacional, pues los giros y peticiones pueden ser inesperadas, y las respuestas, por consecuencia, también.

¿Qué aspectos son clave a la hora de diseñar una conversación entre una persona y un asistente virtual?

Varios estudios de diversas áreas han analizado el amplio mundo de las conversaciones y proveen información que es de mucho valor en el momento de investigar qué se debe tener en cuenta para diseñar conversaciones que puedan llevarse a cabo entre un ser humano y una máquina.

Por ejemplo, en 2013 el filósofo del lenguaje Paul Grice desarrolló unos principios que deben seguir las conversaciones para que se lleven a cabo con éxito, basándose en que una de las partes involucradas siempre espera que la otra parte se comprometa totalmente al tema que se esté discutiendo. Es decir, cuando nosotros hablamos con alguien

esperamos que esa persona nos escuche, nos conteste, que se genere un diálogo en lugar de un monólogo donde hay un único emisor. A estos principios de Grice se les conoce como las *Máximas Conversacionales de Grice*, y se resumen a continuación:

Calidad	Se enfoca en que lo que se diga es certero y se base en bastante evidencia .
Cantidad	Trata de que se provea suficiente información, pero sin excederse .
Relación	Se centra en que aquello que se diga tenga relevancia con el tema tratado.
Manera	Denota la importancia de evitar la ambigüedad, y promover la claridad y el orden .

Tabla: Máximas conversacionales de Paul Grice (2013).

Otro elemento sobre las conversaciones, interesante para considerar, es la *Teoría de la Relevancia*, generada por los científicos cognitivos Dan Sperber y Deidre Wilson, mientras interpretaban las expresiones que se dan dentro de una conversación, con el fin de identificar cómo logramos transmitir un mensaje más allá del sentido literal de las palabras que utilizamos. En esta teoría se presenta que dentro de la comunicación verbal hay dos intenciones por parte del comunicador, por un lado, la de expresar algo a una audiencia (intención informativa), y por el otro, la de manifestar la intención que tiene este al expresarse (intención comunicativa). Esta teoría es especialmente relevante cuando consideramos la idea de la ironía y de lo complejo que es entrenar a las máquinas para que la entiendan. Incluso entre personas, a veces, estos matices de sarcasmo, ironía, dobles sentidos no son capaces de ser captados, así que no es de extrañar que los bots estén ahora todavía en esos primeros estadios de comprensión literal.

Una tercera fuente muy valiosa de información sobre el diseño conversacional, y que además se enfoca en aquellas que ocurren con máquinas, es Cathy Pearl, quien describe una serie de constructos necesarios para lograr crear conversaciones exitosas entre personas e interfaces de voz (VUIs). Estos cinco *Principios de la Conversación* son:

Desambiguación	Hacer preguntas de seguimiento para entender qué se está diciendo.
Contextualización	Utilizar el conocimiento de la situación para proveer información más precisa.
Cooperación	La importancia de dar la cantidad necesaria de información y su relevancia con el tema.
Aprendizaje	Aprender de lo que el usuario expresa y aplicar ese conocimiento para hacerle la vida más fácil.
Demostración	Dejar claro qué sabe el sistema cuando no ha entendido lo que el usuario quiere decir.

Tabla: Principios Conversacionales de Cathy Pearl (2016).

Hablando con las máquinas: ¿Qué esperan las personas de una conversación con un Asistente Virtual?

En Telefónica también hemos explorado en profundidad el campo de las conversaciones entre seres humanos y máquinas, más específicamente con los asistentes virtuales y con Aura, el Asistente Virtual de Movistar. A través de nuestras investigaciones en varios países y con perfiles de usuarios muy diferentes, tanto a nivel demográfico como psicográfico, hemos observado cómo son estas conversaciones y cómo se articulan.

En este apartado ahondaremos en las expectativas, primeras experiencias, ventajas buscadas y mucho más en el mundo del diseño conversacional.

Ventajas que buscamos

El punto de partida debemos considerar cuáles son los beneficios que buscan los usuarios a la hora de mantener una conversación con un asistente virtual en lugar de utilizar una app, web, un *call center*, o cualquier otro canal para interactuar con una empresa o un servicio. En nuestra experiencia, hemos identificado dos beneficios principales: el primero es cómo estas conversaciones permiten el tan codiciado *multitasking*, poder realizar dos tareas al mismo tiempo. Por ejemplo, el usuario puede preguntarle a su asistente que realice una llamada o ponga una alarma mientras está cocinando. Además de permitir realizar dos acciones a la vez, estas conversaciones con los asistentes virtuales permiten que se realicen por voz, manteniendo las manos libres para cualquier otra acción al mismo tiempo. En segundo lugar, los usuarios buscan que estas conversaciones les hagan la vida más fácil. De forma natural, estamos diseñados para luchar contra el cambio, aunque sea positivo nos suele generar una dificultad, por lo que los AVs deben dejar claro desde el principio cómo van a hacer la vida del usuario más fácil y las tareas más sencillas.

Primeras experiencias conversacionales

Por otro lado, es importante resaltar que cuando los usuarios comienzan a tener conversaciones con los asistentes virtuales tienen unas experiencias y esperanzas comunes que hemos visto repetidas de forma global: (I) Altas expectativas, puesto que los usuarios esperan tener conversaciones muy complejas y expresarse de la forma que les plazca y que los AVs les puedan comprender y saber qué están tratando de decir. Prácticamente que sean capaces de leer sus mentes; (ii) La tendencia a imaginar a una persona detrás de la máquina, detrás de la respuesta de esa conversación. A este fenómeno se le llama antropomorfismo, y se basa precisamente en esto, en atribuir cualidades humanas a lo que no lo es, en este caso a la tecnología; y (iii) la experimentación con el asistente, considerando una multiplicidad de escenarios que no se limitan a las funcionalidades más evidentes del asistente. Este tipo de comportamiento tiene como objetivo para los usuarios testar el nivel de inteligencia del asistente, así como su capacidad para improvisar, enfrentarse a preguntas aleatorias y demostrar el potencial que tiene.

Funcionamiento de la mente humana

Cuando se trata de diseño de conversaciones, también se debe considerar la forma en la que la memoria de las personas funciona a la hora de enfrentarse a una conversación. Al igual que cuando hablamos con una persona, los seres humanos solo somos capaces de retener una cantidad de información, así como permanecer concentrados durante un periodo de tiempo al hablar con una máquina.

En el caso de la **capacidad de atención** o *attention span*, la memoria de los seres humanos solo es capaz de retener información de palabras conocidas y sencillas durante 20-30 segundos de forma seguida con un asistente virtual. Esta cantidad de tiempo puede parecer corto, pero si pensamos en nuestra experiencia escuchando un programa de radio, seguramente nos cueste reproducir lo que el locutor ha dicho hace un minuto, mientras que tendremos frescas las últimas frases que ha mencionado.

Por otro lado, tenemos la **carga cognitiva**, es decir, la cantidad de información por respuesta que somos capaces de gestionar. En el caso de las conversaciones entre usuarios y sistemas digitales, la carga cognitiva ideal se encontraría entre 200 y 300 caracteres por respuesta. Si ofrecemos respuestas más cortas el usuario las puede absorber de forma muy sencilla, pero puede que se queden cortas, y si excedemos esta cantidad, se puede generar estrés y frustración en el usuario, quien no es capaz de seguir la conversación o las instrucciones que se le están ofreciendo. Es por ello, que podemos enriquecer una conversación con otros elementos como las imágenes, los videos o la información escrita, ya que pueden ejercer una función facilitadora, aliviando la carga cognitiva. Sin embargo, así mismo, combinar la información que exponemos (hablada o escrita e ilustrada) en la experiencia podría tener un incremento del esfuerzo cognitivo de decodificación de los mensajes, por lo que el diseño del contenido que ofrecemos a través de las interfaces es algo fundamental. Nuevos profesionales y una nueva atención al cuidado del contenido se volverán más relevantes con la proliferación de esta tecnología o modelo de relación.

Es importante señalar, que el tema del que se esté recibiendo la información impacta en la capacidad de atención de la persona, aumentando o disminuyendo la carga cognitiva que somos capaces de gestionar. No es extraño considerar que cuando un tema realmente nos interesa, nos concentramos para poder comprender y prestar atención a esa información que estamos recibiendo, mientras que cuando el tema no nos interesa demasiado, tenemos una tendencia a “desconectar” más fácilmente.

Cómo hablan las personas realmente

En Aura, a través de multitud de tests con usuarios, así como el análisis que se hace de los *logs*, hemos podido observar las diversas maneras en las que la gente habla con los asistentes virtuales y los patrones que se repiten de forma consistente. En primer lugar, tendríamos a las personas que hablan con **palabras clave**, sin enunciar una frase completa, sino usando solo las palabras que resumen de un golpe lo que están buscando. Por ejemplo, decir “Woody Allen” cuando el usuario busca una película de este director para ver en la tele, o “Rosalía” cuando quieren escuchar música de esta artista. Un segundo escenario es el de las **frases correctas completas**, donde los usuarios enuncian de una forma normal y natural la petición que desean hacer, como si estuvieran hablando con otra persona “Quiero ver una peli de Brad Pitt” “Quiero poner una alarma en el calendario para mañana a las 10 de la mañana”. Finalmente, hemos observado un tercer escenario que se suele dar al comienzo del uso de este tipo de aplicaciones de inteligencia artificial, en el que los usuarios tienen dificultades para enunciar la petición que desean hacer en voz alta y terminan por enunciar **frases muy complejas** que no tienen demasiado sentido para la comprensión del receptor, en este caso el asistente virtual, o que hacen uso de un contexto, conocimiento, o recuerdos del propio usuario. Algunos ejemplos serían “¿Me podrías por favor poner la peli de no sé qué director que va sobre un hombre que se encuentra con....?” O “ Me gustaría escuchar la canción que dice algo sobre ... del artista”. En este caso es importante la función de descubrimiento que tienen este tipo

de sistemas, donde se puede aclarar el tipo de lenguaje, aún natural, que mejor comprende el asistente, clarificando a los usuarios que no deben alterar su forma de hablar para referirse a las peticiones o tareas que desean realizar.

“¿Qué están echando?” (Persona preguntando qué programa está actualmente en la TV)

“Encuentra esa sobre un aeropuerto” (Persona refiriéndose a la película La Terminal)

“Dame PepaPi” (Niño pidiendo ver Peppa Pig)

“Ehh... Tengo... ehhhh, mi factura diferente este mes” (Una persona con dificultades para enunciar la petición de un cargo extra el su factura)

Verbatims: Ejemplos de peticiones complejas de descifrar por parte de los usuarios.

A esto tenemos que sumarle que la forma en la que hablan las personas difiere enormemente no solo en torno a como lo hacen con la tecnología, sino como lo hacen en general con otras personas, y esto se debe a diferentes aspectos. Por ejemplo, nos podemos encontrar con **diferentes niveles de madurez cognitiva**, como pueden ser la de un niño versus la de un adulto. En este caso, la forma de hablar, de expresarse, de usar el vocabulario, etc. es diferente, y el asistente virtual debe considerar estas diferencias. También debemos considerar la **mezcla de acentos** a la que nos vamos a enfrentar, puesto que, aunque compartamos un idioma común, dependiendo del lugar del que seamos o residamos nos encontramos con notables diferencias a la hora de expresarnos e incluso vocalizar los mismos términos. Esto también nos sucede respecto a los **localismos** que los usuarios pueden utilizar y que deben ser entendidos por el asistente virtual. En un tercer nivel nos encontramos con **peculiaridades propias de cada persona**: tratar de encontrar la forma de expresar una petición por voz que nunca antes se ha realizado, diferentes maneras de llamar al mismo objeto, una forma de hablar o de hacer una petición poco concreta o ambigua... Teniendo en cuenta todo lo anterior, se puede ilustrar claramente la complejidad a la que nos enfrentamos a la hora de diseñar conversaciones y de darle a todo tipo de usuarios una experiencia motivadora y relevante con un asistente virtual.

Conclusiones

En resumen, en este artículo hemos querido plasmar la importancia de adentrarnos en las experiencias reales de los usuarios con este tipo de asistentes virtuales para ofrecer una variedad de aspectos y principios que debemos considerar a la hora de diseñar conversaciones: la maduración de la experiencia, comportamientos habituales, expectativas, diversidad a la hora de expresarnos en lenguaje natural, etc.

A pesar de que el diseño conversacional tiene un componente tecnológico muy importante, como una empresa que pone al cliente en el centro de su estrategia, a la hora de enfrentarnos a esta disciplina debemos tener en cuenta la experiencia que queremos ofrecerle al usuario para poder generar y construir una relación relevante y basada en la confianza.

Aura, la IA de Telefónica, crea un nuevo modelo de relación con los clientes

Aura, la Inteligencia Artificial de Telefónica, nace como consecuencia del proceso de transformación digital de la Compañía. Su misión es crear un nuevo modelo de relación con el cliente basado en la confianza, que permita nuevas formas de interacción con los usuarios adaptándose a sus expectativas para que puedan obtener respuestas personalizadas e inmediatas, utilizando un lenguaje natural. Aura debe ser capaz, en primer lugar, de entender al usuario y, en segundo lugar, de entender los datos. Finalmente, Aura debe poder conectar los dos aspectos anteriores para saber qué le puede interesar al usuario y cuál es la mejor manera de proporcionar esa información.

Telefónica lleva muchos años trabajando para poder poner a disposición de sus clientes un hogar cada vez más inteligente, en el que el usuario pueda interactuar con Aura de forma fácil y rápida. Telefónica cuenta con un ecosistema propio, que consiste en productos, servicios y experiencias, que forman el [Hogar Digital](#).

2020 © Telefónica Digital España, S.L.U. Todos los derechos reservados.

La información contenida en el presente documento es propiedad de Telefónica Digital España, S.L.U. ("TDE") y/o de cualquier otra entidad dentro del Grupo Telefónica o sus licenciantes. TDE y/o cualquier compañía del Grupo Telefónica o los licenciantes de TDE se reservan todos los derechos de propiedad industrial e intelectual (incluida cualquier patente o copyright) que se deriven o recaigan sobre este documento, incluidos los derechos de diseño, producción, reproducción, uso y venta del mismo, salvo en el supuesto de que dichos derechos sean expresamente conferidos a terceros por escrito. La información contenida en el presente documento podrá ser objeto de modificación en cualquier momento sin necesidad de previo aviso.

La información contenida en el presente documento no podrá ser ni parcial ni totalmente copiada, distribuida, adaptada o reproducida en ningún soporte sin que medie el previo consentimiento por escrito por parte de TDE.

El presente documento tiene como único objetivo servir de soporte a su lector en el uso del producto o servicio descrito en el mismo. El lector se compromete y queda obligado a usar la información contenida en el mismo para su propio uso y no para ningún otro.

TDE no será responsable de ninguna pérdida o daño que se derive del uso de la información contenida en el presente documento o de cualquier error u omisión del documento o por el uso incorrecto del servicio o producto. El uso del producto o servicio descrito en el presente documento se regulará de acuerdo con lo establecido en los términos y condiciones aceptados por el usuario del mismo para su uso.

TDE y sus marcas (así como cualquier marca perteneciente al Grupo Telefónica) son marcas registradas. TDE y sus filiales se reservan todos los derechos sobre las mismas.

Telefonica
